



Study on advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species

MED_UNITS

Final Report

Project partners: Consorzio Nazionale Interuniversitario per le Scienze del Mare – CoNISMa, Agencia Estatal Consejo superior de Investigaciones Científicas – CSIC, COISPA Tecnologia & Ricerca, Consiglio Nazionale delle Ricerche – IRBIM, Consorzio per il Centro Interuniversitario di Biologia Marina ed Ecologia Applicata “G. Bacci” di Livorno (CIBM), Hellenic Centre of Marine Research – HCMR, IFREMER Département Ressources Biologiques et Environnement - Boulogne sur mer, Instituto Español de Oceanografía – IEO, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale – OGS.

*Report by: Spedicato M.T., R. Cannas, K. Mahé, B. Morales, C. Tsigenopoulos, L. Zane, S. Kavadas, I. Maina, G. Scarcella, P. Sartor, V. Bandelj, T. Russo, F. Fiorentino.
November 2021*

EUROPEAN COMMISSION

European Climate, Infrastructure and Environment Executive Agency
Unit D.3 — Sustainable Blue Economy

Contact: CINEA EMFAF CONTRACTS

E-mail: cinea-emfaf-contracts@ec.europa.eu

*European Commission
B-1049 Brussels*

**Study on advancing fisheries
assessment and management
advice in the Mediterranean by
aligning biological and
management units of priority
species**

MED_UNITS

Final Report

***EUROPE DIRECT is a service to help you find answers
to your questions about the European Union***

Freephone number (*):
00 800 6 7 8 9 10 11

(* The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

LEGAL NOTICE

This document has been prepared for the European Commission, however it only reflects the views of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the Internet (<https://ec.europa.eu>).

Luxembourg: Publications Office of the European Union, 2022

PDF	ISBN 978-92-95225-33-6	doi: 10.2926/909535	HZ-01-22-151-EN-N
-----	------------------------	---------------------	-------------------

© European Union, 2022

Reproduction is authorised provided that the source is acknowledged.

TABLE OF CONTENTS

MAIN CONTRIBUTORS TO THE PROJECT	iii
ACKNOWLEDGEMENTS	v
LIST OF TABLES	vi
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	xi
EXECUTIVE SUMMARY	13
RESUME EXECUTIF	19
RESUMEN EJECUTIVO	26
1. INTRODUCTION	33
1.1 OBJECTIVES OF THE PROJECT AND PROJECT SCOPE	36
2. PROJECT STRUCTURE	38
3. METHODOLOGICAL APPROACH AND BREAKDOWN IN WORKING PACKAGES AND TASKS WITH RELEVANT RESULTS AND CONCLUSIONS	40
3.1 WP0 - PROJECT MANAGEMENT AND COORDINATION	40
3.1.1 Reporting activities	40
3.1.2 Meetings with EASME/MARE	41
3.1.3 Meetings with the project partners	41
3.1.4 Data Call	41
3.1.5 Sampling activities coordination	42
3.2 WP1 - POPULATION GENETICS AND PHYLOGEOGRAPHIC STUDIES FOR IDENTIFICATION OF BIOLOGICAL UNITS OF PRIORITY SPECIES	49
3.2.1 Task 1.1. Literature review	49
3.2.2 Task 1.2. Review bioinformatic methods and tools	53
3.2.3 Task 1.3. Genetic sampling design	56
3.2.4 Task 1.4. Pilot genetic analyses	59
3.2.5 Task 1.5. Comprehensive genetic studies at Mediterranean Sea scale	66
3.2.6 Task 1.6. Deliver a consistent and detailed protocol based on genetic analysis for routine sampling	76
3.2.7 Genetic Workshops	79
3.3 WP2 - OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES	81
3.3.1 Task 2.1. Literature review on otolith shape and microchemistry	81
3.3.2 Task 2.2. Otolith shape analysis of European hake and red mullet	83
3.3.3 Task 2.3. European hake and red mullet microchemistry analyses	87
3.3.4 Task 2.4. Compilation of matrices and data interpretation	91
3.4. WP3 – DELINEATE FISHING GROUNDS AND STOCK ASSESSMENT	93
3.4.1 Task 3.1. Fishing grounds delineation	93
3.4.2 Task 3.2. Combine the results on spatial identification of fishing grounds with the spatial distribution of the target species	95
3.4.3 Task 3.3 Perform stock assessment	97
3.5. WP4 – SYNTHESIS AND PROPOSALS	101
3.5.1 Task 4.1 Perform a SWOT analysis	101
3.5.2 Task 4.2 Integrating results by the different WPs and proposals of new management units	117
3.5.3 Task 4.3 Future improvements for developing adaptive spatial fisheries management	122

4. CONCLUDING REMARKS	127
5. LESSON LEARNT	129
ANNEX 1A: References	130
LIST OF ELECTRONIC ANNEXES TO THE MED_UNITS FINAL REPORT	136
WP0 - ANNEXES	136
WP1 - ANNEXES	136
WP2 - ANNEXES	136
WP3 - ANNEXES	136
WP4 - ANNEXES	137

MAIN CONTRIBUTORS TO THE PROJECT

WP, Task and Deliverable leaders

WP0 - Maria Teresa Spedicato (COISPA), project coordinator

WP1 - Rita Cannas, Lorenzo Zane, Alessia Cariani (CoNISMa) and Costas Tsigenopoulos (HCMR)

WP2 - Kélig Mahé (IFREMER), Beatriz Morales Nin, Sílvia Pérez, Ignacio Catalán, Miquel Palmer (IMEDEA-CSIC), Marco Stagioni (CoNISMa), Andrea Massaro (CIBM), Pierluigi Carbonara (COISPA)

WP3 - Stefanos Kavadas, Irida Maina (HCMR), Tommaso Russo (CoNISMa), Giuseppe Scarcella, Anna Nora Tasseti, Silvia Angelini (IRBIM-CNR), Alessandro Ligas, Mario Sbrana (CIBM), Beatriz Guijarro (IEO), Isabella Bitetto (COISPA)

WP4 - George Tserpes (HCMR), Vinko Bandelj (OGS), Paolo Sartor (CIBM), Fabio Fiorentino (IRBIM-CNR), Tommaso Russo (CoNISMa)

Main contributors from the Consortium:

CoNISMa: Luca Bargelloni, Laura Carugati, Leonardo Congiu, Rachele Corti, Lorenzo D'Andrea, Alice Ferrari, Maria Cristina Follesa, Simone Franceschini, Rafaella Franch, Silvia Iori, Porzia Maiorano, Chiara Manfredi, Ilaria Marino, Francesco Maroso, Riccardo Melis, Federica Piattoni, Elisabetta Piazza, Chiara Papetti, Paola Pesci, Fausto Tinti;

CIBM: Claudia Musumeci; Ilaria Rossetti, Marina Sartini;

CNR-IRBIM: Enrico Nicola Armelloni, Federico Cali, Sabrina Colella, Fortunata Donato, Fabio Falsone, Carmen Ferrà, Vita Gancitano, Michele Luca Geraci, Enrico Armeli Minicante, Monica Panfili, Anna Perdichizzi, Adriana Profeta, Jacopo Pulcinella, Paola Rinelli Giacomo Sardo, Danilo Scannella;

COISPA: Marilena Donnaloia, Giuseppe Lembo, Andrea Pierucci, Walter Zupa;

CSIC-IMEDEA and ICM: Monserrat Demestre, Alfredo García de Vinuesa Gutiérrez;

HCMR: Katerina Anastasopoulou, Leila Bordbar, Petroula Botsidou, Katerina Ekonomaki, Panagiotis Kasapidis, Tereza Manousaki, Danai Mantopoulou, Eirini-Slavka Polovina, Xenia Sarropoulou, Vasso Terzoglou, Alexandros Tsakogiannis, Dimitrios Tsaparis;

IOF: Igor Isajlovic; Nedo Vrgoc;

IFREMER: Geoffrey Bled Defruit, Gregoire Certain, Romain Elleboode, François Garren, Angelique Jadaud, Thibaut Kersaudy, Didier Leroy, Solene Telliez;

OGS: Fabrizio Gianni, Simone Libralato, Cosimo Solidoro.

Main contributors from other research institutions of the Mediterranean and the Atlantic Ocean:

Annaba University, Algeria: Hichem Kara;

Arab Academy for Science, Technology & Maritime Transport, Egypt: Hatem Hanafy, Reda M. Fahim;

DFMR: Charis Charilaou, Ioannis Thasitis for the collection and provision of biological samples;

METU Institute of Marine Sciences, Turkey: Meltem Ok.

Institut National Des Sciences et Technologies de la Mer de Salammbô, Tunisia: Adel Gaamour;

Palestine: Abdelnasser Madi;

National Council for Scientific Research, Lebanon (CNRS-L), National Centre for Marine Sciences, Batroun, Lebanon: Sharif Jemaa, Myriam Lteif;

University of Tunis El Manar: Sabine Mekni;

External expert: Anna Bozzano.

Ecole Pratique des Hautes Etudes: Sébastien Couette, Camille Lutet-Toti;

Thünen Institute: Christoph Stransky;

DTU Aqua: Julia Olivia Davies;

Corsica University: Christelle Paillon, Eric Durieux.

ACKNOWLEDGEMENTS

We acknowledge the EU Commission, DG MARE and CINEA together with the partner institutions involved in this project under the Tender EASME/EMFF/2016/032 for their financial support. We are grateful to Fernando Nieto and Silvia Scalco, CINEA, and to Christos Maravelias and Giacomo Chato Osio, DG MARE, for their helpful comments and support throughout the project and to Venetia Kostopoulou, DG MARE, for the support in the Data Call process.

We would like to thank all the colleagues who contributed to the completion of this project, in particular the work package, task and deliverable leaders who took on the responsibility of delivering substantial sections of the project, and the core teams, for their significant contribution.

We are grateful to the FAO Regional Projects: EastMed, MedSudMed, AdriaMed, and CopeMedII, for their support to the collection of samples from the South Mediterranean. Special thanks to: Marcelo Vasconcellos, Stefano Lelli, Enrico Arneri, Luca Ceriola, Pilar Hernandez who took care of this liaison.

LIST OF TABLES

Table 0.1 Number of geopositions covered by species for the sampling of genetic tissues and otoliths. The geopositions of the meaningful Principal Components Analysis (PCA) of genetic and otolith analyses are also reported, as well as the geopositions referred to the concentration of microelements.....	48
Table 1.1 - Overall evaluation of methods (blue lines) and relative tools, the evaluation is based on the literature data and researchers' expertise.....	54
Table 1.2 – List of the proposed and realized sites for the pilot genetic study. The cells in pale green indicate the sites/species foreseen and sampled in the proposed list, the cell in dark green the site sampled but not foreseen or oversampled, while the cells in red indicate the sites foreseen but not available due to the sampling failure.	58
Table 1.3 Responsibility in task 1.4 for the different partners by marker and species	59
Table 1.4 – List of the DNA extracts for the different species and areas. The cells in green indicate the number of extracts of Category 1- good quality, in light green the extracts of Category 2- medium quality, in orange the extracts of Category 3 the extracts of poor quality, not usable for ddRAD. The last row represents the specimens finally included in the ddRADseq library.	61
Table 2.1 Results of Linear Discriminant Analysis according to tested different hypothesis presenting the composition of subunits and the correct classification rate for the red mullet dataset.	85
Table 2.2 Results of Linear Discriminant Analysis according to tested different hypothesis presenting the composition of subunits and the correct classification rate for the European hake dataset.	86
Table 2.3. List of the Mediterranean and Atlantic areas sampled for otolith microchemistry analyses.....	89
Table 3.1 Scheme of the workplan of methods to be applied in the analyses of the fishing footprint.....	93
Table 3.2. Distribution of the target stocks in the GSAs according to the outputs of the Genetic (G) and Otolith (O) studies (WP1 and WP2).	99
Table 3.3. Stock configurations of the assessments carried out and methodologies employed.	99
Table 3.4. Summary of the stock assessments' outputs and comparison with the most recent assessments at GFCM and STECF (last column).	100
Table 4.1 Criteria and qualitative indicators set for the SWOT analysis	101
Table 4.2 – Some guiding examples of interpretation key for evaluating the indicators selected to assist the SWOT. Strengths and Weaknesses.....	102
Table 4.3 - Some guiding examples of interpretation key for evaluating the indicators selected to assist the SWOT. Opportunities and Threats.	103
Table 4.4 – The traffic light approach applied to synthesize the SWOT results.....	104
Table 4.5 – Results of SWOT analysis as regards the GENETIC ANALYSIS (WP1).	105
Table 4.6 – Results of SWOT analysis as regards the OTOLITH SHAPE and MICROCHEMISTRY ANALYSIS (WP 2).	109
Table 4.7 – Results of SWOT analysis as regards the FISHING GROUNDS ANALYSIS (WP 3).	112

Tab 4.8 – Traffic light evaluation for the population genetics approach.	115
Tab 4.9 – Traffic light evaluation for the otolith shape analysis approach.	115
Tab 4.10 – Traffic light evaluation for the otolith microchemistry analysis approach.	115
Tab 4.11 – Traffic light evaluation for the fishing ground identification.	116
Table 4.12 Case studies selected for the scenario modelling in the Task 4.3.....	123

LIST OF FIGURES

Figure 0.1 Sampling area with the representation of subdivision in sampling sub-units. The number in the circles indicate the GSA, the letter the sub-unit within a GSA.	44
Figure 0.2 European hake (<i>M. merluccius</i>). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of the collected samples, while the green triangles the subunits for which results of the Principal Components Analysis (PCA) carried out in WP1 are available.	45
Figure 0.3 Red mullet (<i>M. barbatus</i>). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.	45
Figure 0.4 Giant red shrimp (<i>A. foliacea</i>). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.	45
Figure 0.5 Blue and red shrimp (<i>A. antennatus</i>). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.	46
Figure 0.6 Deep-water rose shrimp (<i>P. longirostris</i>). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.	46
Figure 0.7 Norway lobster (<i>N. norvegicus</i>). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.	46
Figure 0.8 European hake (<i>M. merluccius</i>). GSA and GSA subunits for which results related to the Principal Components Analysis (PCA) carried out in WP2 (Task 2.4) are available. Geopositions are represented.	47
Figure 0.9 Red mullet (<i>M. barbatus</i>). GSA and GSA subunits for which results related to the Principal Components Analysis (PCA) carried out in WP2 (Task 2.4) are available. Geopositions are represented.	47
Figure 1.1 <i>A. foliacea</i> . Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).	50
Figure 1.2 <i>A. antennatus</i> . Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).	51
Figure 1.3 <i>M. merluccius</i> . Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).	51

Figure 1.4 <i>M. barbatus</i> . Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).	52
Figure 1.5 <i>N. norvegicus</i> . Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).	52
Figure 1.6 <i>P. longirostris</i> . Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).	53
Figure 1.7 The designed pipeline components for analysing the raw data and assessing the population structure of MED_UNITS species.	55
Figure 1.8 a) Sampling sites analysed for <i>A. foliacea</i> . b, c) Maps showing with different colors the best scenarios tested with AMOVA: the 3-groups scenario (b) is the one that maximizes the inter-group difference followed by the 5-groups scenario (c). Both scenarios were statistically significant.	69
Figure 1.9 a) Sampling sites analysed for <i>A. antennatus</i> . b, c) Maps showing with different colors the best scenarios tested with AMOVA: the 3-groups scenario (b) is the one that maximizes the inter-group difference followed by the 6-groups scenario (c). Both scenarios were statistically significant.	70
Figure 1.10 a) Sampling sites analysed for <i>M. merluccius</i> . b, c) Maps showing with different colors the best scenarios tested with AMOVA (the Atlantic samples is not shown, it always represents a separate cluster): the 4-groups (1 ATO+ 3 Med) scenario (b) is the one that maximizes the inter-group difference followed by the 8-groups (1 ATO+ 7 Med) scenario (c). Both scenarios were statistically significant.....	71
Figure 1.11 a) Sampling sites analysed for <i>M. barbatus</i> . b) Map showing with different colors the best scenario tested with AMOVA that is the one that maximizes the inter-group difference.	73
Figure 1.12 a) Sampling sites analysed for <i>N. norvegicus</i> . b, c, d) Maps showing with different colors the best scenarios tested with AMOVA (the 2-groups scenario (b) is the one that maximizes the inter-group difference followed by the 3-groups scenarios (c, d). All scenarios were statistically significant	74
Figure 1.13 a) Sampling sites analysed for <i>P. longirostris</i> . b) Map showing with different colors the best scenarios tested with AMOVA: the 3-groups scenario that is the one that maximizes the inter-group difference.....	76
Figure 2.1. European hake: length ranges and trace elements determined by frequency in the previous studies.....	82
Figure 2.2 Available otoliths number for otolith shape analysis of European hake (A) and red mullet (B).	84
Figure 2.4 Identification of boundaries of stocks for hake from otolith shape analysis.	87
Figure 2.5. Mediterranean Sea circulation	89
Figure 2.6 Classification of <i>Merluccius merluccius</i> otolith edge microcomposition by GSA and Mediterranean area.	91
Figure 2.7 <i>Merluccius merluccius</i> groupings by GSA according to otolith shape and microchemistry and both indicators together.	92
Figure 3.1 Fishing footprint for trawlers from AIS data for the EU Mediterranean countries (fishing days) (a) and estimated by MCDA for the non-EU Mediterranean countries (fishing pressure index) (b).	95

Figure 3.2. Flowchart of the methodological steps applied to identify and describe fishing grounds.....	96
Figure 3.3. Estimated hot and cold spots of the <i>Merluccius merluccius</i> for bottom trawling. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate the absence of fishing activity or absence of species.	97
Figure 3.4. Bottom trawler fishing grounds - aggregated hot spots for the six species under investigation. The map indicates the number of species for which a specific cell has been found to be a hot spot.	97
Figure 4.1– Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of European hake samples to the three stocks identified through the analysis of both Genetic data and Otolith shape data. The color of each GSA was obtained mixing green (stock #1), red (stock #2) and blue (stock #3) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1); Bottom right panel: Fishing footprint (total fishing effort in days, for the year 2017) for longliners from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).	119
Figure 4.2 – Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of Norway lobster samples to the four stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing yellow (stock #1), black (stock #2), cyan (stock #3), and magenta (stock #4) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1). ...	120
Figure 4.3 – Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of Deep water rose shrimp samples to the two stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing red (stock #1), and blue (stock #2) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).	121

LIST OF ABBREVIATIONS

Term	Description
AFLP	Amplified Fragment Length Polymorphism
AIS	Automatic Identification System
AMOVA	Analysis of MOlecular VAriance
Ba	Barium
bp	bais pairs
CA	Contracting Authority
Ca	Calcium
CFP	Common Fisheries Policy
CMEMS	Copernicus Marine Environment Monitoring Service
CRM	Certified Reference Material
CR	Control Region
DAPC	Discriminant Analysis of Principal Components
db-MEM	distance-based Moran's Eigenvector Maps
ddRAD	double-digest Random Amplified DNA
EPI C-PCR	Emulsion, Paired Isolation and Concatenation PCR
DG MARE	Directorate-General for Maritime Affairs and Fisheries
DNA	deoxyribonucleic acid
EFA	Elliptical Fourier Analysis
FAO	Food and Agriculture Organization
F_{CT}	relative Fixation Indexes
FDR	False Discovery Rate
FKM	fuzzy k-means
FM	final meeting
FpM	first progress report meeting
FpR	First progress Report
F_{ST}	Fixation index
Gb	Gigabyte
GBS	Genotyping By Sequencing
GC	content of guanine-cytosine
GFCM	General Fisheries Commission for the Mediterranean
GSA	Geographical Sub-Area
GIS	Geographic Information System
HMW	High Molecular Weight
KoM	kick-off meeting
ICES	International Council for the Exploration of the Sea
IM	interim meeting
InR	Inception Report
IR	Interim Report
ISSR	Inter Simple Sequence Repeats
LA-ICPMS	laser ablation inductively coupled plasma mass spectrometry
LDA	Linear Discriminant Analysis
MDS	Multi Dimensional Scaling
MEDITS	MEDIterranean Trawl Survey
Mg	Magnesium
Mn	Manganese
mtDNA	mitochondrial DNA
mtDNA CR	mitochondrial DNA control region
NG	New Generation
NGS	Next Generation Sequencing

PCA	Principal Components Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
pg	picograms
pRDA	Partial Redundancy Analysis
RAPD	Random Amplification of Polymorphic DNA
RDA	Redundancy Analysis
RE	Restriction Enzymes
RFLP	restriction fragment length polymorphism
RNA	Ribonucleic Acid
SC	Steering Committee
SpM	second progress report meeting
SpR	Second progress Report
SNP	Single nucleotide polymorphism
Sr	Strontium
SSR	Single Sequence Repeats
STECF	Scientific, Technical and Economic Committee for Fisheries
SWOT	Strengths, Weaknesses, Opportunities, Threats
VMS	Vessel Monitoring System
1PPM	1st Project Plenary Meeting
2PPM	2nd Project Plenary Meeting
3PPM	3rd Project Plenary Meeting
395 EST-derived	395 expressed sequence tag-derived

EXECUTIVE SUMMARY

Stock identification provides a basis for understanding population dynamics and makes the stock assessment process more robust, thereby developing fisheries management strategies. Multiannual Management Plans under the Common Fishery Policy (EU Reg. 1380/2013) are tools for managing shared stocks in the long term and thus this requires improving our knowledge on biological stock units and fishery management units. Methods for delineating stocks advanced considerably in recent years and include genetic techniques, otolith shape and chemistry, acoustic telemetry, tagging, demographic analysis and meristic data. **The integration of multiple techniques that operate over different temporal and spatial scales** makes it possible to overcome many of the limitations of single technique approaches and **strengthens the inference available from stock structure studies** (Cadrin et al., 2013).

The **identification of fishing grounds is an essential information to delineate the fishing footprints** on the fish and shellfish stocks. To identify fishing grounds different methods are available, based on Automatic Identification System (AIS) and Vessel Monitoring System (VMS). **Linking information on stock boundaries with the one on the localization of the fishing grounds is a key step for the identification of spatial units for fishery management.**

The overall objective of the MED_UNITS project is to identify and match biological and management stock units of several important demersal species in the Mediterranean: **European hake** (*Merluccius merluccius*), **red mullet** (*Mullus barbatus*), **deep water rose shrimp** (*Parapenaeus longirostris*), **giant red shrimp** (*Aristaeomorpha foliacea*), **blue and red shrimp** (*Aristeus antennatus*) and **Norway lobster** (*Nephrops norvegicus*). **The study covers the Geographical Sub-Areas (GSAs) 1-27 (Mediterranean Sea).**

The project structure consists of **5 Work-Packages (WPs)** and 16 Tasks. **WPO-Project management and coordination; WP1-Population genetics and phylogeographic studies for identification of biological units of priority species; WP-2 Otolith shape and microchemistry analyses; WP3-Delineate fishing grounds and stock assessment; WP4-Synthesis and proposals.**

The analyses undertaken in WP1 and WP2 delineate the population units from a biological perspective. WP3 defines the fisheries footprints not necessarily within the boundaries of the current GSAs. The integration of this information takes place in WP4, supported by the explanatory role of ecological/environmental profiles at spatial scale. Overall, this approach is expected to advance fisheries assessment and improve the management advice, reducing the bias associated with the assumption of a given stock unit, when instead multiple stocks are assessed as a single unit or only a portion of a stock is assessed as a closed unit.

WPO - Project management and coordination managed the reporting, institutional meetings, internal meetings and the coordination among the several project activities. Virtual tools were made available to the project for supporting the work organization and sharing documents and data. The coordination of the samplings for the activities of WP1 and WP2 was also carried out, establishing a Sampling Procedure with a **Genetic Hub** to manage the exchange of samples for the genetic analyses and an **Otolith Hub** for the exchange of otoliths. These two hubs **managed the transfer among laboratories of approximately 10500 samples of tissues for genetics and 3700 otolith samples.** Sampling protocols were distributed to the project partners and to the cooperating Institutes. **The sampling strategy was designed to provide a wide coverage also balancing the geographical locations of samples.** Most of the 27 Mediterranean GSAs were divided in subunits (overall 63). **MEDITS survey was the main source of samples for the European Member States.**

A liaison between the Directorate-General for Maritime Affairs and Fisheries (DG MARE), the General Fisheries Commission for the Mediterranean (GFCM)

and the Food and Agriculture Organization (FAO) Regional Projects was established to support the collection of samples also from south and east Mediterranean Countries through the Data Collection Regulation Framework (DCRF). **Samples were collected from Algeria, Tunisia, Egypt, Palestine, Lebanon and Turkey.** Data collected at individual level were stored in a database and complementary information on environmental variables from Copernicus products as well. Overall, 1984 individuals were sampled for *M. merluccius*, 2209 for *M. barbatus*, 1470 for *A. antennatus*, 1693 for *A. foliacea*, 1537 for *N. norvegicus* and 1751 for *P. longirostris*.

WP1 - Population genetics and phylogeographic studies for the identification of biological units of priority species aimed to: review the genetic data available for the six target species; acquire information for the sampling design; deliver a comprehensive review of genetic methods and tools; perform experimental studies on the genetic distribution among and within Mediterranean sub-basins, and provide a detailed protocol for routine sampling and genetic monitoring. **The review of population genetic studies identified preliminary considerations on the background related to stock boundaries, connectivity patterns and population genetic structure of the target species.** This information constituted the basis for the sampling design and the new genetic analyses. The review of genetic methods and tools examined each method/tool for its peculiarities, appropriateness, robustness and accuracy in stock identification. Following the aforementioned evaluation, a highly reproducible pipeline for the analysis of the six species of MED_UNITS was proposed and used within the project. **A first batch of analyses of tissues was carried out in the Pilot Genetic studies' part** for the optimization of protocols. Based on the knowledge from the Pilot Genetic studies, **the totality of the specimens sampled for each species was analysed in the Comprehensive genetic studies' part, allowing an unprecedented biogeographic analysis.**

The genomic method applied was a Genotyping-by-Sequencing (GBS) methodology, constructing reduced-representation libraries in each species. Single Nucleotide Polymorphism markers (SNPs) newly isolated following the double-digest Random Amplified DNA (ddRAD) sequencing were used. The main results by species can be summarized as follows.

For the **giant red shrimp** (*Aristaeomorpha foliacea*), the final dataset was composed of 771 samples (for 30 localities) and 443 high quality SNPs. The results point out an evident lack of genetic differentiation and are generally in agreement with previous studies conducted at smaller geographical scales and less extended sampling points in the Mediterranean Sea. However, using a statistical tool as the hierarchical Analysis of Molecular Variance (AMOVA), **we detected the existence of very weak differentiation between Western, Eastern and Central Mediterranean samples.**

For the **blue and red shrimp** (*Aristeus antennatus*), a total of 1253 SNPs were retained for 886 samples. As for *A. foliacea*, much of the genetic variation was distributed among individuals in the populations, with a **slight support for three groups corresponding to Western Mediterranean, Central Mediterranean, and Eastern Mediterranean.**

For **European hake** (*Merluccius merluccius*) a total of 665 high quality SNPs was retained (for 1,667 samples) and used for all downstream differentiation analyses. **The strongest differentiation was between Atlantic and Mediterranean, but also within the Mediterranean populations following a West to East pattern.**

For **red mullet** (*Mullus barbatus*), the final dataset was composed of 771 samples (for 30 localities) and 853 high quality SNPs. Very low genetic differentiation values were measured, indicating the **absence of clear population structure in the Mediterranean.**

In **Norway lobster** (*Nephrops norvegicus*), the final dataset was composed of 890 samples (for 27 localities) and 730 high quality SNPs. The results showed a significant differentiation of the samples eastern of the Adriatic Sea (GSA17) against the others from the central and Western Mediterranean. Additionally, relatively high and **significant values were also encountered for the separation of the Adriatic Sea**

(GSA17 to 19) from the neighbouring basins to the west (GSA1 to 11) and the east (GSA22).

For deep-water rose shrimp (*Parapenaeus longirostris*), a total of 1,225 SNPs was retained for 782 individuals. Genetic clustering methods and AMOVA analyses indicated the existence of differentiation, and support for a **three-groups scenario: a “western-central” group including samples from Western and Central Mediterranean up to the Strait of Sicily, a “central” group including the remaining samples from the Central Mediterranean except the easternmost Ionian sample, and an “eastern” group including the samples from Eastern Mediterranean and the remaining Ionian.**

Overall, the MED_UNITS sampling was satisfactory for the spatial coverage and number of specimens collected with a total of 10,670 specimens for all six species. In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium quality or even unusable DNAs for the methodology applied. The genotype success of the ‘good DNA quality’ tissues was high (82%), while only 61% of the ‘medium quality’ DNAs included in the libraries ended up producing genotype data. We concluded that the sampling should be designed in order to fulfil the needs of ‘genomics’ requirements (in terms of quality of samples, procedures, storage, timing). Under task 1.6 it was highlighted that **the scientific surveys at sea (as MEDITS or other similar surveys) are a good opportunity to implement proper sampling design for genomic analyses. However, the sampling for genomic analyses cannot be a collateral activity, and adequate resources should be allocated for this activity.** The timeframe for collecting tissues should be carefully defined, and the timing for the experimental phases cannot be too strict. **The collection of tissues should preferentially be realized from alive/freshly caught individuals and the sampling realized as soon as possible. A specific protocol addressing these points was delivered in this project.**

WP2 - Otolith shape and microchemistry analyses conducted a literature review scrutinising about 600 papers to obtain a general overview on applications and methodologies and their potential use. **Ten studies were retained that distinguish *Merluccius* spp. stocks in different areas.** Mediterranean population of *M. merluccius* was evaluated in one study combining a multiple approach (chemical and shape). With reference to *M. barbatus*, **five studies were performed and only one was carried out in the Mediterranean Sea.** For both species, the protocols are common for image acquisition process, extraction of the external outline information and otolith shape analysis as well as for otolith preparation, processing and analysis for microchemistry trace element detection. For each species, the first sexual maturity and the sex could be potential sources of variation. Consequently, only 1 life stage and 1 sex were used by each species.

For **otolith shape analysis**, the used data sets (left and right otoliths) were composed by 1845 otoliths of red mullet from 37 geographical subunits and by 1868 of European hake otoliths from 39 geographical subunits. In addition to the geographical subunits in the Mediterranean, samples were also gathered from 4 ICES areas for comparison purposes. For both species, two complementary analyses were performed using the Linear Discriminant Analysis (LDA) with Jackknifed prediction (Supervised Machine Learning) and the hierarchical clustering (Unsupervised Machine Learning). **For European hake, the optimised classification (correct classification rate = 39.61%) presented 4 groups, distributed as follows: Atlantic Ocean (from ICES IV to ICES VIII), Western Mediterranean Sea (from GSA1 to GSA13), Adriatic Sea with Central Mediterranean Sea (from GSA16 to GSA20) and Eastern Mediterranean Sea (from GSA22 to GSA27).** **For red mullet, the optimised classification (correct classification rate = 37.56%) presented three groups, distributed as follows: Western Mediterranean Sea (from GSA1 to GSA16), Adriatic Sea with Central Mediterranean Sea (from GSA17 to GSA20) and Eastern Mediterranean Sea (from GSA22 to GSA27).**

The analysis of **otolith microchemistry** was based on a subset of the otolith samples used for the otolith shape analysis: 279 otoliths for European hake from 10 Mediterranean subunits, plus two additional areas in the NE Atlantic added for comparison, and 250 otoliths from 10 different Mediterranean GSAs for red mullet. For European hake, 25 isotopes corresponding to the otolith core and otolith edge for each individual were obtained and 19 isotopes for red mullet. **In spite that GSAs differed in otolith microchemistry, the attribution of individuals to the GSA of origin has been correctly predicted from only 30% of the European hake individuals** when using otolith edge data. **The percentage of correct classification increased to 63% when using only Western, Central and Eastern pooled areas** but this increase should be interpreted with caution, given the small number of sampled areas for microchemistry analyses. **For red mullet, correct allocation was 63% for edges and 66% for cores**, albeit a high individual variability that decreased the classification power. Overall, these results can be explained by at least three compatible hypotheses: (1) otolith microchemistry may only reflect water mass features at another spatial scale; (2) the limits of biological populations may include several management units; and (3) alternative processes related with growth rate may also be affecting the microchemical composition and mask the link between water mass features and otolith composition.

The outcomes of otolith shape and chemical composition analyses were combined under the hypothesis that this improves the capability of predicting GSA membership of a given fish from its otolith features. As regards European hake, data for both otolith shape and chemistry of 159 fish from 10 GSA subunits were used, while for red mullet 237 fish from 10 GSA subunits were used. Cross-validated correct predictions of population membership inferred from Linear Discriminant Analysis (LDA) increased to 42.2% after merging shape and chemistry data for European hake (with only shape data it reached 34.6%), while for red mullet it was 44.7% after merging shape and chemistry data, a bit less than using only the chemical composition of the otolith edge (47.2%). Therefore, **combining the two sources of data implied a slight improvement of accuracy for hake but a slight decline for red mullet. All the trials showed moderate success and it is necessary to increase the spatial coverage and the total number of individuals to improve the stock identification using otolith shape and chemical composition. Analysis of otolith shape alone was performed on much larger samples for both species. The time and cost of otolith microchemistry analysis precluded the same sampling coverage.**

WP3 - Delineate fishing grounds and stock assessment aimed at identifying and characterizing fishing grounds over the Mediterranean waters, including where possible non-EU fleets (which is an aspect of great importance), even combining different data sources (e.g. Automatic Identification System - AIS, Vessel Monitoring System - VMS, experimental data from surveys, fisheries statistics data), processing and modelling approaches and devoting part of the work to the quantitative comparison of results. The main steps were: (i) estimation of the potential fishing grounds by processing AIS data and related hot spot analysis; (ii) Multi-Criteria Decision Analysis (MCDA) applied on environmental and fleet data for both large and small-scale components of the fleets and (iii) Cascaded Multilayer Perceptron Network (CMPN) applied on environmental and fleet data for the large-scale component of the fleets. The produced outcomes emphasize the importance of applying and comparing different methods. **The effort dataset was reconstructed by processing AIS data and complemented using estimated (modelled) effort data in the southern/eastern basin, where the AIS is still poorly adopted and received.** Outputs were validated against those freely provided by other projects (EMODnet, Global Fishing Watch) and, in some areas, by comparing the pattern obtained using only AIS data with the one obtained integrating AIS and VMS data (H2020 Project MINOUW). It must be acknowledged that **this “new” fishing dataset covers for the first time the whole highly productive Mediterranean basin and may inform the wider scientific community, as well**

as those involved with policy and management, on fishing footprints in Mediterranean ecosystem.

WP3 also delineated fishing grounds by species or group of species. A methodological framework, based on spatial analytical techniques, for combining the fishing effort (for bottom trawl, longlines, gillnets and trammel nets) and the spatial distribution of biomass for the investigated species, was applied. The main steps followed were: (i) estimation of the potential fishing grounds by species; (ii) hot spot analysis and (iii) the production of aggregated hot and cold spots. **The outcomes demonstrate the possibility for fishers to harvest a certain species in a specific area, while providing spatial information about the number of exploited species in several fishing grounds. The main results revealed important fishing grounds over the entire Mediterranean Sea**, namely in certain areas of the Adriatic, Tyrrhenian, Strait of Sicily, Aegean, eastern Ionian, Balearic, Alboran, Libyan and Levantine Seas.

Within WP3 **thirteen novel stock assessments were carried out** and compared, when possible, with the assessments routinely carried out at the level of single GSA or combination of GSAs by GFCM or STECF. The new stock configurations for the six target species were based on WP1 and WP2 outputs. The results of the analyses show that, in most cases, the new stock assessments do not present particular improvements in term of diagnostics and model accuracy. The lack of improvement can be due to several reasons, apart from the new stock configurations (increased data heterogeneity when the number of aggregated GSAs is increased, model settings, etc.). However, it must be acknowledged that **these trials represent a first and promising approach to the assessment of the new stock configurations and further investigation shall be implemented** before scientific advice can be provided in a reliable and robust way.

WP4 - Synthesis and proposals aimed at synthesizing and combining the results obtained in the WPs 1-3. **A SWOT analysis provided an evaluation of the internal (Strengths/Weaknesses) and external (Opportunities/Threats) factors related to the methods applied in the WPs 1-3.** The results stressed that the genome-wide methods can provide an accurate picture of the genetic differentiation, but if sample processing is not properly followed the quality of results can be greatly reduced. Thus, dedicated samplings are needed. The time to achieve all the process is rather long, about 1 year. The transfer of genetic samples can pose some difficulties, owing to the differences in the legal/bureaucratic aspects among countries. Otolith shape is not much demanding in terms of equipment and costs, while microchemistry is expensive. Samples of otoliths can be obtained from existing data collection campaigns, but otolith shape and microchemistry analyses are applicable only to fish. The identification of fishing grounds has fast progressed and the methodology is consolidated. Access to AIS time series needs a dedicated budget and the access to VMS data is not easy. **The SWOT analysis highlighted the presence of advantages in each method, while at the same time limitations emerged for each approach, pointing out to the need for a continuous integration process with an exchange of knowledge and achievements among the research groups. A data collection for stock identification needs the strengthening of the cooperation and the necessity of investments in capacity building. It would be useful to develop a Regional Sampling Plan in cooperation between the Regional Coordination Group of the DCF and the GFCM.**

The explanatory role of environmental variables at spatial scale in delineating the population structures that emerged from the genetic and otolith analyses was explored using environmental data from Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>). The spatial origin of each individual was considered to reconstruct the spatial structure of the stocks configuration and to investigate the relationship of these stock structures with respect to the main environmental drivers. Fuzzy clustering was used in a first step. Different numbers of potential management units emerged. **For European hake, three stocks were identified on the combination of the genetic and otolith shape data**, one in the

Western Mediterranean, one in the Adriatic-Ionian-Tyrrhenian basins and one Eastern Mediterranean. **For Norway lobster four stocks** were identified; one was characterizing the Adriatic Sea and the subareas of the Ionian and Aegean Sea, while the other three were distributed in the Western Mediterranean. For **deep-water rose shrimp two stocks** were identified and the **discontinuity between the eastern and the western stocks was positioned in the Ionian Sea south of Italy. No stock was clearly identified for either *Aristaeomorpha foliacea* or *Aristeus antennatus*** as the relation with the spatial and environmental variables was low and non-significant. **For red mullet the conclusion is similar**, even though there were some rather inconclusive evidences for the existence of 2 to 3 sub-populations in the Mediterranean. A denser sampling design, perhaps restricted to smaller areas, might help in confirming or rejecting this hypothesis. The most reliable and comprehensive stock configurations by species were selected and **a series of interrelated maps prepared to illustrate a synthesis on the spatial correspondence between stocks, management areas, and fishing grounds.**

Finally, we **tested the potential development of an adaptive spatial fisheries management** through a simulation approach using three case studies. The new stock configurations of giant red shrimp and blue and red shrimp were examined in the Western Mediterranean, giant red shrimp in the Central Mediterranean and deep-water rose shrimp in the Adriatic-Ionian region. These species are covered by a Multiannual Management Plans. **Management scenarios were run using bio-economic simulation models (BEMTOOL and SMART)**, considering the reduction of fishing effort, the improvement of the gear selectivity and spatial closures in areas critical for biological cycles of the targeted species (Essential Fish Habitats). Despite BEMTOOL and SMART assessed the same fisheries under the same scenarios, they were quite different in terms of results. These differences are mainly due to dissimilarities in modelled processes and assumptions. However, **both models evaluated as the best management strategy**, both in terms of gain in Spawning Stock Biomass (SSB) and improvement or light decrease of the current yield, **the scenario characterized by a 10% reduction of the current fishing effort, coupled with an improvement of trawl net selectivity and protection of Essential Fish Habitats (EFHs).** Finally, **an “ideal” roadmap for developing adaptive spatial fishery management aimed to reduce uncertainty in assessment and management procedures was explored**, considering the development of operating models, the simulation of alternative management strategies and evaluation of their performance (against biological, economic, and social objectives). The need for accurate knowledge of populations units and their connectivity was highlighted. Furthermore, **an adaptive management aimed at protecting the spatial structure of the stocks should be considered as a main tool for the identification and protection of nurseries and spawning areas, in order to mitigate both growth and recruitment overfishing.**

RESUME EXECUTIF

L'identification des stocks fournit une base pour comprendre la dynamique des populations et rend plus robuste le processus d'évaluation des stocks, ce qui permet de développer des stratégies de gestion des pêches. Les plans de gestion pluriannuels prévus par la politique commune des pêches (règlement UE n° 1380/2013) sont des outils de gestion à long terme des stocks partagés, ce qui nécessite d'améliorer nos connaissances sur les unités de stocks biologiques et les unités de gestion de la pêche. Les méthodes de délimitation des stocks ont considérablement progressé ces dernières années et comprennent les techniques génétiques mais aussi la forme et la chimie des otolithes, la télémétrie acoustique, le marquage, l'analyse démographique et les données méristiques. **L'intégration de multiples techniques opérant à différentes échelles temporelles et spatiales** permet de surmonter bon nombre des limites des approches mono-marqueur et **renforce les inférences disponibles dans les études de structure des stocks** (Cadrin et al. 2013).

L'identification des lieux de pêche est une information essentielle pour délimiter l'empreinte de la pêche sur les stocks de poissons et de crustacés. Pour identifier les lieux de pêche, différentes méthodes sont disponibles, basées sur les données, Système d'identification automatique (AIS); and Système de surveillance des navires (VMS). **De ce fait, il est possible de relier les informations sur les limites des stocks et celles sur la localisation des lieux de pêche pour aider à l'identification des unités spatiales pour la gestion de la pêche.**

L'objectif global du projet MED_UNITS est d'identifier et de faire correspondre les unités de stock biologiques avec les unités de stock de gestion de plusieurs espèces démersales importantes en mer Méditerranée: **merlu européen** (*Merluccius merluccius*), **rouget barbet de vase** (*Mullus barbatus*), **crevette rose d'eau profonde** (*Parapenaeus longirostris*), **crevette rouge géante** (*Aristaeomorpha foliacea*), **crevette bleue et rouge** (*Aristeus antennatus*) et **langoustine** (*Nephrops norvegicus*). **L'étude couvre les sous-zones géographiques (appelées GSA) de 1 à 27 (couvrant toute la mer Méditerranée).**

La structure du projet consiste en **5 workpackages (WP)** et 16 tâches. **WP0 - Gestion et coordination du projet; WP1 - Génétique des populations et études phylogéographiques pour l'identification des unités biologiques des espèces prioritaires; WP-2 - Analyses de la forme des otolithes et de leur microchimie; WP3 - Zones de pêche délimitées et évaluation de stocks; WP4 - Synthèse et propositions.**

Les analyses prévues dans les WP1 et WP2 délimitent les unités de population d'un point de vue biologique. Le WP3 délimite les zones de pêche qui ne sont pas nécessairement confinées dans les limites des GSA actuelles. L'intégration de ces informations a eu lieu dans le WP4, soutenu par le rôle explicatif des profils écologiques/environnementaux à l'échelle spatiale. Globalement, cette approche devrait faire progresser l'évaluation des pêcheries et améliorer les conseils de gestion, en réduisant le biais associé à l'hypothèse d'une unité de stock donnée, alors qu'au contraire plusieurs stocks sont évalués comme une seule unité ou seulement une partie d'un stock est évaluée comme une seule unité.

WP0 - Gestion et coordination du projet a géré les rapports, les réunions institutionnelles, les réunions internes et la coordination entre les différentes activités du projet. Des outils virtuels ont été mis à la disposition du projet pour soutenir l'organisation du travail et le partage des documents et des données. La coordination des échantillonnages pour les activités des WP1 et WP2 a également été réalisée, en établissant une procédure d'échantillonnage avec **un Hub génétique** pour gérer l'échange d'échantillons pour les analyses génétiques et **un Hub otolithe** pour l'échange d'otolithes. Ces deux hubs **ont géré le transfert entre laboratoires d'environ 10500 échantillons de tissus pour les analyses génétiques et 3700 échantillons d'otolithes.** Les protocoles d'échantillonnage ont été distribués aux partenaires du projet et aux instituts coopérants. **La stratégie d'échantillonnage a été conçue pour fournir une large couverture en équilibrant également les**

emplacements géographiques des échantillons. La plupart des 27 GSA de la Méditerranée a été divisé en sous-unités (63 au total). **La campagne scientifique MEDITS a été la principale source d'échantillons pour les États Membres européens.**

Une liaison entre la direction générale des affaires maritimes et de la pêche (DG MARE), la Commission générale des pêches pour la Méditerranée (CGPM) et les projets régionaux de l'Organisation pour l'alimentation et l'agriculture (FAO) a été établie pour soutenir la collecte d'échantillons également dans les pays du sud et de l'est de la Méditerranée par le biais du Data Collection Regulation Framework (DCRF). **Des échantillons ont été collectés en Algérie, Tunisie, Egypte, Palestine, Liban et Turquie.** Les données collectées au niveau individuel ont été stockées dans une base de données et des informations complémentaires sur les variables environnementales provenant des produits Copernicus. Au total, 1984 individus ont été échantillonnés pour *M. merluccius*, 2209 pour *M. barbatus*, 1470 pour *A. antennatus*, 1693 pour *A. foliacea*, 1537 pour *N. norvegicus* et 1751 pour *P. longirostris*.

Le WP1 - Génétique des populations et études phylogéographiques pour l'identification d'unités biologiques d'espèces prioritaires vise à: passer en revue les données génétiques disponibles pour les 6 espèces cibles ; acquérir des informations pour le plan d'échantillonnage; fournir une revue complète des méthodes et outils génétiques ; réaliser des études expérimentales sur la distribution génétique parmi et dans les sous-bassins méditerranéens, et fournir un protocole détaillé pour l'échantillonnage de routine et le suivi génétique. **L'examen des études de génétique des populations a permis d'identifier les premières considérations sur le contexte lié aux limites des stocks, aux modèles de connectivité et à la structure génétique des populations des espèces cibles.** Ces informations ont constitué la base du plan d'échantillonnage et des nouvelles analyses génétiques. La revue des méthodes et outils génétiques a examiné chaque méthode/outil pour ses particularités, sa pertinence, sa robustesse et sa précision dans l'identification des stocks. Suite à l'évaluation susmentionnée, un protocole hautement reproductible pour l'analyse des six espèces de MED_UNITS a été proposé et utilisé dans le cadre du projet. **Un premier lot d'analyses de tissus a été réalisé dans le cadre des Études Génétiques Pilotes** afin d'optimiser les protocoles. Sur la base des connaissances acquises lors des études génétiques pilotes, **la totalité des spécimens échantillonnés pour chaque espèce a été analysée dans le cadre des Études Génétiques Globales, ce qui a permis une analyse biogéographique sans précédent.**

La méthode génomique appliquée était une méthodologie de Génotypage par Séquençage (GBS) construisant des bibliothèques à représentation réduite pour chaque espèce. Des marqueurs de polymorphisme nucléotidique unique (SNP) nouvellement isolés après le séquençage de l'ADN amplifié aléatoire à double digestion (ddRAD) ont été utilisés. Les principaux résultats par espèce peuvent être résumés comme suit.

Pour la **crevette rouge géante** (*Aristaeomorpha foliacea*), le jeu de données final était composé de 771 échantillons (pour 30 zones) et de 443 SNP de meilleure qualité. Nos résultats mettent en évidence un manque évident de différenciation génétique et sont généralement en accord avec les études précédentes menées à des échelles géographiques plus petites et des points d'échantillonnage moins étendus en mer Méditerranée. Cependant, à partir d'une Analyse de la variance moléculaire (AMOVA) hiérarchique, **l'existence d'une très faible différenciation a été démontrée pour la séparation de la Méditerranée en 3 sous parties : occidentale, orientale et centrale.**

Pour la **crevette rouge** (*Aristeus antennatus*), un total de 1253 SNP a été retenu pour 886 échantillons. Comme pour *A. foliacea*, une grande partie de la variation génétique était distribuée parmi les individus des populations, avec une **légère contribution pour trois groupes correspondant à la Méditerranée occidentale, la Méditerranée centrale et la Méditerranée orientale.**

Pour le **merlu européen** (*Merluccius merluccius*), un total de 665 SNP de haute qualité a été retenu (pour 1667 échantillons) et utilisé pour toutes les analyses de différenciation en aval. **La différenciation la plus forte a été observée entre l'Atlantique et la Méditerranée, mais également au sein de la Méditerranée, selon un schéma Ouest-Est.**

Pour le rouget barbet de vase (*Mullus barbatus*), le jeu de données final était composé de 771 échantillons (pour 30 zones) et 853 SNP de haute qualité. Des valeurs de différenciation génétique très faibles ont été mesurées, indiquant **l'absence de structure de population claire en Méditerranée.**

Chez la **langoustine** (*Nephrops norvegicus*), le jeu de données final était composé de 890 échantillons (pour 27 zones) et de 730 SNP de haute qualité. Les résultats ont montré une différenciation significative des échantillons à l'est de la mer Adriatique (GSA17) par rapport aux autres échantillons de la Méditerranée centrale et occidentale. De plus, **des valeurs relativement élevées et significatives ont également été rencontrées pour la séparation de la mer Adriatique (GSA17 à 19) des bassins voisins à l'ouest (GSA1 à 11) et à l'est (GSA22).**

Pour la **crevette rose d'eau profonde** (*Parapenaeus longirostris*), un total de 1 225 SNP a été retenu pour 782 individus. Les méthodes de regroupement génétique et les analyses AMOVA ont indiqué l'existence d'une différenciation, et le soutien de **trois potentiels groupes: un groupe "occidental-central" comprenant les échantillons de la Méditerranée occidentale et centrale jusqu'au détroit de Sicile, un groupe "central" comprenant les échantillons restants de la Méditerranée centrale, à l'exception de l'échantillon ionien le plus à l'est, et un groupe "oriental" comprenant les échantillons de la Méditerranée orientale et le reste de la mer ionienne.**

Dans l'ensemble, l'échantillonnage MED_UNITS a été satisfaisant pour la couverture spatiale et le nombre de spécimens collectés, avec un total de 10 670 spécimens pour les six espèces. En général, l'ADN de bonne qualité a été facilement obtenu à partir de tissus frais, tandis que les tissus congelés ont généré une plus grande fraction d'ADN de qualité moyenne, voire inutilisable pour la méthodologie appliquée. Le succès du génotypage des tissus de "bonne qualité" était élevé (82%) alors que seulement 61% des ADN de "qualité moyenne" inclus dans les bibliothèques ont produit des données génotypiques. Nous pouvons donc conclure que l'échantillonnage doit être conçu de manière à répondre aux exigences de la "génomique" (en termes de qualité des échantillons, de procédures, de stockage, de calendrier). Dans le cadre de la tâche 1.6, il a été souligné que **les études scientifiques en mer (comme MEDITS ou d'autres campagnes similaires) sont une bonne occasion de mettre en œuvre un plan d'échantillonnage approprié** pour les analyses génomiques. **Cependant, l'échantillonnage pour les analyses génomiques ne peut pas être une activité collatérale, et des ressources adéquates doivent être allouées à cette activité.** Le calendrier de collecte des tissus doit être soigneusement défini, et le calendrier des phases expérimentales ne doit pas être trop strict. **La collecte de tissus doit être réalisée de préférence sur des individus vivants ou fraîchement capturés et l'échantillonnage doit être réalisé le plus rapidement possible. Un protocole spécifique abordant ces points a été défini dans ce projet.**

Le WP2 - Analyses de la forme des otolithes et de la microchimie a effectué une **revue de la littérature** en examinant environ 600 articles pour obtenir une vue d'ensemble des applications et des méthodologies et de leur potentiel. A la fin, **dix études ont été trouvées qui distinguent les stocks de merlus dans des différentes zones.** La population méditerranéenne de *M. merluccius* a été évaluée dans une étude combinant une approche multi-marqueurs (chimie et forme de l'otolithe). En ce qui concerne *M. barbatus*, **cinq études ont été réalisées et une seule a été effectuée en mer Méditerranée.** Pour les deux espèces, les protocoles sont communs pour le processus d'acquisition d'images, l'extraction des informations du contour externe et l'analyse de la forme des otolithes ainsi que pour la préparation, le traitement et l'analyse des otolithes pour la détection des éléments traces en microchimie. Pour

chaque espèce, la première maturité sexuelle et le sexe peuvent être des sources potentielles de variation. Par conséquent, un seul stade de vie et un seul sexe ont été utilisés pour chaque espèce.

Pour **l'analyse de la forme des otolithes**, les ensembles de données utilisés (otolithes gauche et droit) étaient composés de 1845 otolithes de rouget barbet de vase provenant de 37 sous-unités géographiques et de 1868 otolithes de merlu européen provenant de 39 sous-unités géographiques. En plus des sous-unités géographiques de la Méditerranée, des échantillons ont également été collectés dans 4 zones CIEM en Atlantique à des fins de comparaison. Pour les deux espèces, deux analyses complémentaires ont été réalisées en utilisant l'analyse discriminante linéaire (LDA) avec prédiction jackknifed (apprentissage automatique supervisé) et la classification ascendante hiérarchique (apprentissage automatique non supervisé). **Pour le merlu européen, la classification optimisée** (taux de bonne classification = 39,61%) **a présenté 4 groupes**, qui sont distribués comme suit : **Océan Atlantique** (de CIEM IV à CIEM VIII), **mer Méditerranée occidentale** (du GSA1 à GSA13), **mer Adriatique avec mer Méditerranée centrale** (du GSA16 à GSA20) et **mer Méditerranée orientale** (du GSA22 à GSA27). **Pour le rouget, la classification optimisée** (taux de bonne classification = 37,56%) **a présenté trois groupes**, qui sont distribués comme suit : **mer Méditerranée occidentale** (du GSA1 à GSA16), **mer Adriatique avec mer Méditerranée centrale** (du GSA17 à GSA20) et **mer Méditerranée orientale** (du GSA22 à GSA27).

L'analyse de la **microchimie des otolithes** a été **basée sur un sous-ensemble d'échantillons d'otolithes utilisés pour l'analyse de la forme des otolithes** : 279 otolithes pour le merlu européen provenant de 10 sous-unités méditerranéennes, plus deux zones supplémentaires dans l'Atlantique Nord-Est ajoutées à des fins de comparaison, et 250 otolithes provenant de 10 GSAs méditerranéennes différentes pour le rouget barbet de vase. Pour le merlu européen, 25 isotopes correspondant au noyau et au bord de l'otolithe de chaque individu ont été obtenus et 19 isotopes pour le rouget. **Malgré le fait que les GSA diffèrent dans la microchimie des otolithes, l'appartenance à un GSA a été correctement prédite pour seulement 30% des individus** de merlu européen en utilisant les données du bord de l'otolithe. **Le pourcentage de classification correcte a augmenté à 63% en utilisant uniquement les zones groupées occidentales, centrales et orientales**, mais cette augmentation doit être interprétée avec prudence, étant donné le petit nombre de zones échantillonnées pour les analyses de microchimie. **Pour le rouget, l'attribution correcte était de 63% pour les bords et de 66% pour les noyaux**, malgré une forte variabilité individuelle qui a diminué le pouvoir de classification. Globalement, ces résultats peuvent être expliqués par au moins trois hypothèses compatibles : (1) la microchimie des otolithes peut ne refléter les caractéristiques de la masse d'eau qu'à une autre échelle spatiale ; (2) les limites des populations biologiques peuvent inclure plusieurs unités de gestion ; et (3) d'autres processus liés au taux de croissance peuvent également affecter la composition microchimique et masquer le lien entre les caractéristiques de la masse d'eau et la composition des otolithes.

Les résultats des analyses de la forme et de la composition chimique des otolithes ont été combinés dans l'hypothèse que cela améliore la capacité de prédire l'appartenance à une zone (ici GSA) d'un poisson donné à partir de ses caractéristiques otolithiques. En ce qui concerne le merlu européen, les données relatives à la forme et à la composition chimique des otolithes de 159 poissons de 10 sous-unités GSA ont été utilisées, tandis que pour le rouget, 237 poissons de 10 sous-unités GSA ont été utilisés. Les prédictions correctes validées par recouplement de l'appartenance à une population déduite de l'analyse discriminante linéaire (LDA) ont augmenté à 42,2 % après la fusion des données de forme et de chimie pour le merlu européen (avec les seules données de forme, elles atteignaient 34,6 %), tandis que pour le rouget, elles étaient de 44,7 % après la fusion des données de forme et de chimie, un peu moins qu'en utilisant uniquement la composition chimique du bord de l'otolithe (47,2 %). Par conséquent, **la combinaison des deux sources de données a entraîné une légère amélioration**

de la précision pour le merlu, mais une légère baisse pour le rouget. Tous les essais ont montré un succès modéré et il est nécessaire d'augmenter la couverture spatiale et le nombre total d'individus pour améliorer l'identification des stocks en utilisant la forme des otolithes et la composition chimique. L'analyse de la forme des otolithes seule a été effectuée sur des échantillons beaucoup plus grands pour les deux espèces. Le temps et le coût de l'analyse de la microchimie des otolithes n'ont pas permis d'atteindre la même couverture d'échantillonnage.

Le WP3 - Délimitation des lieux de pêche et évaluation des stocks vise à identifier et à caractériser les lieux de pêche dans les eaux méditerranéennes, y compris si possible les flottes non européennes (ce qui est un aspect de grande importance), même en combinant différentes sources de données (par exemple, AIS, VMS, données expérimentales provenant de campagnes, données statistiques sur la pêche), des approches de traitement et de modélisation et en consacrant une partie du travail à la comparaison quantitative des résultats. Les principales étapes ont été : (i) l'estimation des zones de pêche potentielles par le traitement des données AIS et l'analyse connexe des points chauds ; (ii) l'analyse décisionnelle multicritères (MCDA) appliquée aux données environnementales et de la flotte pour les composantes à grande et à petite échelle des flottes et (iii) le réseau de perceptron multicouche en cascade (CMFN) appliqué aux données environnementales et de la flotte pour la composante à grande échelle des flottes. Les résultats obtenus soulignent l'importance de l'application et de la comparaison de différentes méthodes. **L'ensemble des données d'effort a été reconstruit en traitant les données AIS et complété en utilisant des données d'effort estimées (modélisées) dans le bassin sud/est où l'AIS est encore peu adopté et reçu.** Les résultats ont été validés par rapport à ceux fournis gratuitement par d'autres projets (EMODENET, Global Fishing Watch) et, dans certaines zones, en comparant le modèle obtenu en utilisant uniquement les données AIS avec celui obtenu en intégrant les données AIS et VMS (projet H2020 MINOUW). Il faut reconnaître que **ce "nouvel" ensemble de données de pêche couvre pour la première fois l'ensemble du bassin méditerranéen hautement productif et peut informer la communauté scientifique au sens large, ainsi que les personnes impliquées dans la politique et la gestion**, sur les empreintes de la pêche dans l'écosystème méditerranéen.

En outre, **le WP3 a délimité les zones de pêche par espèce ou groupe d'espèces.** Un cadre méthodologique, basé sur des techniques d'analyse spatiale, a été appliqué pour combiner l'effort de pêche (pour le chalut de fond, les palangres, les filets maillants et les trémails) et la distribution spatiale de la biomasse pour les espèces étudiées. Les principales étapes suivies ont été les suivantes (i) l'estimation des zones de pêche potentielles par espèce ; (ii) l'analyse des points chauds et (iii) la production de points chauds et froids agrégés. **Les résultats produits démontrent la possibilité pour les pêcheurs de récolter une certaine espèce dans une zone spécifique, tout en fournissant des informations spatiales sur le nombre d'espèces exploitées dans plusieurs lieux de pêche. Les principaux résultats ont révélé des lieux de pêche importants sur l'ensemble de la mer Méditerranée**, qui se trouvent dans certaines zones de l'Adriatique, de la Tyrrhénienne, du détroit de Sicile, de la mer Égée, de la mer Ionienne orientale, des Baléares, de la mer d'Alboran, de la Libye et du Levant.

Dans le cadre du WP3, **13 nouvelles évaluations de stocks ont été réalisées** et comparées, lorsque cela était possible, aux évaluations réalisées régulièrement au niveau d'un seul GSA ou d'une combinaison de GSA par la CGPM ou le CSTEP. Les nouvelles configurations de stocks pour les six espèces cibles étaient basées sur les résultats des WP1 et WP2. Les résultats des analyses montrent que dans la plupart des cas, les nouvelles évaluations des stocks ne présentent pas d'améliorations particulières en termes de diagnostic et de précision des modèles. Ce manque d'amélioration peut être dû à plusieurs raisons, en dehors des nouvelles configurations de stocks (hétérogénéité accrue des données lorsque le nombre de GSAs agrégés est augmenté, paramètres du modèle, etc). Cependant, il faut reconnaître que **ces essais**

représentent une première approche prometteuse de l'évaluation des nouvelles configurations de stocks et qu'il faudra poursuivre les recherches avant de pouvoir fournir un avis scientifique fiable et solide.

Le WP4 - Synthèse et propositions vise à synthétiser et combiner les résultats obtenus dans les WP1-3. **Une analyse SWOT a fourni une évaluation des facteurs internes (forces/faiblesses) et externes (opportunités/menaces) liés aux méthodes appliquées dans les WP1-3.** Les résultats ont souligné que les méthodes génomiques peuvent fournir une image précise de la différenciation génétique, mais si le traitement des échantillons n'est pas correctement suivi, la qualité des résultats peut être considérablement réduite. Ainsi, des échantillonnages spécifiques sont nécessaires. Le temps nécessaire pour réaliser l'ensemble du processus est assez long, environ 1 an. Le transfert d'échantillons génétiques peut poser quelques difficultés, en raison des différences entre les aspects juridiques/bureaucratiques des différents pays. La forme des otolithes n'est pas très exigeante en termes d'équipement et de coûts, alors que la microchimie est coûteuse. Des échantillons d'otolithes peuvent être obtenus à partir des campagnes de collecte de données existantes, mais les analyses de la forme des otolithes et de la microchimie ne sont applicables qu'aux poissons. L'identification des lieux de pêche a rapidement progressé et la méthodologie est consolidée. L'accès aux séries temporelles AIS nécessite un budget dédié et l'accès aux données VMS n'est pas facile. **L'analyse SWOT a mis en évidence la présence d'avantages dans chaque méthode, alors que dans le même temps, des limites sont apparues pour chaque approche, soulignant la nécessité d'un processus d'intégration continu avec un échange de connaissances et de réalisations entre les groupes de recherche. Une collecte de données pour l'identification des stocks nécessite le renforcement de la coopération et des investissements dans le renforcement des capacités. Il serait utile de développer un plan d'échantillonnage régional en coopération entre le groupe de coordination régionale du DCF et la CGPM.**

Le rôle explicatif des variables environnementales à l'échelle spatiale dans la délimitation des structures de population qui ont émergé des analyses génétiques et des otolithes a été exploré en utilisant les données environnementales du Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>). L'origine spatiale de chaque individu a été prise en compte pour reconstruire la structure spatiale de la configuration des stocks et pour étudier la relation de ces structures de stocks avec les principaux facteurs environnementaux. Un regroupement flou a été utilisé dans un premier temps. Différents nombres d'unités de gestion potentielles sont apparus. **Pour le merlu européen, trois stocks ont été identifiés sur la base de la combinaison des données génétiques et de la forme des otolithes, un en Méditerranée occidentale, un dans les bassins Adriatique-Ionien-Tyrrhénien et un en Méditerranée orientale. **Pour la langoustine, quatre stocks ont été identifiés** ; l'un caractérisait la mer Adriatique et les sous-zones de la mer Ionienne et de la mer Égée, tandis que les trois autres étaient répartis dans la Méditerranée occidentale. **Pour la crevette rose d'eau profonde, deux stocks ont été identifiés et la discontinuité entre les stocks de l'est et de l'ouest a été positionnée dans la mer Ionienne au sud de l'Italie. Pour les deux crevettes rouges, la relation avec les variables spatiales et environnementales était faible et non significative. Pour le rouget barbet de vase, la conclusion est similaire**, même s'il y avait des preuves plutôt peu concluantes de l'existence de 2 à 3 sous-populations en Méditerranée. Un plan d'échantillonnage plus dense, peut-être limité à de plus petites zones, pourrait aider à confirmer ou à rejeter cette hypothèse. Les configurations de stocks les plus fiables et les plus complètes par espèce ont été sélectionnées et **une série de cartes interdépendantes ont été préparées pour montrer une synthèse sur la correspondance spatiale entre les stocks, les zones de gestion et les lieux de pêche.****

Nous avons testé le développement potentiel d'une gestion spatiale adaptative des pêches par une approche de simulation en utilisant trois cas d'études. Les

nouvelles configurations des stocks de crevette rouge géante et de crevette bleue et rouge ont été examinées en Méditerranée occidentale, la crevette rouge géante en Méditerranée centrale et la crevette rose d'eau profonde dans la région Adriatique-Ionienne. Ces espèces font l'objet de plans de gestion pluriannuels. **Des scénarios de gestion ont été exécutés à l'aide de modèles de simulation bio-économiques (BEMTOOL et SMART)**, en tenant compte de la réduction de l'effort de pêche, de l'amélioration de la sélectivité des engins et des fermetures spatiales dans les zones critiques pour les cycles biologiques des espèces ciblées. Bien que BEMTOOL et SMART aient évalué les mêmes pêcheries dans le cadre des mêmes scénarios, leurs résultats étaient très différents. Ces différences sont principalement dues à des différences dans les processus et les hypothèses modélisés. Cependant, **les deux modèles ont évalué comme meilleure stratégie de gestion**, à la fois en termes de gain de biomasse de reproducteurs (SSB) et d'amélioration ou de légère diminution du rendement actuel, **le scénario S4, caractérisé par une réduction de 10% de l'effort de pêche actuel couplé à une amélioration de la sélectivité des chaluts et à la protection des Habitat essentiel (EFHs)**. Enfin, **une feuille de route "idéale" pour le développement d'une gestion spatiale adaptative des pêches visant à réduire l'incertitude des procédures d'évaluation et de gestion a été élaborée**, en considérant le développement de modèles d'exploitation, la simulation de stratégies de gestion alternatives et l'évaluation de leurs performances (par rapport aux objectifs biologiques, économiques et sociaux). La nécessité d'une connaissance précise des unités de populations et de leur connectivité a été mise en évidence. En outre, **une gestion adaptative visant à protéger la structure spatiale des stocks devrait être considérée comme un outil principal pour l'identification et la protection des nurseries et des zones de frai, afin d'atténuer à la fois la surpêche de croissance et de recrutement**.

RESUMEN EJECUTIVO

La identificación de poblaciones proporciona una base para comprender la dinámica de la población y hace más sólido el proceso de evaluación de las poblaciones, desarrollando así estrategias de gestión pesquera. Los planes de gestión plurianuales en el marco de la política pesquera común (Reglamento UE n.1380/2013) son herramientas para la gestión de las poblaciones compartidas a largo plazo y, por lo tanto, esto requiere mejorar nuestros conocimientos sobre las unidades biológicas de poblaciones y las unidades de gestión pesquera. Los métodos para delinear las poblaciones avanzaron considerablemente en los últimos años e incluyen técnicas genéticas, forma y química de otolitos, telemetría acústica, etiquetado, análisis demográfico y datos merísticos. **La integración de múltiples técnicas que operan en diferentes escalas temporales y espaciales** permite superar muchas de las limitaciones de los enfoques de una sola técnica y **fortalece la inferencia disponible de los estudios de estructura de stock** (Cadrin et al., 2013).

La identificación de los caladeros es una información esencial para delimitar las huellas pesqueras en las poblaciones de peces y mariscos. Para identificar los caladeros, se dispone de diferentes métodos, basados en datos del Sistema de Identificación Automática (AIS) y del Sistema de Monitoreo de Buques (VMS). **Vincular la información sobre los límites de las poblaciones con la de la localización de los caladeros es un paso clave para la identificación de unidades espaciales para la gestión pesquera.**

El objetivo general del proyecto MED_UNITS es identificar y hacer coincidir las unidades biológicas con las unidades de gestión de población de varias especies demersales importantes en el Mediterráneo: **Merluza europea** (*Merluccius merluccius*), **salmonete** (*Mullus barbatus*), **gamba rosa de aguas profundas** (*Parapenaeus longirostris*), **gamba roja gigante** (*Aristaeomorpha foliacea*), **gamba azul y roja** (*Aristeus antennatus*) y **cigala** (*Nephrops norvegicus*). **El estudio abarca las subáreas geográficas (GSA) 1-27 (Mar Mediterráneo).**

La estructura del proyecto consta de **5 paquetes de trabajo (WPs)** y 16 tareas: **WP0-Gestión y coordinación de proyectos; WP1-Genética de poblaciones y estudios filogeográficos para la identificación de unidades biológicas de especies prioritarias; WP2-Análisis de forma y microquímica de otolitos; WP3-Delinear los caladeros y evaluación de las poblaciones; WP4-Síntesis y propuestas.**

Los análisis previstos en WP1 y WP2 están delineando las unidades de población desde una perspectiva biológica. WP3 delinea las huellas pesqueras no necesariamente confinadas dentro de los límites de las GSA actuales. La integración de esta información tuvo lugar en el WP4, apoyada por el papel explicativo de los perfiles ecológicos/ambientales a escala espacial. En general, se espera que este enfoque avance en la evaluación de la pesca y mejore el asesoramiento y gestión, reduciendo el sesgo asociado con la suposición de una unidad de población determinada, cuando en su lugar se evalúan varias poblaciones como una sola unidad o solo una parte de una población se evalúa como una unidad cerrada.

WPO - Gestión y coordinación del proyecto gestionó los informes, reuniones institucionales, reuniones internas y la coordinación entre las diversas actividades del proyecto. Se puso a disposición del proyecto herramientas virtuales para apoyar la organización del trabajo y compartir documentos y datos. También se llevó a cabo la coordinación de los muestreos para las actividades de WP1 y WP2, estableciendo un Procedimiento de Muestreo con un **Centro Genético** para gestionar el intercambio de muestras para los análisis genéticos y un **Hub de Otolitos** para el intercambio de otolitos. Estos dos centros **gestionaron la transferencia entre laboratorios de aproximadamente 10500 muestras de tejidos para muestras genéticas y 3700 muestras de otolitos.** Se distribuyeron protocolos de muestreo a los socios del proyecto y a los institutos cooperantes. **La estrategia de muestreo se diseñó para proporcionar una amplia cobertura que también equilibrara las ubicaciones geográficas de las muestras.** La mayoría de las 27 GSA mediterráneas se dividieron

en subunidades (en total 63). **Las campañas MEDITS fue la principal fuente de muestras para los países europeos.**

Se estableció un enlace entre la Dirección General de asuntos Marítimos y Pesca (DG MARE), la Comisión General para la Pesca del Mediterráneo (CGPM) y los proyectos regionales de la Organización para la Agricultura y Alimentación (FAO) para apoyar la recogida de muestras también de los países del Mediterráneo meridional y oriental a través del Marco de Regulación de la Recopilación de Datos (DCRF). Se recogieron muestras de Argelia, Túnez, Egipto, Palestina, Líbano y Turquía. Los datos recopilados a nivel individual se almacenaron en una base de datos con información complementaria sobre las variables ambientales de los productos Copernicus. En general, se muestreó 1984 individuos para *M. merluccius*, 2209 para *M. barbatus*, 1470 para *A. antennatus*, 1693 para *A. foliacea*, 1537 para *N. norvegicus* y 1751 para *P. longirostris*.

WP1 - Genética de poblaciones y estudios filogeográficos para la identificación de unidades biológicas de especies prioritarias tiene como objetivo: revisar los datos genéticos disponibles para las 6 especies objetivo; la adquisición de información para el diseño del muestreo; la realización de un examen exhaustivo de los métodos e instrumentos genéticos; realizar estudios experimentales sobre la distribución genética entre y dentro de las subcuencas mediterráneas, y proporcionar un protocolo detallado para el muestreo de rutina y el seguimiento genético. **En el examen de los estudios genéticos de poblaciones se identificaron las primeras consideraciones sobre los antecedentes relacionados con los límites de las poblaciones, los patrones de conectividad y la estructura genética de las poblaciones de las especies objetivo.** Esta información representó la base para el diseño del muestreo y los nuevos análisis genéticos. En la revisión de los métodos y herramientas genéticas se examinó cada método/herramienta en cuanto a sus peculiaridades, idoneidad, solidez y precisión en la identificación de las poblaciones. **Se realizó un primer lote de análisis piloto de tejidos para los estudios genéticos.** Con base en los resultados de los estudios genéticos piloto, **la totalidad de los especímenes muestreados para cada especie se analizaron en los estudios genéticos finales, lo que permitió un análisis biogeográfico sin precedentes.**

El método genómico aplicado fue una metodología de Genotipado por Secuenciación (GBS) que construyó bibliotecas de representación reducida en cada especie. Se utilizaron marcadores del polimorfismo de un único nucleótido (SNP) recién aislados tras la secuenciación de ADN amplificado aleatorio (ddRAD) de doble digestión. Los principales resultados por especies se pueden resumir de la siguiente manera.

Para la **gamba roja gigante** (*Aristaeomorpha foliacea*), el conjunto de datos final se compuso de 771 muestras (para 30 localidades) y 443 SNP de mayor calidad. Nuestros resultados señalan una evidente falta de diferenciación genética y generalmente están de acuerdo con estudios previos realizados a escalas geográficas más pequeñas y puntos de muestreo menos extendidos en el Mar Mediterráneo. Sin embargo, utilizando el análisis jerárquico de la varianza molecular (AMOVA), **se demostró la existencia de una diferenciación muy débil entre el Mediterráneo occidental, oriental y central.** Para la **gamba azul y roja** (*Aristeus antennatus*), se retuvieron un total de 1253 SNP para 886 muestras. En cuanto a *A. foliacea*, gran parte de la variación genética se distribuyó entre los individuos de las poblaciones, con una **ligera evidencia de tres grupos correspondientes al Mediterráneo occidental, central y oriental.**

Para la **merluza europea** (*Merluccius merluccius*) se mantuvo un total de 665 SNP de alta calidad (para 1.667 muestras) y se utilizaron para todos los análisis de diferenciación posteriores. **La diferenciación más fuerte fue entre Atlántico y Mediterráneo, pero también en el Mediterráneo las poblaciones siguieron un patrón de oeste a este.**

Para el salmonete (*Mullus barbatus*), el conjunto de datos finales se compuso de 771 muestras (para 30 localidades) y 853 SNP de alta calidad. Se midieron valores de diferenciación muy bajos, lo que indica la **ausencia de una estructura poblacional clara en el Mediterráneo.**

En la **cigala** (*Nephrops norvegicus*), el conjunto de datos finales se compuso de 890 muestras (para 27 localidades) y 730 SNP de alta calidad. Los resultados mostraron una diferenciación significativa de las muestras del este del mar Adriático (GSA17) frente a las otras del Mediterráneo central y occidental. Además, también se encontraron **valores relativamente altos y significativos para la separación del mar Adriático (GSA17 a 19) de las cuencas vecinas al oeste (GSA1 a 11) y al este (GSA22).**

Para las **gambas rosadas de aguas profundas** (*Parapenaeus longirostris*), se retuvo un total de 1.225 SNP para 782 individuos. Los métodos de agrupamiento genético y los análisis AMOVA indicaron la existencia de diferenciación y apoyo para **tres escenarios de grupos: uno "occidental-central" que incluye muestras del Mediterráneo occidental y central hasta el Estrecho de Sicilia, uno "central" que incluye las muestras restantes del Mediterráneo central, excepto la muestra jónica más oriental, y un grupo "oriental" que incluía las muestras del Mediterráneo oriental y el Jónico restante.**

En general, el muestreo MED_UNITS fue satisfactorio para la cobertura espacial y el número de especímenes recolectados, con un total de 10.670 especímenes para las seis especies. En general, el ADN de buena calidad se obtuvo fácilmente de tejidos frescos, mientras que los tejidos congelados generaron una fracción mayor de ADN de calidad media o incluso inutilizable. El éxito del genotipo de los tejidos de "buena calidad de ADN" fue alto (82%) mientras que solo el 61% de los AGN de "calidad media" incluidos en las bibliotecas terminaron produciendo datos de genotipos. Por lo tanto, podemos concluir que el muestreo debe diseñarse para satisfacer las necesidades de los requisitos de «genómica» (en términos de calidad de las muestras, procedimientos, almacenamiento, tiempo). En el marco de la tarea 1.6 se destacó que **los estudios científicos en el mar (como MEDITS u otros estudios similares) son una buena oportunidad** para implementar un diseño de muestreo adecuado para los análisis genómicos. **Sin embargo, el muestreo para los análisis genómicos no puede ser una actividad colateral, y se deben asignar recursos adecuados para esta actividad.** El plazo para la recolección de tejidos debe definirse cuidadosamente, y el momento para las fases experimentales no puede ser demasiado estricto. **La recolección de tejidos debe realizarse preferentemente de individuos vivos / recién capturados** y el muestreo debe realizarse lo antes posible. **Un protocolo recogiendo estos temas fue entregado como resultado del proyecto.**

WP2 - Los análisis de forma y microquímica de otolitos se realizó una **revisión previa de la literatura** examinando alrededor de 600 artículos para obtener una visión general sobre las potenciales aplicaciones y metodologías. **Se encontraron diez estudios que determinaron las poblaciones de *Merluccius spp.* en diferentes áreas.** La población mediterránea de *M. merluccius* se evaluó en un estudio que combinó un enfoque múltiple (químico y de forma). Con referencia a *M. barbatus*, **se encontraron cinco estudios y sólo uno se realizó en el mar Mediterráneo.** Para ambas especies, los protocolos son comunes para el proceso de adquisición de imágenes, la extracción de la información del contorno externo y el análisis de la forma del otolito, así como para la preparación, el procesamiento y el análisis de otolitos para la detección de oligoelementos microquímicos. Para cada especie, la primera madurez sexual y el sexo podrían ser fuentes potenciales de variación. En consecuencia, solo una etapa de la vida y un sexo fueron utilizados por cada especie.

Para el **análisis de la forma de los otolitos**, los conjuntos de datos utilizados (otolitos izquierdo y derecho) estaban compuestos por 1845 otolitos de salmonete de 37 subunidades geográficas y por 1868 otolitos de merluza europea de 39 subunidades geográficas. Además de las subunidades geográficas en el Mediterráneo, también se recogieron muestras para la merluza de 4 zonas ICES con el objeto de mostrar mayor discriminación entre grupos. Para ambas especies, se realizaron dos análisis complementarios utilizando el Análisis Discriminante Lineal (LDA) con predicción Jackknifed (Aprendizaje Automático Supervisado) y la agrupación jerárquica (Aprendizaje Automático No Supervisado). **Para la merluza europea, la clasificación optimizada**

(tasa de clasificación correcta = 39,61%) **presentó 4 grupos**, que se distribuyen de la siguiente manera: **Océano Atlántico** (de CIME IV a CIME VIII), **Mar Mediterráneo Occidental** (de GSA1 a GSA13), **Mar Adriático con Mar Mediterráneo Central** (de GSA16 a GSA20) y **Mar Mediterráneo Oriental** (de GSA22 a GSA27). Para el **salmonete**, la clasificación optimizada (tasa de clasificación correcta = 37,56%) **presentó tres grupos**, que se distribuyen de la siguiente manera: **Mar Mediterráneo Occidental** (de GSA1 a GSA16), **Mar Adriático con Mar Mediterráneo Central** (de GSA17 a GSA20) y **Mar Mediterráneo Oriental** (de GSA22 a GSA27).

El análisis de la **microquímica de otolitos se basó en un subconjunto de las muestras de otolitos utilizadas para el análisis de la forma de los otolitos**: 279 otolitos para la merluza europea de 10 subunidades mediterráneas, más dos áreas adicionales en el Atlántico NE agregadas para la comparación, y 250 otolitos de 10 GSA mediterráneas diferentes para salmonetes. Para la merluza europea, se obtuvieron 25 isótopos correspondientes al núcleo y al borde del otolito para cada individuo y 19 isótopos para el salmonete. **A pesar de que las GSA diferían en la microquímica de otolitos, la pertenencia a cada GSA se ha predicho correctamente para el 30% de los individuos de merluza europea** cuando se utilizan datos de bordes de otolitos. **El porcentaje de clasificación correcta aumentó al 63% cuando se utilizaron las áreas del Mediterráneo occidental, central y oriental. Para el salmonete rojo la asignación correcta fue del 63% para los bordes y del 66% para los núcleos**, aunque con una alta variabilidad individual que disminuyó la potencia de clasificación. En general, estos resultados pueden explicarse por al menos tres hipótesis compatibles: (1) la microquímica de otolitos solo puede reflejar las características de la masa de agua a otra escala espacial; (2) los límites de las poblaciones biológicas pueden incluir varias unidades de gestión; y (3) los procesos alternativos relacionados con la tasa de crecimiento también pueden estar afectando la composición microquímica y enmascarando el vínculo entre las características de la masa de agua y la composición del otolito.

Los resultados de los análisis de forma de otolitos y su composición química se combinaron bajo la hipótesis de que esto mejora la capacidad de predecir la pertenencia a GSA de un pez dado a partir de sus características del otolito. Por lo que se refiere a la merluza europea, se utilizaron datos tanto para la forma de otolito como para la química de 159 peces de 10 subunidades GSA, mientras que para salmonete se utilizaron 237 peces de 10 subunidades GSA. Las predicciones correctas validadas cruzadamente de la pertenencia a la población inferida del Análisis Discriminante Lineal (LDA) aumentaron al 42,2% después de fusionar los datos de forma y química para la merluza europea (con solo datos de forma alcanzó el 34,6%), mientras que para el salmonete fue del 44,7% después de fusionar datos de forma y química, un poco menos que usar solo la composición química del borde otolito (47,2%). Por lo tanto, **la combinación de las dos fuentes de datos implicó una ligera mejora de la precisión de la merluza, pero una ligera disminución para el salmonete. Todos los ensayos mostraron un éxito moderado y es necesario aumentar la cobertura espacial y el número total de individuos para mejorar la identificación de la población utilizando la forma de otolito y la composición química. El análisis de la forma del otolito solo se realizó en muestras mucho más grandes para ambas especies. El tiempo y el costo del análisis de microquímica de otolitos impidieron la misma cobertura de muestreo.**

WP3 - Delinear los caladeros y la evaluación de las poblaciones tiene como objetivo identificar y caracterizar los caladeros en aguas del Mediterráneo, incluidas, cuando sea posible, las flotas no pertenecientes a la UE (que es un aspecto de gran importancia), incluso combinando diferentes fuentes de datos (por ejemplo, AIS, VMS, datos experimentales de encuestas, datos estadísticos de pesca), enfoques de procesamiento y modelización y dedicar parte del trabajo a la comparación cuantitativa de resultados. Las principales medidas fueron: i) la estimación de los caladeros potenciales mediante el procesamiento de datos AIS y el análisis de puntos calientes conexo; ii) Análisis de decisión multi criterio (MCDA) aplicado a datos medioambientales

y de flota para componentes grandes y pequeños de las flotas y iii) Cascaded Multilayer Perceptron Network (CMPN) aplicado sobre datos ambientales y de flota para el componente a gran escala de las flotas. Los resultados producidos enfatizan la importancia de aplicar y comparar diferentes métodos. **El conjunto de datos de esfuerzo se reconstruyó procesando datos (AIS) y se complementó utilizando datos de esfuerzo estimados (modelados) en la cuenca sur / este, donde el AIS aún está mal adoptado y recibido.** Los resultados se validaron con los proporcionados libremente por otros proyectos (EMODNET, Global Fishing Watch) y, en algunas áreas, comparando el patrón obtenido utilizando solo datos AIS con el obtenido integrando datos AIS y VMS (Proyecto H2020 MINOUW). Hay que reconocer que **este "nuevo" conjunto de datos de pesca cubre por primera vez todo el conjunto altamente productivo de la cuenca mediterránea y puede informar a la comunidad científica en general, así como a aquellos involucrados en la política y la gestión,** sobre las huellas de pesca en el ecosistema Mediterráneo.

Además, **el WP3 delineó los caladeros por especie o grupo de especies.** Un marco metodológico, basado en técnicas analíticas espaciales, para combinar el esfuerzo pesquero (para redes de arrastre de fondo, palangres, redes de enmalle y trasmallo) y se aplicó la distribución espacial de la biomasa para las especies investigadas. Los principales pasos seguidos fueron: i) estimación de los caladeros potenciales por especies; ii) el análisis de puntos calientes y iii) la producción de puntos calientes y fríos agregados. **Los resultados producidos demuestran la posibilidad de que los pescadores cosechen una determinada especie en un área específica, al tiempo que proporcionan información espacial sobre el número de explotadas especies en varios caladeros. Los principales resultados revelaron importantes caladeros en todo el mar Mediterráneo,** que se encuentran en ciertas zonas de los mares Adriático, Tirreno, Estrecho de Sicilia, Egeo, Jónico oriental, Baleares, Alborán, Libia y Levantina.

En el WP3 **se llevaron a cabo trece nuevas evaluaciones de poblaciones** y se compararon, cuando fue posible, con las evaluaciones realizadas rutinariamente a nivel de GSA única o combinación de GSA por la CGPM o el STECF. Las nuevas configuraciones de stock para las seis especies objetivo se basan en los resultados de WP1 y WP2. Los resultados de los análisis muestran que, en la mayoría de los casos, las nuevas evaluaciones de las poblaciones no presentan mejoras particulares, en términos de diagnóstico y precisión del modelo. La falta de mejora puede deberse a varias razones, aparte de las nuevas configuraciones de stock (aumento de la heterogeneidad de los datos cuando se aumenta el número de GSA agregadas, configuración del modelo, etc.). No obstante, debe reconocerse que **estos ensayos representan un primer y prometedor enfoque para la evaluación de las nuevas configuraciones de las poblaciones y que se llevarán a cabo nuevas investigaciones** antes de que puedan proporcionarse asesoramientos científicos de forma fiable y sólida.

WP4 - Síntesis y propuestas tiene como objetivo sintetizar y combinar los resultados obtenidos en los WP 1-3. **Un análisis DAFO proporcionó una evaluación de los factores internos (Fortalezas/Debilidades) y externos (Oportunidades/Amenazas) relacionados con los métodos aplicados en los WP 1-3.** Los resultados enfatizaron que los métodos genómicos pueden proporcionar una imagen precisa de la diferenciación genética, pero si el procesamiento de la muestra no se sigue adecuadamente, la calidad de los resultados puede reducirse considerablemente. Por lo tanto, se necesitan muestreos dedicados. El tiempo para lograr todo el proceso es bastante largo, aproximadamente 1 año. La transferencia de muestras genéticas puede plantear algunas dificultades, debido a las diferencias en los aspectos jurídicos/burocráticos entre los países. La forma del otolito no es muy exigente en términos de equipos y costos, mientras que la microquímica es costosa. Se pueden obtener muestras de otolitos a partir de campañas de recopilación de datos existentes, pero los análisis de forma de otolitos y microquímica son aplicables solo a los peces. La identificación de los caladeros ha progresado rápidamente y la metodología se ha consolidado. El acceso a las series temporales de AIS necesita un presupuesto dedicado

y el acceso a los datos de VMS no es fácil. **El análisis DAFO destacó la presencia de ventajas en cada método, mientras que al mismo tiempo surgieron limitaciones para cada enfoque, señalando la necesidad de un proceso de integración continua con un intercambio de conocimientos y logros entre los grupos de investigación. Una recopilación de datos para la identificación de stock requiere el fortalecimiento de la cooperación y la necesidad de inversiones en el desarrollo de capacidades. Sería útil elaborar un plan regional de muestreo en cooperación entre el Grupo de Coordinación Regional del DCF y la CGPM.**

El papel explicativo de las variables ambientales a escala espacial en la delineación de las estructuras poblacionales que surgieron de los análisis genéticos y otolíticos se exploró utilizando datos ambientales del Servicio de Monitoreo del Medio Marino de Copernicus (CMEMS, <http://marine.copernicus.eu>). Se consideró el origen espacial de cada individuo para reconstruir la estructura espacial de las poblaciones e investigar la relación de estas estructuras de poblaciones con respecto a los principales impulsores ambientales. En un primer paso se utilizó el agrupamiento difuso. Surgieron diferentes números de unidades de gestión potenciales. En el caso de la merluza europea, se identificaron tres poblaciones sobre la combinación de los datos genéticos y de forma de otolito, una en el Mediterráneo occidental, una en las cuencas Adriático-Jónica-Tirrena y una en el Mediterráneo oriental. En el caso de la cigala, se identificaron cuatro poblaciones; una caracterizada en el mar Adriático y las subáreas del mar Jónico y del mar Egeo, mientras que los otros tres se distribuían en el Mediterráneo occidental. Para la gamba rosa de aguas profundas se identificaron dos poblaciones y la discontinuidad entre las poblaciones oriental y occidental se posicionó en el Mar Jónico al sur de Italia. Para las gambas rojas, ninguna agrupación mostró una distribución espacial reconocible. Para el salmonete la conclusión es similar, a pesar de que hubo algunas evidencias bastante inconclusas de la existencia de 2 a 3 sub-poblaciones en el Mediterráneo. Un diseño de muestreo más denso, tal vez restringido a áreas más pequeñas, podría ayudar a confirmar o rechazar esta hipótesis. Se seleccionaron las configuraciones de poblaciones más fiables y completas por especie y se prepararon una serie de mapas interrelacionados para mostrar la correspondencia espacial entre poblaciones, áreas de gestión y caladeros.

Probamos el desarrollo potencial de una gestión espacial adaptativa de la pesca a través de un enfoque de simulación utilizando tres estudios de caso. Las nuevas configuraciones de poblaciones de gamba roja gigante y gamba azul y roja se examinaron en el Mediterráneo occidental, gamba roja gigante en el Mediterráneo central y gamba rosa de aguas profundas en la región del Adriático-Jónico. Estas especies están sujetas a Planes de Manejo Plurianuales. **Los escenarios de gestión se ejecutaron utilizando modelos de simulación bio-económicos (BEMTOOL y SMART)**, considerando la reducción del esfuerzo pesquero, la mejora de la selectividad de los artes y los cierres espaciales en áreas críticas para los ciclos biológicos de las especies objetivo (Hábitats Esenciales de Peces). A pesar de que BEMTOOL y SMART evaluaron las mismas pesquerías en los mismos escenarios, fueron bastante diferentes en términos de resultados. Estas diferencias se deben principalmente a las diferencias en los procesos y supuestos modelados. Sin embargo, **ambos modelos evaluaron como la mejor estrategia de gestión**, tanto en términos de ganancia en la biomasa del stock reproductor (SSB) como de mejora o disminución ligera del rendimiento actual, **el escenario caracterizado por la reducción del 10% del esfuerzo pesquero actual junto con una mejora de la selectividad de la red de arrastre y la protección de los Hábitats esenciales para Peces (EFHs)**. Finalmente, **se exploró una hoja de ruta "ideal" para desarrollar la gestión espacial adaptativa de la pesca destinada a reducir la incertidumbre en los procedimientos de evaluación y gestión**, considerando el desarrollo de modelos operativos, la simulación de estrategias de manejo alternativas y la evaluación de su desempeño (frente a objetivos biológicos, económicos y sociales). Se destacó la necesidad de un conocimiento preciso de las unidades de población y su conectividad. Además, **una**

gestión adaptativa destinada a proteger la estructura espacial de las poblaciones debe considerar como herramienta principal la identificación y protección de viveros y zonas de desove, con el fin de mitigar tanto la sobrepesca de crecimiento como la sobrepesca de reclutamiento.

1. INTRODUCTION

The Common Fisheries Policy (CFP) aims to ensure that fishing activity is environmentally, economically and socially sustainable in the long-term. Fisheries management can achieve sustainability only when the necessary information is available for each fish and shellfish stock in terms of stock size, geographical distribution, stock identity and structure, fishing pressure, catches and biology (e.g., growth and reproduction).

Knowledge of ecological stock structure of a species is considered essential for effective management. The reason for such importance is related to the fact that most commonly used stock assessment methods rely on the assumption that stocks are discrete and self-recruiting units with homogeneous life history parameters (Begg et al., 1999; Begg and Waldman, 1999). Fisheries management provides single species advice for individual stock units (Kerr et al., 2016). It is assumed that stocks are discrete units with homogeneous vital rates that can be exploited independently of each other and that catches can be assigned to their stock of origin (Cadrin et al., 2013). Violation of the unit stock assumption (i.e., misclassification of the appropriate spatial scale of management) could introduce significant problems affecting stock assessment and fisheries management that can impact sustainability of the resource, profitability of the fishery, resilience of fishing communities, and impede conservation or biodiversity goals (Cadrin et al., 2020).

Traditionally, exploited stocks have been assessed and managed in the Mediterranean Sea according to geographical and political features or regions, for instance GFCM subdivisions (GSAs). There may also be, depending on the geographic location, legal, cultural, and social pressures that prevent revision of stock boundaries or adding complexity to stock assessments. As more information becomes available, stock identification methods reveal inconsistencies between the spatial structure of biological population and the definition of stock units used in assessment and management. The mismatch between the scale of a biological population and that of a management unit or 'stock' can bias stock assessment and hamper the development and implementation of sustainable fisheries management (Cope & Punt, 2009). For example, short-term recommendations (e.g., total allowable catch, TAC, or effort reduction regimes) and long-term strategies (e.g., rebuilding targets and harvest control rules) produced from the stock assessment may be based on an erroneous perception of stock structure and do not account for differentiation in productivity among population components. Despite increased recognition of complex population structure and stock mixing, the disparities between population structure and current management units have not been reconciled (Reiss et al., 2009).

Stock identification provides a basis for understanding population dynamics, makes more robust the stock assessment process and thereby developing fisheries management strategies. Multiannual Management Plans under the CFP (EU reg. 1380/2013) are one of the tools for managing shared stocks in the long term and thereby this requires improving our knowledge on biological stock units and fishery management units. Methods for delineating stocks have advanced considerably in recent years and include genetic techniques, acoustic telemetry, tagging, otolith chemistry, demographic analysis, otolith shape and meristic data. The integration of multiple techniques that operate over different temporal and spatial scales makes it possible to overcome many of the limitations of single technique approaches and greatly strengthens the inference available from stock structure studies.

Under the “Marea” Framework, DG MARE funded in 2015 the project entitled: “Identification of distinct biological units (stock units) for different fish and shellfish

species and among different GFCM-GSA – STOCKMED”¹. The project aimed at identifying stock units and related boundaries for a group of demersal and small pelagic species, which are considered important fishery resources in the Mediterranean Sea. STOCKMED used only already available information from literature, and attempted stock identifications without collecting new samples. This was viewed as a major limitation of the study alongside the non-conclusive results obtained. Filling these gaps may contribute to reduce the current uncertainty in stock unit identification across the Mediterranean, even for the priority species in the basin, and in turn may improve the assessment and management of the resources.

In this regard, a relevant role was played by the FP7 EU project FishPopTrace² that used deoxyribonucleic acid (DNA) single-nucleotide polymorphism (SNP) analysis (single genetic variants) and otolith microchemistry and morphometrics (shape) to test their power for tracing fish from four species (including hake in the Mediterranean Sea) back to population/area of origin (https://cordis.europa.eu/result/rcn/56107_en.html).

Several other studies have investigated **European hake** (*Merluccius merluccius*) population genetics within the Mediterranean Sea using different methods (i.e., allozymes, microsatellites and SNPs -Single nucleotide polymorphism). However, these studies were based on different numbers and different markers, and different sampling efforts, leading sometimes to contrasting results and difficulty to simply merge and compare the outputs. In general, the use of allozymes and microsatellites (Castillo et al., 2004; Cimmaruta et al., 2005) allowed the identification of significant genetic differentiation among samples, suggesting a subdivision of Atlantic and Mediterranean hake stocks. Milano et al. 2014, using SNPs (395 EST-derived) with significantly higher resolving power, confirmed the genetic break between Atlantic and Mediterranean populations and described a finer-scale significant genetic population structure. In particular, in the Mediterranean outlier SNPs revealed a strong differentiation among Western, Central and Eastern Mediterranean geographical samples.

Studies that have investigated **red mullet** (*Mullus barbatus*) population genetics within the Mediterranean Sea using the same method (i.e., nuclear microsatellite markers) were based on different numbers, markers and sampling efforts resulting in sometimes contrasting results making the merging and comparability of the outputs problematic (Maggio et al., 2009; Galarza et al., 2009; Felix-Hackradt et al., 2013; Matic-Skoko et al., 2018).

Only two studies have also investigated **deep-water rose shrimp** (*Parapenaeus longirostris*) population genetics within the Mediterranean Sea using mitochondrial DNA markers and AFLP (Garcia-Merchan et al., 2012; Lo Brutto et al., 2013). The first one analysed the variability of the mtDNA gene COI along the Spanish coasts, and found a substantial homogeneity among locations while the second one using mtDNA (CR) and AFLP, identified four clusters significantly differentiated: Tyrrhenian vs. Strait of Sicily vs. Adriatic vs. Aegean Sea). The greatest contribution to the differences among the four Mediterranean sub basins depended on Aegean and Tyrrhenian areas, which represented the most divergent groups. However, the use of different markers and different sampling efforts (complementary areas were investigated), makes impossible to merge the outputs.

¹Fiorentino F., E. Massuti, F. Tinti, S. Somarakis, G. Garofalo, T. Russo, M.T. Facchini, P. Carbonara, K. Kaporis, P. Tugores, R. Cannas, C. Tsigenopoulos, B. Patti, F. Colloca, M. Sbrana, R. Mifsud, V. Valavanis, and M.T. Spedicato, 2015. Stock units: Identification of distinct biological units (stock units) for different fish and shellfish species and among different GFCM-GSA. STOCKMED Deliverable 03: FINAL REPORT. January 2015, 310 p. https://www.coispa.it/index.php?option=com_content&view=article&id=4&Itemid=107&lang=it

²Carvalho G., E. Mac Aoidh and J. Martinsohn (eds.) 2011. Traceability of Fish Populations and Fish Products: Advances and Contribution to Sustainable Fisheries <https://fishpoptrace.jrc.ec.europa.eu/>

Only five studies have investigated **giant red shrimp** (*Aristaeomorpha foliacea*) population genetics within the Mediterranean Sea (e.g., Cannas et al., 2012a) using different methods (i.e., ISSR, microsatellites, mitochondrial and nuclear genes). Two based on microsatellites data (Cannas et al 2012a; Marcias et al., 2010) did not allow identifying significant differentiation among samples from Sardinia and Sicily, suggesting the lack of genetic differentiation among these two areas, a pattern identified within the Mediterranean using ISSR and nuclear genes (Fernandez et al 2011a; Fernandez et al 2013b). Using mitochondrial DNA (COI gene sequences), Fernandez et al (2013a) identified the occurrence of genetic differences within the Mediterranean, suggesting that a certain degree of genetic differentiation was present among local samples, although when these Mediterranean samples were grouped in western and eastern basins, the variation among groups was not statistically significant. Different sampling efforts and markers used in these studies produced sometimes contrasting results that were not easy to be merged and compared.

Several studies have investigated **blue and red shrimp** (*Aristeus antennatus*) population genetics within the Mediterranean Sea using different methods (i.e., allozymes, AFLP, microsatellites, mitochondrial DNA). The vast majority of studies using all kind of markers did not allow identifying significant differentiation among samples, suggesting a substantial genetic homogeneity within the Mediterranean (e.g., Cannas, 2012b; Heras et al., 2016). In contrast, Fernandez et al. 2011b, using mitochondrial DNA, indicated the occurrence of genetic differentiation among geographical regions.

A few studies have investigated **Norway lobster** (*Nephrops norvegicus*) population genetics within the Mediterranean Sea using different methods (i.e., allozymes, mtDNA RFLP, mtDNA CR) and even in this case results were contrasting.

The advent of Next Generation Sequencing (NGS) technologies has fundamentally changed the way in which genetic sequence data are generated (Hemmer-Hansen et al., 2014). NGS has not only enabled massive increases in number of sequences attained per effort but also facilitated rapid and cost-effective genetic marker discovery and high-throughput genotyping that can be used for population genetics studies (Davey et al., 2011). This new approach is known as ‘Genotyping By Sequencing’ (GBS) and its primary advantage for population genetic studies is the generation of increased quantities of data with improved statistical power and higher genome representation (Narum et al., 2013).

Otolith structural analysis is used in stock identification research worldwide. Otoliths are calcified structures located in the inner ear cavity of all teleost fish. They are isolated within a semi-permeable membrane and bathed in an endolymphatic fluid. These structures serve as a balance organ and also aid in hearing (Campana, 1999; Campana and Thorrold, 2001). Inner ears (left and right), contain three different pairs of otic sacs (sacculus, utriculus and lagena), each containing a different otolith (sagitta, lapillus and asteriscus) (Panfili et al., 2002). Fish otoliths are metabolically inert aragonite structures with a composition that is influenced by the physical and chemical properties of the environment (Elsdon et al., 2008). Furthermore, otoliths grow continuously and record information on life histories in a chronological manner, making it possible to retrieve information on environmental conditions experienced by individual fish from hatching to capture (Campana, 1999). Otoliths show incremental structures with daily and seasonal periodicity over time as they grow throughout the life of the fish and, unlike scales and bones, are metabolically inert (i.e., once deposited, otolith material is unlikely to be resorbed or altered; Casselman, 1987).

Consequently, otoliths have been primarily used as tool for age determination in many fish species thanks to the ability to track growth periodicity, from the daily to annual growth increments. Moreover, otolith shape remains unaffected by the short-term changes in fish condition (Campana and Casselman, 1993) or environmental variations (Campana, 1999). Accordingly, the shape of the otolith has been used as a tool to identify the species, to reconstruct the composition of the diet of predators (fish, seabird, seal etc.) and to discriminate fish stocks. Since Campana and Casselman

(1993), many fishery scientists have developed this type of analysis for stock discrimination studies, as a base for understanding fish population dynamics and achieving reliable assessments for fishery management (Reiss et al., 2009).

Variations in otolith shape of fish from different geographic areas are at least partly expressed during the life history, thus representing a phenotypic measure of stock identification. However, the spatial scales of fisheries differ from the degree of resolution reached by otolith shape studies. For example, Morat et al. (2012) highlighted that several local and close populations can be distinguished within the stock of *M. barbatus* in the Gulf of Lions (westwards from the Rhône River mouth) that was instead identified as a unique entity by Gaertner et al. (1997).

Otoliths are composed by aragonite laid down in an organic matrix and incorporate trace elements along their growth depending of fish biology and water mass composition (Panfili et al. 2002). Therefore, otoliths are considered as natural tags, their chemistry has been used successfully to address issues related to stock identity and fish movements (i.e., Campana et al. 2000). For instance, *Merluccius merluccius* otolith composition has been studied both in the Atlantic and in the Western-Central Mediterranean (Morales-Nin et al., 2005; Swan et al., 2006; Morales-Nin et al., 2014; Tanner et al., 2012). Therefore, some relevant information is available although not the same suite of isotopes has been analysed in all studies. In contrast, no previous analyses have been performed in red mullet.

STECF 17-03 noted that there is a full body of literature dealing with stock identification and reported that state of the art methodologies includes a combination of methods and data sources (including genetic, tagging, otoliths shape and microstructures, chemical markers, surveys, drift modelling, etc.). Combining genetic characterization with otolith shape and microchemistry can corroborate the outcomes from the latter and may take stock mixing into account, providing a more robust stock discrimination, which is particularly useful in designing fishery management strategies based on stock productivity (Hüssy et al., 2016; ICES, 2020).

The identification of the fishing grounds is an essential information to delineate the fishing footprints on the fish and shellfish stocks. To identify fishing grounds, different methods are available. Despite the promising methodological developments and increasing number of applications and tools, the extensive use of VMS data for scientific purposes is hindered by the difficulty of accessing control data for scientific purposes (Natale et al., 2015). An alternative to the VMS is represented by the Automatic Identification System (AIS), which in the EU became compulsory in May 2014 for all fishing vessels of length above 15 meters. AIS data is not related to control purposes and is exchanged also in public domains that expands its availability in respect of the VMS data and offers a very useful opportunity to analyse fishing behaviour at very detailed scale (e.g., James, 2018) and to extending the analysis at supra national level.

Hence, linking information on stock boundaries with the one on the use of fishing grounds is a fundamental step to identify spatial units for fishery management.

1.1 OBJECTIVES OF THE PROJECT AND PROJECT SCOPE

The ultimate objective of the study is to identify and match the biological with the management stock units of several of the most important demersal species in the Mediterranean, included also among the GFCM priority species. The project has the following specific objectives:

1. Identify biogeographical boundaries and population structure of priority species.

2. Investigate genetic distribution among and within sub-basins (e.g., higher genetic structure versus more homogenous pattern; panmictic vs geographical partitioning; isolation vs population connectivity with other adjacent Mediterranean regions).
3. Verify fish stock boundaries, identify and delineate stocks using the ‘state of the art’ approach in stock identification i.e., the application of multiple approaches, to the same biological samples, with comparison of results to achieve an interdisciplinary perspective and consensus (Cadrin et al., 2013).
4. Combine the information on fish stock boundaries and delimitation of spatial units for fisheries management.
5. Identify and delimit the more important fishing grounds in the Mediterranean with associated main origin of the operating fleet in order to define spatial units for fisheries management purposes in the Mediterranean.
6. Perform stock assessments based on the updated list of stock units for the examined priority species as emanating from the present study and compare with previous assessments.

Target species of the study are: **European hake** (*Merluccius merluccius*), **red mullet** (*Mullus barbatus*), **deep water rose shrimp** (*Parapenaeus longirostris*), **giant red shrimp** (*Aristaeomorpha foliacea*), **blue and red shrimp** (*Aristeus antennatus*) and **Norway lobster** (*Nephrops norvegicus*).

The study covers the Geographical Sub-Areas (GSAs) 1-27 (Mediterranean Sea).

2. PROJECT STRUCTURE

The project is organised into 5 inter-correlated Work-Packages (WPs) and 16 Tasks.

WP0: Project management and coordination.

WP1 Population genetics and phylogeographic studies for identification of biological units of priority species

- Task 1.1. Literature review
- Task 1.2. Review bioinformatic methods and tools
- Task 1.3. Genetic sampling design
- Task 1.4. Pilot genetic analyses
- Task 1.5. Comprehensive genetic studies at Mediterranean Sea scale
- Task 1.6. Deliver a consistent and detailed protocol based on genetic analysis for routine sampling

WP2 Otolith shape and microchemistry analyses

- Task 2.1. Literature review on otolith shape and microchemistry
- Task 2.2. Otolith shape analysis of European hake and red mullet
- Task 2.3. European hake and red mullet microchemistry analyses
- Task 2.4. Compilation of matrices and data interpretation

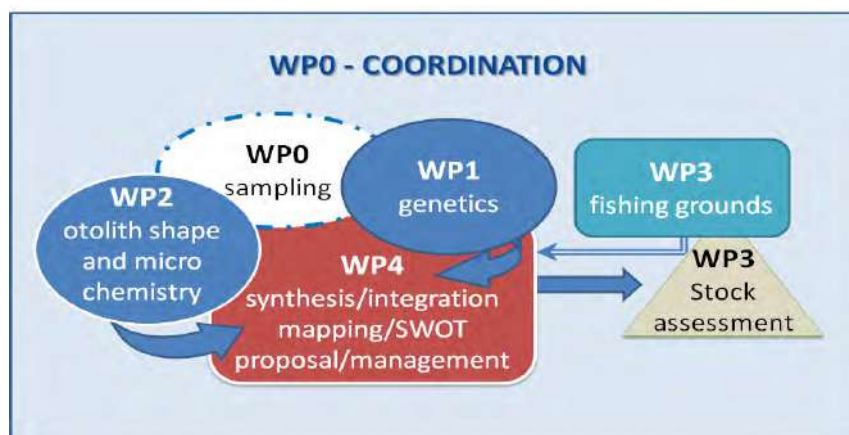
WP3 Delineate fishing grounds and stock assessment

- Task 3.1. Fishing grounds delineation
- Task 3.2. Combine the results on spatial identification of fishing grounds with the spatial distribution of the target species
- Task 3.3. Perform stock assessment
 - Sub-task 3.3.1. European hake case study
 - Sub-task 3.3.2. Red mullet case study
 - Sub-task 3.3.3. Norway lobster case study
 - Sub-task 3.3.4. Deep water rose shrimp case study
 - Sub-task 3.3.5. Giant red shrimp case study
 - Sub-task 3.3.6. Blue and red shrimp case study

WP4 Synthesis and proposals

- Task 4.1 Perform a SWOT analysis
- Task 4.2 Integrating results by the different WPs and proposals of new management units
- Task 4.3 Future improvements for developing adaptive spatial fisheries management

The following diagram summarizes the project structure



In agreement with the project objectives, the analyses included in WP1 and WP2 are delineating the population units from a biological perspective, taking also into account the environmental signatures for fish, by investigating otolith shape and microchemistry. The analyses foreseen in WP3 delineate the fisheries footprints not necessarily confined within the boundaries of the current GSAs. The integration of these two key sources of information took place in WP4, as far as possible supported by the explanatory role of ecological/environmental profiles at spatial scale. This approach is expected to advance fisheries assessment and improve the management advice, reducing the bias associated with the assumption of a given stock unit, when instead multiple stocks are assessed as a single unit or only a portion of a stock is assessed as a closed unit.

Under the coordination of DG MARE, the project relied upon the liaisons with the FAO Regional Projects and GFCM, the former for a successful gathering of biological samples, the latter to gather advice on the alignment of biological and management units of priority species, especially as regards non-European waters. In this context, the project welcomes the participation of researchers from non-EU Countries, in the framework of the collaboration with GFCM and FAO Regional Projects, to the sampling activities, analyses and project meetings.

Partners, timetable, **milestones, deliverables and core teams are reported in the Annex 1** to this report, while **critical steps, possible risks and solutions in the Annex 2**. It should be considered that the workplan of the project and consequently the activities of the various WPs have had a delay caused by the impact of the COVID_19 pandemic. Thus, the Annex 1 is reporting the updated Timetable of the Milestones and Deliverables following the **project extension granted by EASME and DG MARE (Ref. Ares (2020)7421633 – 08.12.2020) and subsequently by CINEA (Ref. Ares (2021)4772243 – 26.07.2021)** for the delivery of the Draft Final Report and the Final Report.

3. METHODOLOGICAL APPROACH AND BREAKDOWN IN WORKING PACKAGES AND TASKS WITH RELEVANT RESULTS AND CONCLUSIONS

Each WP/Task had one/two persons in charge as chair/task leader and a core team, to carrying out the majority of the work. The work plan was then defined with a set of milestones and deliverables. Additional experts within each involved institute also actively supported the work performed under each Task/WP of the project.

Most of the work comprised collection of biological samplings, i.e., genetic tissues and otoliths, genetic and otolith analyses, desk-based studies and meetings (both face to face, especially with EASME/MARE, and web based).

Parts of the project work was expected to be conducted at dedicated meetings/practical sessions (e.g., in WP3 stock assessment), that owing to the COVID-19 pandemic impact were carried out virtually.

This Final Report includes 23 electronic Annexes both Deliverables and few relevant Milestones, as well as protocols. In addition to these electronic annexes, the **ANNEX 1A reports the references**.

3.1 WPO - PROJECT MANAGEMENT AND COORDINATION

Coordinator: Maria Teresa Spedicato (COISPA)

Partners involved: all partners

The aim of this WP is to ensure a smooth running of all the project activities and the successful accomplishment of the project goals. The activities developed in WPO are:

- 1) Co-ordination of the WPs activities in collaboration with WP and Task leaders ensuring connectivity among them;
- 2) Preparation and submission of the required reports;
- 3) Preparation and participation to the meetings with formerly EASME, currently CINEA/DGMARE;
- 4) Co-ordination of the project Plenary meetings and other steering meeting with the project partners;
- 5) Sampling activities coordination.

Considering the need of exchanging data and files, firstly a repository on OneDrive was created by the project coordinator and then, considering the limitations to travelling and the need of organizing meetings and workshops, a Teams platform was set up to virtually supporting all these activities, from file upload to the meetings. On this platform the project documents are stored.

3.1.1 Reporting activities

The **Inception Report (InR)** was first delivered on February 1, then revised following the discussion in the KoM and finally approved by EASME/MARE on March 27, 2019.

The **First progress Report (FpR)** was delivered on June 2019 and approved by EASME/MARE on August 23, 2019.

The **Interim Report (IR)** was delivered on December 30, 2019, discussed during the Interim Meeting with the EASME and MARE on January 14, 2020 and approved by EASME/MARE on January 31, 2020.

The **Second Progress Report (SpR)** was delivered on June 26, 2020 and approved on August 2020.

3.1.2 Meetings with EASME/MARE

The **kick-off meeting (KoM)** was held on 12/02/2019 at EASME premises to discuss the **Inception Report**. The meeting was attended by the coordinators of the MED_UNITS Work Packages and by the project coordinator. EASME and DGMARE officers attended the meeting and a virtual connection was established with JRC scientists.

The **first progress meeting (FpM)** was held on 02/07/2019 at EASME premises to discuss the 1st Progress Report sent to EASME on 7th June 2019.

The **Interim Meeting (IM)** was held on 14.01.2020 at EASME premises to discuss the Interim Report.

The **second progress meeting (SpM)** was held virtually on August 2, 2020

3.1.3 Meetings with the project partners

Virtual meetings have been organised with the WP chairs and also bilaterally within WPs, for the preparation of the Inception Report and the implementation of the project activities linked to the different WPs, milestones and deliverables.

A **1st Project Plenary Meeting (1PPM)** was held on **March 28, 2019**, just after the approval of the Inception report.

The meeting focused in particular on the preparation of the sampling activities, including the sampling protocols (genetic and otolith samplings), the first results from the review process and the allocation of the samples (genetic and otolith) to the GSAs, considering both the pilot and the full genetic study foreseen in the project. This because the MEDITS survey, one of the main potential sources of samples, takes place between May and July and thus it would be useful to take advantage of this survey.

In the intersession phase to the 2nd Project Plenary Meeting, several **virtual meetings** have been held among the project partners and within each WP. These meetings discussed the work organization, especially focusing on the sampling activities, the problems arising, and identifying possible solutions.

The **2nd Project Plenary Meeting (2PPM)** in the week 11-15 November 2019, back-to-back with the 1st Workshop of WP1, discussed the progress of the project, the issues that may have arisen, highlighting possible solutions and planning the project activities for the 2nd year. The meeting and workshop outputs were presented and discussed during the Interim Meeting (month 13, January 2020) with EASME/MARE.

Several SC meetings were held on 2020 and 2021, especially considering that due the impact of COVID-19 pandemic physical meetings were not possible. Six SC meetings were held on 13.03.20, 14.05.20, 03.12.20, 29.03.21, 28.04.21 back-to-back to the 2nd Genetic Workshop, and on 08.06.21.

The **3rd Plenary meeting** was held on June 18 2021. The meeting discussed the results achieved and the finalization of the deliverables.

Some preliminary results of the project achievements were presented to the SubRegional Committee for the Adriatic Sea (SRC-AS) on 20 – 23 April 2021 (online) and to the SubRegional Committee for the Eastern Mediterranean (SRC-EM) 25 – 27 May 2021 (online).

3.1.4 Data Call

Initially, the terms of reference of the MED_UNITS specific contract did not foresee to issuing a specific data call. However, for some of the project objectives, as those in WP3 (stock assessment), it was beneficial to gather official data of the scientific trawl survey

(MEDITS) and fishery dependent data. In addition, economic data at GSA level were needed for the implementation of scenarios in the Task 4.3.

Thus, thanks to the support of DG MARE and CINEA a data call was issued on January 2021. Consensus to access data already on the JRC database was received by all countries. Economic data at GSA level were made available by Italy and Greece. For some other countries the data of the Annual Economic Report (AER) were used, given to the unique alignment between the country and the GSA (e.g., Croatia, Malta, Slovenia, France), in case of Spain suitable proxies were used for the case studies of Task 4.3.

3.1.5 Sampling activities coordination

The samplings of biological material for genetic analyses and otoliths related to the activities of WP1 and WP2 were coordinated by Paolo Sartor (CIBM).

A Sampling Procedure was established and shared with MARE and EASME at the project start. It included the creation of a Genetic Hub (responsible Alessia Cariani, CoNISMA) to manage the exchange of samples for the genetic analyses and an Otolith Hub (responsible Pierluigi Carbonara, COISPA) for the exchange of otoliths. This sampling procedure has been shared also in the 1st Project Plenary Meeting and definitively formalised in the deliverable *D.02 Protocol of biological sampling coordination (ANNEX O.1 to this report)*.

The two hubs carried out a very intensive work to ensure the monitoring of the samplings, the transfer of the samples from the sampling areas to the hubs and from these to the different laboratories in charge of the genetic and otolith analyses. Vials and tubes were prepared and shipped to each sampling units, and cross-checked with the excel sheets of individual data once back at the hubs. A scheme is given below.



Approximately 10500 samples of tissues for genetic and 3700 otolith samples were transferred.

During the MEDITS coordination meeting held in Sète (France) on April 16-17, 2019 the MED_UNITS project, was presented highlighting the importance of the MEDITS survey to provide relevant samples for the genetic and otolith analysis. Participants expressed interest to the project and availability to cooperate.

Sampling protocols for genetic tissue and otolith extraction were distributed to the project partners and to the cooperating Institutes, in order to facilitate the sampling activities. The excel sheet to register the individual data and the Milestone 1.2 (definition of sites for pilot genetic study) were disseminated as well.

Samplings for the genetic pilot study were expected to be accomplished by month 5, May 2019, while the sampling for the genetic full study by month 12, December 2019, in order to have enough time for carrying out laboratory analyses, processing data and assessing results. The former taking advantage from the biological sampling in the frame of the Data Collection Framework and partly from the MEDITS survey, the latter taking advantage almost exclusively from the MEDITS survey, during which it was also possible to sample onboard the target species, at least in some particularly favourable logistic conditions onboard the survey vessels.

At the beginning of June 2019 many research teams had already collected samples (in GSA7, 8, 9, 11, 18, 19, 23) while in other GSAs the sampling was ongoing. During June-July most of the teams involved in MEDITS started the survey, during which sample collection both for pilot and full genetic studies was carried out.

Regarding the samplings in GSAs 1 and 6 the CSIC-ICM (Montserrat Demestre) and Girona University (Joseph Lloret) made available laboratory facilities for the processing of samples. For the GSA5, thanks to the project partner CSIC-IMEDEA, samples were collected from the commercial catches in almost all the subareas and species planned and then processed in cooperation between CSIC-IMEDEA team and a researcher from COISPA.

Regarding the eastern part of the Mediterranean, given that the MEDITS survey was carried out late, samples were not available for the pilot study. These samples were delivered by November 2019. Similarly in Cyprus, where, for the two fish species (European hake and red mullet), suitable samples were not available on time to be included in the pilot study. The samples of the two fish were collected in November and shipped in early January 2020. For Malta there were not samples available.

In particular, the following European Institutes cooperated for the collection of the biological samples: CIBM, COISPA, CSIC-IMEDEA, CNR-IRBIM Mazara del Vallo, CNR-IRBIM Ancona, CNR-IRBIM Messina, CoNISMa UNIBA, CoNISMa UNIBO, CoNISMa UNICA, DFMR Cyprus, HCMR Greece, IFREMER Sète, IMB Montenegro, IOF Split.

In some GSA subunits, such as in GSA19b, GSA18c, GSA 6c and GSA6b resampling was carried out, because a poor DNA quality was found in the first samplings.

As concerns the non-EU Countries, a liaison managed by DG MARE with GFCM, and in particular with the FAO Regional Projects (CopeMedII, MedSudMed, AdriaMed, EastMed) was established in order to support the collection of samples also from south and east Mediterranean Countries. Fishery monitoring under DCRF (Data Collection Regulation Framework, a tool regulating the collection of fishery data for all the Mediterranean countries) was considered as the platform to allow the samples' collection.

MED_UNITS project was thus presented at the FAO Sub-Regional Committee of the Eastern Mediterranean (SRC-EM). The participants expressed their interest and availability to cooperate.

Subsequently, the coordinators of the FAO Regional Projects were contacted by the MED_UNITS coordinator to introduce the sampling documentation. All the FAO Regional Projects coordinators expressed interest to cooperate with the MED_UNITS. The first feedback from the scientists of the different countries was that the otolith sampling is more feasible than the genetic one, because legal constraints in some countries would impede to allow the exit of genetic material (likely as a consequence of the application of the Nagoya Protocol on Access to Genetic Resources). In addition, not all the target species could be sampled by the scientists of the different countries, given the species distribution and the sampling feasibility. Sampling protocols for genetic tissue and otolith extraction were distributed to the responsible of FAO regional projects, in order to facilitate the sampling activities. The excel sheet to register the individual data was distributed as well.

Besides the contacts with the FAO Regional projects, liaisons were also established with laboratories in Turkey, Tunisia and Algeria by IFREMER of Boulogne sur Mer (Kélig Mahé).

Thus, samples were collected in the southern Mediterranean from Algeria, Tunisia, Egypt, Palestine, Lebanon and Turkey.

The species and biological material collected varies, for example in Turkey only red mullet otoliths were collected, while in Egypt several species were sampled for genetic tissue, except *N. norvegicus*, and the fish for the otoliths. As the sample collection and shipments from these countries required some time, the shipment of the samples was coincident with the lockdown due to the COVID-19 pandemic. The genetic tissue collected in Egypt arrived in Rome just at the lock-down day and cannot be shipped to

the genetic hub. This was possible on late May 2020. The same situation regarded the samples from Algeria. Last samples from Tunisia and of few re-samplings from Spain arrived by middle of July 2020.

In brief, 15% of the samples were collected by June 1st 2019, 65% by August 2019, 90% by the end of December 2019, a result close to the target foreseen by the project. The remaining 10% was obtained by July 2020, soon after the reopening of some activities after the lockdown.

The Sampling strategy for the Mediterranean was designed to provide a wide coverage and to balance the locations of samples in order to avoid the concentration of several samples in few areas, while at the same time collecting samples in sites functional to the objectives of WP1 and WP2. Most of the 27 Mediterranean GSA were divided in subunits that were overall 63. The sampling stations during the MEDITS or similar surveys at sea or during the fishing trip were elementary units within such geographical subunits (Fig. 0.1). For example, GSA11 was subdivided in 5 subunits, GSA17 in 3 subunits, as well as GSA22, etc. It was not always possible to get samples from each of these subunits, or in each subunit to allocate more than 1 sampling station.

It is important to mention that samples of European hake from the Bay of Biscay and the English Channel have also been collected (both genetic tissues and otoliths) in order to include in the analysis areas with likely different populations.

Figures from 0.2 to 0.7 summarize the sampling coverage by subarea and species for the genetic studies. Figures 0.8 and 0.9 represent the coverage of the otolith sampling. A subset of these stations was selected for the otolith microchemistry study (details are reported in the WP2 section). Regarding the coverage by species, it should be noted that species such as Norway lobster are rather scant in the southern-eastern side of the Mediterranean.

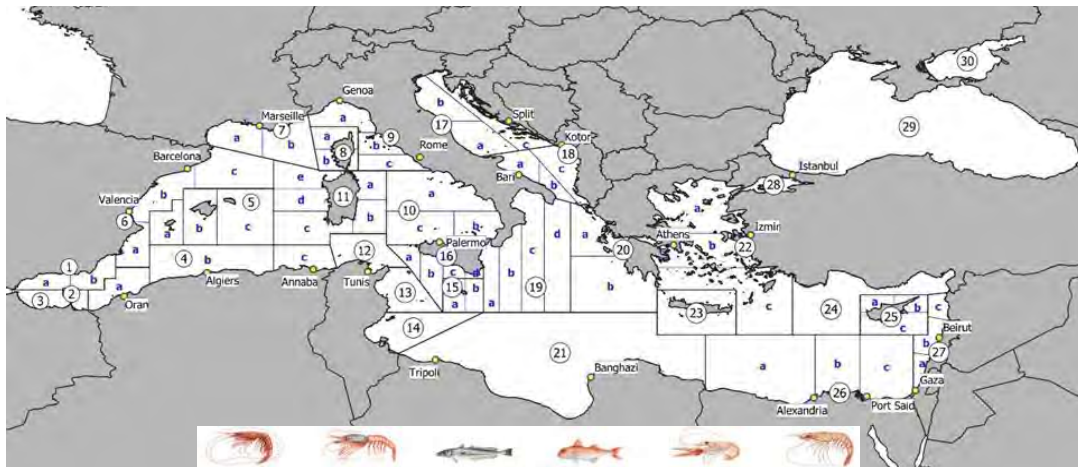


Figure 0.1 Sampling area with the representation of subdivision in sampling sub-units. The number in the circles indicate the GSA, the letter the sub-unit within a GSA.

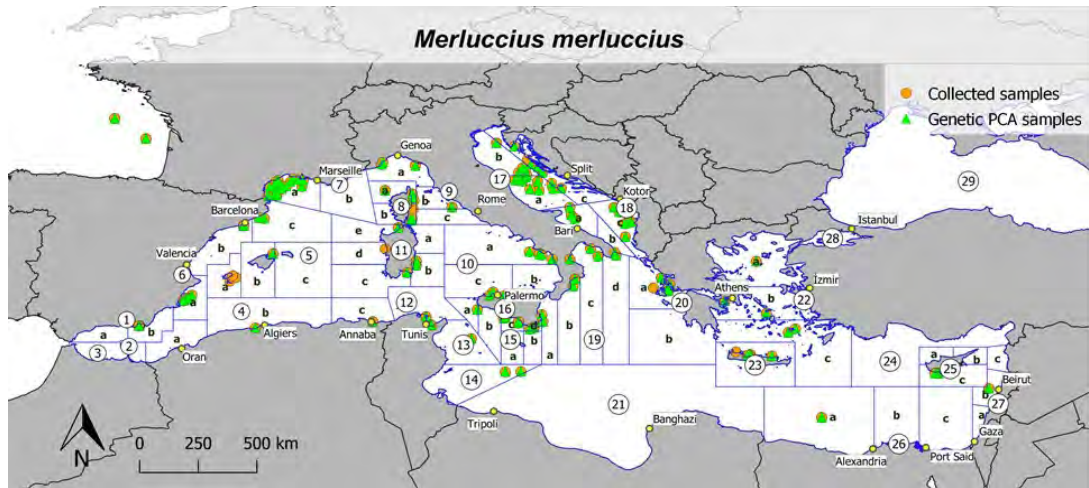


Figure 0.2 European hake (*M. merluccius*). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geositions of the collected samples, while the green triangles the subunits for which results of the Principal Components Analysis (PCA) carried out in WP1 are available.

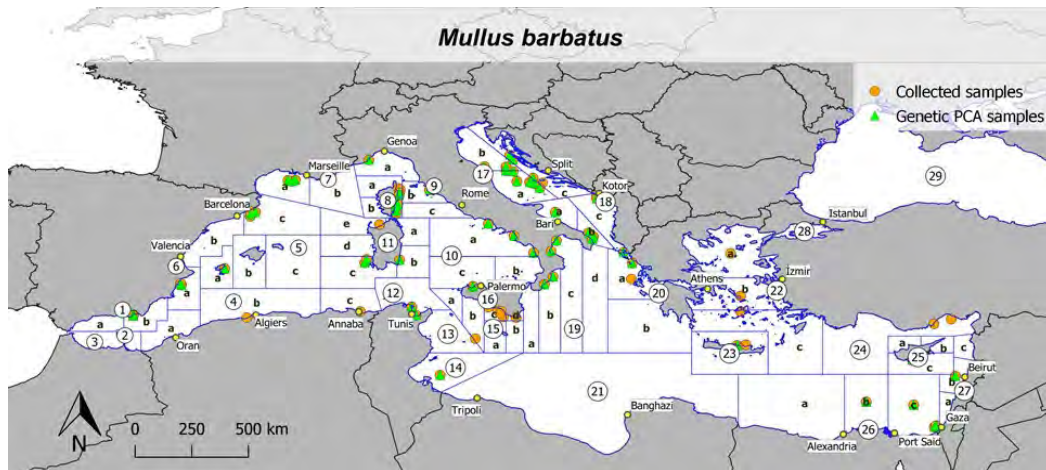


Figure 0.3 Red mullet (*M. barbatus*). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.

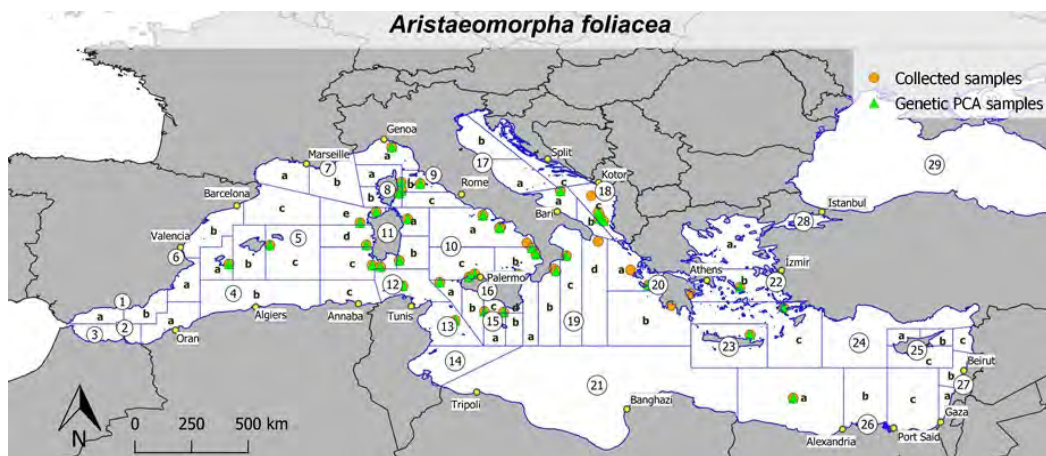


Figure 0.4 Giant red shrimp (*A. foliacea*). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.

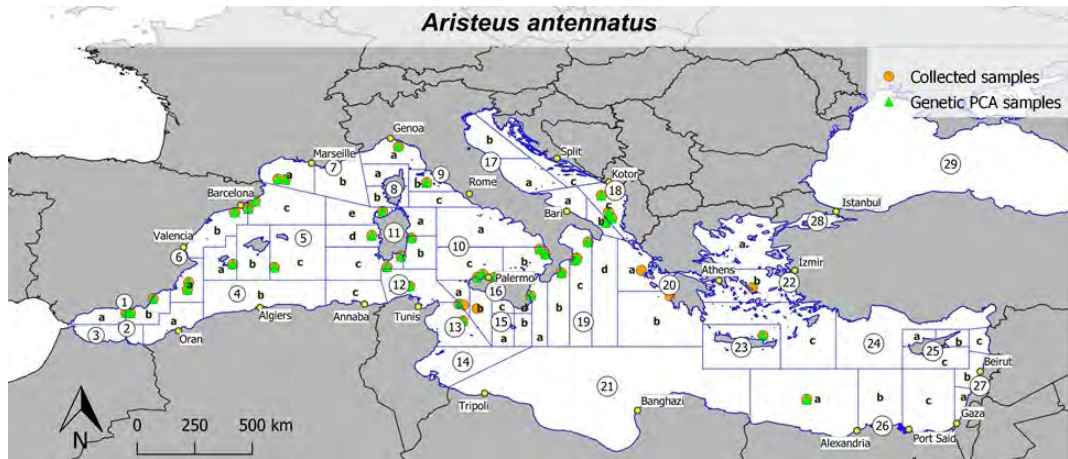


Figure 0.5 Blue and red shrimp (*A. antennatus*). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.

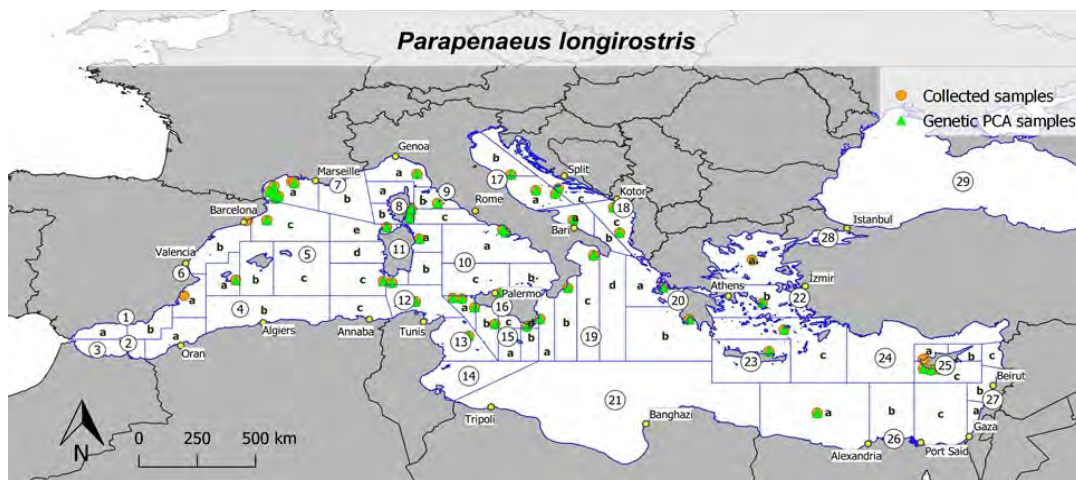


Figure 0.6 Deep-water rose shrimp (*P. longirostris*). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.

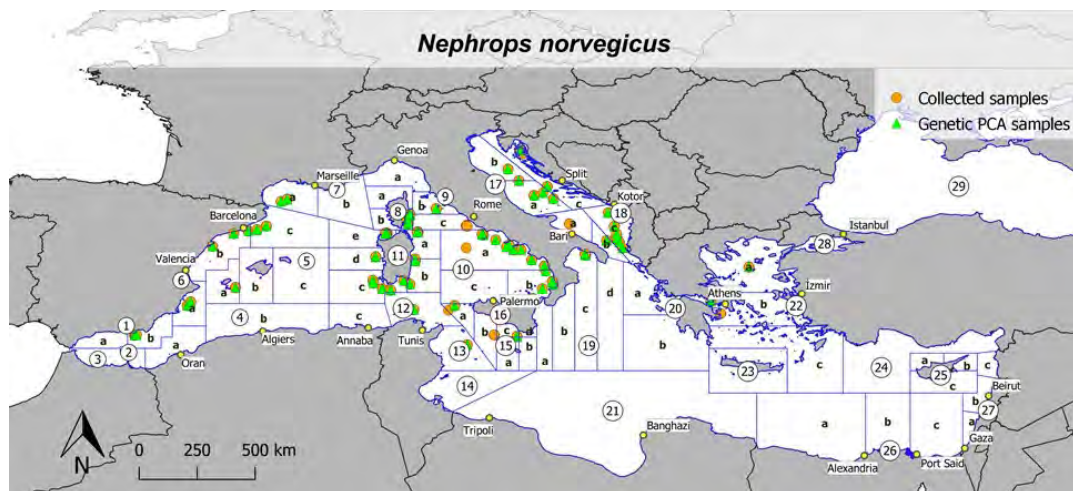


Figure 0.7 Norway lobster (*N. norvegicus*). GSA and GSA subunits covered by samplings for the genetic study. Orange bubbles represent the geopositions of collected samples, while the green triangles the subunits for which results related to the Principal Components Analysis (PCA) carried out in WP1 are available.

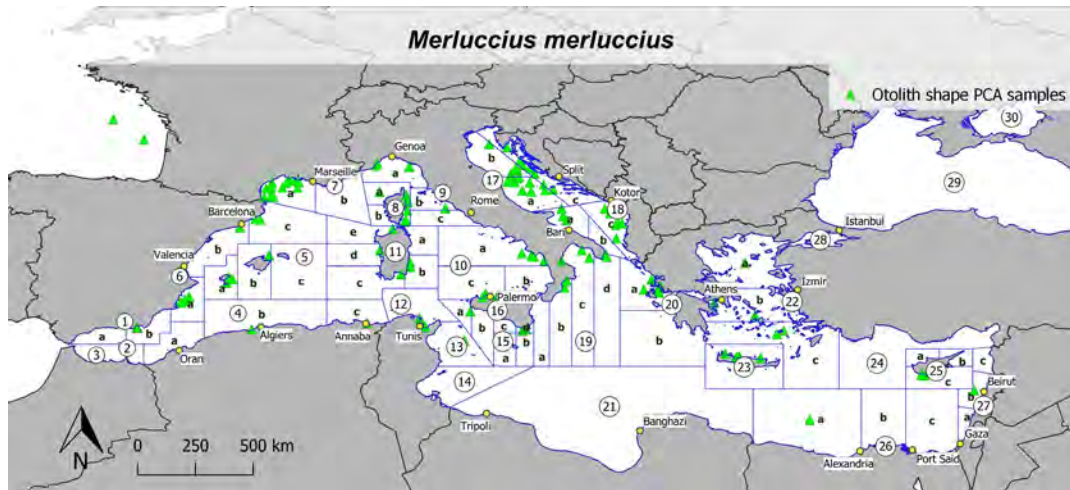


Figure 0.8 European hake (*M. merluccius*). GSA and GSA subunits for which results related to the Principal Components Analysis (PCA) carried out in WP2 (Task 2.4) are available. Geopositions are represented.

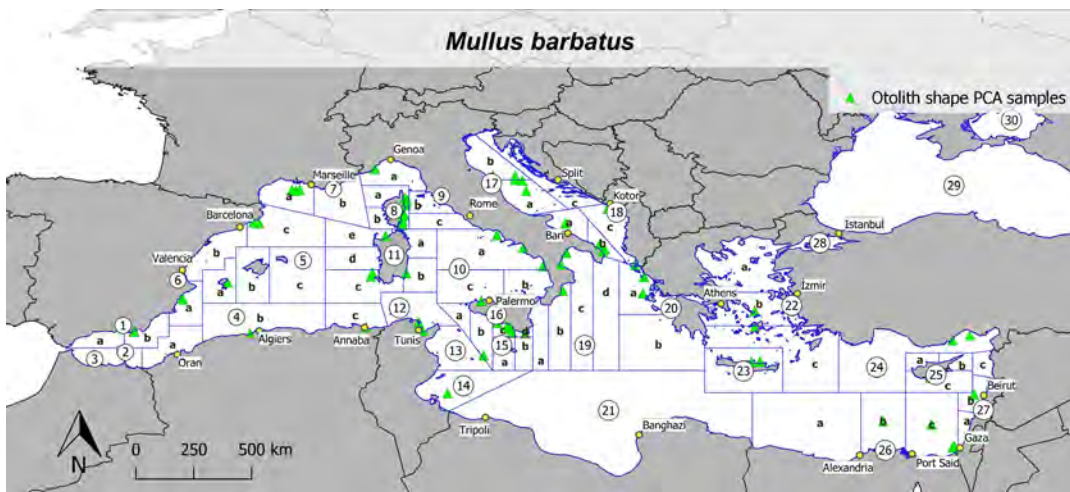


Figure 0.9 Red mullet (*M. barbatus*). GSA and GSA subunits for which results related to the Principal Components Analysis (PCA) carried out in WP2 (Task 2.4) are available. Geopositions are represented.

The number of geopositions covered by species for the sampling of genetic tissues and otoliths are reported in table 0.1 that also reports the geopositions of the meaningful Principal Components Analysis (PCA) of genetic and otolith analysis, as well as the geopositions referred to the concentration of microelements from the otolith microchemistry. These data are presented considering the aggregation at level of FAO Statistical Divisions and FAO Subareas.

All the data collected at individual level were stored in Excel files and then in a Data Base containing information on codifications, biometric measures (e.g., length, sex, maturity, etc.), geoposition of the catch, status of the tissue (fresh or frozen), qualitative information (e.g., DNA extracted), a number of meaningful Principal Component Analysis, depending from the species and type of analysis (i.e., genetics or otolith) and the concentration of microelements from otolith microchemistry. Overall, the following number of individuals was sampled by species: 1984 for European hake, 2209 for red mullet, 1470 for blue and red shrimp, 1693 for giant red shrimp, 1537 for Norway lobster and 1751 for deep-water rose shrimp. Complementary information on environmental variables were also collected from Copernicus products. This data set was shared and used for the data analysis in WP4 (Deliverable 4.2).

Table 0.1 Number of geopositions covered by species for the sampling of genetic tissues and otoliths. The geopositions of the meaningful Principal Components Analysis (PCA) of genetic and otolith analyses are also reported, as well as the geopositions referred to the concentration of microelements.

FAO SUBAREA	FAO STATISTICAL DIVISIONS	ARA		ARS		HKE			
		N. geop.	N. geop. genetic PCA	N. geop.	N. geop. genetic PCA	N. geop.	N. geop. genetic PCA	N. geop. otolith PCA	N. geop. micro conc.
Western	Balearic	12	11	2	2	23	18	21	5
	Gulf of Lion	2	2			17	17	13	
	Sardinia	18	17	22	20	28	25	28	2
Central	Adriatic	6	6	6	5	32	31	31	2
	Ionian	11	7	11	7	25	24	17	5
Eastern	Aegean	2	1	5	3	11	8	11	1
	Levant	2	1	3	3	6	5	6	6
ATO						2	2	2	
Tot.		53	45	49	40	144	130	129	21
FAO SUBAREA	FAO STATISTICAL DIVISIONS	NEP		DPS		MUT			
		N. geop.	N. geop. genetic PCA	N. geop.	N. geop. genetic PCA	N. geop.	N. geop. genetic PCA	N. geop. otolith PCA	N. geop. micro conc.
Western	Balearic	11	11	4	2	13	9	11	6
	Gulf of Lion	2	2	10	8	3	3	3	
	Sardinia	40	36	15	15	21	17	20	4
Central	Adriatic	17	14	7	7	18	16	10	4
	Ionian	8	6	11	11	16	7	15	1
Eastern	Aegean	2	1	4	3	6	2	4	1
	Levant			8	5	12	10	12	3
ATO									
Tot.		80	70	59	51	89	64	75	19

3.2 WP1 - POPULATION GENETICS AND PHYLOGEOGRAPHIC STUDIES FOR IDENTIFICATION OF BIOLOGICAL UNITS OF PRIORITY SPECIES

Chairs: Rita Cannas (CoNISMa - UNICA); Costas Tsigenopoulos (HCMR)

Partners involved: CoNISMa; HCMR

WP1 main aims are below summarized:

- reviewing the genetic data available for the 6-target species: European hake (*Merluccius merluccius*), red mullet (*Mullus barbatus*), deep-water rose shrimp (*Parapenaeus longirostris*), red shrimp (*Aristaeomorpha foliacea*), blue and red shrimp (*Aristeus antennatus*), Norway lobster (*Nephrops norvegicus*);
- performing experimental studies in order to investigate genetic distribution among and within Mediterranean sub-basins;
- delivering a comprehensive review of genetic methods and tools, detailed protocol for routine sampling and genetic monitoring, including bioinformatic aspects.

The genetic data are used to help delineating populations units, in combination with other biological data (i.e., otolith shape and microchemistry). The newly generated genetic data represent highly valuable information, updating and complementing those obtained in previous EU-funded projects such as FishPopTrace and STOCKMED. The former because: a) the increased number of species in the present project (six species compared to 4), for which the SNPs approach is attempted, and b) the geographical scope is expanded. The latter because the present project is based on genetic and otolith analyses and not only on literature review, thus outcomes are supported by experimental evidences.

WP1 is structured in six Tasks with 7 milestones and 7 deliverables.

3.2.1 Task 1.1. Literature review

Task leader: Rita Cannas (CoNISMa)

Participants: CoNISMa; HCMR

Task 1.1 aims was the reviewing of peer-review papers, grey literature, and national and European research activities regarding the genetic structure of the 6-target species in the Mediterranean, including examples of successful application of genetic methods in stock identification and connectivity patterns, and contributing to identify biogeographical boundaries and population structure of the target species.

The data base of the relevant literature, foreseen at the month 1 (M1.1, January 2019) was built. The deliverable **D1.2. - Report of available information about genetics studies on stock identification and connectivity (ANNEX 1.1)** reports all the elements of the review, based on peer-review papers, grey literature, national and European research activities. Deliverable 1.2 compiles and reviews the state of the art at 2019 of all the available genetic information for the 6-target species in the Mediterranean Sea: giant red shrimp (*Aristaeomorpha foliacea*), blue and red shrimp (*Aristeus antennatus*), European hake (*Merluccius merluccius*), red mullet (*Mullus barbatus*), Norway lobster (*Nephrops norvegicus*) and deep-water rose shrimp (*Parapenaeus longirostris*). Examples of successful application of genetic methods in stock identification and connectivity patterns have been included in D1.2.

The review contributed to identify preliminary considerations on the background related to stock boundaries, connectivity patterns and population genetic structure of the target species. This information represented the basis for the sampling design developed in Task 1.3, and the new genetic analyses performed in the Tasks 1.4 at pilot level and Task 1.5 at comprehensive level.

The sampling was planned in order to cover the whole Mediterranean but special attention was devoted to the sampling of areas where previous studies were realized, to test the persistence of the genetic configuration over the years (black dots in the figures from 1.1 to 1.6), especially where the occurrence of genetic differentiation, genetic breaks or intermediate genetic configuration among the main adjacent divergent areas were described (blue arrows in the figures from 1.1 to 1.6).

A synthesis by species is reported.

Aristaeomorpha foliacea- Most of the papers (Cannas et al 2012a; Marcias et al., 2010; Fernandez et al 2011a; Fernandez et al 2013b) failed in identifying genetic differentiation. Fernandez et al 2013a identified the occurrence of genetic differences within the Mediterranean. In particular, differences in the frequencies of mitochondrial haplogroups were detected between the Western (Ibiza, Tyrrhenian Sea and Sicily) and the Central-Eastern Mediterranean locations (Ionian and Aegean). The locality of MAZ (Mazara, Strait of Sicily) displayed an intermediate differentiation (Fig. 1.1).

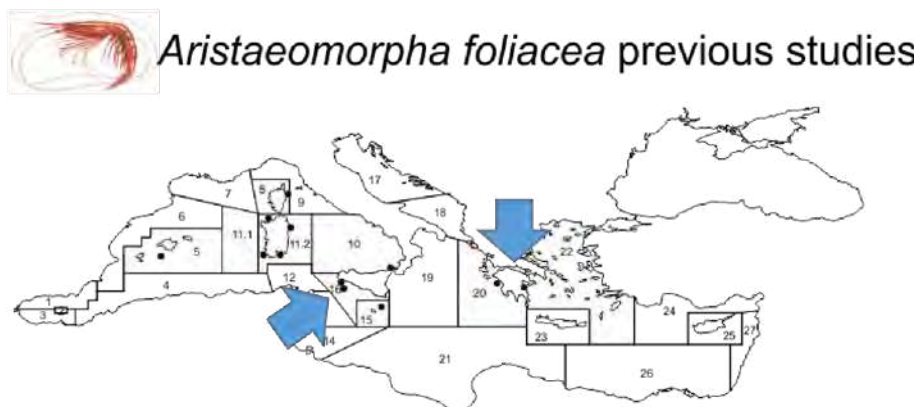


Figure 1.1 *A. foliacea*. Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).

Aristeus antennatus - The vast majority of studies (Sarda et al. 1998; Cannas et al. 2012b, Fernández et al. 2013b, Maggio et al. 2009, Marra et al. 2015; Roldán et al. 2009; Sardà et al. 2010; Lo Brutto et al. 2012) did not allow identifying significant differentiation among areas. On the contrary, Fernandez et al., 2011b indicated the occurrence of genetic differentiation among geographical regions (AO Atlantic Ocean, AS Alboran Sea, WM Western Mediterranean, EM Eastern Mediterranean, and IO Indian Ocean), where the results clearly pointed out the effectiveness of Gibraltar Strait and Sicily Strait geographical constrictions in reducing gene flow and, therefore, producing genetic differentiation between regions (Fig. 1.2).

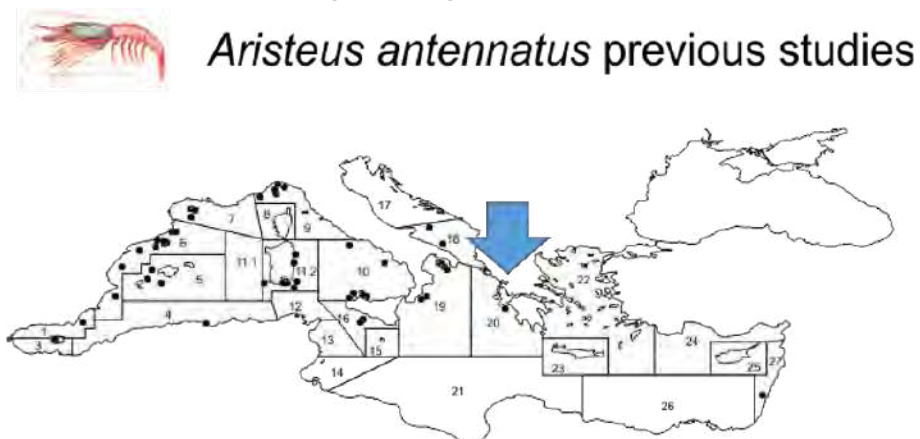


Figure 1.2 *A. antennatus*. Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).

Merluccius merluccius - Several studies identified a substantial homogeneity within the Mediterranean but suggested a strong subdivision of Atlantic and Mediterranean European hake stocks (Lo Brutto et al. 1998; Roldan et al. 1999; Lundy et al. 1999; Castillo et al. 2004; Levi et al. 2004; Lo Brutto et al., 2004; Pita et al. 2010; Tanner et al. 2014). On the contrary, other studies (Castillo et al. 2004, Cimmaruta et al. 2005) described the occurrence of genetic heterogeneity within the Mediterranean Sea. Similarly, Nielsen et al. (2012) and Milano et al. (2014) confirmed the genetic break between Atlantic and Mediterranean populations and described a finer-scale significant genetic population structure. In particular, in the Mediterranean outlier SNPs revealed a strong differentiation among Western, Central and Eastern Mediterranean geographical samples (Fig. 1.3).

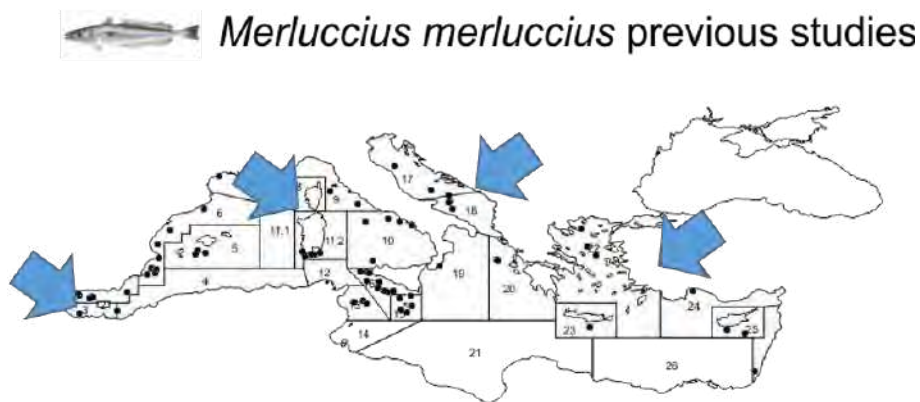


Figure 1.3 *M. merluccius*. Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).

Mullus barbatus - Several studies failed to identify differentiation among samples or described feeble genetically different populations within the Aegean Sea (Mamuris et al. 1998 a, b; Arculeo et al. 1999; Mamuris et al. 2001; Apostodolis et al. 2009). Distant locations like the Western and the Eastern Mediterranean were found differentiated in distinct clusters by Galarza et al. (2009). The Adriatic Sea was found to be differentiated from the other investigated areas (Maggio et al. 2009), whereas red mullet populations along the Mediterranean Spanish coast were genetically homogeneous (Felix-Hackradt et al. 2013). Finally, Matic-Skoko et al. (2018) identified three different genetic clusters coexisting in the Adriatic Sea (Fig. 1.4).

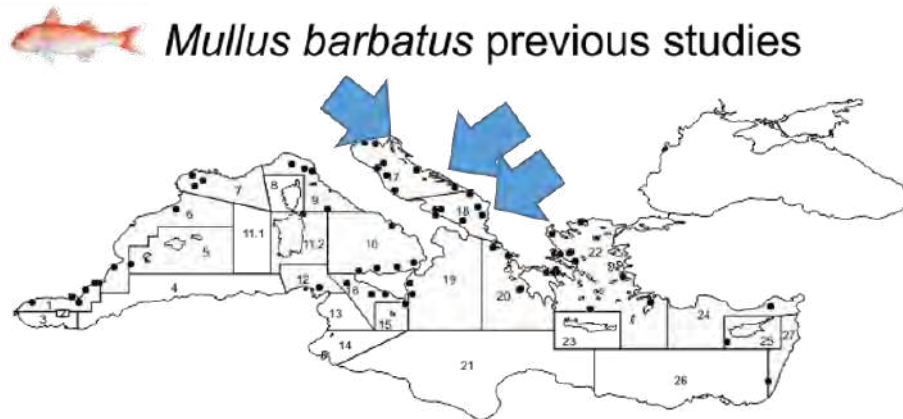


Figure 1.4 *M. barbatus*. Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).

Nephrops norvegicus - Only few and old studies are available. Allozyme data (Passamonti et al., 1997; Maltagliati et al., 1998) revealed low or moderate genetic differentiation between geographical regions (Atlantic vs. Mediterranean) but no geographical pattern of genetic differentiation, thus genetic variability seems to be randomly distributed among populations (Fig. 1.5).

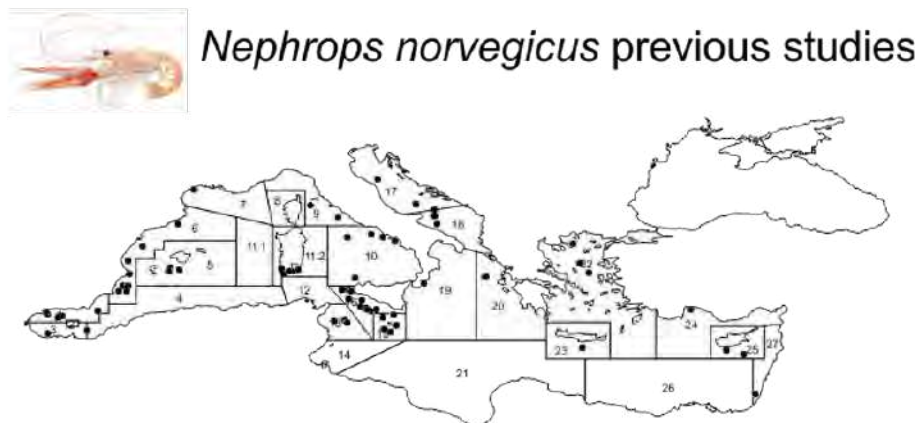


Figure 1.5 *N. norvegicus*. Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).

Parapeneus longirostris - Very few studies (n=2) have investigated the deep-water rose shrimp population genetics within the Mediterranean Sea. Garcia-Merchan et al. (2012) found a substantial homogeneity among locations along the Spanish coasts. Lo Brutto et al. (2013) identified four clusters (according to the Mediterranean sub basins: Tyrrhenian vs. Strait of Sicily vs. Adriatic vs. Aegean Sea) significantly differentiated. The greatest contribution to the differences among the four Mediterranean sub basins depended on Aegean and Tyrrhenian areas, which represented the most divergent groups (Fig. 1.6).

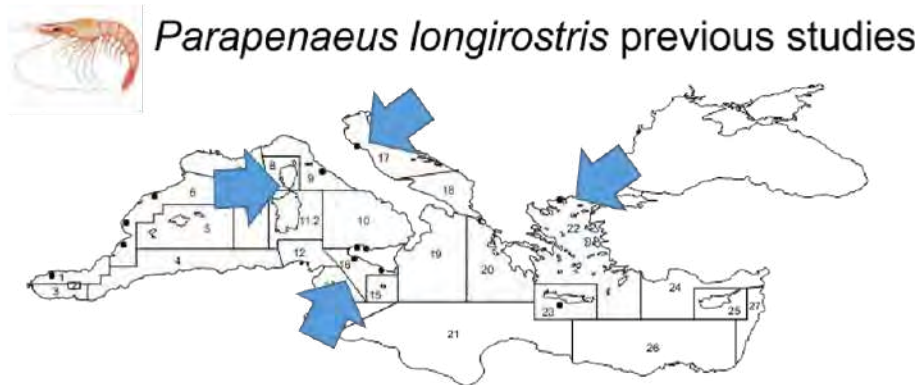


Figure 1.6 *P. longirostris*. Visualization of the sampling areas from previous genetic studies (black dots) and where the occurrence of divergences was noticed (blue arrows).

3.2.2 Task 1.2. Review bioinformatic methods and tools

Task leader: Lorenzo Zane (CoNISMa) and Costas Tsigenopoulos (HCMR)

Participants: CoNISMa; HCMR

The objective of Task 1.2 was a systematic and comprehensive review to illustrate the state of the art of the genetic methods used for stock identification and boundaries delimitation in fisheries. Bioinformatics methods and tools used for the analysis and interpretation of genetic/genomic data, scientific literature databases, such as PubMed, ISI Web of Knowledge, Scopus, as well as grey literature from past and ongoing projects, and reports of working group (i.e., ICES, GFCM and STECF) are part of such review. The aims were also to carefully examine each method/tool for its peculiarities, appropriateness, robustness and accuracy in stock identification, considering, inter alia, the necessary equipment and expertise, the molecular markers needed, the costs, the annual/seasonal replication requirements, the effectiveness, the associated risks, the transferability, and the user-friendliness of the methods.

Under this task the Deliverable **D1.3 - Report with the review bioinformatic methods and tools (ANNEX 1.2)** was issued, illustrating applicability, advantages and disadvantages of different methodologies and techniques, as well as their limitations and strengths. It focuses on bioinformatics methods and tools used for the analysis and interpretation of genetic/genomic data in the context of stock assessment and identification of populations of marine organisms. Through a systematic literature review, the deliverable aims to identify bioinformatic methods used to infer stock boundaries, connectivity patterns and population genetic structure of the target species. This information represented the background for the new genetic analyses performed in Tasks 1.4 and Task 1.5.

D1.3 builds up on a recently published review (Cuellar-Pinzon et al. 2016), which focused on identifying the kind and trend of use of genetic markers in genetic fisheries and extensively considered all the papers published on the topic in the period 2004-2014; the authors identified a switch toward NGS (Next Generation Sequencing) derived markers in fishery research starting from 2011, accompanied by a reduction in the use of “classical” genetic markers, such as microsatellites and mitochondrial DNA-based markers.

A similar search strategy has been followed in D1.3 to identify papers published from 2015 to early 2019, using genetic approaches to study marine populations and stocks and, for each study, the following characteristics were recorded: the kind of marker used, the taxonomic group, the data analysis and methods used. Once the available

bioinformatic methods and tools have been identified, these have been examined, highlighting their peculiarities, appropriateness, robustness and accuracy in stock identification. To this end, the reference is to published literature, whenever available, and to the authors' expertise. The review confirms the trend toward the use of NGS methods but also highlights the fact that microsatellites and mtDNA are still the most commonly used markers in genetic fisheries studies.

With regard to the markers used, though 13% of the papers used SNPs, most of them were still based on microsatellites (47%) and, noticeably, on mtDNA markers (37%); the few other studies with different methodologies included allozymes, AFLP, RAPD, EPIC-PCR and nuclear DNA sequencing. Importantly, the percentage of studies using SNPs increases to 37.5% when considering only the studies published so far in 2019. Most of the papers were characterized by the use of different bioinformatic methods, clearly depending on the markers used. Most of the studies were characterized by the use of different bioinformatic methods, clearly depending on the markers used. With regard to mtDNA markers, the data analysis strategy is clearly delineated, and relies mainly on AMOVA (Analysis of MOlecular VAriance) or pairwise F_{st} (Fixation Index). With regard to the use of microsatellites, in addition to AMOVA and F_{st} , clustering between individuals follows, often accompanied (but rarely substituted by) other, less model dependent, representations of relationships between individuals such a PCoA (Principal Coordinate Analysis), or DAPC (Discriminant Analysis of Principal Components). With regard to the use of SNPs, the situation seems more variegated, likely due to the challenges imposed by much bigger dataset. Preliminary comparisons between population samples are still often performed, in this case also with recent software/packages. Individual clustering is still used, it is often replaced by DAPC, PCoA and MDS (Multi Dimensional Scaling), even performed with specialized software. As expected for genome-wide markers, the use of outlier detection test becomes very common, while, probably due to the complexity of SNPs dataset, inference of migration patterns or assignment test seems rather uncommon.

The different bioinformatic methods and tools have been examined individually based on published literature and our own research experience. The overall evaluation is reported in the Table 1.1; the following aspects have been considered: equipment needs, expertise, molecular markers (number), costs, annual/seasonal replication requirements, effectiveness, associated risks, transferability, and user-friendliness.

Table 1.1 - Overall evaluation of methods (blue lines) and relative tools, the evaluation is based on the literature data and researchers' expertise.

Method/tool	Equipment (software)§	expertise	N° markers	costs	replication requirements	effectiveness	risks	transferability	user-friendliness
mtDNA	*	*	*/**	*	*	*	*	****	****
AMOVA	Arlequin	*	*	*	*	*	*	****	***
pairwise FST	Arlequin	*	*	*	*	*	*	****	***
PCoA	Genalex	*	*	*	*	*	*	****	***
haplotype clustering	TCS/Network	*	**	*	*	*	*	**	**
genetic stock analysis	mixstock in R	**	**	*	*	*	*	**	**
microsatellites	**	**	**/**	**	**	**	**/**	*	**
AMOVA	Arlequin	*	**	**	*	**	**	****	***
pairwise FST	Arlequin	*	**	**	*	**	**	****	***

PCoA	Genalex, Adegenet	*	**	**	*	**	**	**	***
genetic stock analysis	mixstock in R	**	***	**	**	**	***	**	**
population differentiation	Genalex, Fstat, Genetix, Genepop	*	**	**	*	**	**	***	***
clustering analysis	Structure, Clumpp, Distruct	***	***	**	**	**	***	**	***
DAPC	Adegenet	*	***	**	**	**	***	**	***
migration patterns	GeneClass, Bayesass, Migrate	***	***	**	**	**	***	**	**
outlier detection	Lositan, Bayescan	*	***	**	**	*	*	**	**
SNPs	****	****	*** / ****	****	**	** / ***	** / ***	*/ ***	*/ ***
SNP identification and genotyping	STACKS	****	****	****	**	***	***	**	*
population differentiation	NGSdist, diveRsity	**	***	***	*	**	**	***	***
clustering analysis	Structure, fastSTRUCTURE, fineSTRUCTURE, ParallelStructure, Structure_threader, parastructure	***	****	****	**	***	***	**	*
DAPC	Adegenet	*	***	***	*	***	**	**	***
PCoA	Adegenet, NGStools, Vegan	*	***	***	*	**	**	**	***
MDS	NGStools, Vegan	*	***	***	*	**	**	**	***
outlier detection	Lositan, Bayescan	*	****	***	*	**	**	**	**
genome-wide analyses	PopGenome	***	****	****	**	***	***	**	*

§ = see main text for references.

* = low; ** = medium; *** = high; **** = very high.

Following the aforementioned evaluation, needs and solutions, we built a highly reproducible pipeline for the analysis of the six focal species of MED_UNITS incorporating multiple of the reviewed software (Figure 1.7), taking into account the computational cost of analysing the dataset.

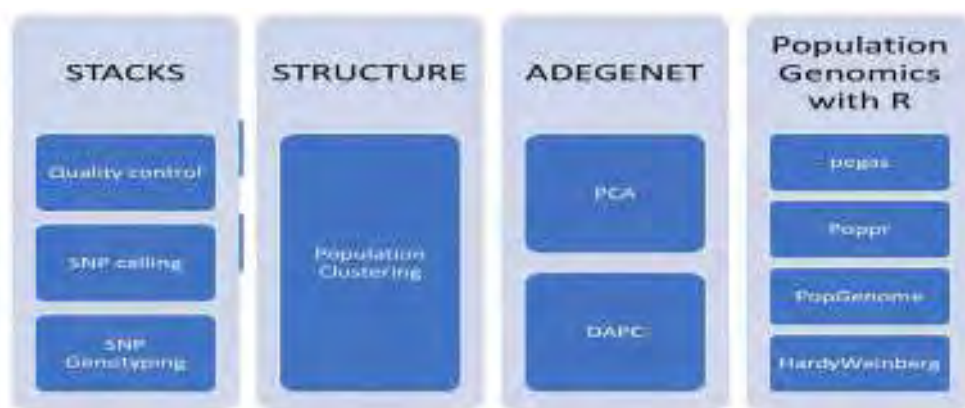


Figure 1.7 The designed pipeline components for analysing the raw data and assessing the population structure of MED_UNITS species.

3.2.3 Task 1.3. Genetic sampling design

Task leader: Rita Cannas (CoNISMa)

Participants: CoNISMa; HCMR

Task 1.3 aims were to define the sites for the genetic study and to deliver the protocol for genetic tissue samplings. Task 1.3 foresaw that the samples for genetic analyses from European Mediterranean waters and non-European countries were obtained as described in WPO. As regards the protocol for sampling genetic tissues (Deliverable 1.1) a first draft was issued by January 2019 and annexed to the Inception Report. Following the inputs received during the kick-off meeting with the Contracting Authority (February 2019) and feedbacks from partners during the first plenary meeting of the project (March 2019) the Deliverable D1.1 was revised, including the excel sheet and a visual protocol to facilitate the sampling activities. In the **ANNEX 1.3** the **Deliverable 1.1 - Sampling and shipping protocol** is reported.

The project included a first batch of analyses of tissues for the pilot study, and a second batch of analyses for the comprehensive study.

A first draft of the Milestone M1.2, describing criteria for sampling sites' selection and including a 'provisional' list of pilot preferred locations (priority 1), and alternative locations (priority 2 and 3) for the sampling of the six species under investigation, was first issued by January 2019 and annexed to the Inception Report, in view of the kick-off meeting (February 2019). Then, to make clearer the approach to the sampling, M1.2 was revised (May 10th, 2019, 1st Progress Report) (**M1.2 v1, ANNEX 1.4 to this report**), including the list of the sampling sites by GSA and species, along with maps with detailed indications of areas where the sampling effort had to be performed, in order to cover the areas in which discontinuities were identified from previous studies. M1.2 included thus also indications for the samplings of the comprehensive genetic study (Task 1.5). To minimize risks and to compensate the eventual failure of sampling in the non-EU countries, 48 additional samples were allocated to European GSAs preferably adjacent to the non-EU GSAs. These extra 48 samples represented spare samples to be used only in case of unavailability of some sampling sites in order to guarantee anyhow, at the end of the sampling phase the planned number of sites for the genetic analyses, i.e., a total of maximum 228 locations. However, the need of using these 'spare samples' was limited because, thanks to the liaison between DGMARE and GFCM, samples were also gathered from some GSAs or GSA subunits, depending from the species, in the southern Mediterranean.

On April 2019, a total of 48 sites were proposed as Pilot genetic sites, i.e., the locations that should have been analyzed as a priority in Task 1.4. The pilot study was a pre-screening phase for testing eight sites corresponding to the different genetic stocks identified in the latest studies with 50 individuals from each location. These sites were chosen among the localities where genetic discontinuities were identified based on previous studies (elements also in D1.2).

However, samplings for the pilot study experienced some difficulties and samples were delivered at the relevant laboratories of CoNISMa and HCMR by the end of July/early August 2019, instead of at month 5, May 2019. The samplings for the comprehensive study were almost completed by the end of October 2019 (87%), except few areas. Thus, considering the availability of samples at month 7 (July 2019), a new list was selected, based on the so-called 'realized' Pilot genetic sampling sites, that is the locations included in the Task 1.4 for genetic analyses. Table 1.2 compares the proposed versus the realized pilot sampling sites. In most cases, the foreseen samples for the Western and Central Mediterranean were available, while there were no samples from the Eastern Mediterranean, the Alboran area, and the Southern Mediterranean. These areas were included in the comprehensive genetic studies (Task 1.5). It should be noted that species such as *A. foliacea* and *P. longirostris* are rather scant in the westernmost GSAs, as *N. norvegicus* is in the southern-eastern ones, where it is almost absent.

In the comprehensive study (Task 1.5), from a minimum of 15 to a maximum of 30 additional locations were thus foreseen to be analysed in the whole Mediterranean Sea. In the end, the total number of investigated locations was foreseen to vary from a minimum of 23 to a maximum of 38, summing up the pilot locations and the additional ones. Then it was foreseen that the total number of sites to be analyzed, considering the 6 target species of the project, would be comprised between 138 and 228.

The final locations were expected to preferably include those not included or under-represented in previous papers and from areas where previous studies indicated the possible existence of genetic discontinuities.

About 35 individuals in each location were foreseen for genotyping with the panel of markers selected in the pilot study. The number of 35 individuals sampled at each location for the comprehensive study is a trade-off found after a literature review made also in the SPELMED project (Specific Contract 02 under the framework contract EASME/EMFF/2016/032), following, among the others, the same order of magnitude suggested by Fumagalli (2013), based on experimental simulations to achieve the highest accuracy in predicting population structure in the SNPs approach; similar numbers are considered a valid sampling size for microsatellites analyses (Hale et al., 2012).

The sampling sites for genetics were planned to be, as much as possible, coincident with those for the otolith samplings.

Actually, at the end of the project, 210 sites have been sampled for the genetic studies.

Table 1.2 – List of the proposed and realized sites for the pilot genetic study. The cells in pale green indicate the sites/species foreseen and sampled in the proposed list, the cell in dark green the site sampled but not foreseen or oversampled, while the cells in red indicate the sites foreseen but not available due to the sampling failure.

Pilot sampling sites – proposed as in the First Progress Report Annex 6 Table 1																		
Species	GSA1	GSA3	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	total
<i>A. foliacea</i>			1				1	1	1	1			1	1	1			8
<i>A. antennatus</i>	1		1	1			1		1	1			1	1				8
<i>M. merluccius</i>	1	1							1	1		1	1		1		1	8
<i>M. barbatus</i>			1				1				2	2			1		1	8
<i>N. norvegicus</i>	1		1	1			1		1	1	1				1			8
<i>P. longirostris</i>							1	1		1	1	1	1		1	1		8
total	3	1	4	2	0	0	5	2	4	5	4	4	4	2	5	1	2	48

Pilot sampling sites – realized based on the samples available at the end of July 2019																		
Species	GSA1	GSA3	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	total
<i>A. foliacea</i>							1	1	2	1		1	2					8
<i>A. antennatus</i>				2	1		1		2			1	1					8
<i>M. merluccius</i>				2			1		2	1		1	1					8
<i>M. barbatus</i>				1			1		2		2	2						8
<i>N. norvegicus</i>				2		1	1		2		1	1						8
<i>P. longirostris</i>				1			1		2	1	1	1	1					8
total	0	0	0	8	1	1	6	1	12	3	4	7	5	0	0	0	0	48

3.2.4 Task 1.4. Pilot genetic analyses

Task leader: Costas Tsigenopoulos (HCMR) and Lorenzo Zane (CoNISMa)

Participants: CoNISMa; HCMR

The main aims of the Pilot Genetic studies can be summarized as follows:

1. optimization of protocols: type of marker to use, sample size, level of sequencing/number of markers;
2. optimization of the final sampling effort by area/species;
3. analysis of the temporal stability, comparing the current results with those of previous genetic studies.

A Genotyping-by-Sequencing (GBS) methodology constructing reduced-representation libraries in each species was selected. SNPs markers newly isolated following the double-digest Random Amplified DNA (ddRAD) sequencing were used.

The analyses in task 1.4 were under the responsibility of HCMR and CoNISMa as described in Table 1.3.

Table 1.3 Responsibility in task 1.4 for the different partners by marker and species

Species	SNPs	SSRs
<i>Aristaeomorpha foliacea</i>	HCMR	HCMR
<i>Aristeus antennatus</i>	CoNISMa	CoNISMa
<i>Merluccius merluccius</i>	CoNISMa	CoNISMa
<i>Mullus barbatus</i>	HCMR	HCMR
<i>Nephrops norvegicus</i>	HCMR	HCMR
<i>Parapenaeus longirostris</i>	CoNISMa	CoNISMa

The activities of Task 1.4 started in April 2019 with preliminary tests, aiming at optimizing the protocols and giving indication on the best procedure to realize the sampling activities. Considering the first samples collected for this test in March 2019 and the following ones, collected until June 2019, for the purposes of the pilot study, 18 sampling locations from nine GSAs (GSA6, GSA7, GSA8, GSA9, GSA11, GSA16, GSA17, GSA18, GSA19 and GSA23) were covered by CoNISMa-UNICA, IFREMER, CIBM, CNR-IRBIM, HCMR and COISPA; tissues for the six species have made available to CoNISMa-UNIPD and HCMR for some preliminary laboratory analyses and then for the pilot study. Not all the samples taken for preliminary laboratory analyses were adequate for the pilot study.

Total genomic DNA has been isolated from these tissues using commercial kits, following manufacturer's protocol or tailor-made extraction protocols modified from already published ones (salt extraction protocols either by Miller et al., 1988 or Cruz et al., 2017). The amount and quality of DNA has been quantified using a spectrophotometer and resulted of good/medium quantity and quality for some species/areas and low-medium for others but, in any case, suitable for the following steps (i.e., restriction with enzymes). Based on these provisional results, indications have been given for the sampling activities in order to have the best quality DNA, stressing in particular the need to collect the tissues immediately after the capture of the animals (within 1-2 hours from the haul), and to strictly avoid the use of any preservative (chemicals) especially in crustaceans. Alternatively, if specimens were frozen and transferred to the laboratories, collectors were advised to dissect and preserve muscle tissues without allowing the samples to thaw.

For the SNP isolation and identification, some individual DNAs available in all six species have been digested each time by two specific high-fidelity Restriction Enzymes (RE). This is the part of the pilot study which required the most careful optimization, because the great difference in the genome size of crustaceans compared to fish and the dissimilar GC content (e.g., less frequent cutting sites) among species result in different performance.

Several Restriction Enzymes (RE) were tested to identify the best combination for the species under investigation (specifically *SbfI*, *SphI*, *NlaIII* and *PstI*). A set of adapters was ordered to construct double-digest Random Amplified DNA (ddRAD) libraries and simultaneously genotype 288 samples in a single run.

Regarding, the protocols for the DNA extraction, given that the quality of DNA is crucial for ddRAD method, mainly because of the need for accurate quantification and equal representation of all samples in the final library, the following 6-steps workflow was adopted in the pilot study:

1. DNA extraction;
2. 1st Quality Control through gel electrophoresis images and Nanodrop spectrophotometer measurements;
3. Post-extraction treatment with RNase;
4. 2nd Quality Control: gel image;
5. DNA quantification: Qubit Fluorometer and Qubit dsDNA BR Assay;
6. Final concentration adjustment with dilution and transferring of samples to 96-well plates for library construction.

Standard salt extraction protocols by Miller et al., 1988 and Cruz et al. 2017 were used in HCMR and CoNISMa, respectively. Unfortunately, alternative extraction protocols based on commercial kits, successfully used in the preliminary phases, failed to provide high molecular weight DNA, probably due to the lower quality of tissues available, the lower efficiency and lower yields of these extraction methods when dealing with problematic samples. The salt extraction protocol is considered a rather cheap approach, but significantly time and labor-consuming, requiring adequate work subdivision among labs and involvement of more time and personnel than previously planned.

In general, after processing of more than 400 samples per species, the salting out method provided HMW DNA without RNA contaminants in most cases.

DNA extracts have been classified into three (3) quality categories ranging from 1 (high quality) to 3 (not usable for ddRAD). Category 1 (good quality) refers to DNA extracts where only a sharp band of HMW DNA is present (no 'smear'). Category 2 (medium quality) refers to DNA extracts where both a band of HMW DNA and some 'smears' of degraded DNA are present, or to cases that the band of HMW is absent and category 3 where the DNA is degraded or absent.

Some samples from specific GSA subareas failed to provide HMW DNA for some species (Table 1.4). Moreover, for all the species high intra-area variability was observed, which raised the question if the indications provided for sampling have been respected and/or some other, not identified, factors that could affect DNA extraction results. In particular, the indication to provide tissues from fresh animals was probably not always strictly followed due to logistic problems, when specimens have to be obtained from commercial landings or from frozen individuals, probably totally unfrozen before processing, despite the opposite clearly required in the protocol. However, in some situations in which the same operator and procedures were applied for the sample processing the results were contrasting, depending on the species.

Among the three species in charge of HCMR (*M. barbatus*, *A. foliacea* and *N. norvegicus*), *M. barbatus* was the one that performed worse since smears of degraded DNA were present in most samples. As a result, only 6 populations were processed for library construction with some of them included although considered of medium or bad quality. Subareas 11de and 18c failed to provide usable DNA and were excluded. For *N.*

norvegicus, the extracted DNA was of high quality in most samples but again in some subareas (GSA6b, 6c, 18a) the quality of DNA was very low or even prohibitive for library construction. Finally, for *A. foliacea* the majority of samples provided DNA of good quality, except for 2 subareas (GSA19b, 19c).

Table 1.4 – List of the DNA extracts for the different species and areas. The cells in green indicate the number of extracts of Category 1- good quality, in light green the extracts of Category 2- medium quality, in orange the extracts of Category 3 the extracts of poor quality, not usable for ddRAD. The last row represents the specimens finally included in the ddRADseq library.

DNA category	Species								
	<i>A. foliacea</i>								
Sub-area GSA	09b(p)	10b	11e	11(p)c	16b	18a	19b	19c	total
1	49	44	50	48	23	10	0	0	224
2	1	2	0	0	27	32	3	47	112
3	0	4	0	2	0	8	47	3	64
ddRAD	50	44	47	50	44	46	0	7	288
	<i>A. antennatus</i>								
Sub-area GSA	06b	06c	07a	09p	11ab	11p	18b	19b	total
1	0	1	32	17	30	26	30	38	174
2	15	15	9	20	7	12	13	11	102
3	35	34	9	13	13	12	7	1	124
ddRAD	15	14	41	36	37	38	43	49	273
	<i>M. barbatus</i>								
Sub-area GSA	06c	09b	11c	11de	17a	17b	18a	18c	total
1	0	6	0	0	0	1	25	0	32
2	46	44	37	0	50	45	25	0	247
3	4	0	13	50	0	4	0	50	121
ddRAD	46	50	46	0	50	50	46	0	288
	<i>M. merluccius</i>								
Sub-area GSA	06b	06c	09bc	11c	11e	16ab	18a	19b	total
1	38	41	47	32	48	45	49	50	350
2	12	6	1	17	1	5	0	0	42
3	0	3	2	1	1	0	1	0	8
ddRAD	33	30	37	29	38	38	41	42	288
	<i>N. norvegicus</i>								
Sub-area GSA	06b	06c	08ab	09(a)b	11e	11p(c)	17ab	18a	total
1	10	14	50	50	50	41	47	0	262
2	19	10	0	0	0	3	3	8	43
3	21	26	0	0	0	6	0	42	95
ddRAD	24	22	50	50	50	46	46	0	288
	<i>P. longirostris</i>								
Sub-area GSA	06c	09b	11de	11p	16b	17ab	18a	19b	total
1	12	27	43	40	18	36	45	38	259
2	15	13	5	8	17	11	3	11	83
3	23	10	2	2	15	3	2	1	58
ddRAD	27	36	39	38	35	36	39	38	288

As concerns the three species in charge of CoNISMa (*M. merluccius*, *A. antennatus*, and *P. longirostris*), European hake (*M. merluccius*) resulted to have high quality DNA with only two extracts of very poor quality. For crustaceans a certain fraction of the samples was not usable for *P. longirostris*, with subareas 06c, 09b and 16b showing about 50% of the specimens with poor quality DNA (cat. 3 and 4) and even larger fraction for *A. antennatus*, with subareas 6b and 6c providing DNA classified only in categories 2 and 3 and less of 50% of specimens providing good or medium quality DNA in all subareas, except 19b.

Regarding, the library preparation, the first step in the library design was to evaluate the digestion and to choose the appropriate pair of enzymes for each species. To improve efficiency and reduce costs HCMR performed also the digestion tests on the three species assigned to CoNISMa. After the digestion tests, the appropriate pair of enzymes for the crustaceans resulted *SbfI* -*NlaIII* {(*SbfI*; CCTGCA[^]GG) 8-base cutter & 4-base cutter)} and *PstI* - *NlaIII* {(6-base cutter & 4-base cutter)}, while for the fish species the pair *SbfI*-*SphI* {(*SphI*; GCATG[^]C) 8-base cutter & 6-base cutter – i.e., a classical pair of RE used in fish studies in the HCMR lab}.

Regarding the timing of Task 1.4, in May-June 2019, all reagents (enzymes, tags, primers) and materials were ordered, and the sequencing center selected and alerted in order to be ready to process the libraries in August- September 2019 (according to the original timetable). Unfortunately, unexpected delays in sampling have led to the shifting of some months the experimental work in the genetic laboratories. In particular, contrary to the availability of samples by early June, most tissues were delivered by the Genetic Hub to the different laboratories involved at the end of July/early August 2019. Therefore, the DNA extractions could have started only in September, after the summer holidays period for both CoNISMa and HMCRR staff. However, given the need of optimization of species-specific protocols, especially challenging for the crustacean species, and the time needed by the sequencing center, the task 1.4 duration was rather longer than foreseen. The period for the completion of D1.4 shifted to February 2020 and then it was impacted by COVID-19 pandemic. The Deliverable was hence issued on 10th June 2020.

Since the end of February, the spreading of COVID-19 determined a progressive slow-down of activities, with a complete stop of all working possibilities and presence in the laboratories and offices from the first week of March till the first week of May 2020. Nevertheless, while the laboratories were closed for two consecutive months, the researchers continued to work remotely, as much as possible. However, the impossibility to promptly access remotely to PCs or servers in the different institutions caused problems even for some online computing activities (bioinformatic and population genetic analyses) needed for the completion of Task 1.4 and the writing of the **Deliverable 1.4. Report with the results of the pilot studies** (see **ANNEX 1.5**).

In June 2020 the laboratory work re-started but with severe limitations imposed by the social distancing policies, with unavoidable difficulties and slowdowns. Therefore, the original plans for the activities foreseen up to the end of the project have been totally reconsidered in order to recover for the delays caused by the forced stop and to re-organize the work based on the new rules imposed by the national institutions for the working places.

During January- May the activities were concentrated on the completion of Task 1.4 (pilot genetic analyses) with the bioinformatic analyses and population genetic analyses for the 4 crustacean species, and the refinement of the analyses for the 2 fish species, for which preliminary results were already available and presented at the Genetic Workshop held in November 2019.

Further results in brief from the pilot study

A total of 48 sites, 8 for each of the 6 target species, have been genetically analyzed during the pilot phase. It is noteworthy highlighting there was a difference between the sites as originally proposed and actually realized (samples available for the analyses at the end of July 2019), as reported in table 1 of task 1.3. In most cases, the foreseen samples for the Western and Central Mediterranean were available, while there were no samples from the Eastern Mediterranean, the Alboran Sea, the Balearic Islands and the Southern Mediterranean. These areas are of priority importance, then they were included anyhow in the comprehensive genetic studies.

As concerns the methodology, while for two fish species the ddRADseq protocol is well-established and tested, we faced major difficulties for the four crustacean species, for which little genomic information exists and no previous attempts have been reported. Additionally, a major impediment was the DNA quality, which is a prerequisite for the downstream ddRADseq library preparation and the correct bioinformatic analysis. However, we have produced one ddRADseq library per species, each including some 288 multiplexed specimens from at least 6 sampling locations.

Regarding the population genetic results and comparisons with previous studies they are briefly summarized in the paragraphs below. Full details can be found in D1.4 (**ANNEX 1.5**). It is important to remind that these results are preliminary and susceptible to change when additional samples are added (enlarging the geographical coverage) and the protocols (especially for the crustacean species) further optimized. In brief, the genetic differentiation among the sampling locations has been investigated using the following analyses (software) and metrics.

- **Discriminant Analysis of Principal Components (henceforth DAPC)**; Jombart, 2008; Jombart and Ahmed, 2011) performed with Adegenet ver. 2.1.1 (R version 3.5.1, R Development Core Team, 2014; <http://www.r-project.org>).
- **Bayesian clustering among the sampled sites (henceforth STRUCTURE analysis)**, performed with the software STRUCTURE v 2.3.4 (Pritchard et al., 2000; Falush et al 2003, 2007; Hubisz et al., 2009).
- Genetic differentiation among ‘populations’ measured by **Fixation Index** (henceforth F_{ST} value). Both pairwise and global F_{ST} values were calculated with Arlequin ver. 3.5.1.2 (Excoffier and Lischer, 2010). To correct for multiple testing the Benjamini–Yekutieli procedure (Benjamini and Yekutieli 2001), controlling for the **False Discovery Rate (FDR)**, was used.
- **AMOVA Analysis of Molecular Variance** (AMOVA) and the calculation of the relative Fixation Indexes of differentiation among groups of populations (henceforth F_{CT}), performed again with Arlequin, grouping the samples on the basis of *a priori* hierarchical structures. The statistical significance of the resulting values was estimated by comparing the observed distribution with a null distribution generated by 10,000 permutations, in which individuals were redistributed randomly into samples.

The following paragraphs describe the main results obtained in D1.4 by species. It is worth pointing out that all these results were preliminary, as the final results of the project are the ones described for Task 1.5. Similarly, the protocols of Task 1.4 were further optimized and successfully used within Task 1.5.

For *M. merluccius*, after filtering and extensive data checking, a panel of 734 high quality SNPs was used for the genetic analyses; filtering required eliminating several bad quality specimens leading to the inclusion of 268 individuals in the final analysis. Pairwise genetic differentiation among sampling sites were small (F_{ST} values ranging from 0.0003 to a maximum value of 0.0114), but highly significant in most of the comparisons (17 out of 28). Both DAPC and STRUCTURE indicated low differentiation between samples, while AMOVA analyses highlighted a strongest differentiation of GSA06b and 06c from the rest of the samples ($F_{CT}=0.00306$, $P=0$), and to a lesser extent, a further differentiation of samples from GSA18a and 19b ($F_{CT}=0.00263$, $P=0$), suggesting the **possible occurrence of three or more different genetic clusters**.

This result represented a significant advance in respect to previous published studies, where differences were not detected between Mediterranean samples, unless using outlier loci (Milano et al. 2014).

For *M. barbatus*, due to low quality of the extracted DNA, we had to exclude samples from GSA11de and 18c, and work on samples from the remaining six locations, although some of them were considered of medium or bad quality. A panel of 580 higher quality SNPs was used for all downstream differentiation analysis. **All samples showed a great similarity between them** (F_{ST} values from 0 to 0.0049) with the highest values,

but not significant after the FDR correction, coming from GSA06c (Spanish coast) (Fig. 1.2a). GSA6c was also divergent from the others in the DAPC plot (Fig. 1.2b); however, this sample performed the worst in the sequencing process and lead to the least number of loci, so its differentiation is mainly due to the amount of data available. STRUCTURE and AMOVA analyses confirmed the **lack of significant differentiation ($F_{CT} < 0$, $P > 0.05$) among all the investigated subareas.**

When these preliminary results were compared to those previously reported in the literature (see D1.2) we can highlight that we did not encounter:

- **the differentiation of the Adriatic Sea (present samples 17a and 17b) from the neighboring regions reported in Maggio et al. (2009) and Matic-Skoko et al. (2018);**
- **the Spanish sample differentiation advanced by Galarza et al. (2009), although the sampling effort was limited, with large areas not covered at all.**

In conclusion, the *M. barbatus* study required further analysis to explore the parameter space to extract robust conclusions on the population structure among the studied populations.

For *P. longirostris*, most of the samples provided high or medium quality DNA extracts, except for locality 06c, where about half of the specimens (23/50) were of poor quality. After filtering, a panel of 1,045 higher quality SNPs was used for all the following population genetics analysis and 226 individuals were included in the final analysis. Pairwise F_{ST} showed a low level of differentiation among samples, with two comparisons involving GSA17ab significant after correction for multiple tests. DAPC and STRUCTURE as well as AMOVA analyses ($F_{CT} = 0.01594$, $P = 0$) pointed out **a significant differentiation not only among GSA17 and the others but also of specimens from the Ionian and Adriatic (GSA18 and GSA19).**

Results of this preliminary analysis:

- **seemed to confirm the existence of significant differentiation among Mediterranean samples, detected with AFLP and mtDNA (Lo Brutto et al. 2013). It was also particularly promising, when considering that Lo Brutto and colleagues (2013) detected the deepest differentiation when comparing Aegean samples, still not included in the pilot study, with other Mediterranean sites.**
- **we judged that a further optimization of the library preparation was needed to provide the highest resolution and the most robust results, given that the pilot study dataset was characterized by a low coverage and high level of missing data. In particular, we decided to modify the size selection step to reduce the total number of loci in the full study (Task 1.5).**

For *A. antennatus*, many of the samples (124 out of 400 tested) provided poor quality DNA extracts, with the majority of specimens from localities 06b and 06c not suitable for ddRAD analysis. ddRAD was performed on 273 individuals from all the 8 localities, but with sample size limited to 15 and 14 specimens (for GSA06b and 06c). After filtering, a panel of 1,253 higher quality SNPs was used for all the population genetics analysis, but only 197 individuals were included in the final analysis. Pairwise F_{ST} showed no differences among samples, similarly to the DAPC, STRUCTURE and AMOVA analyses ($F_{CT} = 0.00097$; $P > 0.05$).

Results of this preliminary analysis:

- **seemed to confirm the very low, if any, genetic differentiation among Mediterranean samples, as already reported using allozymes (Sardà et al. 1998), microsatellites (Cannas et al. 2012b), mtDNA sequences (Fernández et al. 2013, Maggio et al. 2009, Marra et al. 2015; Roldán et al. 2009; Sardà et al. 2010) and AFLP (Lo Brutto et al. 2012).**

- **much higher resolution and more robust results was obtained by further optimization of the ddRAD protocol.** In fact, even more than for deep-water rose shrimp, the pilot study dataset was characterized by a low coverage and a high level of missing data. It was finally improved by introducing a modification of the size selection step to reduce the total number of loci of the catalogue, and therefore increase the average coverage and the number of retained SNPs.

For *N. norvegicus*, the majority of samples provided DNA of high quality in most samples; in only two cases (GSA06b, 06c), approximately half of the specimens were used, and in only one (GSA18a) the DNA quality was very low, and this sample was excluded from the library construction. After several steps, a panel of 1,393 higher quality SNPs was used for all downstream differentiation analysis. In the Norway lobster, samples showed a moderate similarity between them (F_{ST} values ranged from 0.006 to 0.0927 with the highest observed between GSA17, North Adriatic). Similarly, DAPC, STRUCTURE and AMOVA analyses clearly indicated a moderately significant differentiation between the GSA17 sample and all others ($F_{CT}= 0.06934$, $P=0$), therefore in the pilot study the occurrence of at least two distinct genetic clusters can be hypothesized.

With practically no recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, pilot results were the first to show

- **a slight but considerable differentiation of the Adriatic sample.**
- **the Norway lobster pilot study was the best shown for the crustacean species and was expected to provide interesting results without any further optimization.**

Finally, for *A. foliacea*, the majority of samples provided DNA of good quality, except for two subareas: GSA19b and 19c (finally not included in the library preparation and analyses). After several steps, a panel of 3,437 higher quality SNPs was used for all downstream differentiation analysis for only 6 sampling locations. All samples showed a great similarity between them (F_{ST} values from negative to 0.0069). DAPC analysis showed two samples quite distant from the others, i.e., GSA09b (N. Tyrrhenian coast) and 11c (south Sardinia). However, in this case these two samples performed the best in the sequencing process and lead to the greatest number of loci with respect to the others, therefore their divergence is due to the different amount of data available. STRUCTURE (Fig. 1.6c) and AMOVA results did not indicate any significant differentiation between group ($F_{CT}<0$; $P>0,05$).

On the overall, the available results of the pilot study:

- **did not provide strong evidence of structuring in distinct genetic clusters, as results referred to samples from the Central and Western Mediterranean**, where the lack of genetic differentiation among locations had already been reported (Cannas et al 2012a; Marcias et al., 2010; Fernandez et al., 2011a; Fernandez et al., 2013a).

A further optimization of the library preparation took place in order from one hand to substantially decrease the total number of SNPs recovered and on the other hand increase the number of common SNPs for downstream genetic analyses.

A pilot rarefaction analysis was performed in order to investigate if a reduction in the sampling effort (i.e., the number of specimens per sample) might substantially influence the population genetics results. The very preliminary results, obtained for the two species for which the ddRAD provided the most reliable results, showed that the values of heterozygosity and population differentiation statistics (pairwise F_{st}) were very similar when calculated on the full dataset and on reduced datasets (reduction to 50% and 75% of its original sample size).

Detailed results can be found in D1.4 (ANNEX 1.5).

Using the approach described in D1.4, a more comprehensive analysis was performed during Task 1.5, when a larger number of individuals per site was available.

As a summary and based on the provisional results from the pilot study, indications had been given for the future sampling activities in order to have the best quality DNA, in particular:

- **the need to collect the tissues immediately after the capture of the animals (within 1-2 hours from the haul);**
- **to strictly avoid the use of any preservative (chemicals) especially in crustaceans, alternatively, if specimens were frozen and transferred to the laboratories, collectors were advised to dissect and preserve muscle tissues without allowing the samples to thaw.**

3.2.5 Task 1.5. Comprehensive genetic studies at Mediterranean Sea scale

Task leader: Lorenzo Zane (CoNISMa) and Costas Tsigenopoulos (HCMR)

Participants: CoNISMa; HCMR

Due to the delays in Task 1.4 activities, the beginning of Task 1.5 shifted to mid-December 2019.

Four laboratories (three CoNISMa labs, in Padova, Cagliari and Ravenna, and the HCMR lab in Heraklion) were involved in extracting the DNA from all tissues available in the Genetic Hub. The responsibility of the analyses in the full study were in charge of the different partners as described in the Task 1.4.

All the other laboratory procedures (library construction, sequencing) and data analysis in Task 1.5 were performed as described for the pilot studies (Task 1.4) and specimens were genotyped following the double-digest Random Amplified DNA (ddRAD) sequencing.

The Deliverable 1.5 foreseen in Task 1.5 was split in two Deliverables, following the need of reshaping some aspects of the work organization due to the COVID-19 pandemic, giving more time for the laboratory work and data processing. Thus, the two Deliverables under Task 1.5 are the **Deliverable 1.5.1 (ANNEX 1.6)**, which refers to the results of three out of the six target species sampled throughout the Mediterranean Sea, namely *M. merluccius*, *A. antennatus* and *A. foliacea*, and the **Deliverable 1.5.2 (ANNEX 1.7)**, which reports the results of the other 3 species, i.e., *M. barbatus*, *P. longirostris* and *N. norvegicus*.

Based on the knowledge acquired from the pilot genetic study (D1.4, **ANNEX 1.5**), in which we optimized the laboratory protocols and bioinformatic pipelines, we processed the totality of the specimens sampled for each of the three species across the Mediterranean Sea, performing an unprecedented biogeographic analysis, always comparing the results for each species to those of previous genetic studies from literature.

The total number of collected specimens from each species was DNA extracted, and after taken into account their DNA quality, the best specimens were included in double-digest Random Amplified DNA (ddRAD) libraries construction and analyzed (genotyped) on the HiSeq4000 Illumina sequencing platform. In general, for the first batch of species, the ddRADseq protocol performed better in the fish species (*M. merluccius*), while we faced some difficulties for the two crustacean species, for which little genomic information exists and no previous attempts have been reported.

For *A. foliacea*, the complete study was conducted with specimens from 32 localities (GSAs and GSA-subunits) (Fig. 1.8). In total, 1,692 specimens were sampled from the Balearic Islands to the Cypriot and Egyptian waters. Unfortunately, DNA quality was

poor for some samples, which finally led us to exclude 21% of them from library preparations and downstream analyses. Using all individuals genotyped, the species "catalogue" comprised 2,393,590 ddRAD loci, and filtering out samples with low number of stack loci (<4,000) and for SNPs present in at least 70% of the samples, we ran all phylogeographic analyses with a dataset composed of 771 samples (for 30 localities) and 443 higher quality SNPs.

Our results point out an evident lack of genetic differentiation and are generally in agreement with previous studies conducted at smaller geographic scales and less extended sampling points in the Mediterranean Sea.

Therefore, using the ddRADseq approach, which provided lots of hundreds of polymorphic markers from an extensive sampling plan from the West (Balearic Islands) to the East of the Mediterranean Sea (Cyprus and Egypt):

- **we measured very low pairwise population differentiation metrics (F_{st} values), and the absence of clear population structure in the Mediterranean;**
- **the greatest part of the identified genetic variation is attributed to differences among individuals in the populations, and much less among groups;**
- **in hierarchical AMOVA, different alternative scenarios of grouping populations based on Mediterranean basins, were tested. Some scenarios indicated the existence of very weak differentiation (statistically significant in 5 or 3 groups) (F_{CT} values 0.00007 to 0.00082);**
- **the highest values encountered are found significant for the three major groups of the Western, the Central and the Eastern Mediterranean.**

For *A. antennatus*, a total of 1,471 specimens were sampled from 31 sampling locations (Fig. 1.9). Quality of DNA extracts was poor and a total of 1043 individuals were used for ddRAD. After filtering, a total of 1253 SNPs were retained for 886 samples.

- **All samples showed a great similarity between them with very small F_{st} values, though the global F_{st} was statistically significant.**
- **Unsupervised genetic clustering, showed slight differentiation, especially when comparing samples from Western and Eastern Mediterranean.**
- **As for *A. foliacea*, much of the genetic variation was distributed among individuals in the populations, with a slight support for three groups corresponding to Western Mediterranean Sea, Central Mediterranean Sea, and Eastern Mediterranean Sea.**
- **Measuring the genetic differentiation between groups of population samples using F_{ct} metrics, homogeneity was found among samples from Balearic, Tyrrhenian and Ionian Sea, while a more pronounced structure was detected among samples from Adriatic, Aegean, and Levantine seas.**
- **Results confirm a very low genetic differentiation among Mediterranean samples detected in published studies but highlight an underlying significant structure due to the higher power of the markers used.**
- **The best scenario explaining population structure, maximizing F_{ct} value, corresponded to three groups, the first including samples from the Western Mediterranean Sea, the second including samples from the Central Mediterranean Sea, and the third one including samples from the Eastern Mediterranean Sea ($F_{ct}=0.00145$, $P<0.001$). Values of differentiation were extremely small ($F_{ct}= 0.00104$, $P<0.001$) but still significant when grouping samples in six groups based on Mediterranean basins.**

For *M. merluccius* a total of 1,728 samples (from 41 sampling locations) were genotyped (Fig. 1.10). A total of 665 high quality SNPs was retained (for 1,667 samples)

and used for all downstream differentiation analyses. DNA quality was good for most samples and finally about 90% of the samples were used in library preparations.

In brief:

- **Comparisons between samples showed a significant genetic differentiation. The strongest differentiation was between Atlantic and Mediterranean samples, but also within the Mediterranean populations were structured following a West to East pattern.**
- **Analysis identified four groups corresponding to main subdivisions, the first including the sample from the Atlantic Ocean, the second including samples from the Western Mediterranean Sea, the third including samples from the Central Mediterranean Sea, and the fourth one including samples from the Eastern Mediterranean Sea.**
- **In addition, arranging population samples in 8 groups according to their geographic origin from West-to-East, provided highly significant evidence of geographic subdivision.**

Therefore, *M. merluccius* full study provides interesting results and the finding of significant genetic structure within Mediterranean represents a **significant advance respect to previous published studies**, where such differences were detected only using outlier genetic loci. **The differentiation between Western (GSA01 to GSA12), Central (GSA13 to GSA20) and Eastern (GSA22 to GSA27) Mediterranean samples is now fully supported, and additional differences should be taken into account and monitored, particularly in the Central and Eastern parts of the basin.**

Details for the three species above reported are in the D1.5.1 (**ANNEX 1.6**).

A. foliacea - sampling sites analysed

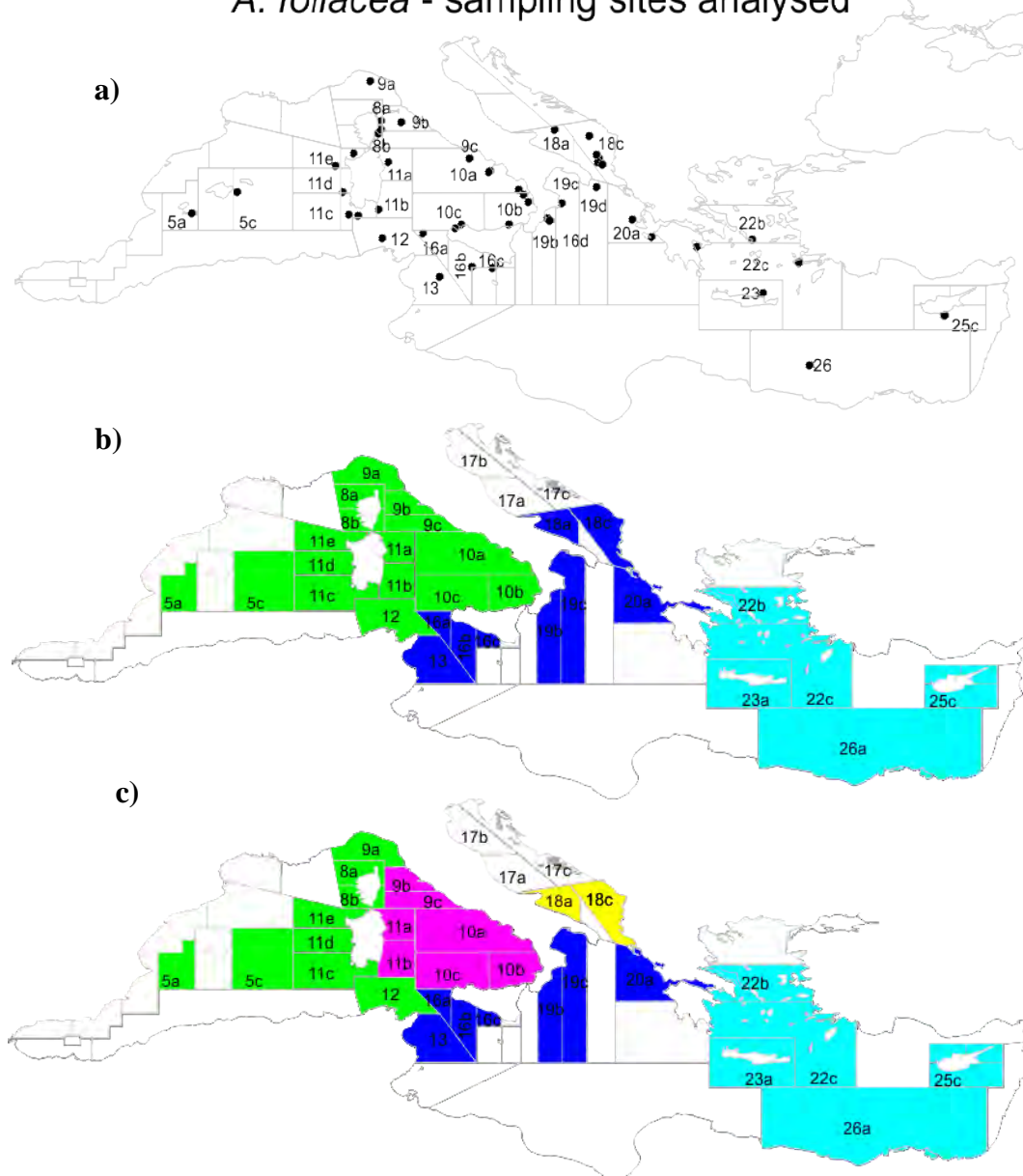


Figure 1.8 a) Sampling sites analysed for *A. foliacea*. b, c) Maps showing with different colors the best scenarios tested with AMOVA: the 3-groups scenario (b) is the one that maximizes the inter-group difference followed by the 5-groups scenario (c). Both scenarios were statistically significant.

A. *antennatus* - sampling sites analysed

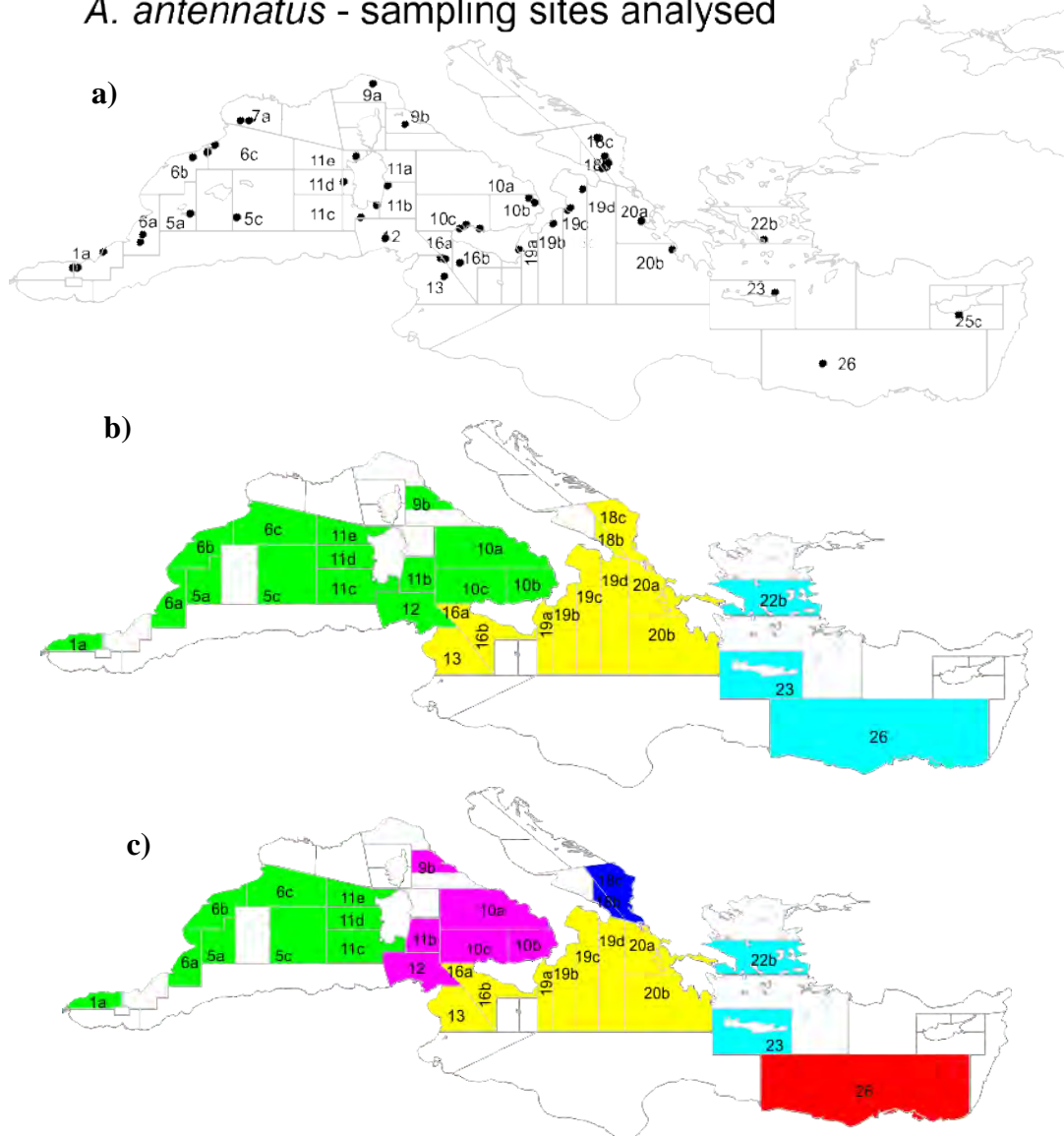


Figure 1.9 a) Sampling sites analysed for *A. antennatus*. b, c) Maps showing with different colors the best scenarios tested with AMOVA: the 3-groups scenario (b) is the one that maximizes the inter-group difference followed by the 6-groups scenario (c). Both scenarios were statistically significant.

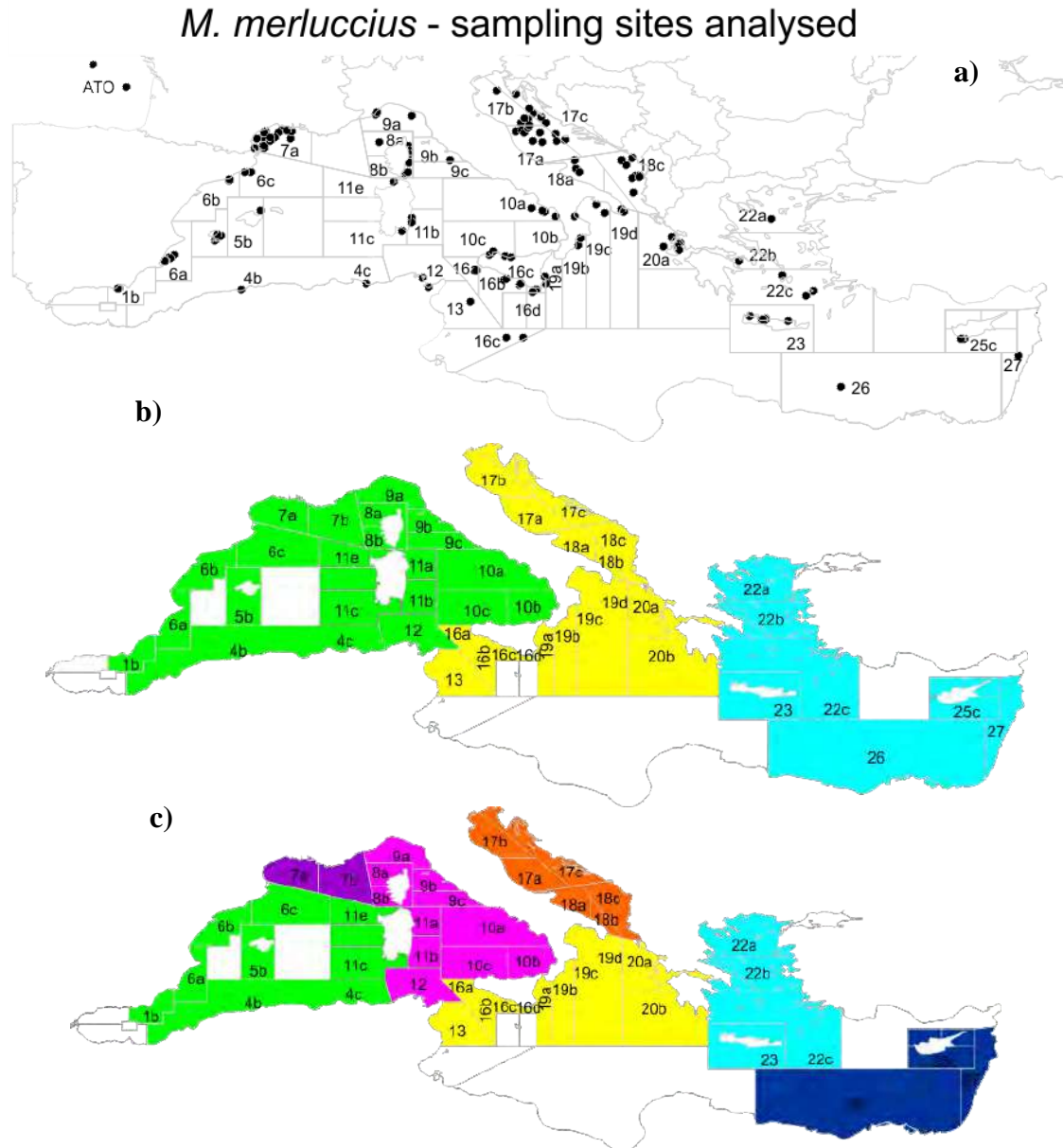


Figure 1.10 a) Sampling sites analysed for *M. merluccius*. b, c) Maps showing with different colors the best scenarios tested with AMOVA (the Atlantic samples is not shown, it always represents a separate cluster): the 4-groups (1 ATO+ 3 Med) scenario (b) is the one that maximizes the inter-group difference followed by the 8-groups (1 ATO+ 7 Med) scenario (c). Both scenarios were statistically significant.

The application of ddRADseq to this comprehensive genetic study of the two aristeids (*A. foliacea* and *A. antennatus*), and European hake (*M. merluccius*) evidenced a much lower level of differentiation of the two crustaceans that is in line with previous studies. This implies that to obtain solid support for these crustaceans a very high number of loci and a big sample is needed.

Our study evidenced two critical steps that can increase the efficiency of the ddRADseq for crustaceans, the first related to DNA quality and the second to the optimization of the protocol for library preparation and sequencing. For DNA quality, a particular care should be taken during sample collection and preservation, to reduce the number of DNA extracts not suitable to ddRAD sequencing. For the optimization of the protocol for library preparation and sequencing we found that particular care should be devoted to

DNA extraction itself, selection of restriction enzymes, size-selection of fragments for high throughput sequencing, level of multiplexing, and/or filtering for bioinformatic analysis. We advise, for future applications of ddRADseq in new species, particularly crustaceans, to perform a small-scale optimization study to address these issues in advance.

The Deliverable 1.5.2 (ANNEX 1.7) comes as a continuation of the results already reported in Deliverable 1.5.1 and finalizes the comprehensive genetic work for the last three out of the six target species. Here, the totality of the locations sampled throughout the Mediterranean Sea was genetically analyzed. The total number of collected specimens from each species was DNA extracted, and after taking into account their DNA quality, the best specimens were included in double-digest Random Amplified DNA (ddRAD) libraries construction and analyzed (genotyped) on the HiSeq4000 Illumina sequencing platform.

Contrary to the deliverable 1.5.1, in which the ddRADseq protocol performed better in the fish species (*M. merluccius*) than in other two crustacean species (giant red shrimp and blue and red shrimp), in this second part of the comprehensive study, we see that the best results were obtained for Norway lobster than for the other two species, that are red mullet and deep-water rose shrimp. For all these three species, little genomic information exists and no previous attempts have been reported.

In *M. barbatus*, the complete study was conducted with specimens from 38 localities and 2,133 specimens sampled from the W. Mediterranean (GSA1b, N. Alboran Sea) to the South-Eastern Mediterranean Sea (GSA25c, 26 and 27) (Fig. 1.11). DNA quality was generally medium, with only one third of the samples providing good DNA quality, whereas another one third was not used since the DNA quality was bad. The latter led us to finally not include into the final sample-set of 1,373 specimens for the ddRAD library preparation, samples from seven GSAs which had low DNA quality. Using all individuals genotyped, the species "catalogue" comprised 462,836 ddRAD loci, and filtering out samples with low number of stack loci (<3,000) and for SNPs present in at least 70% of the samples, we ran all phylogeographic analyses with a dataset composed of 771 samples (for 30 localities) and 853 high quality SNPs.

Our results point out an **evident lack of genetic differentiation** and are generally in agreement with previous studies conducted at smaller and little more extensive sampling points range in the Mediterranean Sea. Therefore, using the ddRADseq approach which provided lots of hundreds of polymorphic markers from an extensive sampling plan from the West (Balearic Islands) to the East of the Mediterranean Sea (Cyprus and Egypt), **we measured:**

- **very low pairwise F_{st} values and the absence of clear population structure in the Mediterranean.**
- **the greatest part of the identified genetic variation was attributed to differences among individuals in the populations, and much less among groups.**
- **the highest values encountered are found significant for the three major groups of the Western, the Central and the Eastern Mediterranean and less for the West to East Mediterranean differentiation; however, this differentiation is explaining only 1.8 and 1.3%, respectively, of the divergence between groups.**

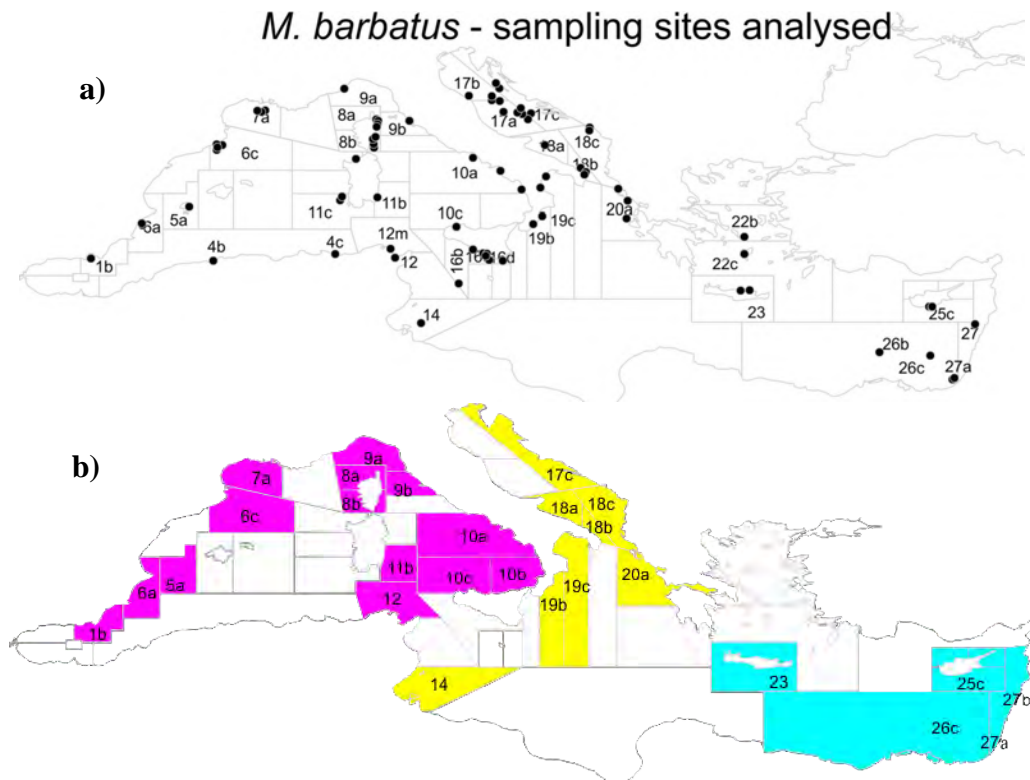


Figure 1.11 a) Sampling sites analysed for *M. barbatulus*. b) Map showing with different colors the best scenario tested with AMOVA that is the one that maximizes the inter-group difference.

In *N. norvegicus*, the comprehensive study included 1,537 specimens from 30 localities sampled from the W. Mediterranean (GSA1b, N. Alboran Sea) to the Aegean Sea (GSA22b) in the Eastern Mediterranean Sea (Fig. 1.12). DNA extraction quality was surprisingly high for this species analyzed in the project (61.1%) and only 18.5% were not used since the DNA quality was bad. Therefore, we finally included 1,152 specimens for the ddRAD library preparation, and unfortunately samples from two GSAs which had low DNA quality were not represented. The species "catalogue" using all individuals genotyped, comprised 1,682,988 ddRAD loci, and filtering out samples with low number of stack loci (<3,000) and for SNPs present in at least 70% of the samples, we ran all phylogeographic analyses with a dataset composed of 890 samples (for 27 localities) and 730 high quality SNPs.

Our results reaffirm the ones we got from the pilot study and showed the following conclusions:

- a significant differentiation of the samples eastern of the Adriatic Sea (GSA17) against the others from the central and Western Mediterranean Sea;
- additionally, when testing for alternative scenarios relatively high and significant values were also encountered for the separation of the Adriatic Sea (GSA17 to 19) from the neighbouring basins to the west (GSA1 to 11) and the east (GSA22).
- with practically few recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, current results are the first to show a considerable differentiation of the Adriatic and Eastern Mediterranean samples.

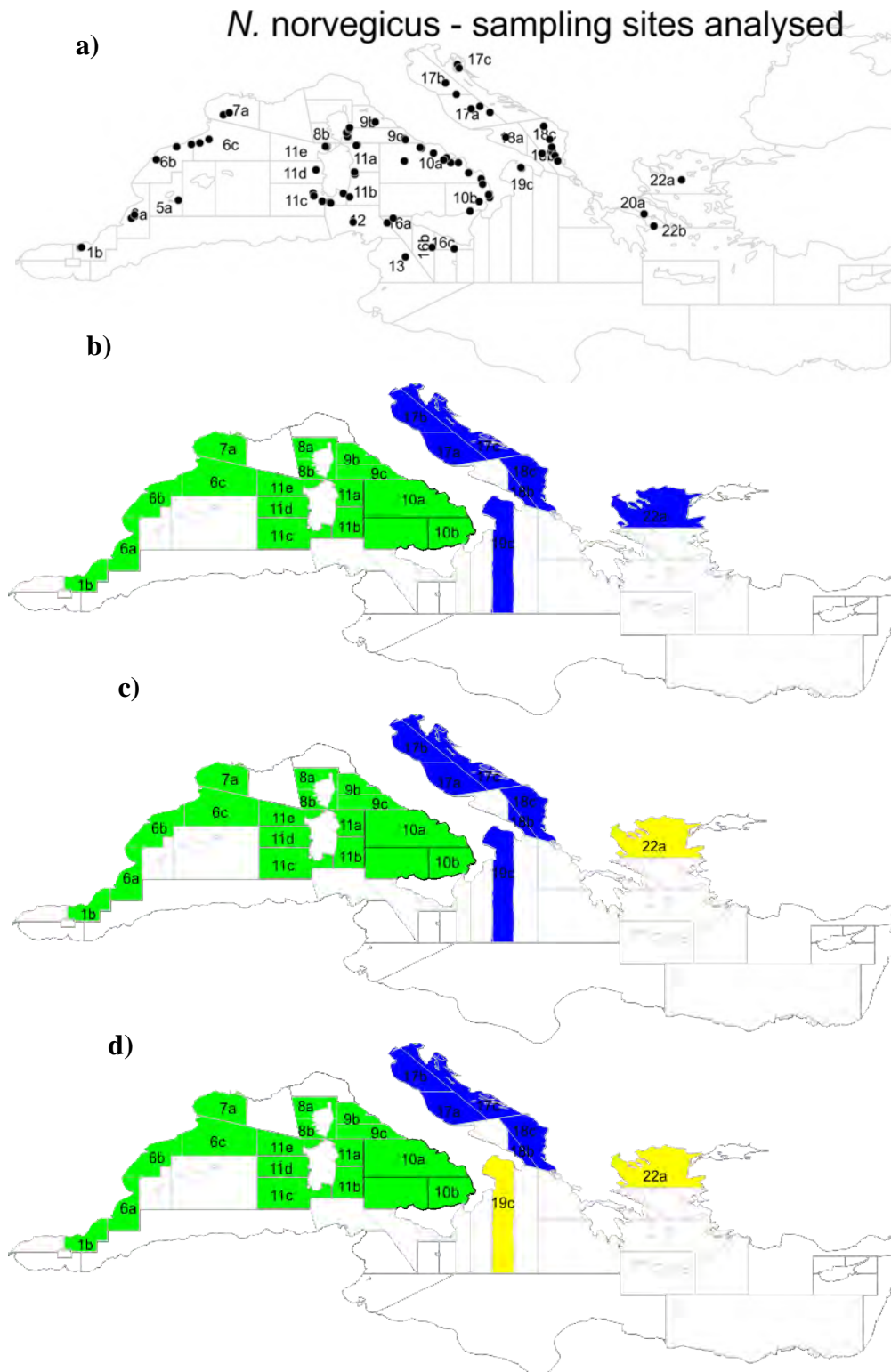


Figure 1.12 a) Sampling sites analysed for *N. norvegicus*. b, c, d) Maps showing with different colors the best scenarios tested with AMOVA (the 2-groups (scenario (b) is the one that maximizes the inter-group difference followed by the 3-groups scenarios (c, d). All scenarios were statistically significant

For *P. longirostris*, a total of 1,750 specimens were sampled from 35 sampling locations (Fig. 1.13). Quality of DNA extracts was poor and a total of 1,008 individuals

from 34 locations were used for ddRAD. After filtering, a total of 1,225 SNPs was retained for 782 individuals.

In brief:

Pairwise F_{st} , a metric of population differentiation among pairs of samples, showed many significant pairwise differentiation, and the global F_{st} value was highly significant.

Genetic clustering methods confirmed the existence of differentiation, indicating potential subdivisions of East Mediterranean samples and a slightly different cluster composition for samples West and East of the Strait of Sicily.

As for other species analysed in this project, much of the genetic variation was distributed among individuals in the populations, but a support for three groups was found: a “western-central” one including samples from Western and Central Mediterranean Sea up to the Strait of Sicily, a “central” one including the remaining samples from the Central Mediterranean Sea except the easternmost Ionian sample, and an “eastern” group that included the samples from Eastern Mediterranean Sea and the remaining Ionian.

The results of the comprehensive study are in line with the results of the only previous study investigating population genetics at the Mediterranean scale, which detected using mitochondrial DNA a deep differentiation between Aegean samples and Tyrrhenian Adriatic ones.

Our results suggest that:

- **the major genetic breakpoint is indeed located in the Strait of Sicily and that a possible further subdivision is present at the boundary between Central and Eastern Mediterranean, though not coinciding with the boundary between Ionian and Aegean Sea.**
- **the existence of significant genetic differentiation should be taken into account in the future and additional investigations are needed particularly in the Sicily Strait area, where deep-water rose shrimp is an important resource shared by different nations being the main target of trawl fisheries.**

The application of ddRADseq to this comprehensive genetic study of the two crustacean species (*P. longirostris* and *N. norvegicus*) and one fish (*M. barbatus*) evidenced different levels of differentiation for the three species: higher in the Norway lobster, which is something novel for the species, medium-high in deep-water rose shrimp, and very low for the red mullet.

Details on the three species above reported are in the D1.5.2 (ANNEX 1.7).

We remind that the previous reported genetic study for the two aristeids (*A. foliacea* and *A. antennatus*) and European hake (*M. merluccius*) also evidenced a much lower level of differentiation for the two crustaceans. This implies that to obtain solid support, especially in species with low genetic diversity and unknown genome size, a very high number of loci and a big sample is needed.

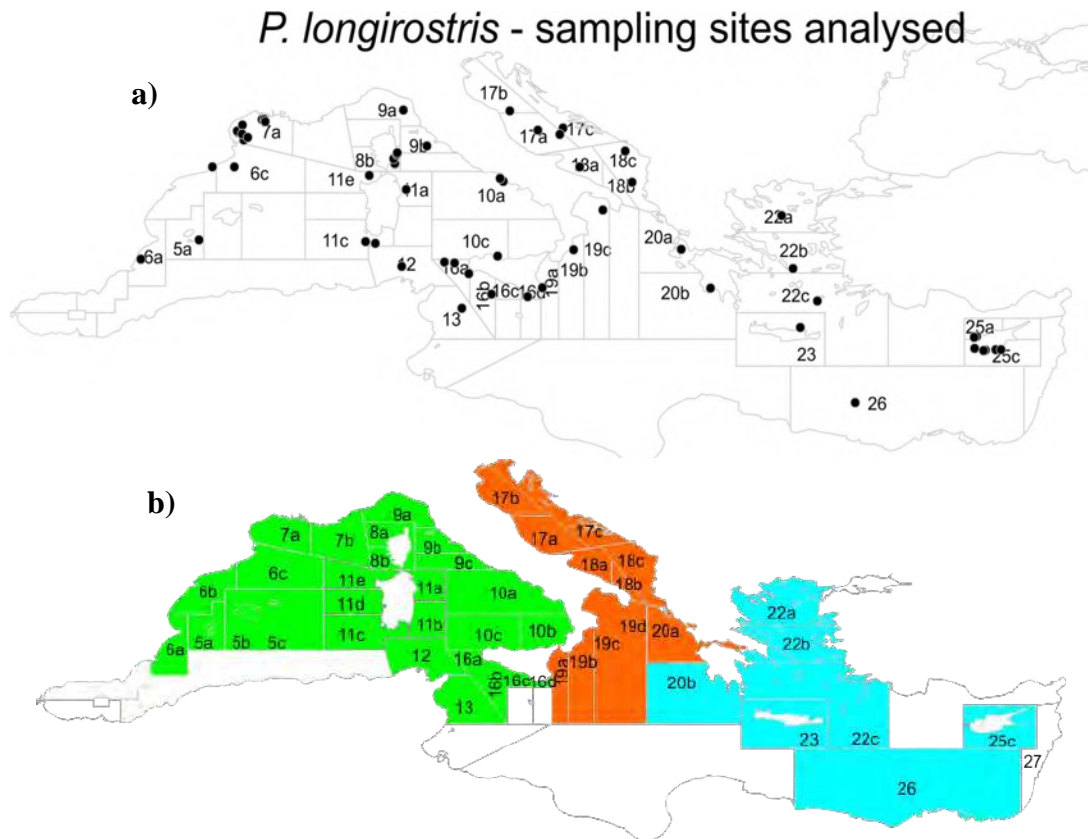


Figure 1.13 a) Sampling sites analysed for *P. longirostris*. b) Map showing with different colors the best scenarios tested with AMOVA: the 3-groups scenario that is the one that maximizes the inter-group difference.

Our study, considering all the six investigated species, highlighted two critical steps that can increase the efficiency of the ddRAD sequencing: the first related to DNA quality, and the second to the optimization of the protocol for library preparation and sequencing.

- For DNA quality, a particular care should be taken during sample collection and preservation, to reduce the number of DNA extracts not suitable to ddRAD sequencing. The quality of DNA was still variable for specimens collected at different sites and, to a lesser extent, for specimens collected at the same site, implying that further optimization is needed.
- For the optimization of the protocol for library preparation and sequencing we found that particular care should be devoted to DNA extraction itself, selection of restriction enzymes, narrower size-selection of fragments for high throughput sequencing, lower level of multiplexing, and/or stricter filtering for bioinformatic analysis.

We advise, for future applications of ddRAD sequencing in new species, particularly crustaceans, to perform a small-scale optimization study to address these issues in advance.

3.2.6 Task 1.6. Deliver a consistent and detailed protocol based on genetic analysis for routine sampling

Task leader: Rita Cannas (CoNISMa)

Participants: CoNISMa, HCMR

Task 1.6 objective was to develop detailed protocol based on genetic data to be used in a routine basis in the future, to help the stock identification. Task 1.6 started as scheduled, but it was completed with the writing of D1.6 following the availability of final results from Task 1.5 and thus delivered on June 2021.

The deliverable **D1.6 - Detailed protocol for routine sampling and genetic monitoring (ANNEX 1.8)** is the output of Task 1.6. It describes the main outcomes of WP1 (Tasks 1.3-1.4-1.5) in terms of samples analysed and data obtained.

In overall, the MED_UNITS sampling was satisfactory for the spatial coverage and number of specimens collected with a total of 10,670 specimens for all six species. Contrarily to the initial expectations, the timely sampling has finally proved to be the most problematic aspect. According to the sampling protocol for genetic analyses, tissues were collected both from freshly caught (28.7%) and frozen specimen (52%). Unfortunately, for a fraction of tissues no indication was given on the sampling conditions and tissue type (19.3%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium quality or even unusable DNAs. All DNAs classified as ‘unusable’ were discarded and were not included in the following step (library preparation for the genotyping). A total of 7,544 individuals were included in libraries, most of them (69.8%) having a ‘good’ DNA quality rating while the remaining nearly one third (30.2%) were ‘medium-quality’ samples.

A total of 5,690 individuals were successfully genotyped (75.4% of those included in the libraries). The genotype success of the ‘good DNA quality’ tissues was high (82%) while only 61% of the ‘medium quality’ DNAs included in the libraries ended up producing genotype data. However, the quality of DNAs obtained by tissue type as well as the genotype success varied considerably by species; therefore, results are described and discussed in detail at the species level.

In brief, for the fish species:

- *M. merluccius* gave the best results. A total of 1,664 individuals were successfully genotyped (97.6% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 98.5% versus 92.9% of ‘medium’ DNAs.
- *M. barbatus* gave the worst results: a total of 771 individuals were successfully genotyped (56.1% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 57.8% versus 54% of ‘medium’ DNAs.

On overall, for the crustacean species:

Similar results were obtained for *A. antennatus* and *P. longirostris* with about 79% of individuals included in libraries successfully genotyped, followed by *N. norvegicus* (72.2%) and *A. foliacea* (60.8%). In detail:

- *A. foliacea*: A total of 770 individuals were successfully genotyped. The genotype success of the ‘good DNA quality’ tissues was 68.4% versus only 53.1% of ‘medium’ DNAs.
- *A. antennatus*: A total of 825 individuals were successfully genotyped. The genotype success of ‘good DNA quality’ tissues was 86.6% versus only 60% of ‘medium quality’ DNAs.
- *N. norvegicus*: A total of 890 individuals were successfully genotyped. The genotype success of ‘good DNA quality’ tissues was 84.2% versus only 50% of ‘medium’ DNA.
- *P. longirostris*: A total of 770 individuals were successfully genotyped. The genotype success of ‘good DNA quality’ tissues was 78.1% versus 71.9% of ‘medium’ DNAs.

Our results confirm what it is known in the literature: the ‘good DNA quality’ tissues produced better results measured as number of genotyped individuals. As the proportion of medium-quality DNA in libraries increased, the whole performance of the sequencing output decreased on overall for all the samples included. This is explained by the fact

that the level of DNA degradation affects the efficiency of reduced representation sequencing, increasing the number of missing data/loci, reducing the number of total reads and SNPs when using a given threshold to finally obtain a solid dataset for measuring population differentiation and stock identity.

Among the main lessons learned in this project we can list:

- **A timely sampling is the crucial step for having reliable genomic data in a short period project.**
- **A clear indication is to rely preferentially on fresh tissues and speed up the sampling and processing procedures.**
- **Larger quantities of samples should be collected in order to have enough samples per area. Extra samples will allow to overcome the unavoidable decrease in numbers due to low DNA quality samples or failures during the experimental steps.**

In more general terms we can conclude that:

- **The sampling is very important. It should be designed in order to fulfil the needs of ‘genomics’ requirements** (in terms of quality of samples, procedures, storage, timing) with special indications for shorter preservation times, or the need to include additional steps that may increase cost and logistical constraints.
- **The scientific surveys at sea (as MEDITS or other similar surveys) are a good opportunity to implement proper sampling design for genomic analyses**, because of the wide geographical coverage, the relatively homogeneous temporal sampling and the sampling locations. **However, the sampling for genomic analyses cannot be a collateral activity** or side-project, opportunistically realized during the standard surveys at sea, and a supplementary task in addition to the multiple activities performed on board by the scientific staff; thus, **adequate resources should be allocated for this activity.**
- If the sampling is performed by multiple groups of scientists operating in different areas, not necessarily involved in the laboratory genetic analyses, **it is highly recommended to set up a central Hub in charge of the coordination of the sampling activities between the samplers in the field and the technicians in the laboratory.**
- **The timeframe for collecting tissues should be carefully defined to address the specific biological problem** (stock identification), taking into consideration the biology (life history cycle) of the species under investigation and the necessity to make meaningful comparisons among populations (areas).
- **In the planning of any project, the timing for the experimental phases cannot be too strict but extensive pilot experiments at the beginning of a project are recommended** to be performed for testing different experimental conditions.
- **The collection of tissues should preferentially be realized from alive/freshly caught individuals and the sampling realized as soon as possible within 1-2 hours from the death of the animal, always kept in optimal conditions (cold temperature).**
- **Crustaceans can be particularly problematic. In this case, it is highly advisable to use only fresh and well-preserved tissues and to allocate the proper time to optimize the molecular protocols for each species under study.** This is especially true for shrimps because their muscle tissues consist of numerous fiber proteins, majorly composed of myofibrillar and sarcoplasmic proteins. Thus, finding the best conditions for digesting these fibrous proteins from the initial step of the DNA extraction method applied may result in higher DNA yields and quality for downstream analytical approaches.

3.2.7 Genetic Workshops

In WP1 two Workshops were foreseen.

- 1st Genetic Workshop, at month 9, September 2019,
- 2nd Genetic Workshop, at month 21, September 2020.

Considering the time required both the first analyses and data processing, the first workshop has been held in November 2019, back-to-back with the 2nd Plenary meeting of the project, as agreed with the Contracting Authority during the 1st Progress Meeting.

The second workshop was delayed as the other activities of Task 1.5, due COVID-19 pandemic, and held on April 2021 back-to-back with the MED_UNITS project meeting.

The objective of the first workshop was to discuss on the markers to choose and how proceed with the subsequent full studies after the completion of the pilot analyses, while the objective of the second one was to discuss on the results obtained and the incoming final report, close the completion of the final studies.

These two meetings have been specifically planned for the representatives of the different institutes involved in WP1 to meet and discuss on the harmonization of protocols, procedures, analyses and results. However, the participation of other scientists from other WPs was welcome, since it could be very useful for the correct interpretation of the genetic results by ‘non-experts’ and their use in combination to the data acquired in the other WPs.

The 1st Med_Units Genetic Workshop was held in Heraklion the 14th of November 2019 and attended by about 30 scientists.

During this first Genetic Workshop the following points were discussed:

- protocol adopted for the tissue sampling, DNA extraction, isolation: results by species, advantages and pitfalls and eventual amendments to be adopted in light of the downstream DNA quality;
- DNA quality assessment and choice of samples to be finally included in library preparation;
- optimization of the library preparation protocol;
- bioinformatic analyses towards the population genomics of the pilot phase: case studies of *M. barbatus* and *M. merluccius*;
- implementation of the MED_UNITS database: specific genetic needs, update and transferring of data to other WPs.

During the First Genetic Workshop it was agreed that the raw sequencing data are temporarily stored in the servers of the partners involved in WP1-bioinformatic tasks until the completion of the full analyses. These data will be made available to the Contracting Authority for permanent storage and future use, if requested.

The Workshop also agreed that the results of the genetic study will regard the identification of populations units with the finest available resolution, given the sampling sites and the markers at disposal. The results on such genetic clusters will be made available to the other WP to integrate the analysis from a biological perspective, if needed. The results on the identification of population units will be also available to WP4 for an integrated thorough analysis. Following the proposal made by the project coordinator during the 1st Genetic Workshop and, it was decided to prepare a common database for storing specific genetic data, in order to make them accessible both internally and to the other WPs. These data are complementary to the ones related to the sampling information already assembled.

It was agreed to insert in the database the following individual data:

- the DNA extraction (quantification of DNA) and quality of DNA (e.g., good, bad, medium, etc., according to the agreed pre-defined categories following specific criteria);

- the library information for each individual (individual included/excluded in libraries, inclusion in library n°);
- data on the genetic profile (individual genotype);
- data on the population genomic results (individual cluster assignment, genetic distances among individuals/sites/clusters).

The **2nd Genetic Workshop** was held virtually on the Zoom platform on the 28th of April 2021, back-to-back with the 3rd project meeting. The meeting was attended by 34 participants to the MED_UNITS project.

The results achieved in the Tasks 1.5.1 and 1.5.2 for the 6 target species of the project were presented and discussed. The workshop also took the commitment to deliver as soon as possible the outcome needed for the project Tasks 3.3 and 4.2, compiling the database at individual level on the SharePoint including the results of the Principal Components Analysis at the individual level.

3.3 WP2 - OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

Chair Otolith shape analysis Kélig Mahé (IFREMER); **chair Otolith microchemistry** Ignacio Catalán (CSIC)

Partners/subcontractors involved: IFREMER, CSIC, COISPA, CIBM, CNR, CoNISMa

3.3.1 Task 2.1. Literature review on otolith shape and microchemistry

Task leaders A. Massaro (CIBM) and I. Catalán/S. Perez (CSIC)

Participants: CIBM, COISPA, CSIC, IFREMER

The literature review aimed at gaining a comprehensive view regarding the results obtained in the Mediterranean basin and adjacent areas on the basis of the used methodology and species. **The literature review (deliverable D2.1 Report on the literature review on the otolith and microchemistry analyses ANNEX 2.1 to this report) examined more than 600 papers** related with otolith shape analysis to obtain a general overview on applications and methodologies of this analytical approach and its potential use. Particular attention was addressed to stock identification studies focusing on *Merluccius* spp. and *Mullus* spp. Otolith shape analysis have applications in different research fields, most of the publications concern shape analysis as tool to distinguish different stocks.

Different approaches have been used to describe and compare the morphology of otoliths, and two main groups can be recognized to describe and compare otoliths shape: univariate methods, using linear measurement, and multivariate methods describing the whole otolith shape from a mathematical point of view. Fourier, Wavelet Transform, Geomorphometric Analysis, Geodesic approach and Curvature Scale Space represent the most used methods to describe the whole otolith shape; while Shape Indices (Roundness, Circularity, Form Factor, Rectangularity, Ellipticity, Aspect Ratio) represents an option to analyze otolith shape differences.

Ten studies were carried out to identify *Merluccius* spp. stocks in different areas. Mediterranean population of *M. merluccius* was evaluated in one study combining a multiple approach analysis (chemical and shape). **With reference to *Mullus barbatus*, five studies were performed.**

In general, otolith shape analysis represents a useful tool in stock identification; moreover, it seems a very low-cost method with high accuracy in the results making this analyses a feasible tool to evaluate the structure of populations at different levels.

The review on otolith microchemistry provided a background for the analysis in the framework of MED_UNITS project and contributed to a better understanding of the methodologies to be applied. The overarching idea behind the review lies in that otolith microchemistry from individual fish, and particularly the dynamics of seven commonly analysed elements (Li, Mg, Mn, Cu, Zn, Sr and Ba), is linked to their concentration in the surrounding water. Ba and Sr are particularly linked to water properties (i.e., salinity, chemical composition) or productivity, and, when analysed in otoliths and coupled to age information, can indicate the environmental history of individual fish. In addition, stable isotope analysis in otoliths (C and O) can be concurrently used to derive additional environmental history records of fish (e.g., temperature).

A search on Web of Science was conducted for data extraction on the published information on otolith chemistry for European hake and red mullet. Information on sampling characteristics, individual fish variables, instrumental procedure, type of analyses, statistical treatment, environmental parameters and reference details were collated. Results were discussed in terms of reliability and usability within the project.

A total of 82 references were found for European hake. From these, a filter based on otoliths and the Mediterranean resulted in 11 manuscripts containing relevant

information for the data review. On the other hand, a total of 30 references were found for red mullet albeit none for otoliths microchemistry; therefore, all microchemistry analyses generated in MED_UNITS will be novel for this species.

The conclusions of the review highlight that, in order to maximize the success of the activities within the project, we must 1) use otoliths from immature fish; 2) follow the procedures established by Morales-Nin et al. (2014) for European hake otoliths; 3) test the procedures more suitable for the red mullet otoliths; 4) analyze otolith cores and otolith edges to obtain the natal and sampling location signatures; 5) cover the widest geographical area inside the Mediterranean to ensure enough differentiation of the populations considered.

Following the results of the review, the rationale for changes in the *M. merluccius* otolith sampling and analysis was pointed out (by B. Morales-Nin & S. Pérez. IMEDEA). Thus, following the discussions of the first plenary meeting held in Rome on March 28th 2019, it was decided to modify the sampling protocol to select immature fish. The total length of first maturity might vary on the different GSAs, therefore the sampling will be adapted to collect immature fish 22-28 cm TL around 1 year of life (Figure 2.1). Because males mature at a size much smaller than females, only females with a total length smaller than 28 cm should be part of the otolith microchemical samples.

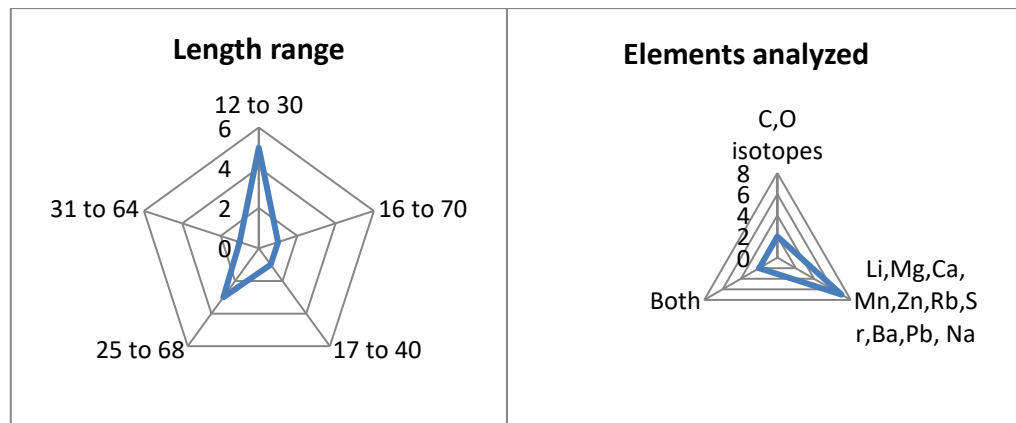


Figure 2.1. European hake: length ranges and trace elements determined by frequency in the previous studies.

Additionally, it was agreed that because the otolith's preparation and microanalysis protocols at IMEDEA do not require clean methods during sampling, European hake otoliths can be removed from the fish using standard routines. It was recommended to use plastic vials for the otoliths storage to avoid breaking during the handling and transport.

Due to the liberation of some funds, IMEDEA increased the number of otoliths to be analysed to 25 otoliths from each sampling location for the pilot genetic study. The foreseen number of otoliths to be analyzed was 200.

Therefore, the Milestone 2.1- the protocol for otoliths sampling has been modified accordingly (**ANNEX 2.2 to this report**) to the above text.

Following the feedback of the responsible for the microchemistry analyses on red mullet (M. Stagoni, UNIBO, CoNISMa) regarding *M. barbatus* there were no specific bibliographic data on the processing of otoliths. These otoliths are very thin and fragile and it is very difficult to obtain intact thin sections by means of a single or double blade cut. Samples were therefore addressed to fish specimens of at least 12 cm in total length to be able to work otoliths sufficiently large and resistant to mounting and polishing

operations. From some tests carried out on different types of assembly and processing the best solution is to glue the otoliths on the sagittal face with the sulcus side up.

A protocol (D.2.3) for the microchemistry analysis was elaborated for both species, analysis were carried out using LA-ICPMS which allows high spatial resolution and precision.

Due to the fragility and thermal susceptibility of otoliths cold-curing epoxy resins with low shrinkage and cyanoacrylic glues were considered to be used to mount the otoliths on glass slide and to minimize cracks.

To reach the core layer polishing took place with aluminium oxide lapping papers from 9 to 1 micron particles size steps on hard flat surface and a final step on microfiber pad impregnated with 0.05 particles.

The sections obtained, after appropriate washing with MilliQ water and drying at low temperature, are ready to be analyzed at LAICPMS transferring the individual sections to an appropriate mount to maximize the number of samples to be analyzed and minimize analysis times.

3.3.2 Task 2.2. Otolith shape analysis of European hake and red mullet

Task leader: K. Mahé (IFREMER) and P. Carbonara (COISPA)

Participants: IFREMER, COISPA

Otolith shape analysis in MED_UNITS aims at contributing to the identification of fish stocks. Otolith shape remains unaffected by the short-term changes in fish condition (Campana and Casselman, 1993) or environmental variations (Campana, 1999). Accordingly, the shape of the otolith has been used as a tool to identify the species, to reconstruct the composition of the diet of predators (fish, seabird, seal, etc.) and to discriminate fish stocks. Since Campana and Casselman (1993), many fishery scientists have developed this type of analysis for stock discrimination studies, as a base for understanding fish population dynamics and achieving reliable assessments for fishery management (Reiss et al., 2009). As a result, more than 90 papers were published from 1993 to 2017 on the identification of marine fish populations or stock structure using otolith shape.

In the milestone *M.2.3 Ready-to apply protocols for otolith shape analysis*, working material on methods and procedures to apply was collected to finalize the deliverable ***D.2.2 Protocol for otolith shape analysis and data-treatment of otoliths of European hake and red mullet (ANNEX 2.3 to this report)***, that reported all the details related to the otolith shape analysis. The *M2.6 Completion of otolith shape analysis* has been delayed, because all otoliths sampled in coordination with WPO were delivered to the IFREMER laboratory by the end of November/early December 2019 and still some otoliths were expected from south Mediterranean countries. The completion of the Deliverable ***D2.4 Report on the results of otolith shape analysis and multivariate analysis of European hake and red mullet*** was thus postponed to October 2020, in order to incorporate the last delivered samples in the analyses. It is the **ANNEX 2.4** to this report.

The used data sets were composed by 1845 adults of red mullets from 37 subunits of geographical subareas and by 1868 juveniles of European hake from 39 subunits of geographical subareas and from 4 ICES areas (Fig 2.2), the latter were included for comparison/contrasting purposes.

In synthesis, after extraction (left and right otoliths) and cleaning, to minimize distortion errors within the normalization process during image analysis, images of the whole left and right sagittal otoliths are scanned under reflected light and stored with high resolution (3200 dpi). Image acquisition process is standardized and automatized to limit the user bias. External Outline information is extracted from TNPC software. With

these data, three methods are used to describe otolith contours: the size parameters (Length; Width; Perimeter; Area); shape indices and the Elliptic Fourier Analysis. To test Elliptical Fourier Descriptors (EFDs) of otolith outline, firstly, Principal Components Analysis (PCA) is applied on EFDs matrix and a subset of the resulting Principal Components (PCs) is then selected as otolith shape descriptors according to the broken stick model. After the pre-processing of EFDs, several mixed-effects models (multivariate for EFDs and univariate for size parameters or shape indices) are fitted with potential factors (or explanatory variables) on the otolith shape (response variable described by size parameters, or shape indices or the PCs matrix for EFDs). These analyses may provide a better understanding of the drivers (and their interaction) which control the otolith shape, as directional asymmetry, ontogenetic effect, environmental effects, and genetic difference and finally the geographical effect.

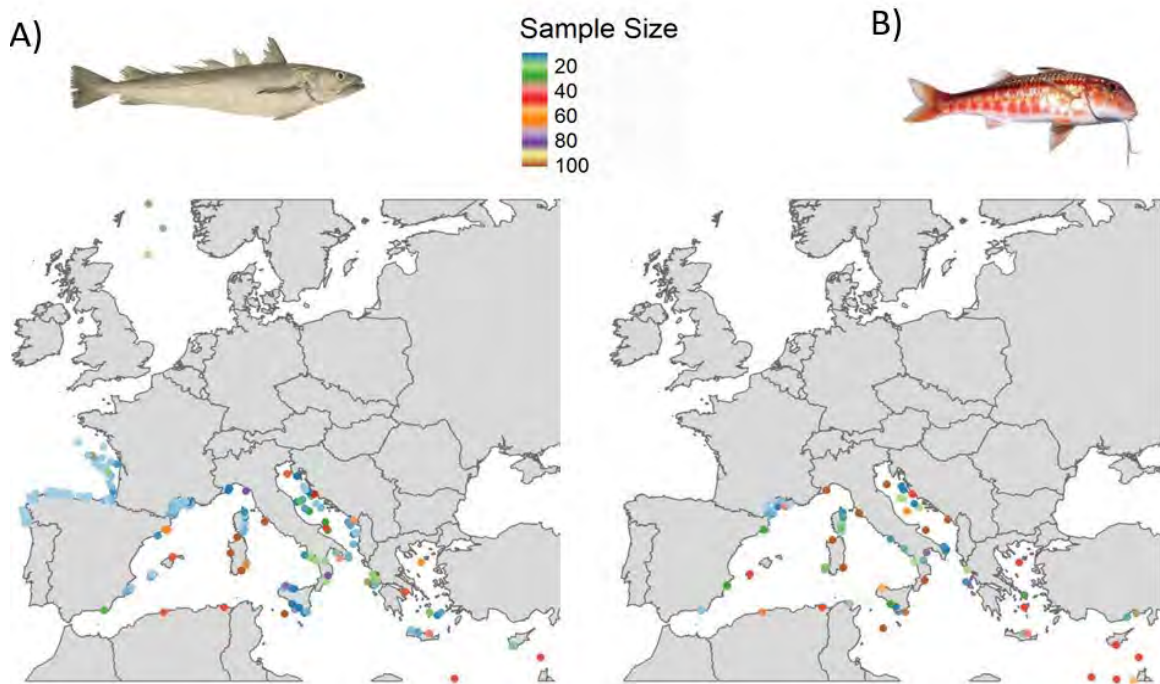


Figure 2.2 Available otoliths number for otolith shape analysis of European hake (A) and red mullet (B).

For each otolith of the study, the first 99 elliptical Fourier harmonics (H) were extracted corresponding to 396 parameters to describe each external shape. Therefore, the first step was to reduce the number of descriptors (Elliptical Fourier Descriptors; EFDs) using the cumulated Fourier Power (F) for each individual otolith and applying Principal Components Analysis (PCA) on EFDs. The maximum number of harmonics (n_k) to reconstruct the otolith shape with a precision of 99.99% was $n_k=29$ for red mullet and $n_k=50$ for European hake. During this step, the outliers corresponding to very characteristic shapes were removed from the analysis. Consequently, the final number of individuals analysed was 1845 for red mullet and 1868 for hake.

Principal Components Analysis (PCA) was applied on EFDs matrix to decrease the number of variables used to describe otolith shape variability, while ensuring the main sources of shape variation. The number of significant Principal Components selected from the broken stick model was 6 for both species. Principal Components Analysis of the first 29 Fourier harmonics for red mullet showed that the first and the second PCs accounted for 33% and 17% of the total variance respectively. For hake, PCA of the first 50 Fourier harmonics showed that the first and the second PCs accounted for 45% and 15% of the total variance respectively.

From PCs, two complementary analyses were performed: the Linear Discriminant Analysis (LDA) with Jackknife prediction (Supervised Machine Learning) and the hierarchical clustering (Unsupervised Machine Learning). Unsupervised Machine Learning (to identify the optimal number of groups) showed 3 groups for red mullet. The second analysis (LDA) is supervised to predict the position of individuals according to their origin geographical area and thus evaluate the percentage of correct classification (predicted area=actual area). Sampling area was used as an explanatory variable in the subsequent LDA. The overall Jackknife classification success was 9.91% and 6.39% at the geographical scale of the GSA and Subunits of GSA, respectively. The analysis showed significant differences among groups of **red mullet** sampled in different areas of Mediterranean Sea (at GSA scale: Wilks' $\lambda = 0.7398$; $\chi^2=551.26$; $p=0.001$; at GSA-Subunit level: Wilks' $\lambda = 0.6421$; $\chi^2=789.69$; $p=0.001$).

By identifying the 3 groups, we tested all possible combinations to optimize the percentage of correct classification (Tab. 2.1). **The optimised classification (correct classification rate = 37.56%) presented 3 groups which are distributed as follows: Western Mediterranean Sea (from GSA1 to GSA16), Adriatic Sea with Central Mediterranean Sea (from GSA17 to GSA20) and Eastern Mediterranean Sea (from GSA22 to GSA27) (Fig. 2.3).**

Table 2.1 Results of Linear Discriminant Analysis according to tested different hypothesis presenting the composition of subunits and the correct classification rate for the red mullet dataset.

Hypotheses	Subregions	GSA Composition	Number	Correct classification rate
1	western med	gsa 1-16	1000	37.56%
	central med	gsa 17-20	398	
	eastern med	gsa 22-27	447	
2	western med	gsa 1-14	828	35.61%
	central med	gsa 16-20	570	
	eastern med	gsa 22-27	447	
3	western med	gsa 1-12	828	33.22%
	adriatic sea	gsa 17-18	246	
	central & eastern med	gsa 14, 16 & 19-27	771	
4	western med	gsa 1-16	1000	35.66%
	adriatic sea	gsa 17-18	246	
	central & eastern med	gsa 19-27	599	
5	western med	gsa 1-16	1000	35.82%
	central med	gsa 17-19	343	
	eastern med	gsa 20-27	502	
6	western med	gsa 1>16	1000	35.77%
	central med	gsa 17>20 without 19b	349	
	central & eastern med	gsa 19b & 22>27	496	
7	western med	gsa 1-16 without 14	954	35.88%
	central med	gsa 17-20	398	
	central & eastern med	gsa 14 & 22-27	493	



Figure 2.3 Identification of boundaries of stocks for red mullet from otolith shape analysis.

Likewise, the analysis was conducted **for European hake**. The results of the hierarchical clustering (Unsupervised Machine Learning) showed an optimized distribution of individuals in 4 clusters that did not seem to indicate clearly geographical patterns. The second analysis was supervised to predict the position of individuals according to their origin geographical area and thus evaluate the percentage of correct classification (predicted area=actual area). Sampling area was used as an explanatory variable in the LDA. The overall Jackknife classification success was 9.26% and 7.01% at GSA and GSA sub-unit level, respectively. By identifying the 4 groups, we tested all possible combinations to optimize the percentage of correct classification (Tab. 2.2). **The optimised classification (correct classification rate = 39.61%) presented 4 groups which are distributed as follows: Atlantic Ocean (from ICES IV to ICES VIII), Western Mediterranean Sea (from GSA1 to GSA13), Adriatic Sea with Central Mediterranean Sea (from GSA16 to GSA20) and Eastern Mediterranean Sea (from GSA22 to GSA27) (Fig. 2.4).**

Table 2.2 Results of Linear Discriminant Analysis according to tested different hypothesis presenting the composition of subunits and the correct classification rate for the European hake dataset.

Hypothesis	Subunit	GSA Composition	Number	Correct classification rate
1	Atlantic	ICES IV-VIII	196	38.54%
	Western Med	GSA1-16	932	
	Adriatic Sea	GSA17-20	393	
	Eastern Med	GSA22-27	347	
2	Atlantic	ICES IV-VIII	196	39.61%
	Western Med	GSA1-13	867	
	Adriatic Sea	GSA16-20	458	
	Eastern Med	GSA22-27	347	
3	Atlantic	ICES IV-VIII	196	39.07%
	Western Med	GSA1-13	951	
	Adriatic Sea	GSA16-22	595	
	Eastern Med	GSA23-27	210	
	Atlantic	ICES IV-VIII	196	

4	Western Med	GSA1-16	932	38.00%
	Adriatic Sea	GSA17-22	530	
	Eastern Med	GSA23-27	210	
5	Atlantic	ICES IV-VIII	196	35.12%
	Western Med	GSA1-12	841	
	Adriatic Sea	GSA13-20	484	
6	Eastern Med	GSA22-27	347	33.46%
	Atlantic	ICES IV-VIII	196	
	Western Med	GSA1-12	841	
	Adriatic Sea	GSA13-22	621	
	Eastern Med	GSA23-27	210	



Figure 2.4 Identification of boundaries of stocks for hake from otolith shape analysis.

Further details on the methods applied and results can be found in the *D2.4 Report on the results of otolith shape analysis and multivariate analysis of European hake and red mullet ANNEX 2.4*.

3.3.3 Task 2.3. European hake and red mullet microchemistry analyses

Task leader: B. Morales/I. Catalán (IMEDEA/CSIC), M. Stagoni (CoNISMa-UNIBO)

Participants: CSIC, CoNISMa

European hake otolith composition has been studied both in the Atlantic and in the Western-Central Mediterranean (Morales-Nin et al. 2005; Swan et al. 2006; Tanner et al., 2012; Morales-Nin et al. 2014). Therefore, some relevant information is available,

although not the same suit of isotopes has been analysed in all studies. In contrast, no previous analyses have been conducted in red mullet.

This task started with the elaboration of common protocols for otolith preparation, processing and analysis for microchemistry trace element detection. The protocols were based on the revision of literature (D.2.1), on previous work on European hake carried out at IMEDEA (UIB/CSIC) and already published (references in D.2.1) and on pilot trials on red mullet carried out at CoNISMa.

Regarding Milestone M.2.5 the two laboratories in charge of the microchemistry analyses (IMEDEA and CoNISMa UNIBO) have finalized the **deliverable D.2.3 - Protocol for analysis and data-treatment of trace elements in otoliths of European hake and red mullet** in due time on January 2020 (**ANNEX 2.5 to this report**).

The otoliths sample collections were optimized in coordination with WPO and WP1. According to Milestone 2.1, the otolith microchemistry study uses the otoliths of *Merluccius merluccius* and *Mullus barbatus* collected for the otolith shape analysis. For each species, the first sexual maturity and the sex could be potential sources of microchemistry variation. Consequently, **only 1 life stage and 1 sex were used by each species. For European hake, they would be juvenile fish (females <28 cm TL and males <16 cm TL); in the case of red mullet, adult fish of both sexes (TL >12 cm). For microchemistry, only females of both species are used.** Based upon the results of the otolith shape study, a subsample of otoliths (at least 25 otolith per target species) were selected for the microchemistry analyses. This sample size has been demonstrated to be suitable in other studies (e.g. Morales-Nin et al. 2005; Morales-Nin et al. 2014).

In several on-line coordination meetings, the geographical sub-areas to be sampled have been discussed.

During the 2nd Plenary meeting a tentative partial list of the sites to be used for European hake microchemistry was proposed: GSA 6b, 11c, 16b, 18a, though it was highlighted that this choice would benefit of the results from the otolith shape and the genetic analyses.

The list of the areas was redefined based upon the origin of the available otoliths, the results of the otolith shape analysis, the preliminary results of the genetic analyses and the water masses circulation at Mediterranean scale.

European hake

For European hake the discussion on the final areas to be sampled focused on the Mediterranean Sea circulation (Fig. 2.5). Additionally, samples from different Atlantic areas are analysed and used in comparative analyses. The sampling areas of both Mediterranean and Atlantic waters are presented in Table 2.3. Apart from this, the microchemistry results of European hake from the NE Atlantic (ICES VIIIc, Galician Coast) carried out by IMEDEA in the framework of the Spanish funded project DREAMER (CTM2015-66676-C2-1-R) are also employed in the comparative analysis.

It was agreed to analyse samples from 11 geographical areas in the case of European hake, including 2 ICES areas, and 10 areas in the case of red mullet (table 2.3).

Thus, the otolith samples received in IMEDEA from IFREMER were processed for laser-ICPMS analysis following the protocol described in **D.2.3 Protocol for analysis and data-treatment of trace elements in otoliths of European hake and red mullet** (**ANNEX A2.5** to this report) first presented in the SpR. In brief, the otoliths are mounted and polished in the sagittal section until ultra-polishing of the surface. The sections are then mounted in random order in petrographic slides for further analysis. The protocol follows ultra-clean methodology.

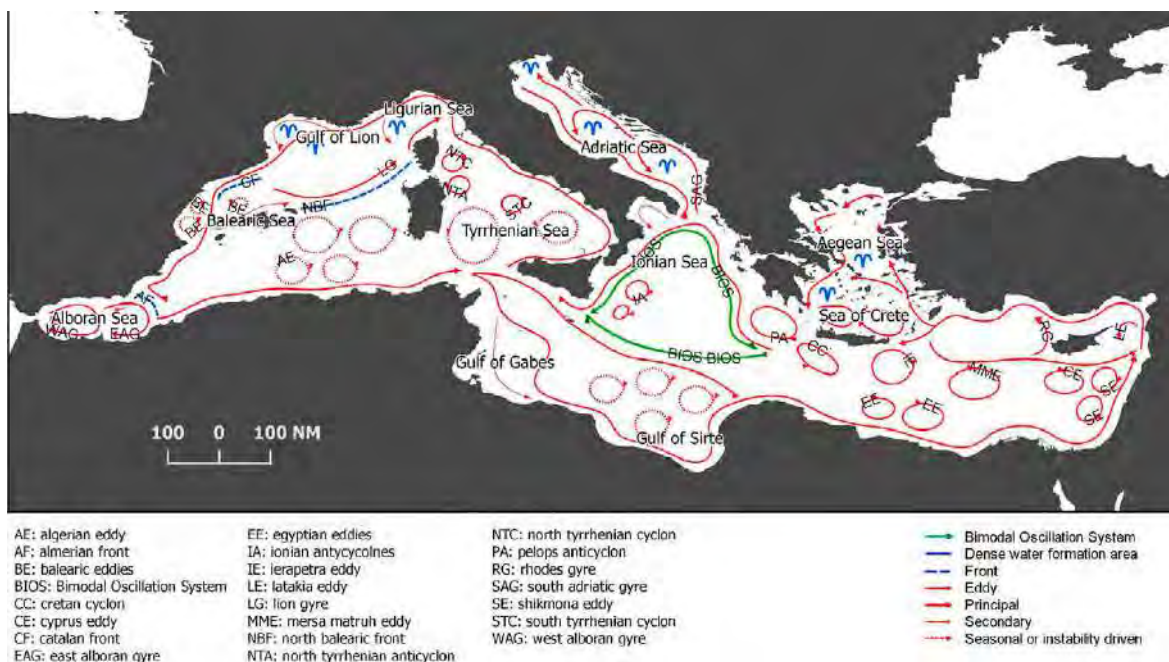


Figure 2.5. Mediterranean Sea circulation

Table 2.3. List of the Mediterranean and Atlantic areas sampled for otolith microchemistry analyses

Species	Mediterranean GSAs	Atlantic ICES
<i>Merluccius merluccius</i>	1b, 9b, 11c, 16b, 18a, 20a, 22b, 25c, 26a, 27b	VIIIa, VIIIc (analysed)
<i>Mullus barbatus</i>	1b, 6a, 9a, 11c, 12m, 17b, 18b, 19b, 22c, 27a.	

Deviation of the planned schedule due to COVID-19 pandemic

The microchemistry analyses of European hake and the related part of the deliverable 2.5 was finalized by January 2021, but the Task duration was extended to June 2021 for completing the analyses related to red mullet that suffered of further delay.

European hake

Spanish laboratories closure and confinement at home has precluded the achievement of the **D.2.5** in June 2020. The on-line work has made feasible to communicate with the project partners and to participate in coordination and discussion on-line meetings. Before the confinement, samples from GSAs 1b, 11c, 16b and 18a, and ICES VIIIa were polished and ready to be mounted to their final petrographic glass slide. However, the rest of the samples (20a, 22b, 26a, and 27b) were not received until 13th May 2020, samples from GSA 25c were received on 24th June 2020 and from GSA 9b arrived to IMEDEA on 30th June 2020. The work at the laboratory facilities necessary for the otolith preparation of these areas could not start until the last week of May 2020 due to pandemic restrictions.

IMEDEA carries out the LA-ICPMS analysis on the Analytical Service at A Coruña University (Galicia), requiring to move from Balearic Islands to Galicia region. Traveling inside Spain and mobility between regions was allowed in early July 2020. So it was envisaged to have all samples processed by the end of July and analysed in early-August 2020. The post-treatment of data, checking quality and conversion to element/Ca concentrations, would require approximately 2 months. Therefore, the part of the

Deliverable 2.5 Report on European hake and red mullet microchemistry analysis (ANNEX 2.6 to this report) related to *M. merluccius* was accomplished according to the project extension, by January 2021.

Mullus barbatus

Italian laboratories closure and confinement at home has precluded the achievement of the **D.2.5** in June 2020. The on-line work has made feasible to communicate with the project partners and to participate in coordination and discussion on-line meetings. Before the confinement, samples from GSAs 9a and 17b were polished ready to be mounted to their final microscope glass slide. However, the rest of the samples (1a, 6a, 11c, 12m, 18b, 19b, 22c, 27a) were not yet received until 9^h June. The work at the laboratory facilities necessary for the otolith preparation of these areas could not start until the second week of June due to pandemic restrictions.

CoNISMa-UNIBO carries out the LA-ICPMS analysis on the CNR - Mass Spectrometry LA-ICP Laboratory hosted by University of Pavia (Italy) However, restrictions imposed by CNR and University rules and other impediments caused by Covid-19 caused further delay, so the analyses and data elaboration were concluded by June 2021.

A first draft of the *Deliverable 2.5. Report on European hake and red mullet microchemistry analysis* was issued on January 2021 reporting the results of the microchemistry analysis of *M. merluccius* otoliths as explained above. The final version of the Deliverable 2.5 (**ANNEX 2.6**) includes the microchemistry analyses for both *M. merluccius* and *M. barbatus* otoliths.

Results

In the case of European hake 279 otoliths were analysed, all of them belonging to 10 Mediterranean management units (hereafter, GSAs), plus two additional areas in the NE Atlantic added for comparison. Red mullet analysis was referred to a subsample of 250 otoliths from 10 different GSAs.

For European hake, a data matrix of concentration (ppm) for 25 isotopes corresponding to the otolith core and otolith edge for each individual was obtained using a methodology developed at IMEDEA and using 6 Certified Reference Materials (CRMs) for isotopic quantification. The data matrix was analysed using classification and linear multivariate models.

In spite that GSAs differed in otolith microchemistry, GSA membership (predict the position of individuals according to their geographical area of origin) has been correctly predicted from only 30% of the fish when using otolith edge data (Figure 2.6). Such a pattern suggests that between-GSA differences in the water mass where a given fish is actually living is not fully reflected in the otolith edge's chemical composition, at least at the spatial scale considered. The lack of links between a GSA and the microchemical composition of the otolith edge precludes to use the core data for safely estimating the natal origin of a given fish or linking specific fishing grounds to nursery areas, at least at the spatial scale considered. The percentage of correct classification increased to 63% when using only Western, Central and Eastern pooled areas but this increase should be interpreted with caution, given the number of sampled subareas.

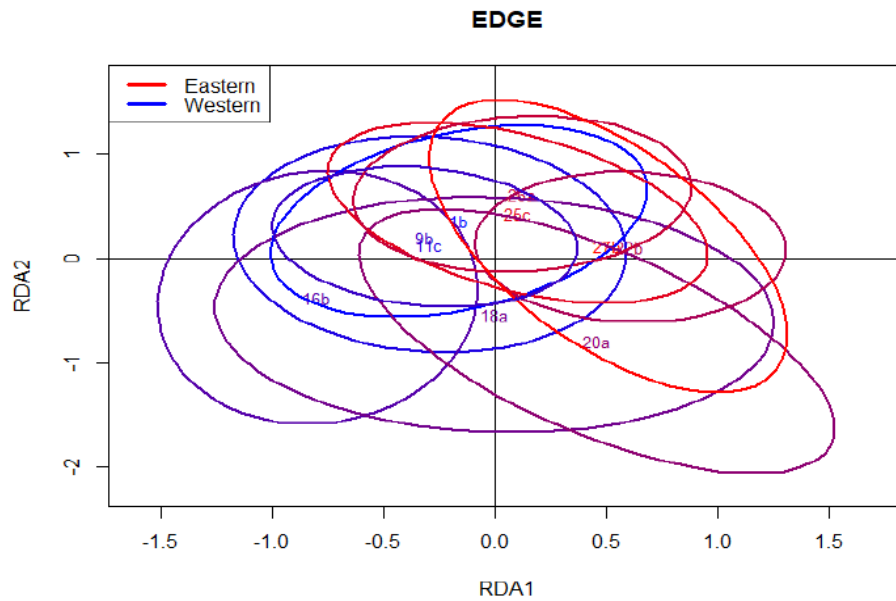


Figure 2.6 Classification of *Merluccius merluccius* otolith edge microcomposition by GSA and Mediterranean area.

Red mullet elemental quantification was obtained for 19 isotopes, with samplings performed at both the core and edge of the otolith. Four (4) CRMs were analysed to obtain isotope concentrations (ppm) using commercial software iolite for data reduction. **Results were analysed using Discriminant Analysis of Principal Components (DAPC) treating core and edge sampling sites independently, in order to assign microchemical information to GSA natal and fishing origin, respectively. Correct allocation was 63% for edges and 66% for cores, albeit a high individual variability that decreased the classification power.**

The results reported here can be explained by at least three compatible hypotheses: (1) otolith microchemistry may only reflect water mass features at another spatial scale, (2) the limits of biological populations may include several management units, and (3) alternative processes related with growth rate may also be affecting the microchemical composition and mask the link between water mass features and otolith composition.

3.3.4 Task 2.4. Compilation of matrices and data interpretation

Task leader: K. Mahé (IFREMER) and M. Palmer (CSIC)

Participants: CSIC, IFREMER, CoNISMa-UNIBO, COISPA

Task 2.4 is fully based on the outcomes of the previous Tasks of WP2 and it was thus affected by the same problems and delay that were described before, in particular these are mainly related to Task 2.3. A first draft of the **Deliverable 2.6, Matrix generation for joint analysis of stock delineation**, including only the analyses related to *M. merluccius* was issued on January 2021. Then it was integrated with the results of *M. barbatus* and the final version (**ANNEX 2.7** to this report) was issued on July 2021.

Otolith markers (i.e., shape and chemical composition) can be used as stock identification methods, including the current delimitation of the Mediterranean management units (named GSA, Geographical Sub-Areas according to GFCM classification). **The hypothesis that otolith chemical composition data and otolith shape data can be combined for improving the capability of predicting GSA**

membership of a given fish from its otolith features has been tested. Concerning **European hake**, data for both shape and chemistry otoliths are available for 159 fish from 10 GSA subunits (1b, 9b, 11c, 16b, 18a, 20a, 22b, 25c, 26a, 27b). Cross-validated correct predictions of population membership inferred from Linear Discriminant Analysis (LDA) when using only shape data reached 34.6%. The same analysis using only the chemical composition of the otolith edge showed 29.9% of correct prediction. Finally, **after merging shape and chemistry data, the percentage of correct prediction of population membership was 42.2%.**

Concerning **red mullet**, data for both shape and chemistry otoliths are available for 237 fish from 10 GSA subunits (1a, 6a, 9a, 11c, 12m, 17b, 18b, 19b, 22c and 27a). Cross-validated correct predictions of population membership inferred from Linear Discriminant Analysis (LDA) when using only shape data reached 23.2%. The same analysis using only the chemical composition of the otolith edge showed 47.2% of correct prediction. Finally, **after merging shape and chemistry data, the percentage of correct prediction of population membership was 44.7%.**

Therefore, with this level of sampling, combining the two sources of data implied a slight improvement of accuracy for hake (Figure 2.7) but a slight decline for red mullet. In any case, all the trials completed showed moderate success and it is necessary to increase the spatial coverage and the total number of individuals to improve the stock identification using otolith shape and chemical composition.

Analysis of otolith shape alone was performed on larger samples for hake (n= 1868 from 39 subunits of GSAs in the Mediterranean Sea and from 4 ICES areas in the Atlantic Ocean) and red mullet (n= 1845 from 37 subunits of geographical subareas). The time and cost of otolith microchemistry analysis precluded to have the same coverage as for otolith shape.

Further details can be found in the Deliverable 2.6.

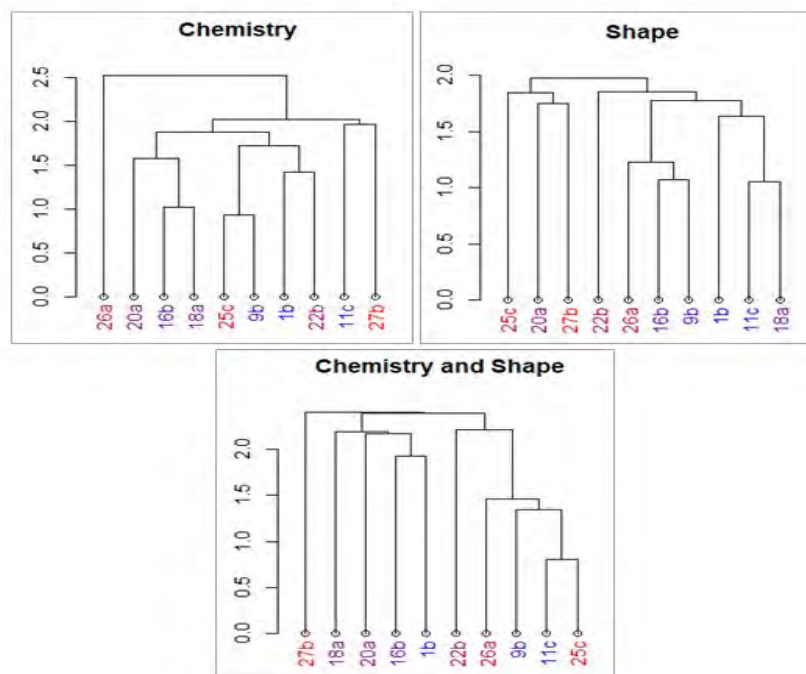


Figure 2.7 *Merluccius merluccius* groupings by GSA according to otolith shape and microchemistry and both indicators together.

3.4. WP3 – DELINEATE FISHING GROUNDS AND STOCK ASSESSMENT

Chairs: Stefanos Kavadas (HCMR); Giuseppe Scarcella (CNR)

Partners involved: CIBM, CoNISMa, CNR, COISPA, HCMR, IEO

WP3 aims at delineating fishing grounds to provide a fishery-based information that is complementary to the ones obtained in WP1 and WP2 (identifying population units) and it is preparatory for the thorough integrated analysis in WP4 (Task 4.2 Integrating results by the different WPs and proposals of new management units).

WP3 is split in 3 tasks:

Task 3.1 Fishing grounds delineation;

Task 3.2 Combine the results on spatial identification of fishing grounds with the spatial distribution of the target species;

Task 3.3 Updated stock assessments, retrospective comparisons, bias estimations and implications for fisheries management advice.

3.4.1 Task 3.1. Fishing grounds delineation

Task leader: Tommaso Russo (CoNISMa – UNITORV)

Participants: HCMR, CoNISMa; COISPA, CNR

The objective of Task 3.1 is to identify and characterize fishing grounds over the Mediterranean waters, including where possible non-EU fleets, which is an aspect of great importance nowadays. Although no explicit and unambiguous definition of “fishing ground” exists, it is largely acknowledged that a fishing ground represents an area where fishing activity is routinely carried out (on a yearly or seasonal scale) as a result of a fisherman strategy aimed at maximizing catches of some target species.

Milestone 3.4 was the basis for the Deliverable *D3.1 Maps on: (a) the fishing grounds in temporal and spatial scale, (b) hot and cold spots* that was finalized in due time, at the beginning of March 2020, just before the lockdown for the Covid_19 pandemic.

The main goal of this deliverable 3.1 Maps on: (a) the fishing grounds in temporal and spatial scale, (b) hot and cold spots (**ANNEX 3.1** to this report) was to estimate the fishing footprint for different fleet segments and countries operating in the Mediterranean Sea, even combining different data sources, processing and modelling approaches, and devoting part of the work to the quantitative comparison of obtained results. A scheme of the workplan of methods applied is summarised in the Table 3.1.

Table 3.1 Scheme of the workplan of methods to be applied in the analyses of the fishing footprint.

Tracking devices	Methods
With AIS/VMS	VMSbase platform (Russo et al. 2014a; Russo et al., 2016, Russo et al., 2019a); The procedure applied in Galdelli et al., 2019, Tasseti et al., 2019, Ferrà et al., 2018
Without AIS/VMS	Multi-Criteria Decision Analysis (MCDA) applied on environmental and fleet data (Kavadas et al., 2015) for both large- and small-scale components of the fleets; Cascaded Multilayer perceptron network applied on environmental and fleet data (Russo et al., 2019b) for the large-scale component of the fleets.

Despite the high significance of small-scale fisheries in the Mediterranean Sea, the vast majority of the small-scale vessels do not meet the criteria to host VMS or AIS technologies onboard. Therefore, there is no actual estimate of their fishing footprint on a spatial scale. Moreover, the coverage of VMS and AIS data for the southern/eastern Mediterranean (non-EU) countries is also not sufficient to provide an estimation of fishing effort for bottom trawlers. Given that primary data on fishing vessels locations (VMS, AIS) were not available for small-scale vessels and bottom trawlers on all spatial scales, a methodological approach based on a Multi-Criteria Decision Analysis (MCDA - Kavadas et al. 2015) has been employed to estimate a fishing pressure index for the small-scale fishery (SSF-FPc) and bottom trawlers (OTB-FPc) in the Mediterranean Sea (OTB-FPc was estimated for non-EU countries and SSF-FPc for both EU and non-EU countries as well).

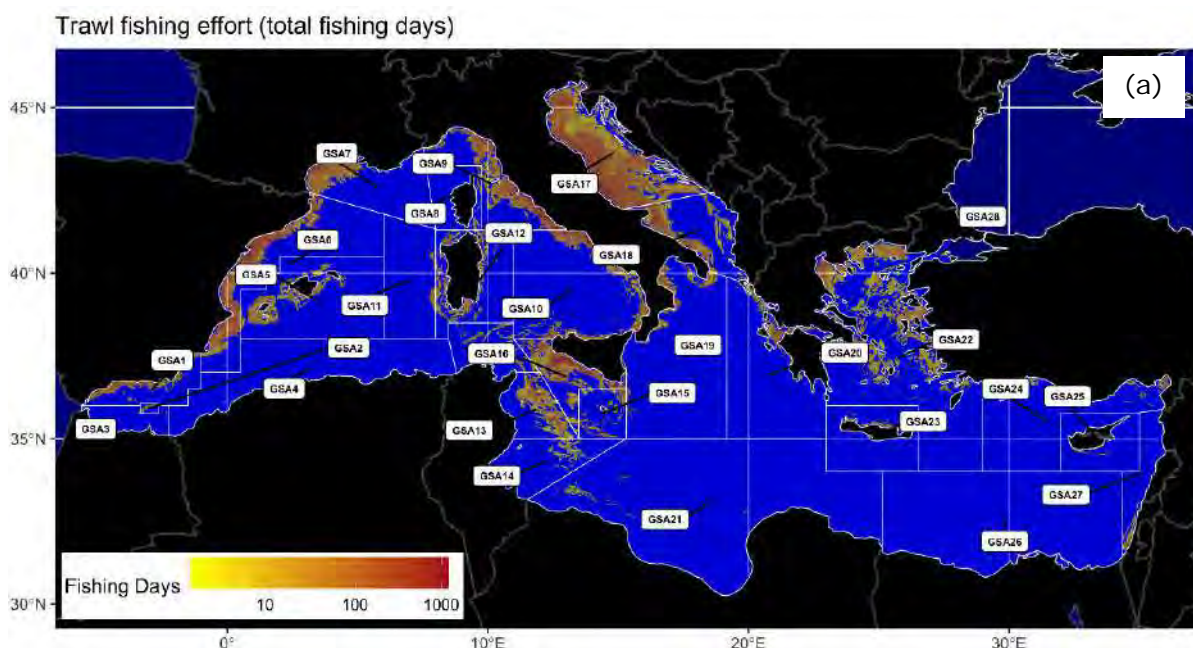
In summary, results reveal both promising findings and key limitations of estimating fishery indicators from AIS data. For monitoring fishing, AIS has some notable limitations: it can be easily turned off and broadcast incorrect identity information, its reception is poor in the Mediterranean southern areas and it is adopted by only a fraction of fishing vessels (above 15 m) that is not evenly distributed between Mediterranean regions making it difficult to compare activity. Despite these limitations, **the use of available, high-frequency and seamless AIS data and performing algorithms, proved useful in identifying and delimiting the more important fishing grounds in the Mediterranean basin and so responding to some of the specific aims of the project.**

The analysis carried out in D3.1 emphasized the need to apply and compare different methods to estimate fishing footprint (ICES 2019).

In order to evaluate potential clusters in the distribution of the fishing activity of each gear and identify statistically significant hot spots, a Hot Spot Analysis was performed using the Getis-Ord G_i^* statistic.

In addition, a modelling approach was applied to reconstruct the fluxes of catches from fishing grounds to harbours using the additional information provided by logbook data or other information about landings (Gerritsen and Lordan 2011; Russo et al. 2018).

A series of maps were produced to visualize the different fishing footprint returned by the application of the different methodologies in order to support a visual inspection and comparison of the outputs. An example is reported in the figure 3.1.



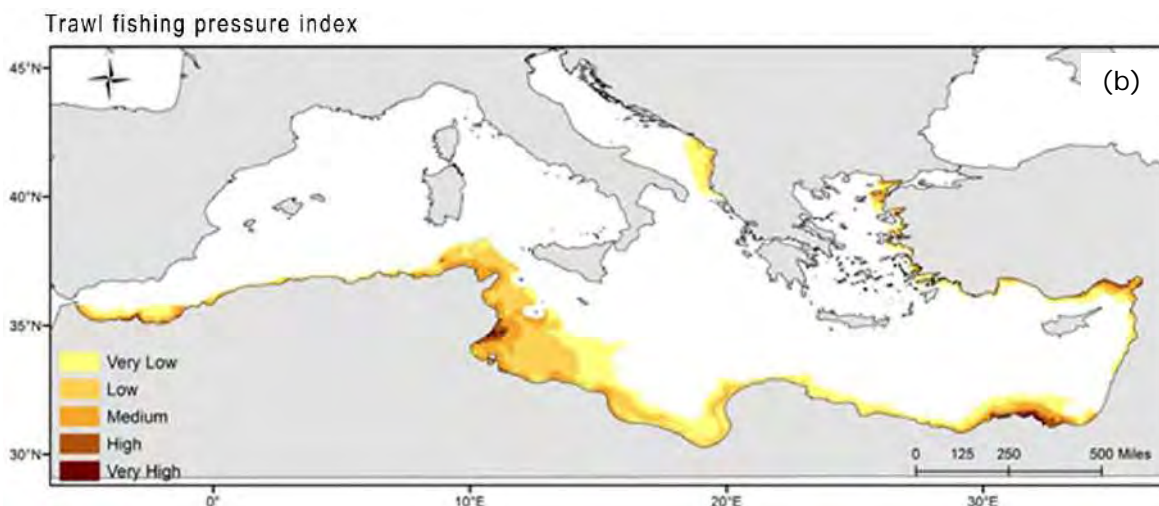


Figure 3.1 Fishing footprint for trawlers from AIS data for the EU Mediterranean countries (fishing days) (a) and estimated by MCDA for the non-EU Mediterranean countries (fishing pressure index) (b).

All details are reported in the ***D3.1 Maps on: (a) the fishing grounds in temporal and spatial scale, (b) hot and cold spots (ANNEX 3.1).***

3.4.2 Task 3.2. Combine the results on spatial identification of fishing grounds with the spatial distribution of the target species

Task leader: Stefanos Kavadas, Irida Maina (HCMR)

Participants: HCMR, CoNISMa, CIBM, COISPA

The aim of Task 3.2 is to synthesize and combine the results of *Task 3.1 Fishing grounds delineation* on the identified fishing pressure with the information from the literature and previous projects on the spatial distribution of six target species studied in the project (European hake, *Merluccius merluccius*; Norway lobster, *Nephrops norvegicus*; deep-water rose shrimp, *Parapenaeus longirostris*; red mullet, *Mullus barbatus*; giant red shrimp, *Aristaeomorpha foliacea*; blue and red shrimp, *Aristeus antennatus*).

The analyses and results are reported in the ***Deliverable 3.2 Maps combining the fishing grounds and the spatial distribution of the target species (ANNEX 3.2*** to this report)

The goal of this synthesis is to delineate the fishing grounds by species or group of species. Those areas are characterized by both fishing activity and species abundance important for fisheries, including the target species and the fishing intensity from different sectors. Based on methods for analysing spatial patterns and mapping clusters, statistically significant hot and cold spots of the combined fishing grounds and species spatial distribution were studied. Aggregated outputs of the species under investigation provided spatial information about the number of exploited species in several fishing grounds.

A methodological framework based on spatial analytical techniques for exploring the patchy distribution of fishing effort and target species was used as a baseline for the investigation and mapping of fishing grounds in the Mediterranean Sea. The approach was based on combining the fishing effort of a certain fishery (bottom trawl, longlines, gillnets and trammel nets) and the potential habitat use of the investigated species. The

steps followed for identifying and analyzing fishing grounds were: (i) estimation of the potential fishing grounds by species, (ii) Hot spot analysis and (iii) the production of aggregated hot and cold spots for the studied species.

The steps for identifying and analyzing fishing grounds are summarized in the flowchart of Figure 3.2. Several datasets on fishing effort and species spatial distribution were used, while the spatial data on species distribution were mainly based on previous studies performed in EU Mediterranean region.

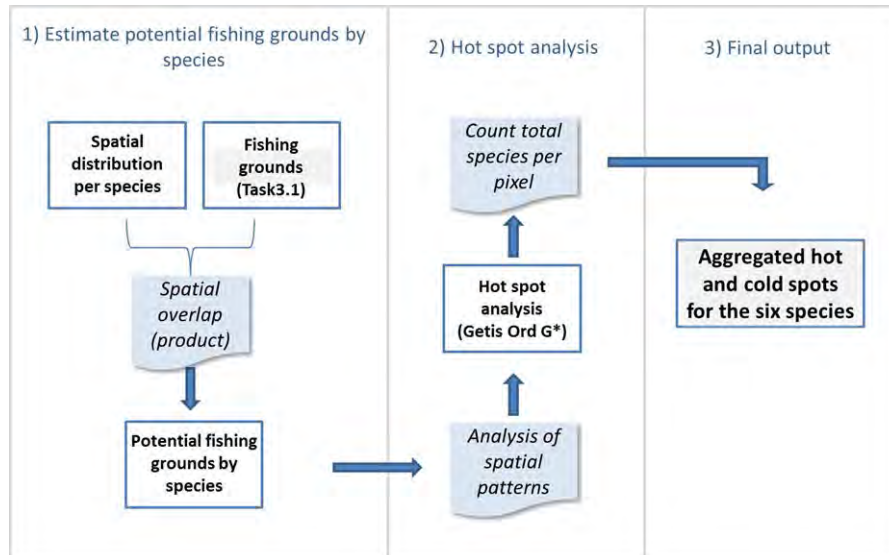


Figure 3.2. Flowchart of the methodological steps applied to identify and describe fishing grounds.

Fishing effort for bottom trawlers, longlines and static nets with length category > 12 m, derived from Task 3.1, were based on the analysis of AIS data. In addition, a fishing pressure index for small scale fisheries <12 m (gillnets and trammel nets) and bottom trawl (particularly estimated for the south Mediterranean fishing fleets) were calculated based on a Multi-Criteria Decision Analysis methodology applied on environmental and fleet data in Task 3.1. The above datasets were used as input to the current analysis process. Spatial data of species distribution were also used and based on review of scientific literature (e.g., papers from the Monograph: Mediterranean demersal resources and ecosystems: 25 years of MEDITS trawl surveys; M.T. Spedicato, G. Tserpes, B. Mérigot and E. Massutí (eds). 2019) and historical scientific projects (e.g., STOCKMED and MEDISEH). The estimations were performed for the entire Mediterranean basin, aiming to provide a more comprehensive view of the fishing grounds by target species.

In order to determine the potential fishing grounds by species, the spatial overlap (as a product) of two types of information was used: (i) the spatial distribution by species and (ii) the fishing effort. The produced outcomes, demonstrate the possibility for fishers to harvest a certain species in a specific area.

To perform hot spot analysis (estimation of hot and cold spots by species and métier), spatial clustering techniques based on the potential fishing grounds of each species were used. **Maps of hot and cold spots for each species** were based on the Getis-Ord G_i^* statistic. An example related to *M. merluccius* is reported in Figure 3.3.

Finally, the aggregated hot spots showed the areas over the Mediterranean with the highest number of commercial species for bottom trawling which are found in the Adriatic, Tyrrhenian, Strait of Sicily, Aegean, eastern Ionian, Balearic, Alboran, Libyan and Levantine Sea (Fig. 3.4).

Details are reported in the **Deliverable 3.2 Maps combining the fishing grounds and the spatial distribution of the target species (ANNEX A3.2)**.

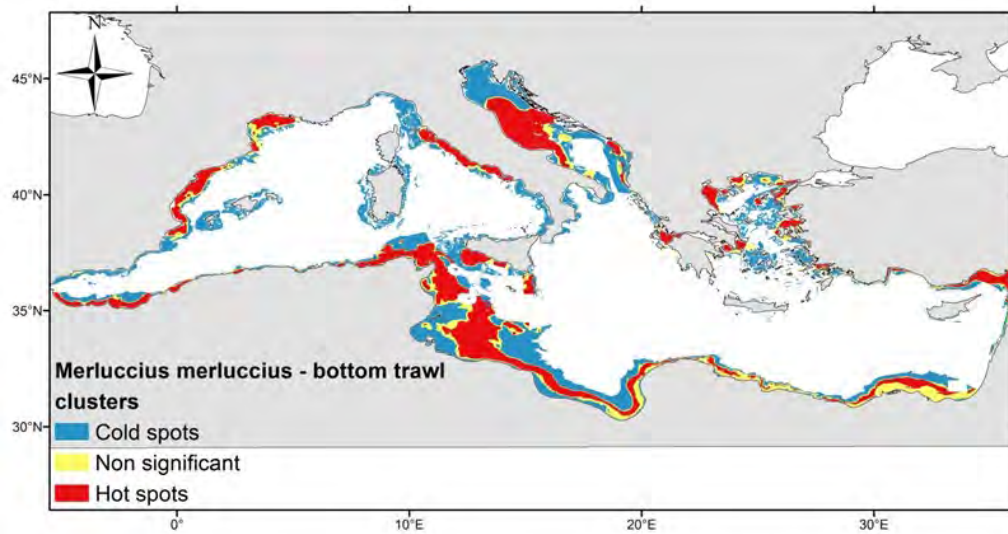


Figure 3.3. Estimated hot and cold spots of the *Merluccius merluccius* for bottom trawling. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate the absence of fishing activity or absence of species.

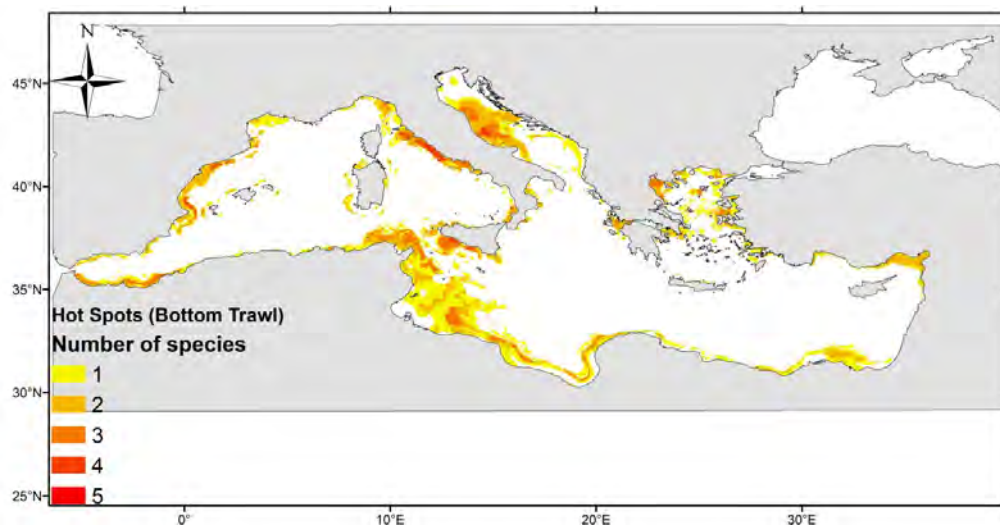


Figure 3.4. Bottom trawler fishing grounds - aggregated hot spots for the six species under investigation. The map indicates the number of species for which a specific cell has been found to be a hot spot.

3.4.3 Task 3.3 Perform stock assessment

Task leader: Giuseppe Scarcella (CNR); Beatriz Guijarro (IEO)

Participants: CIBM, COISPA, CoNISMa, CNR, HCMR, IEO

This task has started on month 13, January 2020 working to the M3.6 Review of stock status of target species and their current stock configuration. A thorough review of the available assessments in the last years was performed. The review was accomplished

considering Reference Year, Assessment Method, Current values of Fishing Mortality, Reference points, the ratio F_{curr}/F_{unique} ; Stock status (exploitation and biomass level), Source (e.g., GFCM, STECF). 20 assessments were reviewed for European hake, since 2017; 15 for red mullet from 2018; 12 for Norway lobster from 2015; 9 for deep water rose shrimp since 2018; 10 for giant red shrimp from 2015 and 20 for the blue and red shrimp since 2015. For species as European hake and deep water rose shrimp the use of combined GSAs in the stock assessment is a more frequent approach, while the rest of the species, in particular red mullet and Norway lobster, are usually assessed by single GSA.

It was foreseen to conduct about 12 assessments, tailored in liaison with DG MARE and CINEA. The assessments and relevant discussion were foreseen in a dedicated workshop, but this was not possible due to the COVID_19 pandemic and the work was conducted virtually.

The main objective of this task is to perform stock assessments taking into consideration the outputs of WP1 and WP2. The present deliverable has been delayed to July 2021 due to the late availability of WP1 and WP2 outcomes and COVID-19 impacts.

Delineating the stock structure of harvested resources is a major pre-requisite for the understanding of their population dynamics, conducting reliable stock assessment, and providing robust scientific advice for an effective management. Nonetheless, stock structure uncertainty is pervasive in fisheries science. This problem is partly due to the intrinsically multifaceted concept of a stock as an ecological unit, and because the definition of stock can vary according to the objective of scientific research (Cadrin et al. 2005; Hammer and Zimmermann, 2005).

The ***Deliverable 3.3 Report on the stock assessments with the new stock configurations for the 6 target species of the study (ANNEX 3.3*** to this report) summarizes and evaluates the new stock configurations explored applying the stock assessment methodologies routinely used for the six target stocks at single GSA or combination of GSAs.

The stock assessments are carried out in six case studies species, based on the updated list of stock unit configurations for the examined target species as emanating from the present study, and results are compared with previous assessments. A final number of thirteen new assessments had been carried out tailored in liaison with CINEA and DGMARE.

It is important to stress that these new assessments are not going to provide any scientific advice in term of status of the stocks compared with relevant reference points. Notwithstanding, the new assessments can represent a useful tool to explore the new stock configurations indicated by WP1 and WP2. The methodologies for each assessment were selected on a case-by-case basis depending on the available data and taking also into consideration, when possible, the outputs of task 3.1 regarding the delineation of fishing grounds by gear and fleet and their link with the spatial distribution of the target species (task 3.2).

In each case study, possible sources of bias in previous estimations of stock status were investigated and identified, together with the potential improvements brought by the new management units proposed by the MED_UNITS.

The results of WP1 and WP2 are schematized in Table 3.2 in term of GSAs distribution of each stock. Taking into account the data availability and in liaison with CINEA/DGMARE, the assessments that have been carried out are summarized in Table 3.3.

The methodologies used for the new stock assessments are the same employed in 2019-2021 joint benchmarks of the scientific bodies in the Mediterranean (GFCM/SAC and STECF) for selected species and areas.

A set of analysis were implemented, when possible, to evaluate if the stock

configurations envisaged in MED_UNITS and presented in Table 4.2, improved the fit of data to model assumption. Different aspects of the assessment process were explored through a step procedure investigating: Cohort consistency, Stock trajectories, Model Diagnostics.

The outcomes of the new analyses are summarized in Table 3.4.

The results of the analyses show that in most of the cases, except for the new configurations of the red shrimps stocks (*Aristaeomorpha foliacea* and *Aristeus antennatus*) assessed in western and central Mediterranean, Norway lobster in the western Mediterranean and deep-water rose shrimp in the GSAs 17-20, the new stock assessments do not present particular improvements of diagnostics; this lack of progresses can be due to several reasons (increased data heterogeneity when the number of aggregated GSAs is increased, model settings, etc...), apart from the new stock configurations. It must be acknowledged that the trials attempted under Task 3.3 of the MED_UNITS project represent a first approach to the assessment of the new stock configurations, and further investigation shall be implemented before scientific advice can be provided in a reliable and robust way.

More details can be found in the ***Deliverable 3.3 Report on the stock assessments with the new stock configurations for the 6 target species of the study (ANNEX 3.3)***.

Table 3.2. Distribution of the target stocks in the GSAs according to the outputs of the Genetic (G) and Otolith (O) studies (WP1 and WP2).

Sea-Basin	W	W	W	W	W	W	W	W	W	W	W	W/C	W/C	W/C	C	C	C	C	E	E	E	E	E	E	
<i>M. merluccius G</i>	1	4	5	6	7	8	9	10	11	12		13			16	17	18	19	20		22	23	25	26	27
<i>M. merluccius O</i>	1	4	5	6	7	8	9	10	11	12		13			16	17	18	19	20		22	23	25	26	27
<i>M. barbatus G</i>	1		5	6	7	8	9	10	11	12			14		17	18	19	20			23	25	26	27	
<i>M. barbatus O</i>	1		5	6	7	8	9	10	11	12				16	17	18	19	20			23	25	26	27	
<i>A. foliacea</i>			5			8	9	10	11	12		13			16		18	19	20		22	23	25	26	
<i>A. antennatus</i>	1		5	6			9	10	11	12		13			16		18	19	20		22	23		26	
<i>P. longirostris</i>			5	6	7	8	9	10	11	12		13			16	17	18	19	20a	20b	22	23	25	26	
<i>N. norvegicus</i>	1			6	7	8	9	10	11						17	18	19			22					

Table 3.3. Stock configurations of the assessments carried out and methodologies employed.

N.	Species	GSAs combination	Method
1	<i>M. merluccius</i>	GSAs 1 - 12	Age based (a4a)
2		GSAs 22 - 27	Production model (SPiCt)
3	<i>M. barbatus</i>	GSAs 1, 5 - 16	Age based (a4a)
4		GSAs 17-20	Age based (a4a)
5		GSAs 23-25	Age based (a4a)
6	<i>P. longirostris</i>	GSAs 1, 5-16	Age based (a4a)
7		GSAs 17-20a	Age based (a4a)
8	<i>N. norvegicus</i>	GSAs 1, 5, 6, 9 – 11	Age based (a4a)
9		GSAs 17-19	Production model (SPiCt)
10	<i>A. foliacea</i>	GSAs 5, 8-12	Age based (a4a)

11		GSA 13, 16, 18-20	Age based (a4a)
12	<i>A. antennatus</i>	GSA 1, 5, 6, 9-12	Age based (a4a)
13		GSA 15, 16, 18-20	Age based (a4a)

Table 3.4. Summary of the stock assessments' outputs and comparison with the most recent assessments at GFCM and STECF (last column).

N.	Species	GSAs combination	Method	Status according to Med_Units configuration	Status according to GFCM and STECF most recent assessments
1	<i>M. merluccius</i>	GSAs 1 - 12	Age based (a4a)	In overfishing	GSAs 1-7 in overfishing GSAs 8-11 in overfishing GSAs 12-16 in overfishing
2	<i>M. merluccius</i>	GSAs 22 - 27	Production model (SPiCt)	In overfishing	-
3	<i>M. barbatus</i>	GSAs 1, 5 - 16	Age based (a4a)	In overfishing	GSA 1 in overfishing GSA 6 in overfishing GSA 7 in overfishing GSA 9 in overfishing GSA 10 in overfishing
4	<i>M. barbatus</i>	GSAs 17-20	Age based (a4a)	Sustainably exploited	GSAs 17-18 in overfishing GSA 20 in overfishing
5	<i>M. barbatus</i>	GSAs 23-25	Age based (a4a)	In overfishing	GSA 22 sustainably exploited GSA 25 in overfishing
6	<i>P. longirostris</i>	GSAs 1, 5-16	Age based (a4a)	In overfishing	GSA 1 in overfishing GSA 5 in overfishing GSA 6 in overfishing GSA 9-11 in overfishing GSA 12-16 in overfishing
7	<i>P. longirostris</i>	GSAs 17-20a	Age based (a4a)	In overfishing	GSA 17-19 in overfishing
8	<i>N. norvegicus</i>	GSAs 1, 5, 6, 9 – 11	Age based (a4a)	In overfishing	GSA 6 in overfishing GSA 9 sustainably exploited
9	<i>N. norvegicus</i>	GSAs 17-19	Production model (SPiCt)	In overfishing	GSA 17-18 in overfishing
10	<i>A. foliacea</i>	GSAs 5, 8-12	Age based (a4a)	In overfishing	GSA 9-12 in overfishing
11	<i>A. foliacea</i>	GSAs 13, 16, 18-20	Age based (a4a)	Sustainably exploited	GSA 18-19 in overfishing
12	<i>A. antennatus</i>	GSAs 1, 5, 6, 9-12	Age based (a4a)	In overfishing	GSA 1 in overfishing GSA 5 in overfishing GSA 6 in overfishing GSA 9-11 in overfishing
13	<i>A. antennatus</i>	GSAs 15, 16, 18-20	Age based (a4a)	In overfishing	-

3.5. WP4 – SYNTHESIS AND PROPOSALS

Chairs: Maria Teresa Spedicato (COISPA); George Tserpes (HCMR)

Partners involved: COISPA, HCMR, CIBM, OGS, CoNISMa– UNITORV, CNR

The activities of the WP4 were affected by some delay as cascade effects of the postponement of the preceding activities related to WP1, partly WP2 (microchemistry of red mullet) and Task 3.3. The completion of the activities of Task 4.2 (Deliverable D4.3), and then the finalization of those related to the deliverable 4.4 were achieved by August 2021.

WP4 aims at synthesizing and combining the results obtained in the WP1-3, in order to provide an integrated view at spatial scale. Information from the different WPs regarding population units (WP1-2) and fishing grounds (WP3) are synthesised to understand, from one side the possible strengths and weakness of the approaches applied, in the perspective of regular data gathering, in support of fish stocks’ monitoring and assessment. From the other side, the objective was to explore the possible explanatory role of environmental profiles at spatial scale in delineating the population structures that emerged from the genetic and otolith analyses. Combining the information on the fishing grounds with those on the stock units would support the identification of spatial units for fisheries management.

WP4 is split in 3 tasks with 4 Deliverables.

3.5.1 Task 4.1 Perform a SWOT analysis

Task leader: Paolo Sartor (CIBM)

Participants: CIBM, COISPA

The aim of Task 4.1 was to evaluate the advantages/disadvantages of the methods applied within Work Packages *WP1-Population genetics*, *WP2-Otolith shape and microchemistry analysis*, and *WP3-Identification of fishing grounds*, to assist in the identification of stocks units and boundaries. The outcomes (**Deliverable 4.1 Report with results of SWOT analysis**) are here reported.

In the task 4.1 a **SWOT analysis**, a method used in several decision-making situation, **was applied evaluating the internal** (Strengths/Weaknesses) **and external** (Opportunities/Threats) **factors related to the methodologies applied and their possible implementation on a routine basis**. For the purposes of this analysis, Strengths were considered as the positive internal factors, while Weaknesses were the negative internal elements of the methodologies/approaches applied. Opportunities and Threats are in relation to a possible adoption of the methodology in a wider and more regular context, to assist in the identification and delineation of stock units. As such, Opportunities represent the prospects and potentialities of that methodology, taking advantage of existing strengths and considering the weaknesses in the context. Threats are obstacles or limitations, which can prevent or impede a wider or more regular use of the methodology.

The analysis was supported by the identification of three specific criteria, namely Structural aspects; Operational aspects; Outcomes and a set of qualitative indicators. The association between indicators and specific criteria is reported in Table 4.1.

Table 4.1 Criteria and qualitative indicators set for the SWOT analysis

CRITERIA	INDICATORS
----------	------------

1. Structural aspects	1.1. Equipment needed 1.2. Expertise needed 1.3. Methodological aspects 1.4. Theoretical aspects
2. Operational aspects	2.1. Collection and storage of samples/data 2.2. Laboratory/analysis (processing phase), costs 2.3. Time needed
3. Outcomes	3.1. Applicability/transferability of results

Elements such as costs, time required, logistic aspects, complexity, reliability, expertise needed and expertise bottlenecks were taken into account. Robustness and applicability of results, as well as replication of approaches, proper of the different methodologies used in this project were evidenced and discussed.

The SWOT exercise was performed with the proactive collaboration of the key scientists (e.g., WP or Task leaders) involved in WP1, WP2 and WP3.

To facilitate and to standardise the collection of comments and suggestions from the experts, guiding examples of an interpretation key for evaluating the indicators selected to assist the SWOT analysis were drafted (Tables 4.2 and 4.3). The suggestions showed in these tables were proposed as a non-exhaustive example of possible Strengths, Weaknesses, Opportunities and Threats. Therefore, the experts were free to argue any other consideration they deemed appropriate.

In addition, an exam of the Project Deliverables and Reports was carried out to gain the key elements helping in the SWOT assessment. This evaluation was then used to complement the feedback from the experts directly involved in the Project. In this evaluation the same criteria and indicators as mentioned above were taken into account, as well as the Project structure and development.

Table 4.2 – Some guiding examples of interpretation key for evaluating the indicators selected to assist the SWOT. Strengths and Weaknesses.

CRITERIA	INDICATORS	STRENGTHS AND WEAKNESSES	
1. Structural aspects	1.1 Equipment needed	<i>The need of a relatively simple and low expensive equipment (low investment costs) is considered a strength.</i>	<i>The need of a highly specialised equipment usually implies high investment costs. It is considered a weakness due to the time and the cost necessary to set up the equipment.</i>
	1.2 Expertise needed	<i>The need for good skills / experiences but not extremely refined, both for collecting samples/data and subsequent analyses, is considered a strength.</i>	<i>The need for extremely refined skill/experience, both for collecting samples/data and subsequent analyses can be a weakness.</i>
	1.3 Methodological aspects	<i>The reliability and the accuracy/precision of the method, as well as its reproducibility in time and space, can be considered a strength, for a regular and standard data gathering.</i>	<i>High uncertainty or poor reproducibility of the results, in case of absence of intercalibration and/or validation techniques can be considered a weakness.</i>
	1.4 Theoretical aspects	<i>A robust technical and scientific background supporting the relevance of</i>	<i>A scarce and uncertain technical and scientific background supporting the</i>

		<i>the method in providing indications for the stock identification is a strength.</i>	<i>relevance of the method in providing indication for the stock identification is considered a weakness.</i>
<u>2. Operational aspects</u>	2.1 Collection and storage of samples/data	<i>The need of rather simple procedures for the collection, storage and transport of samples can be a strength.</i>	<i>The need of complex and sensitive procedures for the collection, storage data and transport of samples can be a weakness.</i>
	2.2 Time needed	<i>The short/limited time needed to achieve all the process (sampling/data collection, analysis, provision of results) can support studies with short term results. This can be considered a strength</i>	<i>A long time needed to achieve all the process (sampling/data collection, analysis, provision of results, replication and validation of the analyses), implies studies with medium-long terms results. This can be a weakness for routine implementation in data collection.</i>
	2.3 Laboratory (processing phase) costs	<i>Low additional costs (other than the investment costs) that must be sustained during the operational phase, can be considered a strength.</i>	<i>Important additional costs, other than the investment costs (e.g., for laboratory chemicals, supplies) that must be sustained during the operational phase can be a weakness that in some case can represent an obstacle for the realization of the study.</i>
<u>3. Outcomes</u>	3.1 Applicability/transferability of results	<i>Optimized protocols already existing that can be easily transferred/adapted to different species/area around the Mediterranean.</i>	<i>Lack of repository or standards for data accessibility and comparability</i>

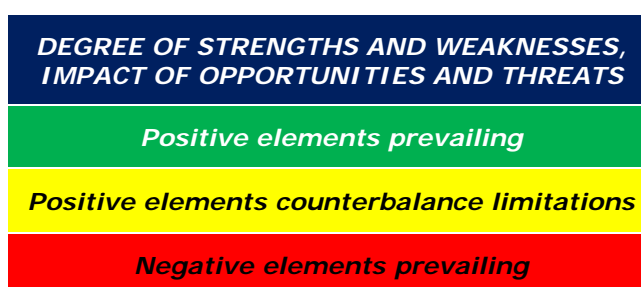
Table 4.3 - Some guiding examples of interpretation key for evaluating the indicators selected to assist the SWOT. Opportunities and Threats.

CRITERIA	INDICATORS	OPPORTUNITIES AND THREATS	
<u>1 Structural aspects</u>	1.1 Equipment needed	<i>Having many laboratories at European level already working with stock identification studies can be an opportunity for future activities.</i>	<i>The sensitivity and needs for routine maintenance of the instrumentation can be considered a threat.</i>
	1.2 Expertise needed	<i>Having teams already trained and with diversified expertise can be an opportunity to face with a wide variety of topics, especially for regular data gathering.</i>	<i>Few experts having specific skill on the topic can be considered a threat in the case of necessity of replacement of personnel or for the implementation of a regular data gathering at wide scale.</i>
	1.3 Methodological aspects	<i>Standards (even commercial) are already available and tested for the analyses or that can be easily adapted to specific needs.</i>	<i>Technology and methods are not yet sufficiently consolidated and this can determine uncertainty in the results and delays in the planned activities.</i>
	1.4 Theoretical aspects	<i>Methodologies recently applied with promising results</i>	<i>A methodology not fully supported by robust evidences</i>

		<i>in the stock identification can be considered an opportunity for the implementation of regular data gathering</i>	<i>can be a threat, due to the risk to obtain uncertain results</i>
<u>2. Operational aspects</u>	2.1 Collection and storage of samples/data	<i>Absence of barriers and/or restrictions for collecting samples or data can be an opportunity.</i>	<i>Complex procedures of sampling or data collection can hamper the study and therefore can be considered a threat.</i>
	2.2 Time needed	<i>A rather quick method can represent several opportunities, also to replicate the study under different conditions.</i>	<i>The possible need of outsourcing some analyses with a possible loose of full control of the process, and possible prolongation of the time needed for task completion. In can be a threat.</i>
	2.3 Laboratory (processing phase) costs	<i>The laboratory processing phases are standardized, costs are affordable.</i>	<i>Lack of a consolidated laboratory routines, need of case by case fine tuning and optimization. High costs for sample collection and laboratory analyses,</i>
<u>3. Outcomes</u>	3.1 Applicability/transferability of results	<i>Examples of applicability/transferability of the results already exist. Availability of genetic markers and/or standards/procedures for other analyses of biological samples (e.g., otoliths).</i>	<i>Difficult transferability of the outcomes given the low level of standardization, limited knowledge transfer and equipment accessibility.</i>

The SWOT analysis has been also supported by a simple and easy to interpret representation, according to a traffic light approach. The “degree of strengths and weaknesses” and the “Impact of opportunities and threats” have been taken into account according to the indicators showing relatively positive/green elements, or neutral/yellow, or occurrence of bottleneck/red responses.

Table 4.4 – The traffic light approach applied to synthesize the SWOT results.



The results of the SWOT analysis are summarised in the Tables 4.5, 4.6 and 4.7.

Table 4.5 – Results of SWOT analysis as regards the GENETIC ANALYSIS (WP1).

CRITERIA	INDICATORS	Strengths	Weaknesses	Opportunities	Threats
1. Structural aspects	1.1 Equipment needed	<i>Materials and reagents required for the sampling of tissues are quite easy to obtain. Standard equipment in molecular biology is sufficient for most of the analyses required, including DNA extraction and ddRAD library preparation. Few dedicated instruments for DNA quantification are needed.</i>	<i>There is a need of a highly specialized equipment for processing the tissues for the genetic analyses, in particular raw data production (NG sequencing platforms) and data analyses from the genomic approach (big data, bioinformatic facilities). Raw data analysis requires adequate computational power. So, if a high performance server, ideally with more than 1Tb RAM is not available or accessible, this can be a bottleneck for data analysis. Moreover, the high performance sequencing facilities are not easy to develop mainly due to cost limitations. Next Generation Sequencing is usually outsourced.</i>	<i>Many laboratories and research centers are fully equipped, to conduct at least the first steps of the genetic analyses, because already involved in stock identification studies. The equipment needed does not change depending on the species.</i>	<i>It is necessary a routine maintenance and constant update of the instruments and laboratory equipment. Sequencing platforms and sub-contracting partners usually have their output limitations and turnaround times to provide results. As concerns some problematic species, there might be the need to use dedicated instruments for DNA size selection when performing ddRAD libraries. This can generate important additional costs and, though they can be accessed from external services, an overall increase of the total budget and the time needed can be produced.</i>
	1.2 Expertise needed	<i>Only some initial trainings are needed, but not of specific previous experience and extremely refined skill, only for collecting samples. Only a general expertise from molecular biology laboratories is needed. Once trained, the personnel will be able to perform the entire process.</i>	<i>Though the level of required expertise is generally low/medium for the laboratory part, a higher level for supervision is needed. In the case of analysis at wide geographical scale involving several partners, a critical issue can be the need to homogenise expertise and to align capability levels and standards. This can be also time consuming. Specific high level expertise is required for data analysis (bioinformatics).</i>	<i>There are many Institutions and Research centers in Mediterranean that have reached a high level of expertise for sample collection and analysis and have already worked with stock identification studies using genetic tools. Once the protocol is optimized for one species, the replication of the study is straightforward. A team with genetic expertise is able to work on several species and topics.</i>	<i>There is the need of routinely update the skills, due to the continuous technological advancement of equipment technologies, and the informatic requirements.</i>

	<p>1.3 Methodological aspects</p>	<p><i>The ddRAD genetic analysis is reliable and accurate. Resolution power is high.</i></p>	<p><i>The uncertainty of the results, depending on the quality/quantity of sampling tissues. Refinements of protocols for non-modeled species (e.g., highly complex genomes). The optimization of such protocols might need more time than initially planned. This can be a weakness and can be tackled by including in any project a pilot preparatory study</i></p>	<p><i>Experimental protocols, dedicated software, bioinformatic pipelines, analytical methods are already available. The possibility to extend the ddRAD genetic analysis to any species of fish or invertebrate is an opportunity, to compare patterns among species. This can be also an opportunity to develop scientific interactions between institutions.</i></p>	<p><i>The lack of routine standards easily applicable to any context and the need of optimization of protocols for specific species can cause uncertainty/delays in the activities.</i></p>
	<p>1.4 Theoretical aspects</p>	<p><i>There is a solid background supporting the relevance of genetic analyses in providing indications for stock identification. Genome-wide genetic methods such as ddRAD provide an accurate picture of the genetic differentiation and are much more reliable than methods based on little genetic information, such as mtDNA sequencing or microsatellite analysis with few loci. Using ddRAD, that provides plenty of genetic loci (SNPs), the resolution power is expected to be adequate to identify also small differences. Genetic results are comparable among labs and highly reproducible. Genetic differences are not affected by environmental or ontogenetic differences.</i></p>	<p><i>Need of refinements of protocols for non-modeled species (e.g., high complex genomes) may be needed. Though differences between stocks or populations identified with genetic methods are robust, some uncertainty remains when no differences are detected, which could be due to small sample size or inadequate resolution power.</i></p>	<p><i>The ddRAD methodology is increasingly being applied in the stock identification in several areas. This methodology has been recently applied on several species, and it provides the opportunity for the implementation of future studies.</i></p>	<p><i>If the methodology has not been used or optimized before in a given species, this can be a threat. Software and new methods appear continuously urging the need for updating the analytical approaches.</i></p>
<p><u>2. Operational aspects</u></p>	<p>2.1</p>	<p><i>The protocols for the collection, storage and transport of samples are well defined,</i></p>	<p><i>Sensitive procedures of sampling, shipping and storage of samples can hamper or slow</i></p>	<p><i>Surveys at sea or monitoring of landings, routinely performed under specific</i></p>	<p><i>The operational conditions onboard commercial vessels can be not suitable. The</i></p>

	Collection and storage of samples/data	<i>standardized and relatively simple, though they require an allocated time and budget.</i>	<i>down the study. If sample processing is not properly followed (i.e., processing fresh tissue and store immediately in ethanol) the quality of genetic data can be greatly reduced and many specimens can be discarded. Storage and transport of samples, can be difficult to control in large scale studies with several partners involved.</i>	<i>programs (e.g. DCF) can represent an opportunity to obtain tissues with an extensive spatial and temporal coverage.</i>	<i>absence of a proper cooperation among technicians and scientists from different partners. Transfer of genetic samples can pose some difficulties, owing to the differences in the legal/bureaucratic aspects among countries, in particular non-EU countries. Compliance with the Protocol of Nagoya Convention on biological diversity (Nagoya, 2010).</i>
	2.2 Time needed	<i>The molecular biology methodology is relatively quick if protocols have been already optimized for a given species (e.g., complex genomes) and good quality samples are collected and delivered to the genetic labs. A dataset of about one thousand individuals can be processed in some months.</i>	<i>In spite of the methodology relatively quick, the time needed to achieve all the process (sampling, sample preparation and analysis, data analysis, provision of results) is rather long. After the completion of the sampling, at least 6 months are needed, more likely 1 year, depending on the species complexity and spatial coverage. The time needed can increase significantly if many specimens are discarded due to low quality of DNA.</i>	<i>Replicated studies are feasible and highly recommended after optimization.</i>	<i>Field activity (sample collections) may require time and have delays due to several factors (e.g. adverse weather conditions, unavailability of tissues at the right time/amount). Additional time can be required for optimizing protocols and procedures in non-model species (taxa) The outsourcing of sequencing can result in delay, and it does not allow to have full control of the process.</i>
	2.3 Laboratory (processing phase) costs	<i>Nowadays, costs are dropping fast and specialized platforms exist to undertake part of the work (sequencing).</i>	<i>DNA extraction can be a time-consuming step of the entire process, and it relies on good quality of the collected samples. This might require scaling up the persons involved in DNA extractions. For some problematic species, there might be the need to use dedicated instruments for DNA size selection when performing ddRAD libraries. This can generate additional costs and,</i>	<i>Constant improvement in technologies leading to the possibility to access larger amount of data. Currently there is a wide cooperation among labs of different Institutions/Countries. This can contribute to share the costs and decrease the overall budget of a given project. .</i>	<i>Costs of chemicals and lab supplies can increase unpredictably due to products' availability. Most Projects are flexible to allow sub-contracting but others not, and changes in the price list should then be absorbed by the Project's partners.</i>

			<p><i>though they can be charged to external services, an increase of the total budget.</i></p> <p><i>Additional costs (e.g. laboratory consumables, chemicals, and outsourcing to specialized services) often must be sustained during the operational phase, especially to optimize the protocols.</i></p> <p><i>Additional costs (staff and consumables) may be needed for the repetition of analysis or the resampling in some sites due to the poor quality of samples and protocol optimization.</i></p> <p><i>Consumables and outsourcing are needed for sequencing, the latter requiring a significant dedicated budget.</i></p>		
<p><u>3. Outcomes</u></p>	<p>3.1 Applicability/transferability of results</p>	<p><i>The results are highly applicable and can be easily transferred. Once the ddRAD general protocol is set up by a given laboratory, this can be transferred and also adapted, with a preliminary work, to other species and geographical contexts of interest.</i></p>	<p><i>Species-specific differentiation patterns cannot be taken for granted for all marine organisms under study.</i></p> <p><i>In case of small genetic differentiation, it is critical to assess the stability and coherence of the genetic structure throughout the time. Big data are not easy to transfer.</i></p>	<p><i>Along with stock identification, the obtained genomic data can possibly be analysed to address several issues and deepen biological aspects (e.g., adaption, resilience to climate change and fishery pressure, connectivity, biodiversity monitoring).</i></p> <p><i>Taking advantage of the excellent scientific profile of research teams around the Mediterranean Sea, there are good perspective for future collaborative studies.</i></p>	<p><i>There can be some limitations if new methods/technologies emerge over time.</i></p> <p><i>The outcomes of the analyses (genetic structuring and level of differentiation) can be comparable among different studies, but with caution; intercalibration and/or validation techniques may be needed.</i></p> <p><i>The not adequate level of biological and/or molecular knowledge for some species might result in a non-optimal interpretation of the data.</i></p>

Table 4.6 – Results of SWOT analysis as regards the OTOLITH SHAPE and MICROCHEMISTRY ANALYSIS (WP 2).

CRITERIA	INDICATORS	Strengths	Weaknesses	Opportunities	Threats
1. Structural aspects	1.1 Equipment needed	<i>Currently there are many laboratories in Europe with the equipment and expertise needed, especially to conduct the otolith shape analysis.</i>	<i>As concerns otolith shape analysis, it is necessary to have “in house” a sclerochronological Lab for the preparation of samples (mounting, polishing). Microchemistry analysis, requires costly equipment and a highly expert technicians; therefore, sometimes there is the necessity of using external services. LAICPMS (Laser Ablation Inductively Coupled Plasma Mass Spectrometry) analysis is a relatively expensive and time-consuming analysis.</i>	<i>The fact that in many fisheries Labs around Mediterranean and Europe there is available equipment and expertise for the otolith shape analysis can be an opportunity for future works. The current experience in many laboratories in ageing studies is also an opportunity to apply investigations based on microchemistry. LAICPMS analysis: in perspective, the spatial resolution, the detection limit and the time needed for the analysis will significantly improve with a reduction in total costs.</i>	<i>As regards LAICPMS analysis, currently the necessary equipment to perform this routine is available only in a few laboratories around Mediterranean.</i>
	1.2 Expertise needed	<i>Currently there are many laboratories in Europe with the necessary expertise to carry out studies based on otoliths, mainly on shape analysis.</i>	<i>There is the need of expertise in all the phases of the study: samples preparation, analysis, and statistical treatment of the raw data for otolith shape and microchemistry. Sample preparation needs individual skills in manual grinding and polishing of small and extremely fragile samples. Long practice is required to achieve good results. In particular, LAICPMS analysis is a highly specialized technique and requires peculiar experience.</i>	<i>Detailed protocols have been elaborated to develop the otolith shape and microchemistry analysis that can be used for further studies.</i>	<i>The handling of samples and analytical procedures may depend of a single person that might have time availability problems. Needs of training new expert and close interactions among scientists that is not always possible.</i>

	<p>1.3 Methodological aspects</p>	<p>Several studies using the otolith shape were performed and consequently there are available procedures with standardized protocol and automatized processing steps for sample preparation and analysis. Some institutions (e.g., IMEDEA and IFREMER) have a long experience in all phases of otolith and microchemical analysis.</p>	<p>For otolith shape before to use the otolith, it is required to measure the asymmetry to include this potential bias. As regards microchemistry, during the sample preparation, there is the risk of possible chemical contamination. For all methods, it is necessary to have all biological and geographical information.</p>	<p>At IMEDEA, IFREMER, CONISMA, in house programs have been developed for the calibration and treatment of the Laser-ICPMS raw data. Several R scripts to analyze the otolith shape were developed. Therefore, sample preparation protocols can be easily transferable to other laboratories.</p>	<p>As regards microchemistry analysis, the analytical results are expressed as trace elements (ppm). This requires careful calibration and filtering. Usually, every laboratory works with its own methodology and this could affect comparisons.</p>
	<p>1.4 Theoretical aspects</p>	<p>There are many scientific evidences showing the utility of the information coming from otoliths shape and microchemistry analyses in supporting stock identification. Otolith shape is the sum of genetic and environmental factors. The otolith microchemistry is the product of the composition of the water masses where the fish has born and where it has dwelled.</p>	<p>The ontogenetic effects on the otolith shape are not always known. The physiological effects upon the incorporation of chemical elements into the otolith are in part unknown. The discriminating force depends on environmental and space / time variability. The methods based on otoliths are limited to bony fish.</p>	<p>Specific studies, e.g., Case Studies can be carried out to better define the possible environmental variability on a spatial / temporal scale. E.g., the comparison between 2D and 3D shape analyses according to different locations, a study that started during MED_UNITS Project.</p>	<p>The extrapolation from individual fish data to stock definition and properties can produce some uncertainty in the final results.</p>
<p><u>2. Operational aspects</u></p>	<p>2.1 Collection and storage of samples/data</p>	<p>In general, both for otolith shape and microchemistry analysis, collection and storage of samples are relatively easy and less expensive. Sampling procedures are well known and standardized.</p>	<p>Though sample handling is not particularly demanding, otoliths are fragile, and so must be handled and stored appropriately. For otolith microchemistry ultracleaning methods are necessary; this are available not in all the laboratories, but in many laboratories involved in MED_UNITS. For the otolith shape analysis it is necessary to clean the otolith before storage too.</p>	<p>Samples can be obtained from existing data collection campaigns (e.g., MEDITS and biological sampling) in the context of DCF, without a big additional effort, given that these types of biological samples are routinely taken by several laboratories in the DCF context</p>	<p>Uncareful treatment of samples could imply a source of contamination or damaging the otoliths, hampering the subsequent analyses.</p>

	<p>2.2 Time needed</p>	<p>Although the works with otoliths are usually time consuming, if high partners' expertise and collaboration (e.g. in the MED_UNITS consortium) are present, this allows for a good optimization of time. In addition, when the otoliths are cleaned immediately during their collection, the subsequent work can be speeded up.</p>	<p>Sclerochronological studies are in general time consuming: a lot of steps are required (sample collection and preparation, data extraction, data treatment). The time needed to prepare or to clean the otolith before the image analysis for otolith shape can be not negligible. The same as regards LAICPMS analysis: long time cannot be reduced due to technical constraints.</p>	<p>Shape analyses can be automated thanks to image treatment computer softwares. Otolith microchemistry preparation protocols are well known by most specialized laboratories that helps for time optimization.</p>	<p>In particular, as regards sample collection in a wide geographical area, several actors can be involved; this could produce notable extension of the time expected, due to several factors, as difficulties in communication, problems in sample shipment, etc..</p>
	<p>2.3 Laboratory (processing phase) costs</p>	<p>Sample preparation is in general low expensive. Otolith preparation requires highly specific tools and consumable, but it is not very expensive. For otolith shape analysis, it is possible to use a scanner, fast and low-cost technology. Work can be conducted in house.</p>	<p>Extra costs can be produced due to the need of externalizing some of the processes. Microchemistry and in particular LAICPMS analysis are highly expensive and require specific instruments.</p>	<p>A joint purchase of consumables can reduce costs for microchemistry. For otolith shape, only scanner can be used. The wider application of these methods in future can reduce the costs.</p>	<p>Extra cost can be produced if samples cannot be provided by official routes (i.e., scientific surveys as MEDITS, biological samplings in DCF, fishermen and fish markets). The costs of technical gases needed for LAICPMS are characterized by strong fluctuations. Microchemistry analyses costs are moderately high</p>
<p>3. Outcomes</p>	<p>3.1 Applicability/transferability of results</p>	<p>The wide bibliography and methodology make the results applicable and robust for both otolith shape and microchemistry.</p>	<p>Lack of open repositories. The results are comparable between laboratories only after an accurate intercalibration.</p>	<p>It is possible to analyze very large samples (e.g., from experimental surveys) especially for otolith shape and develop common methodologies based on existing achievements.</p>	<p>The need of time-consuming trainings and specialized equipment can make transferability rather complicated. Intercalibration is not always possible.</p>

Table 4.7 – Results of SWOT analysis as regards the FISHING GROUNDS ANALYSIS (WP 3).

CRITERIA	INDICATORS	Strengths	Weaknesses	Opportunities	Threats
1. Structural aspects	1.1 Equipment needed	<i>Low investment costs: the equipment needed to perform the analysis already exists in many labs (e.g., PC, ArcGIS license, open source software-R). AIS data are rather easy to obtain (e.g., downloadable after payment), even in a short time; they have high spatial and temporal resolution.</i>		<i>The integration and comparison with other data (e.g., catch data, environmental parameters) can be achieved and thus analyses performed in a wider context. Open data from Global Fishing Watch are also freely available and can be useful for quick application/comparison</i>	<i>Challenge of data harmonization when combining different source of information. Data from the south of the Mediterranean are scant or absent.</i>
	1.2 Expertise needed	<i>There is the need for a specific skill/experience from fisheries scientists, but not extremely refined, similar to that needed for combining the existing information (e.g., spatial distribution of biomass indices).</i>	<i>To perform refined and accurate studies, an advanced know how about processing of spatial data, ecological modelling, statistics and computer programming is needed.</i>	<i>Frontier field for methods development. The expertise in spatial analytical methods is available in many Research Institutions around the Mediterranean.</i>	<i>The potential risk of unrealistic or misleading results when data are handled by non experts.</i>
	1.3 Methodological aspects	<i>Several state-of-the-art methods and procedures for data processing are already available. The methods used in MED_UNITS, provided flexible and reliable options for handling differences in data availability. They have been previously validated and evaluated with data from specific Case Studies (e.g., Greece, Italy). Several outcomes were integrated, through a spatial overlap analysis, they are covering the whole Mediterranean Sea.</i>	<i>The uncertainty related to the classification of fishing activities. The intercalibration of fishing pressure metrics (fishing time, swept area, etc.) is challenging.</i>	<i>The currently available methodological framework provided reliability and flexibility and could be used as a baseline, on which future investigations on the spatial identification of fishing grounds, in the entire Mediterranean or other areas could be based.</i>	<i>Some problems can occur if other data (e.g CPUE, catch) are not available or are available at different scale of the effort data.</i>
	1.4	<i>The methods applied and validated/tested on specific case</i>	<i>The methodology, although rather consolidated and</i>	<i>A robust technical and scientific background is</i>	<i>If applied at wide spatial scale there is the need to</i>

	Theoretical aspects	<i>studies provided promising results.</i>	<i>validated, needs some adjustments, especially when very different exploitation patterns are occurring in different areas.</i>	<i>present in many research institutions around Mediterranean.</i>	<i>sharing knowledge and expand training possibilities, especially in the southern Mediterranean</i>
<u>2. Operational aspects</u>	2.1 Collection and storage of samples/data	<i>Information on fishing effort spatial and temporal pattern is already available (e.g. from specific archives or repositories, DCF and DCRF data.</i>	<i>Fishing effort spatial and temporal data are not readily available and data collation is time consuming.</i>	<i>The exploitation of fishing effort data at spatial scale is becoming more and more used by different teams, opening new perspectives for methods improvement and scientific cooperation.</i>	<i>Some barriers and/or restrictions prevent the use of already collected information. VMS data are not easily accessible.</i>
	2.2 Time needed	<i>In the case of already existing information, the entire process is rather fast. Analysing and integrating the information might be achieved in a medium time scale.</i>	<i>Medium-long time is needed to re-analyze data and achieve advanced results e.g., the temporal identification of fishing grounds. Expert judgment is often needed for some steps (e.g., vessel identification): this is rather time consuming.</i>	<i>The study could be repeated under different conditions, in a relatively medium time frame.</i>	<i>When AIS or VMS data are not available the time needed for the analysis can increase.</i>
	2.3 Laboratory (processing phase) costs	<i>In general, the costs are rather low. It is an in-house processing.</i>	<i>There is a need of appropriate resources for data storage.</i>		<i>Costs are generally related to the expert worktime, but the access to AIS time series need a dedicated budget.</i>
<u>3. Outcomes</u>	3.1 Applicability/transferability of results	<i>Possibility to identify country/fleet specific fishing ground. The outcomes can provide an impressive reconstruction of the fishing footprint.</i>	<i>Misinterpretation of results (scale-related estimation of the Swept area ratio).</i>	<i>Maps of fishing grounds per species (or group of species) and fishing gear are available at large spatial scale, as well as maps of hot and cold spots per species or aggregated for species assemblages.</i>	<i>Unbalanced representativeness of fishing footprint</i>

As regards **the genetic component of the study, it can be summarised that genome-wide methods such as ddRAD provide an accurate picture of the genetic differentiation, plenty of genetic loci (SNPs)**, and the resolution power is expected to be adequate to identify also small differences. Considering future applications, if protocols have been already optimized and samples are of good quality, the process can be relatively quick, but the finalization of the protocols' optimization for a new species can require some time. However, if sample processing is not properly followed, i.e., processing fresh tissue and store immediately in ethanol, then keep the same at -20°C, the quality of genetic data can be greatly reduced and many specimens can be discarded. Thus, **dedicated samplings are needed, as the operational conditions onboard commercial vessels can be not suitable.**

Standard equipment in molecular biology is sufficient for most of the genetic analyses required, including DNA extraction and ddRAD library preparation. Few dedicated instruments for DNA quantification are needed, but **Next Generation Sequencing is usually outsourced** and, despite the costs are decreasing with time, it is still a quite costly service. In addition, the outsourcing of sequencing can produce some delays, and it does not allow to have full control of the process. Raw data analysis in the bioinformatics genetic applications requires adequate computational power are not easy to transfer.

Methods such as ddRAD have been recently applied on several species, and this provides the opportunity for the implementation of future studies. In spite the methodology is relatively quick, the **time needed to achieve all the process** (sampling, sample preparation and analysis, data analysis, provision of results) **is rather long**. For the studies that can provide results in medium-terms, it is needed at last 6 months, but more likely – 1 year, depending on the species complexity and spatial coverage of the study, after the completion of the sampling.

Species-specific differentiation patterns cannot be taken for granted for all marine organisms under study. In case of small genetic differentiation, it is critical to assess the stability and coherence of the genetic structure throughout the time.

Taking advantage of the excellent scientific profile of research teams around the Mediterranean Sea, there are good perspectives for future collaborative studies. However, a close cooperation is needed.

The transfer of genetic samples can pose some difficulties, owing to the differences in the legal/bureaucratic aspects among countries, in particular non-European countries of the Mediterranean (e.g., compliance with the Protocol of the Nagoya Convention on the biological diversity).

There are many scientific evidences showing the utility of the information coming from otoliths shape and microchemistry analyses in supporting stock identification. Otolith shape is the sum of genetic and environmental factors. The otolith microchemistry is the product of the composition of the water masses where the fish has born and where it has dwelled. Currently there are many laboratories in Europe with the equipment and expertise needed to conduct the **otolith shape analysis that is not much demanding in terms of equipment and costs**, while the resolution power for stock differentiation can be considered quite accurate. For otolith shape analysis, it is possible to use a scanner, a fast and low-cost technology, the work can be conducted in house. It is possible to analyse very large samples especially for otolith shape and develop common methodologies based on the existing achievements. However, the time needed to prepare or to clean the otolith before the image analysis for otolith shape can be not negligible.

Microchemistry, and in particular LAICPMS analysis, instead require specific instruments and are highly expensive, extra costs can be produced due to the need of externalizing some of the processes. The need of time-consuming trainings and specialized equipment can make transferability rather complicated. Intercalibration is not always possible.

Sampling procedures of otoliths are well known and standardized. Samples can be obtained from existing data collection campaigns (e.g., MEDITS and biological sampling from commercial catches) in the context of Data Collection Framework, without a big additional effort, given that these types of biological samples are routinely taken by several laboratories in the DCF/DCRF context. Where “historical” otolith collections of certain species are present, these can be also used for assessments of past stock profiles. **It should be noted, however, that the methods of otolith shape and microchemistry are applicable only to fish.**

The SWOT analysis included also elements related to the identification of fishing grounds, despite not linked to the stock identification issue, because the information on the fishing grounds and fishing footprint can be useful for the stock management. **The identification of fishing grounds has fast progressed in the last decade**, it requires skilled experts as regards spatial analysis and modelling. **The methodology, although rather consolidated and validated, needs some adjustments, especially when very different exploitation patterns are occurring in different areas.** Time needed for the analysis is not negligible when time series with different gaps should be analysed. Costs are generally affordable and related to the expert worktime, but the access to AIS time series need a dedicated budget, and the access to VMS data is not easy. **The available methodological framework provides reliability and flexibility and can be used as a baseline**, on which future investigations on spatial identification of fishing grounds in the whole Mediterranean can be based.

The results of the SWOT analysis summarised by means of the traffic light approach are reported in the tables 4.8 - 4.11; the evaluation related to the otolith analysis has been split here between otolith shape and microchemistry, given that the two techniques, albeit based on the same biological materials, have different characteristics, advantages and bottlenecks.

Tab 4.8 – Traffic light evaluation for the population genetics approach.

POPULATION GENETICS	<i>Degree of strengths and weaknesses</i>	<i>Impact of opportunities and threats</i>
1. Structural aspects		
2. Operational aspects		
3. Outcomes		

Tab 4.9 – Traffic light evaluation for the otolith shape analysis approach.

OTOLITH SHAPE ANALYSIS	<i>Degree of strengths and weaknesses</i>	<i>Impact of opportunities and threats</i>
1. Structural aspects		
2. Operational aspects		
3. Outcomes		

Tab 4.10 – Traffic light evaluation for the otolith microchemistry analysis approach.

OTOLITH MICROCHEMISTRY ANALYSIS	<i>Degree of strengths and weaknesses</i>	<i>Impact of opportunities and threats</i>
1. Structural aspects		
2. Operational aspects		
3. Outcomes		

Tab 4.11 – Traffic light evaluation for the fishing ground identification.

<i>FISHING GROUNDS</i>	<i>Degree of strengths and weaknesses</i>	<i>Impact of opportunities and threats</i>
1. Structural aspects		
2. Operational aspects		
3. Outcomes		

One of the key points in the delimitation of stock unit is the ‘state of the art’ approach i.e. the application of multiple approaches, to the same biological samples, with comparison of results to achieve an interdisciplinary perspective and consensus (Cadrin et al., 2013; Cadrin, 2020).

STECF 17-03 noted that there is a full body of literature dealing with stock identification and reported that state of the art methodologies includes a combination of methods and data sources (including e.g. genetic, tagging, otoliths shape and microstructures, chemical markers, surveys, drift modelling, etc.). Recently ICES (2020) provided annual updates on recent applications of stock identification methods to species assessed by ICES and on advances in stock identification methods.

The insights from these scientific bodies point out that combining genetic characterization with otolith shape and other sources of information related to stock identification can corroborate the outcomes and may take stock mixing into account, providing a more robust stock discrimination, which is particularly useful in designing fishery management strategies based on stock productivity (Hüssy et al., 2016). **It should be noticed that, indeed, the results from the genetics and the otolith shape analyses gathered in MED_UNITS showed a good level of agreement in the identification of stock configurations for European hake and red mullet.**

It is also useful to link stock boundaries to the use of fishing grounds, to identify spatial units for fishery management.

The SWOT analysis highlighted the presence of positive elements and advantages in each method, while at the same time bottlenecks and limitations emerged for each approach, pointing out to the need for a continuous integration process with an exchange of knowledge, and achievements among the different research groups and laboratories.

The consideration of a dedicated data collection comprehensive of the approaches used in MED_UNITS, in particular for genetic and otolith analyses, should take into account the following elements. In particular, for the genetic studies, supplementary resources, in addition to the current ones, are needed also for sampling biological tissues that is not the case for otolith. This type of analyses, however is restricted to fish. A data collection to pursue the objective of stock identification and its periodic replication should be based on a strong cooperation at regional level, among the European and non-European countries, considering also the difficulties in transferring these biological samples from one area to the other for genetic studies. Thus, besides the need of strengthening the cooperation between the research teams of the Mediterranean, and the need of investments in building capacity and sharing knowledge, it would be necessary to develop **a Regional Sampling Plan that should involve the different actors, in a framework that should see the cooperation of the Regional Coordination Groups under DCF and the GFCM.**

3.5.2 Task 4.2 Integrating results by the different WPs and proposals of new management units

Task leader: Vinko Bandelj (OGS)

Participants: COISPA, CoNISMa – UNITORV, OGS, HCMR

The main objectives of Task 4.2 were to identify possible management units for the species under study by combining the data coming from other WPs (genetic data from WP1 and otolith data from WP2) with environmental data collected in this Task and explicit spatial predictors. In particular, the Workplan envisaged the use of fuzzy clustering methods (Bezdek, 1981), explicit spatial predictors, and direct gradient analysis methods such as the Redundancy analysis (RDA) (van den Wollenberg, 1977). The spatial origin of each individual was considered to reconstruct the spatial structure of the stocks configuration and to investigate the relationship of these stock structures with respect to the main environmental drivers.

In the activities of Task 4.2 two Deliverables were foreseen to support the definition of possible management units, the **Deliverable 4.2 Report with the maps of population units and discontinuities** (ANNEX 4.1 to this report) and the **Deliverable 4.3 Report with the maps combining population and management units** (ANNEX 4.2 to this report). The activities of this task had some delay and the effective duration was prolonged to August 2021.

The first activity of this task was to collect the environmental data from the products available on the Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>), in particular from the physical and biogeochemical reanalysis. The data were extracted at the bottom and for the euphotic zone (0-200 m) in the depth range of each species, and several statistics were computed over different time intervals (5 years, 3 years, 1 year). For the needs of Task 4.2 only the median values of the extracted environmental variables were used. Maps of these environmental variables are the Annexes from 1 to 6 of this D4.2 and are reported as **ANNEX 4.1a** to this report.

The D4.2 used as fuzzy clustering method the fuzzy k-means (FKM), a general case of the well-known partitive algorithm k-means (Bezdek, 1984), since it provides the needed flexibility: it does identify discontinuities in the data (i.e., clusters), but it also assesses the level of fuzziness in the association of data to clusters. In this way, FKM represents an intermediate solution between clustering methods and gradient analysis methods. As explicit spatial method the Moran Eigenvector Maps (MEM) (Borcard & Legendre, 2002; Dray et al., 2006) were selected since they can model all-scales spatial patterns by computing independent continuous variables from the coordinates of the sampling locations. Both, FKM membership grades, as variables to be explained, and MEMs, as predictors, can be used in the Redundancy Analysis (RDA) along with environmental variables.

In Task 4.2, with FKM we first searched for possible clusters existing in genetic data (fish and shrimp species), otolith shape and microchemistry data (fish species only), and combinations of them. Internal cluster validation indexes were applied in order to assess if the clusters found were well separated and internally homogeneous. Spatial and environmental variables were used as external validations for the fuzzy clusters obtained: the more variance the spatial and environmental variables were able to explain, the more those clusters do show a spatially structured and environmentally based distribution. Spatial analysis was performed by using depth, geographic coordinates of the sampling locations, and the MEMs computed from the geographic coordinates of the sampling locations. Environmental analysis was performed by using variables over the three-time intervals and the two-depth of extractions. For the clusterizations that were better explained by both, the spatial and the environmental predictors, a variation partitioning procedure (Borcard & Legendre, 1992) and a forward

selection procedure (Blanchet et al., 2008) were applied in order to disentangle the mutual relationships between predictors and to select the best final models.

Maps of the distribution of the fuzzy clusters and of the habitat characteristics of each species were produced as GIS layers and used to figure production and for other activities of Task 4.2 (see D.4.3, **ANNEX 4.2** to this report).

For the fish species, where more different datasets were available on which to base the sub-population identification, **the results showed that the combination of different data, e.g., the genetic and the otolith shape data, or the otolith shape and the otolith microchemistry data, or all three together, is a sound strategy** as it addresses the differences emerging between different stocks due to genetic isolation or habitat characteristics. It should be noticed that, indeed, the results from the genetics and the otolith shape analyses showed a good level of agreement in the identification of stock configurations for European hake and red mullet. **For both fish species the combined datasets showed strengthen relationships with the spatial and environmental variables**, as if similar patterns in both datasets would more clearly emerge, thus increasing the possibility to detect and interpret the observed differences between sub-populations.

Three stocks of *M. merluccius* were identified on the combination of the genetic and otolith shape data, using spatial and environmental variables: one in the Western Mediterranean, one in the Tyrrhenian-Ionian-Adriatic basins one in the Eastern Mediterranean.

For *N. norvegicus* one stock was characterizing the Adriatic Sea and the sampled subareas of the Ionian and Aegean Sea, while the other three were distributed in the Western Mediterranean.

For *P. longirostris*, the discontinuity between the western and the eastern stocks was positioned in the Ionian Sea south of Italy.

These results depend critically on the sampling design, which was based on the GSA subareas. Even if in several GSA subareas there were more than one sampling location, a denser sampling design would be needed to consistently reconstruct possible sub-populations at a scale smaller than that of the GSA subareas.

No stock was clearly identified for either *Aristeomorpha foliacea*, *Aristeus antennatus* or *Mullus barbatus*. For the two red shrimps the relation with the spatial and environmental variables was low, non-significant and did not show any pattern related to the number of clusters. The only possible conclusion is that for these two species, based on the data produced in this study and with the methods applied here, only one population could be identified in the sampled areas of the Mediterranean. For *M. barbatus* the conclusion is much the same, even though there were some rather inconclusive evidences for the existence of 2 to 3 sub-population in the Mediterranean. Other evidences pointed to the possibility of a divergence of *M. barbatus* populations at scales smaller than those sampled here, but also these were inconclusive since limited by the scale of the sampling design. A denser sampling design, perhaps restricted to smaller areas, might help in confirming or rejecting this hypothesis.

In D4.2 the spatial origin of each individual was considered to reconstruct the spatial structure of the stocks configuration and to investigate the relationship of these stock structures with respect to the main environmental drivers.

In D4.3 the most reliable and comprehensive stock configurations by species, as identified in the Deliverable 4.2, were selected and the obtained stock structures were projected on the management partitioning of the Mediterranean Sea (i.e., FAO Geographical Sub Areas – GSA -or GSA subunits), in order to obtain a series of maps combining population, management units and the fishing footprint (i.e., the gear-specific patterns of the fishing effort). For some species, it was not possible to identify a clear and reliable stock structure (e.g., when differences do not appear since the populations resulted homogenous for the investigated variables), whereas, for other

species, some relevant differences arose and suggested the partitioning of the Mediterranean Sea in some sub-areas occupied by different stocks.

As highlighted above the results did show that different numbers of potential stock units can be identified in the Mediterranean for the different species under study. In particular, for *Merluccius merluccius* three stocks were identified; for *Nephrops norvegicus* four stocks; and for *Parapenaeus longirostris* two stocks.

Thus, a series of interrelated maps were prepared to show a synthesis of the spatial correspondence between stocks, management areas, and fishing grounds, while taking into account the probability of membership of the different subareas to the stocks (Figs. 4.1-4.3).

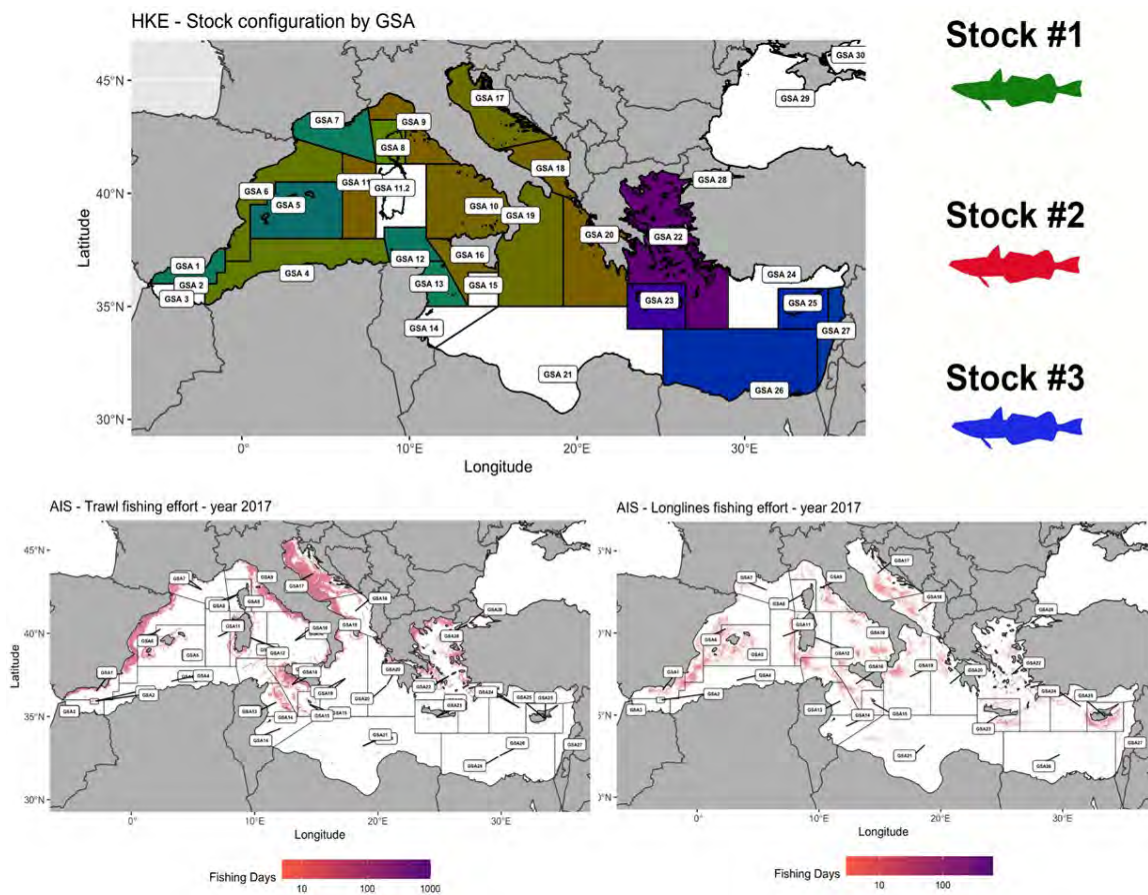


Figure 4.1– Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of European hake samples to the three stocks identified through the analysis of both Genetic data and Otolith shape data. The color of each GSA was obtained mixing green (stock #1), red (stock #2) and blue (stock #3) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1); Bottom right panel: Fishing footprint (total fishing effort in days, for the year 2017) for longliners from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).

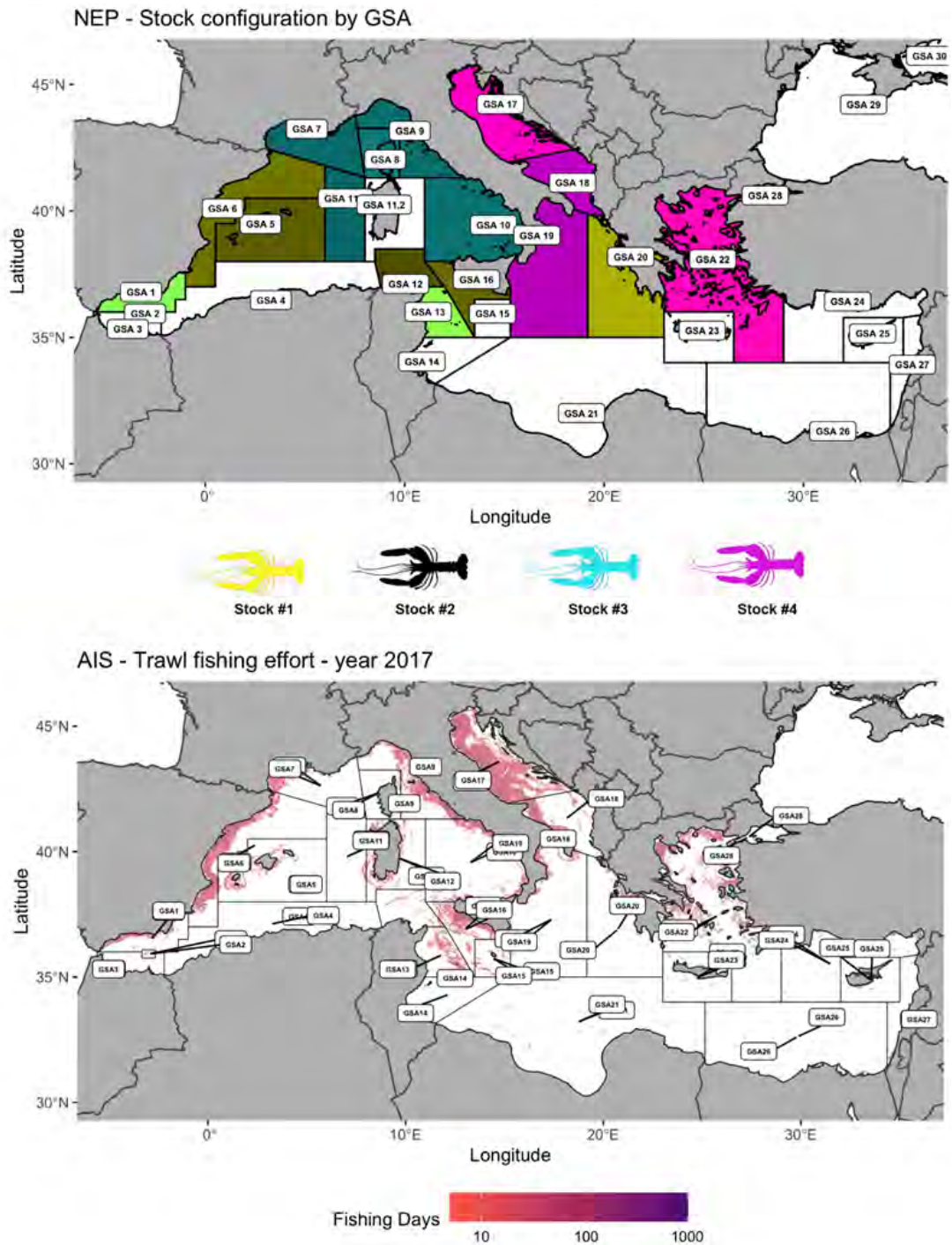


Figure 4.2 – Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of Norway lobster samples to the four stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing yellow (stock #1), black (stock #2), cyan (stock #3), and magenta (stock #4) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).

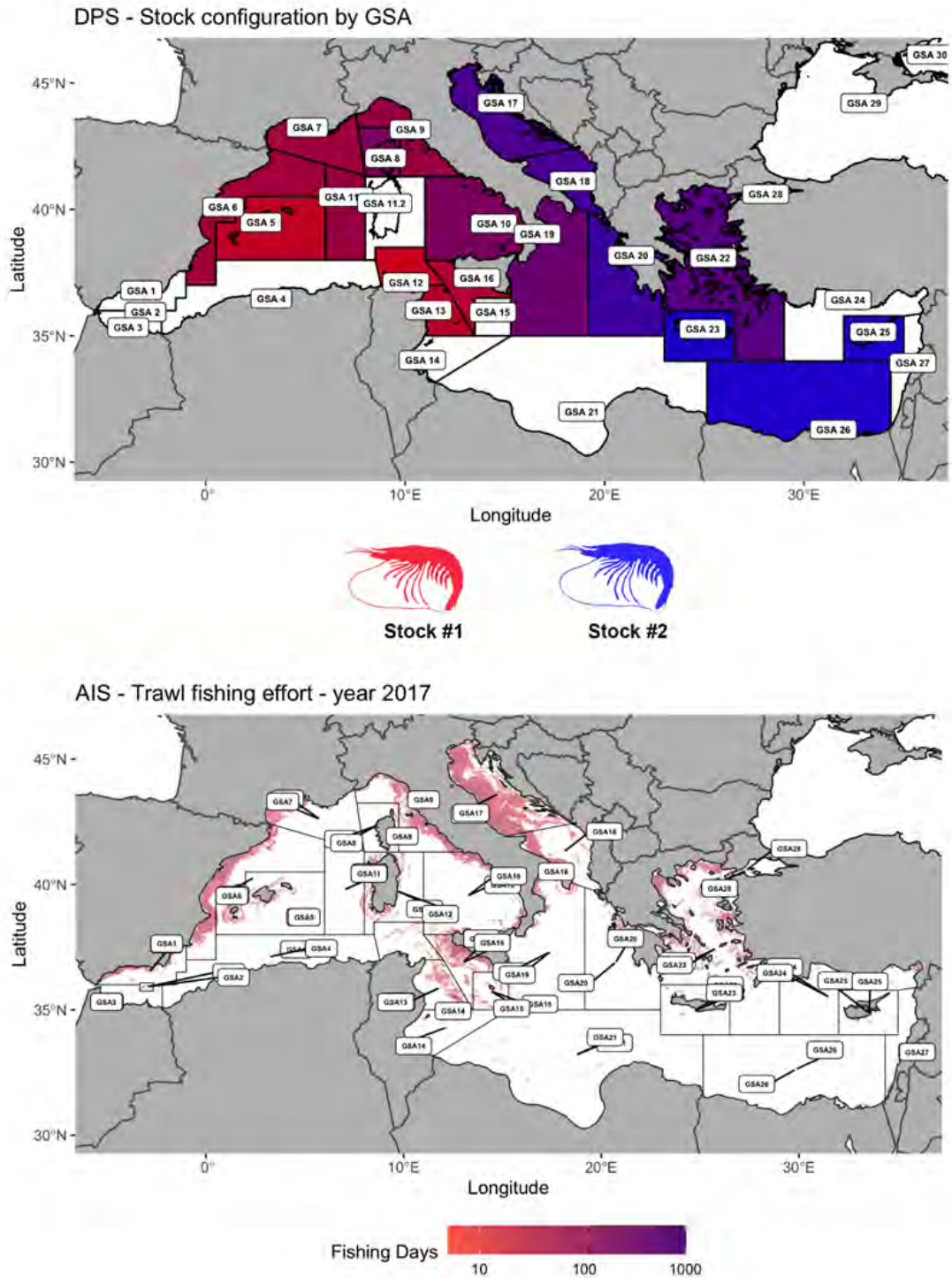


Figure 4.3 – Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of Deep water rose shrimp samples to the two stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing red (stock #1), and blue (stock #2) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).

3.5.3 Task 4.3 Future improvements for developing adaptive spatial fisheries management

Task leader: Fabio Fiorentino (CNR)

Participants: COISPA, CNR, CoNISMa

The activities of task4.3 have been delayed due to the availability of outcomes from Task 3.3 that was affected by cascade effect of delay from the preceding WPs, due to COVID-19 difficulties that impacted most of the project activities. The work of task 4.3 has been carried out remotely soon after the availability of WP3 outputs by target stock configurations.

Task 4.3 aims at examining potential development of adaptive spatial fisheries management through simulation approaches.

According to the project, three (3) case studies were selected, considering the stock configurations emerging from WP1 and in turn considered in Task 3.3 for stock assessment. In agreement with DG MARE, the case studies of Task 4.3 took into account stock configurations related to three high-value shrimps, that is *Aristeus antennatus* (ARA), *Aristaeomorpha foliacea* (ARS), and *Parapenaeus longirostris* (DPS), which are also targets of the fisheries in specific basins of the Mediterranean and subject to Management Plans. Indications on the case study, stock configuration, justification for case study selection, scenarios to be implemented and model to be applied are synthesized in Table 4.12

Details on the methodological approach, assumptions, results and discussion are reported in the **Deliverable 4.4 Report on the scenario modelled to explore the potential of developing adaptive spatial fisheries management (ANNEX 4.3** to this report).

In particular, ARA and ARS were assessed in the western Mediterranean (GSAs from 1 to 12) (Case study CS1), ARS in the Central Mediterranean (ARS; GSA 15, 16, 18, 19 and 20) (Case study CS2), and DPS (GSA 17, 18, 19 and 20) (Case Study CS3). **Different management scenarios agreed with DGMARE were evaluated by using bio-economic simulation models (BEMTOOL and SMART). These management scenarios considered reduction of fishing effort, improvement of the gear selectivity and spatial closures in areas critical for biological cycles of the targeted species (Essential Fish Habitats).**

For CS1 and CS2 five scenarios have been implemented: i) S0: status quo (baseline), ii) S1: 10% reduction in 2021 + closure areas, iii) S2: 10% reduction in 2021 + 20% reduction in 2022 + closure areas, iv) S3: 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023+ closure areas, and v) S4: 10% reduction in 2021+gear selectivity change in 2023+ closure areas.

For CS3 three scenarios have been investigated: i) S0: status quo (i.e., no variations respect to 2021 (transition phase), ii) S1: linear reduction to the ref. point until 2026 +closures, and iii) S2: 10% reduction of fishing day in 2022+Selectivity change in 2023 +closures.

BEMTOOL is a multi-species and multi-gear bio-economic simulation model for mixed fisheries. It follows a multi-fleet approach simulating the effects of management scenarios on stocks and fisheries on a fine time scale (month). Two case studies have been assessed using BEMTOOL:

- CS1 - ARS GSAs 1-12 and GSAs ARA 1-12;
- CS3 – DPS GSAs 17-18-19-20.

BEMTOOL results showed the benefits of improving the exploitation pattern, through spatial closures and/or selectivity improvements.

Table 4.12 Case studies selected for the scenario modelling in the Task 4.3

Case Study	Stock Config.	Justifications	Scenarios	Model
CS 1	ARS GSAs 1-12 and ARA 1-12	Both stocks are part of the Western MAP and target of specific fisheries, both stocks are fished unsustainably, A with a higher F_{curr}/F_{msy} ratio	<ol style="list-style-type: none"> 1. 10% reduction in 2021 + closures*; 2. 10% reduction in 2021+20% reduction in 2022 + closures* 3. 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023 +closures* 4. 10% reduction in 2021+Selectivity change (e.g. Gorelli et al. 2017) in 2023 +closures* 	BEMTOOL SMART
CS 2	ARS GSAs 13, 15, 16, 18, 19, 20	ARS in GSAs12-16 is a target of the Recommendation GFCM/43/2019/6 on management measures for sustainable trawl fisheries targeting giant red shrimp and blue and red shrimp in the Strait of Sicily (geographical subareas 12, 13, 14, 15 and 16), and ARS is one of the most important species targeted by the fisheries in the Strait of Sicily. SAC should provide to GFCM specific advice to implement a MAP by 2022. In addition, the Recommendation GFCM/42/2018/4. Recent assessment for GSAs 18-19 resulted in an F_{curr}/F_{msy} of 1.1, no recent assessments are available for the Strait of Sicily on a multiannual management plan for sustainable trawl fisheries targeting giant red shrimp and blue and red shrimp in the Ionian Sea (geographical subareas 19, 20 and 21).	<ol style="list-style-type: none"> 1. 10% reduction in 2021 + closures*; 2. 10% reduction in 2021+20% reduction in 2022 + closures* 3. 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023 +closures*** 4. 10% reduction in 2021+Selectivity change in 2023 +closures* 	SMART
CS 3	DPS GSAs 17-18-19-20a.	DPS in GSA17-18 is a target of the Recommendation GFCM/43/2019/5 related to the MAP for demersal stocks in the Adriatic. It has become an important target of the fishery in the whole area. The last assessment (2020) indicates that the stock is exploited well beyond F_{msy} .	<ol style="list-style-type: none"> 1. <i>status quo</i>, i.e. no variations respect to 2021 (transition phase) 2. linear reduction to the ref. point until 2026 +closures** 3. 10% reduction of fishing day (following the formula in the GFCM Recommendation) in 2022+Selectivity change in 2023 +closures** 	BEMTOOL
<i>Notes:</i>				
<i>Reference years Western 2015-2017; Transition year 2020; closures* are the ones of EWG 20-13</i>				
<i>Reference years Adriatic MAP 2015-2018; closures* are the ones of GFCM Recommendation</i>				
<i>Reference years for Central Mediterranean 2016-2018; closures *** are those based on in Fiorentino et al., 2013 and MEDISEH project</i>				

Firstly, the discard fraction in the catch is remarkably reduced and secondly, it allows to limit the effort reduction, providing similar economic outcomes and higher biological performances.

The implementation of the two case studies with **BEMTOOL**, using the stock assessment results of the new stock configurations, showed how the available knowledge on spatial dynamics of population components living in a study area can be integrated in a bio-economic model to accommodate the development of an adaptive spatial fishery management.

For both ARS and ARA in CS1, S4, combining an effort reduction of 10% with a change in the exploitation pattern, is the best performing scenario, allowing to reduce the F of around 35-40%; this percentage is quite close to the percentage required by ARS but only half of the reduction required by ARA. Although S2 and S3 scenarios are quite equivalent in terms of F, the 20% of effort reduction applied in 2022 in S2 produces increase in SSB and total landing slightly higher than S3, that distributes the 20% reduction in two years. The projections show that S4 increases the SSB of ARS and ARA of 63% and 144 % respectively, with a decrease in landing of 5% for ARS and an increase of 61% for ARA.

From an economic point of view S2 and S3 are very similar and allow to reach revenues well above S0 and S1. Although in the short term S4 projections show a sharp decrease in revenues and CR/BER, due to the change in selectivity, this scenario allows to reach, in the long term, values quite comparable to the ones of S2 and S3, reducing the effort of 10% instead of 30% overall.

As regards the DPS, the decrease of the overall F is higher for S1, while the combination of effort reduction and change in exploitation pattern of S2 scenario returns a decrease of 16% that does not allow to reach the F_{MSY} range, but allows to increase the SSB of 40% respect to S0 in 2026.

In terms of SSB, S1 allows to obtain the best performance, followed by S2; on the other hand, scenario S2 returns the higher overall catch than S1 in the long term, followed by S0. Although both S1 and S2 decrease remarkably the discard respect to S0 in the long term, S2 allows to reduce the discard fraction already from 2022 and to reach the smallest value.

From an economic point of view S2 is the scenario best performing in terms of total revenues, allowing to reach values in line with S0, after an initial reduction in the short term; the increase on the overall fleet revenues is of 3%, but varies among the different fleet segments between 2% and 23%. On the other hand, S1, after the reduction in revenues until 2026, due to implementation of the management measures, allow to reach values slightly lower than S0.

SMART is a spatially- explicit multi-species individual-based model allowing to evaluate the effects on stock dynamics and fishery performance under different time and spatial based management scenarios. The model needs spatially-resolved data on catch rates by age and fishing effort; in this application the implicit-spatially SMART model was implemented. Two case studies have been assessed by a modified version of SMART:

- CS1 - ARS GSAs 1-12 and GSAs ARA 1-12;
- CS2 – ARS GSAs 13-16,18-20.

The trends of SSB for the ARS in the GSAs 1-12 simulated by SMART indicate that a decreasing of the stock is expected in the present situation and that only some of the explored scenarios are likely to stop or reduce this decline.

The most promising approach is represented by the Scenario S4, which integrates spatial closures, effort reduction and changes in the selectivity. The management measures in the Scenarios S1 and S3 seem adequate to slow down the decreasing trend, while those in the Scenario S2 are likely to stabilize the SSB values after 2 years. In the case of ARA in the GSAs 1-12, the model suggested that management measures in the Scenario S1 are not

doi: 10.2926/909535
ISBN 978-92-95225-33-6

enough to stop the expected decline observed in the S0, whereas those in the Scenario S2 should allow to stop the decline and stabilize the SSB. However, the management measures in the Scenario S3 seem more effective, at least in the middle term, and those in the Scenario S4 are expected to determine a partial recovery of this stock. The predicted trends of SSB for the ARS in the GSAs 15-16 and 18-20 indicate that an increasing of the stock is expected in the present situation and that only the set of management measures considered in S4 could modify this dynamic of the stock and improve its condition. S1-3, however, are associated with a declining phase in the last part of the forecast period.

Despite BEMTOOL and SMART assessed the same fisheries under the same scenarios, they were quite different in terms of results. These differences are mainly due to differences in modelled processes and assumptions in MSE. However, both models evaluated as the best management strategy, both in terms of gain in SSB and improve or light decrease of the current yield, the scenario S4. This scenario is characterized by reducing of the 10% the current fishing effort coupled with an improvement of trawl net selectivity and protection of EFHs.

These results are in line with literature drawing the importance to adopt management strategies targeted not only to decrease the current fishing effort but also to improve the selectivity and, more in general, the exploitation pattern of demersal stocks to increase the sustainability of trawl fisheries in the Mediterranean (Colloca et al., 2013; Russo et al., 2014b; Gorelli et al., 2017; Brčić et al, 2018; Khoukh & Maynou, 2018; Russo et al., 2019a; Lucchetti et al., 2021).

Finally, an “ideal” roadmap for developing adaptive spatial fishery management aimed to reduce uncertainty in assessment and management procedures was explored.

On the basis of the review by Kerr et al. (2017), spatial structure and connectivity (both in terms of larval dispersal and fish migrations) within and between populations are considered as main factors affecting productivity (spawning-stock biomass, SSB), stability (variation in SSB), resilience (time to rebuild SSB after environmental disturbance), and sustainability (maximum sustainable fishing mortality and yield) of exploited populations. Therefore, uncertainty on stock spatial structure and impact of fishing effort on spatial distribution of exploited populations should be considered one of the main factors that can make fisheries management ineffective (McGilliard et al., 2015; Goethel et al., 2016; Carson et al., 2017; Kerr et al., 2017; Khoukh & Maynou, 2018).

Within an adaptive spatial based fishery framework, it should be relevant understanding the degree of spatial isolation or overlap between populations and harvest stocks to identify the appropriate strategy, to assure the renewability of exploited population and reduce uncertainty in assessment and management.

Taking into account Kerr and Goethel (2014) suggestions, an ideal key steps in the development of MSE for assessing the implications of spatial stock structure in assessment and management could include:

- i. Development of operating models that represent the most probable configuration of population structure of the fishery resource, using the best available data and considering the trade-off between completeness and resolution of information (in terms of space and population structure/dynamic);
- ii. Simulation of alternative management strategies;
 - a. Generation of spatial detailed data on stock demography from operating models and application of stock assessment methods which include spatial dimension of fisheries;
 - b. Application of alternate management strategies that integrate information on population structure and spatial impact of fishing fleet;
 - c. Projection for a fixed time period of the operating model given the advice from management strategies on stock status indicators and related catch;
- iii. Evaluation of performance of alternative management strategies against performance criteria (including biological, economic, and social objectives) to determine the best alternative taking into account biological, economic, and social

trade-offs.

The necessity of accurate knowledge of populations units and their connectivity was highlighted and the main obstacles to the effective adoption of adaptive spatial fishery management was discussed. Furthermore, in an adaptive management aimed at protecting the spatial structure of the stock the identification and protection of nurseries together spawning areas should consider as a main tool, in order to mitigate both growth and recruitment overfishing.

4. CONCLUDING REMARKS

- All specimens of each target species sampled in MED_UNITS were analysed in a comprehensive genetic study, allowing an **unprecedented biogeographic analysis and a genetic characterization at Mediterranean scale. Some species as Norway lobster were investigated at such scale for the first time.**
- Additionally, **for the two fish species, the European hake and the red mullet, otolith samples were used for an otolith shape analysis at Mediterranean scale, complemented by otolith microchemistry analysis in selected sites. An unprecedented coverage at spatial scale and for the number of individuals analysed was achieved.**
- The use of the same individuals for genetic and otolith analyses allowed, **for the first time, to start the implementation of a ‘state-of-the-art’ methodology** (Cadrin et al., 2013) **in the Mediterranean.**
- **The differentiation between Western, Central and Eastern Mediterranean populations of European hake is now fully supported** and the combined datasets showed meaningful relationships with the spatial and environmental variables. Findings from genetic and otolith studies showed coherent results, with well differentiated Atlantic and Mediterranean European hake stocks, as already known from the literature, but also within the Mediterranean populations.
- **For red mullet a lack of genetic differentiation was observed, in agreement with previous studies conducted at smaller scale in the Mediterranean Sea. A very low, yet significant, difference was found for the three major groups of the Western, the Central and the Eastern Mediterranean. A difference detected also in the otolith shape analysis.** The microchemistry analysis showed a high individual variability. The relation with the spatial and environmental variables was low and non-significant. Possibly a divergence of red mullet populations at scales smaller than those sampled could be hypothesized, probably a denser sampling design might help in confirming or rejecting this hypothesis.
- Regarding crustaceans the situation is diverse, according to the species.
- **For blue and red shrimp, a slight differentiation was observed, especially when comparing samples from Western and Eastern Mediterranean.** The best scenario explaining population structure corresponded to three groups, the first including samples from the Western Mediterranean Sea, the second from the Central Mediterranean Sea, and the third one from the Eastern Mediterranean Sea. However, values of differentiation, though significant, were extremely small.
- **Likewise, for giant red shrimp a lack of genetic differentiation was observed,** in agreement with previous studies conducted at smaller geographic scales. Different alternative scenarios of grouping populations based on Mediterranean basins were tested. **Some scenarios indicated the existence of very weak differentiation with the highest values found significant for the three major groups of the Western, Central and Eastern Mediterranean.**
- **For the two red shrimps the relation with the spatial and environmental variables was low and non-significant,** thus no differentiation among units could be estimated with the support of these variables.
- **For deep water rose shrimp genetic clustering methods confirmed the existence of differentiation,** indicating **potential subdivisions of East Mediterranean samples and a slightly different cluster composition for samples West and East of the Strait of Sicily, that is identified as the major genetic breakpoint.** This discontinuity was also identified in the analysis using spatial and environmental variables.
- **In Norway lobster relatively high and significant values were encountered for the separation of the Adriatic Sea (GSA17 to 19) from the neighbouring**

basins to the west (GSA1 to 11) and the east (GSA22). Current results are the first to show a considerable differentiation of the Adriatic and Eastern Mediterranean samples, a pattern confirmed when using also the environmental and spatial variables.

- **Regarding the identification of the fishing grounds, the “new” fishing dataset covers for the first time the whole highly productive Mediterranean basin** and may inform the wider scientific community, as well as those involved with policy and management, on fishing footprints in the Mediterranean ecosystem.
- **Several data sources and modelling were integrated for the first time at Mediterranean scale** and the outcomes demonstrate the possibility for fishers to harvest a certain species in a specific area, while **providing spatial information about the number of exploited species in several fishing grounds.**
- Thirteen **novel stock assessments were carried out, using the new stock configurations. They represented a first and promising approach to the assessment of the new stock units,** although not in all the cases, for several reasons, an improvement was detected with the new configurations. **Further investigation shall be implemented** before scientific advice can be provided in a reliable and robust way.
- The evaluation of an adaptive spatial fisheries management using **bioeconomic models and simulation approaches estimated as the best management strategy,** both in terms of gain in SSB and improve or decrease weakly the current yield, **the scenario characterized by reducing of 10% the current fishing effort coupled with an improvement of trawl net selectivity and protection of Essential Fish Habitats (EFH).**

5. LESSON LEARNT

- The **SWOT analysis**, evaluating the internal factors related to the methodologies applied and their possible implementation on a routine basis, highlighted the presence of positive elements and **advantages in each method applied in MED_UNITS**, while at the same time **bottlenecks and limitations emerged for each approach, pointing out to the need for a continuous integration process with an exchange of knowledge and achievements among the different research groups** and laboratories.
- Spatial structure and connectivity (both in terms of larval dispersal and fish migrations) within and between populations are considered as main factors affecting productivity, stability resilience and sustainability of exploited populations. **Further progress on the ‘state-of-the-art’ approach is thus a key point that should comprise additional methods for delineating stocks, including population connectivity studies.**
- **For DNA quality, particular care should be taken during sample collection and preservation, to reduce the number of DNA extracts not suitable for high-throughput genotypic methods** like the RAD-seq applications. The collection of tissues should preferentially be realized from alive/freshly caught individuals and the sampling realized as soon as possible within 1-2 hours from the death of the animal, always kept in optimal conditions (low temperatures).
- **Scientific surveys at sea (as MEDITS or other similar surveys) are a good opportunity** to implement proper sampling design for genomic analyses **but this cannot be a collateral activity as it requires the allocation of adequate resources.** Larger quantities of samples should be collected in order to have enough samples per area. Extra samples will allow to overcome the unavoidable decrease in numbers due to low DNA quality samples or failures during the experimental steps.
- It is highly recommended to set up **central Hubs**, both for genetic or otolith collection of samples, **in charge of the coordination of the sampling activities** between the samplers in the field and the technicians in the laboratory.
- For the optimization of the protocol for library preparation and sequencing, **particular care should be devoted to DNA extraction itself, selection of restriction enzymes, narrower size-selection of fragments for high throughput sequencing, lower level of multiplexing, and/or stricter filtering for bioinformatic analysis.** For future applications of ddRAD sequencing in new species, particularly crustaceans, we advise to **perform a small-scale optimization study** to address these issues **in advance.**
- **A data collection to pursue the objective of stock identification and its periodic replication should be based on a strong cooperation at regional level**, among the European and non-European countries, thus including several actors in the development of a Regional Plan.

ANNEX 1A: REFERENCES

- Apostolidis, A. P., K. A. Moutou, C. Stamatis & Z. Mamuris, 2009. Genetic structure of three marine fishes from the Gulf of Pagasitikos (Greece) based on allozymes, RAPD, and mtDNA RFLP markers. *Biologia* 64(5):1005-1010 doi:10.2478/s11756-009-0161-0
- Arculeo, M., S. Lo Brutto, M. Cammarata, M. Scalisi & N. Parrinello, 1999. Genetic variability of the Mediterranean Sea red mullet, *Mullus barbatus* (Pisces, Mullidae). *Russian Journal of Genetics* 35(3):292-296
- Balloux, F., 2001. EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations. *Journal of Heredity* 92(3):301-302 doi:10.1093/jhered/92.3.301.
- Begg, G. A., Hare, J. A., and Sheehan, D.D. 1999. The role of life history parameters as indicators of stock structure. *Fisheries Research*. 43: 141-163.
- Begg, G. A., Waldman, J. R. 1999. A holistic approach to fish stock identification. *Fisheries Research*. 43: 35-44.
- Benjamini Y., Yekutieli D., 2001. The control of false discovery rate under dependency. *Ann. Stat.*, 29, 1165–1188.
- Bezdek, J.C., 1981. Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York, NY, USA.
- Bezdek, J.C., Ehrlich, R., Full, W.E., 1984. FCM: the fuzzy c-means clustering algorithm. *Computers and Geosciences* 10, 192–203.
- Blanchet, F.G., Legendre, P., Borcard, D., 2008. Forward selection of explanatory variables. *Ecology* 89, 2623–2632.
- Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model* 153, 51–68.
- Campana, S.E., 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Mar. Ecol. Prog. Ser.* 188, 263–297.
- Brčić J., Herrmann B., Sala A., 2018. Predictive models for codend size selectivity for four commercially important species in the Mediterranean bottom trawl fishery in spring and summer: Effects of codend type and catch size. *PLoS ONE* 13(10): e0206044. <https://doi.org/10.1371/journal.pone.0206044>.
- Cadrin, S. X., Friedland, K. D., and Waldman, J. R. 2005. Stock identification methods: An overview. In *Stock identification methods: Applications in fishery science*. Edited by S.X. Cadrin, K.D. Friedland and J.R. Waldman. Elsevier Academic Press, pp. 3-6.
- Cadrin S., L.A. Kerr, S. Mariani., 2013. *Stock Identification Methods. Applications in Fishery Science*. 2nd Edition, eBook ISBN: 9780123972583. 588 pp.
- Cadrin, S. X., Maunder, M. N., & Punt, A. E., 2020. Spatial Structure: Theory, estimation and application in stock assessment models. Doi: 10.1016/j.fishres.2020.105608.
- Campana, S.E., 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Mar. Ecol. Prog. Ser.* 188, 263–297.
- Campana S.E., Casselman J.M., 1993, Stock discrimination using otolith shape analysis. *Can. J. Fish. Aquat. Sci.* 50, 1062–1082.
- Campana, S.E., Chouinard, G.A., Hanson, J.M., Fréchet, A., Bratley, J., 2000. Otolith elemental fingerprints as biological tracers of fish stocks. *Fish. Res.* 46, 343–357.
- Cannas, R., S. Marcias, F. Sacco, A. Cau & A. M. Deiana, 2012a. First isolation and characterization of genomic SSR markers for the giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827). *Genet Mol Res* 11(3):2745-8 doi:10.4238/2012.August.17.1.
- Cannas, R., F. Sacco, M. C. Follesa, A. Sabatini, M. Arculeo, S. Lo Brutto, T. Maggio, A. M. Deiana & A. Cau, 2012b. Genetic variability of the blue and red shrimp *Aristeus antennatus* in the Western Mediterranean Sea inferred by DNA microsatellite loci. *Marine Ecology* 33(3):350-363 doi:10.1111/j.1439-0485.2011.00504.x.
- Carson, S., Shackell, N., Mills Flemming, J., 2017. Local overfishing may be avoided by examining parameters of a spatio-temporal model. *PloS one*, 12(9), e0184427.
- Casselman, J. M. 1987. Determination of age and growth. In *The Biology of Fish Growth*, pp. 209-242. Ed. by A. H. Weatherley and H. S. Gill. Academic Press, New York. 443 pp.
- Castillo A.G.F., Martinez J.L., Garcia-Vazquez E., 2004. Fine spatial structure of Atlantic hake (*Merluccius merluccius*) stocks revealed by variation at microsatellite loci. *Marine Biotechnology*, 6, 299–306.
- Catchen J.M., Amores A., Hohenlohe P.A., Cresko W.A., Postlethwait J.H., 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes Genomes Genetics*.;1:171–182. doi: 10.1534/g3.111.000240.
- Cimmaruta R., Bondanelli P., Nascetti G., 2005. Genetic structure and environmental heterogeneity in the European hake (*Merluccius merluccius*). *Molecular Ecology*, 14, 2577–2591.

- Colloca F., M. Cardinale, F. Maynou, M. Giannoulaki, G. Scarcella, K. Jenko, J. M. Bellido, F. Fiorentino, 2013 Rebuilding Mediterranean fisheries: a new paradigm for ecological sustainability. *Fish & Fisheries* 14, 1: 89–109. DOI: 10.1111/j.1467-2979.2011.00453.x.
- Cope, J. M., and Punt, A.E. 2009. Drawing the lines: Resolving fishery management units with simple fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences*. 66: 1256-12.
- Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E. and Orlando, L., 2017., Experimental conditions improving in-solution target enrichment for ancient DNA . *Mol Ecol Resour*, 17: 508-522. doi:10.1111/1755-0998.12595
- Dray, S., Legendre, P., Peres-Neto, P.R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>.
- Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12: 499–510.
- Elsdon, T.S., Wells, B.K., Campana, S.E., Gillanders, B.M., Jones, C.M., Limburg, K.E., Secor, D.H., Thorrold, S.R., Walther, B.D., 2008. Otolith chemistry to describe movements and life-history parameters of fishes: hypotheses, assumptions, limitations and inferences. *Oceanogr. Mar. Biol. Annu. Rev.* 46, 297–330.
- Excoffier, L. & Lischer, H. E. L. 2010. Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10, 564–567 .
- Falush, D., M. Stephens & J. K. Pritchard, 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567-1587.
- Falush, D., M. Stephens & J. K. Pritchard, 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7(4):574-578 doi:10.1111/j.1471-8286.2007.01758.x.
- Félix-Hackradt, F. C., Hackradt, C. W., Pérez-Ruzafa, Á. & García-Charton, J. A. 2013. Discordant patterns of genetic connectivity between two sympatric species, *Mullus barbatus* (Linnaeus, 1758) and *Mullus surmuletus* (Linnaeus, 1758), in south-western Mediterranean Sea. *Mar. Environ. Res.* 92, 23–34.
- Fernández, M. V., F. Maltagliati, F. G. Pannacciulli & M. I. Roldán, 2011a. Analysis of genetic variability in *Aristaeomorpha foliacea* (Crustacea, Aristeidae) using DNA-ISSR (Inter Simple Sequence Repeats) markers. *Comptes Rendus Biologies* 334(10):705-712 doi:<https://doi.org/10.1016/j.crv.2011.07.005>.
- Fernández, M. V., S. Heras, F. Maltagliati, A. Turco & M. I. Roldán, 2011b. Genetic structure in the blue and red shrimp *Aristeus antennatus* and the role played by hydrographical and oceanographical barriers. *Marine Ecology Progress Series* 421:163-171 doi:10.3354/meps08881.
- Fernandez MV, Heras S, Maltagliati F, Roldan MI. 2013a. Deep genetic divergence in giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827) across a wide distributional range. *Journal of Sea Research* 76:146-153.
- Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán, 2013b. Multilocus comparative phylogeography of two Aristeid shrimps of high commercial interest (*Aristeus antennatus* and *Aristaeomorpha foliacea*) reveals different responses to past environmental changes. *PLOS ONE* 8(3): e59033 doi:10.1371/journal.pone.0059033
- Ferrà C., Tassetti A.N., Grati F., Pellini G., Polidori P., Scarcella G., Fabi G., 2018. Mapping change in bottom trawling activity in the Mediterranean Sea through AIS data, *Marine Policy*, 94:275-281. <https://doi.org/10.1016/j.marpol.2017.12.013>.
- Fiorentino F., Vitale S., 2021. How Can We Reduce the Overexploitation of the Mediterranean Resources? *Front. Mar. Sci.* 8:674633. doi: 10.3389/fmars.2021.674633.
- Fumagalli, M., 2013. Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. *PLOS ONE* 8(11): e79667 doi:10.1371/journal.pone.0079667.
- Gaertner J.C., Chessel D., Bertrand J.A., 1997, Stability of spatial structures of demersal assemblages: a multitable approach. *Aquat. Living Resour.* 11, 75–85
- Galarza, J. A., G. F. Turner, E. Macpherson & C. Rico, 2009. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). *Canadian Journal of Fisheries and Aquatic Sciences* 66(9):1478-1490 doi:10.1139/f09-098.
- Galdelli A., Mancini A., Zingaretti P., Tassetti A.N., Ferrà C., Armelloni E.N., Scarcella G., Fabi G., 2019. A cloud computing architecture to map trawling activities using positioning data. *Proceedings of the ASME 2019 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference* (in press).

- García-Merchán, V.H., Robainas-Barcia, A., Abelló, P., Macpherson, E., Palero, F., García-Rodríguez, M., Gil de Sola, L., Pascual, M., 2012. Phylogeographic patterns of decapod crustaceans at the Atlantic Mediterranean transition. *Molecular Phylogenetics and Evolution* 62, 664e672.
- Garoia, F., I. Guarniero, C. Piccinetti & F. Tinti, 2004. First microsatellite loci of red mullet (*Mullus barbatus*) and their application to genetic structure analysis of Adriatic shared stock. *Marine Biotechnology* 6(5):446-452 doi:10.1007/s10126-004-3045-x.
- Gerritsen H, Lordan C., 2011. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J. Mar. Sci.* 68, 245–252, doi:10.1093/icesjms/fsq137.
- Goethel, D. R., Kerr, L. A., and Cadrin, S. X., 2016. Incorporating spatial population structure into the assessment-management interface of marine resources. In *Management Science in Fisheries: An Introduction to Simulation-based Methods*, pp. 319–347. Ed. by C. T. T. Edwards, and D. J. Dankel. Routledge, New York. 460 pp.
- Gorelli, G., Company J.B., Bahamón, N., Sardà, F., 2017. Improving codend selectivity in the fishery of the deep-sea red shrimp *Aristeus antennatus* in the northwestern Mediterranean Sea. *Sci. Mar.* 81 (3): 381-386.
- James M., T. Mendo, E. Jones, K. Orr, A. McKnight, J. Thompson 2018. AIS data to inform small scale fisheries management and marine spatial planning. *Marine Policy*, 91: 113 - 121 10.1016/j.marpol.2018.02.012.
- Jombart T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi: 10.1093/bioinformatics/btn129.
- Jombart T. and Ahmed I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr52.
- Hale, M. L., T. M. Burg & T. E. Steeves, 2012. Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. *PLOS ONE* 7(9): e45170 doi:10.1371/journal.pone.0045170.
- Hammer, C. Zimmermann, C. 2005, The role of stock identification in formulating fishery management Advice. In *Stock identification methods: Applications in fishery science*. Edited by S.X. Cadrin, K.D. Friedland and J.R. Waldman. Elsevier Academic Press, pp. 631-658.
- Hemmer-Hansen, J., Therkildsen, N.O. & Pujolar, J.M. 2014. Population genomics of marine fishes: Next- generation prospects and challenges. *The Biological Bulletin*, 227: 117-132.
- Heras, S., L. Planella, I. Caldarazzo, M. Vera, J. L. García-Marín & M. I. Roldán, 2016. Development and characterization of novel microsatellite markers by Next Generation Sequencing for the blue and red shrimp *Aristeus antennatus*. *PeerJ* 2016(7) doi:10.7717/peerj.2200.
- Hüssy, K., Mosegaard, H., Albertsen, C. M., Eg Nielsen, E., Hansen, J. H., & Eero, M., 2016. Evaluation of otolith shape as a tool for stock discrimination in marine fishes using Baltic Sea cod as a case study. *Fisheries Research*, 174, 210-218. DOI: 10.1016/j.fishres.2015.10.010.
- Kalish J.M. 1989. Otolith microchemistry: validation of the effects of physiology, age and environment on otolith composition. *J. Exp. Mar. Biol. Ecol.* 132: 151-178.
- Kavadas S, Maina I., Damalas D, Dokos I, Pantazi M. et al., 2015. Multi-criteria decision analysis as a tool to extract fishing footprints: application to small scale fisheries and implications for management in the context of the maritime spatial planning directive. *Mediterranean Marine Science*, 16 (2), 294-304.
- Kavakiotis, I., P. Samaras, A. Triantafyllidis & I. Vlahavas, 2017. FIFS: A data mining method for informative marker selection in high dimensional population genomic data. *Computers in Biology and Medicine* 90: 146-154 doi:https://doi.org/10.1016/j.combiomed.2017.09.020.
- Kavakiotis, I., A. Triantafyllidis, D. Ntelidou, P. Alexandri, H.-J. Megens, R. P. M. A. Crooijmans, M. A. M. Groenen, G. Tsoumakas & I. Vlahavas, 2015. TRES: Identification of Discriminatory and Informative SNPs from Population Genomic Data. *Journal of Heredity* 106(5):672-676 doi:10.1093/jhered/esv044.
- Kerr, L. A., and Goethel, D. R. 2014. Simulation modeling as a tool for synthesis of stock identification information. In *Stock Identification Methods: Applications in Fishery Science*, 2nd edn, pp. 501–534. Ed. by S. Cadrin, L. Kerr, and S. Mariani. Elsevier Academic Press, Burlington. 566 pp.
- Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L., Worsøe Dickey-Collas, M., Goethel, D. R., Hatfield, E. M.C., Kritzer, J. P., and Nash, R.D.M. 2017. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. *ICES Journal of Marine Science*, 74: 1708–1722.
- Khokh, M., Maynou, F., 2018. Spatial management of the European hake *Merluccius merluccius* fishery in the Catalan Mediterranean: Simulation of management alternatives with the InVEST model. *Sci. Mar.* 82, 175-188.
- Kroodsma D. A., J. Mayorga, T. Hochberg, N. A. Miller, K. Boerder, F. Ferretti, A. Wilson, B. Bergman, T. D. White, B. A. Block, P. Woods, B. Sullivan, C. Costello, B. Worm 2018. Tracking the global footprint of fisheries. *Science* 359, 904–908.

- ICES. 2020. Stock Identification Methods Working Group (SIMWG). ICES Scientific Reports. 2: 94. 32 pp. <http://doi.org/10.17895/ices.pub.7485>.
- Levi, D., B. Patti, P. Rizzo, S. Lo Brutto, N. Parrinello & M. Arculeo, 2004. Genetic and morphometric variations of Mediterranean hake, *Merluccius merluccius*, in the Strait of Sicily (central Mediterranean): implications for stock assessment of shared resources. *Italian Journal of Zoology* 71(2): 165-170 doi: 10.1080/11250000409356568.
- Lo Brutto, S., M. Arculeo, A. Mauro, M. Scalisi, M. Cammarata & N. Parrinello, 1998. Allozymic variation in Mediterranean hake *Merluccius merluccius* (Gadidae). *Italian Journal of Zoology* 65: 49-52.
- Lo Brutto, S., M. Arculeo & N. Parrinello. 2004. Congruence in genetic markers used to describe Mediterranean and Atlantic populations of European hake (*Merluccius merluccius* L. 1758). *Journal of Applied Ichthyology* 20(2): 81-86 doi: 10.1046/j.1439-0426.2003.00514.x.
- Lo Brutto, S., T. Maggio, A. M. Deiana, R. Cannas & M. Arculeo, 2012. Further investigations on populations of the deep-water blue and red shrimp *Aristeus antennatus* (Risso, 1816) (Decapoda, Dendrobranchiata), as inferred from Amplified Fragment Length Polymorphism (AFLP) and mtDNA analyses. *Crustaceana* 85(11): 1393-1408 doi: 10.1163/15685403-00003131.
- Lo Brutto, S., T. Maggio & M. Arculeo, 2013. Isolation by distance (IBD) signals in the deep-water rose shrimp *Parapenaeus longirostris* (Lucas, 1846) (Decapoda, Panaeidae) in the Mediterranean sea. *Marine Environmental Research* 90: 1-8 doi: 10.1016/j.marenvres.2013.05.006.
- Lucchetti, A., Virgili, M., Vasapollo, C., Petetta, A., Bargione, G., Li Veli, D., Brčić, J., Sala, A., 2021. An overview of bottom trawl selectivity in the Mediterranean Sea. *Mediterranean Marine Science*, 22(3): 66-585. doi: <https://doi.org/10.12681/mms.26969>.
- Lundy, C. J., P. Moran, C. Rico, R. S. Milner & G. M. Hewitt, 1999. Macrogeographical population differentiation in oceanic environments: a case study of European hake (*Merluccius merluccius*), a commercially important fish. *Molecular Ecology* 8(11): 1889-1898 doi: 10.1046/j.1365-294x.1999.00789.x.
- McGilliard, C. R., Punt, A. E., Methot Jr, R. D., & Hilborn, R., 2015. Accounting for marine reserves using spatial stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(2), 262-280.
- Maggio, T., S. Lo Brutto, F. Garoia, F. Tinti & M. Arculeo, 2009. Microsatellite analysis of red mullet *Mullus barbatus* (Perciformes, Mullidae) reveals the isolation of the Adriatic Basin in the Mediterranean Sea. *ICES Journal of Marine Science* 66(9): 1883-1891 doi: 10.1093/icesjms/fsp160
- Maltagliati, F., L. Camilli, F. Biagi & M. Abbiati, 1998. Genetic structure of Norway lobster, *Nephrops norvegicus* (L.) (Crustacea: Nephropidae), from the Mediterranean Sea. *Scientia Marina* 62: 91-99.
- Mamuris, Z., A. P. Apostolidis, A. J. Theodorou & C. Triantaphyllidis, 1998a. Application of random amplified polymorphic DNA (RAPD) markers to evaluate intraspecific genetic variation in red mullet (*Mullus barbatus*). *Marine Biology* 132(2): 171-178 doi: 10.1007/s002270050383.
- Mamuris, Z., A. P. Apostolidis & C. Triantaphyllidis, 1998b. Genetic protein variation in red mullet (*Mullus barbatus*) and striped red mullet (*M. surmuletus*) populations from the Mediterranean Sea. *Marine Biology* 130(3): 353-360 doi: 10.1007/s002270050255
- Mamuris, Z., C. Stamatis, K. A. Moutou, A. P. Apostolidis & C. Triantaphyllidis, 2001. RFLP analysis of mitochondrial DNA to evaluate genetic variation in striped red mullet (*Mullus surmuletus* L.) and red mullet (*Mullus barbatus* L.) populations. *Marine Biotechnology* 3(3): 264-274 doi: 10.1007/s101260000075
- Manousaki T., A. Tsakogiannis, J. B. Taggart, C. Palaiokostas, D. Tsaparis, J. Lagnel, D. Chatziplis, A. Magoulas, N. Papandroulakis, C.C. Mylonas, C.S. Tsigenopoulos 2016. Exploring a nonmodel teleost genome through rad sequencing—linkage mapping in Common Pandora, *Pagellus erythrinus* and comparative genomic analysis. *G3: Genes, genomes, genetics*, 6 (3): 509-519
- Marcias, S., F. Sacco, A. Cau & R. Cannas, 2010. Microsatellite markers for population genetic studies of the giant red shrimp *Aristaeomorpha foliacea* (Crustacea, Decapoda). *Rapp Comm int Mer Médit* 39: 386.
- Marra, A., S. Mona, R. M. Sa, G. D'Onghia & P. Maiorano, 2015. Population Genetic History of *Aristeus antennatus* (Crustacea: Decapoda) in the Western and Central Mediterranean Sea. *Plos One* 10(3): 16 doi: 10.1371/journal.pone.0117272.
- Matic-Skoko, S., T. Segvic-Bubic, I. Mandic, D. Izquierdo-Gomez, E. Arneri, P. Carbonara, F. Grati, Z. Ikica, J. Kolutari, N. Milone, P. Sartor, G. Scarcella, A. Tokac & E. Tzanatos, 2018. Evidence of subtle genetic structure in the sympatric species *Mullus barbatus* and *Mullus surmuletus* (Linnaeus, 1758) in the Mediterranean Sea. *Sci Rep* 8: 14 doi: 10.1038/s41598-017-18503-7
- Milano, I., M. Babbucci, F. Panitz, R. Ogden, R. O. Nielsen, M. I. Taylor, S. J. Helyar, G. R. Carvalho, M. Espiñeira, M. Atanassova, F. Tinti, G. E. Maes, T. Patarnello & L. Bargelloni, 2011. Novel tools

- for conservation genomics: Comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. *PLoS ONE* 6(11) doi:10.1371/journal.pone.0028008.
- Milano I, Babbucci M, Cariani A, Atanassova M, Carvalho GR, Espiñeira M, Fiorentino F, Garofalo G, Geffen AJ, Helyar S et al. 2014. Outlier SNP markers reveal fine-scale population genetic structuring across European hake (*Merluccius merluccius*). *Molecular Ecology* 23: 118-135
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16(3):1215.
- Morales-Nin, B.; Swan, S.C.; Gordon, J.D.M.; Palmer, M.; Geffen, A.J.; Shimmield, T.; Sawyer, T. (2005). Age-related trends in otolith chemistry of *Merluccius merluccius* from the north-eastern Atlantic Ocean and the western Mediterranean Sea. *Marine and Freshwater Research*, 56: 599-607.
- Morales-Nin, B.; Pérez-Mayol, S.; Palmer, M.; Geffen, A.J. (2014). Coping with connectivity between populations of *Merluccius merluccius*: An elusive topic. *Journal of Marine Systems*, 138: 211-219.
- Morat F., Y. Letourneur, D. Nerini, D. Banaru, Ie. Batjakas. 2012. Discrimination of red mullet populations (Teleostean, Mullidae) along multi-spatial and ontogenetic scales within the Mediterranean basin on the basis of otolith shape analysis. *Aquatic Living Resources*, EDP Sciences, 25, pp.27-39.
- Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R. & Hohenlohe, P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22: 2841–2847.
- Natale F, Gibin M, Alessandrini A, Vespe M, Paulrud A (2015) Mapping Fishing Effort through AIS Data. *PLoS ONE* 10(6): e0130746. <https://doi.org/10.1371/journal.pone.0130746>
- Nielsen, E. E., A. Cariani, E. Mac Aoidh, G. E. Maes, I. Milano, R. Ogden, M. Taylor, J. Hemmer-Hansen, M. Babbucci, L. Bargelloni, D. Bekkevold, E. Diopere, L. Grenfell, S. Helyar, M. T. Limborg, J. T. Martinsohn, R. McEwing, F. Panitz, T. Patarnello, F. Tinti, J. K. J. Van Houdt, F. A. M. Volckaert, R. S. Waples, G. R. Carvalho, J. E. J. Albin, J. M. V. Baptista, V. Barmintsev, J. M. Bautista, C. Bendixen, J.-P. Berge, D. Blohm, B. Cardazzo, A. Diez, M. Espinera, A. J. Geffen, E. Gonzalez, N. Gonzalez-Lavin, I. Guarniero, M. Jerome, M. Kochzius, G. Krey, O. Mouchel, E. Negrisolo, C. Piccinetti, A. Puyet, S. Rastorguev, J. P. Smith, M. Trentini, V. Verrez-Bagnis, A. Volkov, A. Zanzi & C. FishPopTrace, 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications* 3 doi:10.1038/ncomms1845
- Panfili, J., Pontual, H. (de), Troadec, H. & Wright, P.J. (Eds.), 2002. *Manual of fish sclerochronology*. Coédition Ifremer-IRD, 464p.
- Passamonti, M., B. Mantovani, V. Scali & C. Frogliia, 1997. Allozymic Characterization of Scottish and Aegean Populations of *Nephrops norvegicus*. *Journal of the Marine Biological Association of the United Kingdom* 77(3):727-735 doi:10.1017/S0025315400036158.
- Pita, A., P. Presa & M. Perez, 2010. Gene flow, multilocus assignment and genetic structuring of the European hake (*Merluccius merluccius*). *Thalassas* 26(2):129-133.
- Pritchard, J. K., M. Stephens & P. Donnelly, 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155(2):945.
- Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, 10: 361–395.
- Roldán, M. I., S. Heras, R. Patellani & F. Maltagliati, 2009. Analysis of genetic structure of the red shrimp *Aristeus antennatus* from the Western Mediterranean employing two mitochondrial regions. *Genetica* 136(1):1-4 doi:10.1007/s10709-008-9330-2.
- Russo T, D'Andrea L, Parisi A, Cataudella S 2014a. VMSbase: An R-Package for VMS and Logbook Data Management and Analysis in Fisheries Ecology. *PLoS ONE* 9(6): e100195. <http://doi:10.1371/journal.pone.0100195>.
- Russo T, Parisi A, Garofalo G, Gristina M, Cataudella S, Fiorentino F. 2014b SMART: A Spatially Explicit Bio-Economic Model for Assessing and Managing Demersal Fisheries, with an Application to Italian Trawlers in the Strait of Sicily. *PLoS ONE* 9(1): e86222. doi:10.1371/journal.pone.0086222
- Russo T, D'Andrea L, Parisi A, Martinelli M, Belardinelli A, Boccoli F, Cignini I, Tordoni M, Cataudella S. 2016. Assessing the fishing footprint using data integrated from different tracking devices: issues and opportunities. *Ecological Indicators* 69: 818–827.
- Russo T., E.B. Morello, A. Parisi, G. Scarcella, S. Angelini, L. Labanchi, M. Martinelli, L. D'Andrea, A. Santojanni, E. Arneri, S. Cataudella 2018. A model combining landings and VMS data to estimate landings by fishing ground and harbor. *Fisheries Research* 199, 218-230. <https://doi.org/10.1016/j.fishres.2017.11.002>
- Russo T, D'Andrea L, Franceschini S, Accadia P, Cucco A, Garofalo G, Gristina M, Parisi A, Quattrocchi G, Sabatella RF, Sinerchia M., 2019a. Simulating the effects of alternative management measures of trawl fisheries in the Central Mediterranean Sea: application of a multi-species bio-economic modelling approach. *Frontiers in Marine Science*, 6: 542

- Russo T, Franceschini S, D'Andrea L, et al. 2019b. Predicting Fishing Footprint of Trawlers From Environmental and Fleet Data: An Application of Artificial Neural Networks. *Front Mar Sci* 6: 670.
- Ryman, N. & S. Palm, 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes* 6(3):600-602 doi:10.1111/j.1471-8286.2006.01378.x.
- Sardà, F., C. Bas, M. I. Roldán, C. Pla & J. Lleonart, 1998. Enzymatic and morphometric analyses in Mediterranean populations of the rose shrimp, *Aristeus antennatus* (Risso, 1816). *Journal of Experimental Marine Biology and Ecology* 221(1): 131-144 doi:10.1016/s0022-0981(97)00119-6.
- Sardà, F., M. I. Roldán, S. Heras & F. Maltagliati, 2010. Influence of the genetic structure of the red and blue shrimp, *Aristeus antennatus* (Risso, 1816), on the sustainability of a deep-sea population along a depth gradient in the Western Mediterranean. *Scientia Marina* 74(3):569-575 doi:10.3989/scimar.2010.74n3569.
- Spedicato M.T., G. Tserpes, B. Mérigot and E. Massutí (eds). 2019. Mediterranean demersal resources and ecosystems: 25 years of MEDITS trawl surveys. *Scientia Marina*.
- Swan, S.C.; Geffen, A.J.; Morales-Nin, B.; Gordon, J.D.M.; Shimmield, T.; Sawyer, T.; Massutí, E. (2006). Otolith chemistry: an aid to stock separation of *Helicolenus dactylopterus* (bluemouth) and *Merluccius merluccius* (European hake) in the Northeast Atlantic and Mediterranean. *ICES Journal of Marine Science*, 63: 504-513.
- Tanner, S.; Henrique N. Cabral H.N., Thorrold, S.R. 2012. Testing an otolith geochemistry approach to determine population structure and movements of European hake in the northeast Atlantic Ocean and Mediterranean Sea. *Fisheries Research* 125– 126: 198– 205.
- Tanner, S. E., M. Perez, P. Presa, S. R. Thorrold & H. N. Cabral, 2014. Integrating microsatellite DNA markers and otolith geochemistry to assess population structure of European hake (*Merluccius merluccius*). *Estuarine Coastal and Shelf Science* 142:68-75 doi:10.1016/j.ecss.2014.03.010.
- Tassetti AN, Ferrà C, Fabi G, 2019. Rating the effectiveness of fishery-regulated areas with AIS data, *Ocean Coast. Manag.* 175: 90-97. ISSN 0964-5691, <https://doi.org/10.1016/j.ocecoaman.2019.04.005>.
- Tassetti A.N., C. Ferra, G. Scarcella, G. Fabi 2016. MEDSEA_CH5_Product_6 / Impact of fisheries on the bottom from AIS data combined with habitat vulnerability. EMODnet Medsea Checkpoint. <https://doi.org/10.12770/56788352-9c7c-47cb-b039-4cb655b6a802>
- van den Wollenberg, A.L., 1977. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* 42, 207–219. <https://doi.org/10.1007/BF02294050>

LIST OF ELECTRONIC ANNEXES TO THE MED_UNITS FINAL REPORT

The following electronic Annexes are also available at: www.coispa.it

WPO - ANNEXES

ANNEX 0.1 - DELIVERABLE D0.2 - PROTOCOL OF BIOLOGICAL SAMPLING COORDINATION

WP1 - ANNEXES

ANNEX 1.1 - DELIVERABLE D1.2 - REPORT OF AVAILABLE INFORMATION ABOUT GENETICS STUDIES ON STOCK IDENTIFICATION AND CONNECTIVITY

ANNEX 1.2 - DELIVERABLE D1.3 - REPORT WITH THE REVIEW BIOINFORMATIC METHODS AND TOOLS

ANNEX 1.3 - DELIVERABLE 1.1 (revised) - SAMPLING PROTOCOL FOR THE GENETIC SAMPLING

ANNEX 1.4 - FOLLOW UP OF THE M1.2 - DEFINITION OF SITES FOR PILOT GENETIC STUDIES

ANNEX 1.5 - DELIVERABLE D1.4 - REPORT WITH THE RESULTS OF THE PILOT STUDIES

ANNEX 1.6 - DELIVERABLE D1.5.1 - DRAFT REPORT WITH THE RESULTS OF THE FULL STUDY FOR THE SPECIES *MERLUCCIVUS MERLUCCIVUS*, *ARISTAEOMORPHA FOLIACEA* AND *ARISTEUS ANTENNATUS*

ANNEX 1.7 - DELIVERABLE D1.5.2 - REPORT WITH THE RESULTS OF THE FULL STUDY FOR ALL THE PROJECT TARGET SPECIES

ANNEX 1.8 - DETAILED PROTOCOL FOR ROUTINE SAMPLING AND GENETIC MONITORING

WP2 - ANNEXES

ANNEX 2.1 - DELIVERABLE D2.1 - REPORT ON THE LITERATURE REVIEW ON THE OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

ANNEX 2.2 - PROTOCOL FOR OTOLITH EXTRACTION

ANNEX 2.3 - DELIVERABLE D2.2 - PROTOCOL FOR OTOLITH SHAPE ANALYSIS AND DATA-TREATMENT OF OTOLITHS OF EUROPEAN HAKE AND RED MULLET

ANNEX 2.4 – DELIVERABLE 2.3 - PROTOCOL FOR ANALYSIS AND DATA-TREATMENT OF TRACE ELEMENTS IN OTOLITHS OF EUROPEAN HAKE AND RED MULLET

ANNEX 2.5 - DELIVERABLE 2.4 - REPORT ON THE RESULTS OF OTOLITH SHAPE ANALYSIS AND MULTIVARIATE ANALYSIS OF EUROPEAN HAKE AND RED MULLET

ANNEX 2.6 - DELIVERABLE D2.5 - REPORT ON EUROPEAN HAKE AND RED MULLET MICROCHEMISTRY ANALYSIS

ANNEX 2.7 – DELIVERABLE D2.6 - MATRIX GENERATION FOR JOINT ANALYSIS OF STOCK DELINEATION

WP3 - ANNEXES

ANNEX 3.1 - DELIVERABLE D3.1 - MAPS ON: (A) THE FISHING GROUNDS IN TEMPORAL AND SPATIAL SCALE, (B) HOT AND COLD SPOTS

ANNEX 3.2 - DELIVERABLE D3.2 - MAPS COMBINING THE FISHING GROUNDS AND THE SPATIAL DISTRIBUTION OF THE TARGET SPECIES

ANNEX 3.3 - DELIVERABLE 3.3 - REPORT ON THE STOCK ASSESSMENTS WITH THE NEW STOCK CONFIGURATIONS FOR THE 6 TARGET SPECIES OF THE STUDY

WP4 - ANNEXES

ANNEX 4.1 - DELIVERABLE D4.2 - REPORT WITH THE MAPS OF POPULATION UNITS AND DISCONTINUITIES

ANNEX 4.1A –MED_UNITS DELIVERABLE 4.2 ANNEXES

ANNEX 4.2 - DELIVERABLE 4.3 - REPORT WITH THE MAPS COMBINING POPULATION AND MANAGEMENT UNITS

ANNEX 4.3 - DELIVERABLE 4.4 - REPORT ON THE SCENARIO MODELLED TO EXPLORE THE POTENTIAL OF DEVELOPING ADAPTIVE SPATIAL FISHERIES MANAGEMENT

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP0 - PROJECT MANAGEMENT AND COORDINATION

Deliverable 0.2
Protocol of biological sampling coordination

Responsible: Paolo Sartor

Date: March 22, 2019

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

Deliverable 1.2 – Report of available information about genetic studies on stock identification and connectivity

LEAD: RITA CANNAS (CONISMA)

PARTICIPANTS: COSTAS TSIGENOPOULOS (HCMR), ALESSIA CARIANI, RICCARDO MELIS, LORENZO ZANE (CONISMA)

Date: March 2nd, 2019

Contents

EXECUTIVE SUMMARY	5
D1.2 Objective	5
D1.2: methodology	5
D1.2: main results	5
Aristaeomorpha foliacea	9
Aristeus antennatus	9
Merluccius merluccius	9
Mullus barbatus	9
Nephrops norvegicus	10
Parapenaeus longirostris	10
OBJECTIVE	11
METHODOLOGY	11
Data sources	11
THE FINAL DATASET	14
THE ANALYSIS: ARISTAEOMORPHA FOLIACEA (GIANT RED SHRIMP)	30
Information from studies including Mediterranean population samples (HP datasets)	30
GDMU_Af01=Af03	30
GDMU_Af02	32
GDMU_Af04	35
GDMU_Af05	37
Information from studies including population samples outside the Mediterranean Sea (P datasets)	39
Comments and evaluation of data: giant red shrimp	39
THE ANALYSIS: ARISTEUS ANTENNATUS	40
Information from studies including Mediterranean population samples (HP datasets)	40
GDMU_Aa03	40
GDMU_Aa05	42
GDMU_Aa06	44
GDMU_Aa07	46
GDMU_Aa08	48
GDMU_Aa09	50
GDMU_Aa10	52
GDMU_Aa11	54
GDMU_Aa12	56
Information from studies including population samples outside the Mediterranean Sea (P datasets)	58
Comments and evaluation of data: BLue and red shrimp	58

THE ANALYSIS: <i>MERLUCCIUS MERLUCCIUS</i>	59
Information from studies including Mediterranean population samples (HP datasets)	59
GDMU_Mm01	59
GDMU_Mm02	61
GDMU_Mm03	63
GDMU_Mm07	65
GDMU_Mm08	67
GDMU_Mm09	69
GDMU_Mm12	71
GDMU_Mm15	74
GDMU_Mm18	77
GDMU_Mm19	80
GDMU_Mm20	83
Information from studies including population samples outside the Mediterranean Sea (P datasets)	85
Comments and evaluation of data: European hake	86
THE ANALYSIS: <i>MULLUS BARBATUS</i>	87
Information from studies including Mediterranean population samples (HP datasets)	87
GDMU_Mb03	87
GDMU_Mb05	90
GDMU_Mb06	93
GDMU_Mb10	95
GDMU_Mb11	97
GDMU_Mb15	99
GDMU_Mb16	101
GDMU_Mb18	103
GDMU_Mb20	106
GDMU_Mb24	108
Information from studies including population samples outside the Mediterranean Sea (P datasets)	111
Comments and evaluation of data: red mullet	111
THE ANALYSIS: <i>NEPHROPS NORVEGICUS</i>	112
Information from studies including Mediterranean population samples (HP datasets)	112
GDMU_Nn02	112
GDMU_Nn03	114
GDMU_Nn06	116
GDMU_Nn07	119
GDMU_Nn14	122
Information from studies including population samples outside the Mediterranean Sea (P datasets)	124
Comments and evaluation of data: Norway lobster	124
THE ANALYSIS: <i>PARAPENAEUS LONGIROSTRIS</i>	125
Information from studies including Mediterranean population samples (HP datasets)	125
GDMU_PI05	125
GDMU_PI06	128

Information from studies including population samples outside the Mediterranean Sea (P datasets)131

Comments and evaluation of data: deep-water rose shrimp131

CONCLUSIONS..... 132

GLOSSARY AND ACRONYMS 134

Executive Summary

D1.2 OBJECTIVE

Deliverable 1.2 compiles and reviews all the available genetic information for the 6-target species in the Mediterranean Sea: giant red shrimp (*Aristaeomorpha foliacea*), blue and red shrimp (*Aristeus antennatus*), European hake (*Merluccius merluccius*), red mullet (*Mullus barbatus*), Norway lobster (*Nephrops norvegicus*) and deep water rose shrimp (*Parapenaeus longirostris*).

The review contributes to identify stock boundaries, connectivity patterns and population genetic structure of the target species. This information represents the basis for the of sampling design (Task 1.3) and the new genetic analyses performed in Tasks 1.4 and Task 1.5.

Applicability, advantages and disadvantages of different methodologies and techniques, as well as their limitations and strengths will be briefly addressed but they will be deeply discussed in the full methodological review included in D1.3.

D1.2: METHODOLOGY

This systematic review is based on peer-review papers, grey literature, and data from national and European research activities. In particular, the main bulk of D1.2 is derived from Deliverable 7 (Cannas et al., 2014¹) produced during the MAREA Stockmed project (Fiorentino et al., 2014²). In MED_UNITS, the data mentioned above have been updated and revised, including genetic information from the papers published in the period 2014 to 2019.

To update the data from published sources, we have followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach (Moher et al., 2009³).

All publications/reports that contained information on genetic differentiation/structuring in the investigated species were considered pertinent to this review. The datasets (reports/papers) were subdivided in three categories:

- **High Priority (HP):** the ones that contained at least a sample population for the target species in the Mediterranean Sea.
- **Priority (P):** the ones that contained at least a sample population for the target species in the Black Sea or adjacent Atlantic Ocean.

D1.2: MAIN RESULTS

¹ Cannas, R., Cariani, A., Ferrari, A., Cau, A., Follesa, M.C., Tsigenopoulos, C. and Tinti, F. (2014) Individual reporting species sheets with genetic estimates of population differentiation indicators/parameters. Deliverable 7 MAREA (MEDITERRANEAN HALIEUTIC RESOURCES EVALUATION AND ADVICE) Specific Contract no 7 - STOCKMED: "Stock units: Identification of distinct biological units (stock units) for different fish and shellfish species and among different GFCM-GSA" (SI2.642234), COMMISSION OF THE EUROPEAN COMMUNITIES Directorate-General for Fisheries.

² Fiorentino, F., Massuti, E., Tinti, F., Somarakis, S., Garofalo, G., Russo, T., Facchini, M.T., Carbonara, P., Kapiris, K., Tugores, P., Cannas, R., Tsigenopoulos, C., Patti, B., Colloca, F., Sbrana, M., Mifsud, R., Valavanis, V. and Spedicato, M.T. (2014) Stock units: Identification of distinct biological units (stock units) for different fish and shellfish species and among different GFCM-GSA. STOCKMED Deliverable 03: FINAL REPORT, COMMISSION OF THE EUROPEAN COMMUNITIES Directorate-General for Fisheries.

³ Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman & P. G. The, 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Medicine 6(7):e1000097 doi:10.1371/journal.pmed.1000097.

In brief, a total of 48 datasets provided useful information on population genetic differentiation for the target species, 41 within the Mediterranean and other 7 from the Black Sea or the adjacent Atlantic Ocean. In general, the fish species have a higher number of studies compared to the crustaceans, with *M. merluccius* being the most extensively studied species and *P. longirostris* the less investigated one (Figure 1). The most recent and high throughput markers (Single Nucleotide Polymorphisms, SNPs) have been applied on the European hake only. A larger number of studies have been based on microsatellite markers which are still considered as highly informative, once developed in the species under study; they have been used in four species, both in multiple studies per species (*M. merluccius* and *M. barbatus*) or in a single study (*A. foliacea* and *A. antennatus*). Old and less resolutive markers (i.e., proteins, RAPD, mtDNA sequences etc.) have been used in all the six species. In overall, the most informative markers (SNPs) are also the most demanding in terms of costs and methodological procedures, thus their inherent constraints could limit the number of specimens that are analysed and the spatial/temporal coverage of the study unless an adequate budget is allocated for the sampling and the analyses.

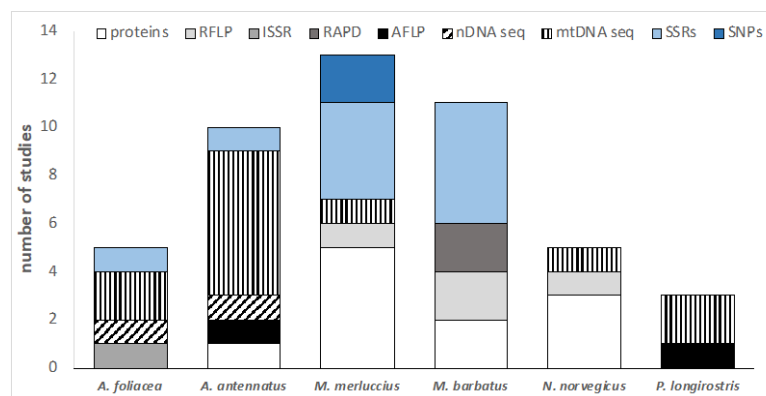


FIGURE 1 POPULATION GENETIC STUDIES AND RELATIVE GENETIC MARKERS USED IN THE 6 TARGET SPECIES

Table 1 summarizes in a concise form all the most relevant information obtained from each dataset: the extent of the area/samples investigated, the strength of the data in support for the genetic clustering, the number of clusters.

TABLE 1 SUMMARY OF THE MAIN RELEVANT INFORMATION FOR THE 6 TARGET SPECIES

species	GDMU	Genetic markers	N GSA	Mean N i/s	Overall differentiation	Genetic structure (clusters)	Number of Mediterranean clusters
<i>A. foliacea</i>	Af01	SSRs (6)	★ ^{WC}	★	-	-	1(WC)
<i>A. foliacea</i>	Af02	ISSR	★ ^{WCE}	★★★	★	-	1(WCE)
<i>A. foliacea</i>	Af04	mtDNA (COI)	★ ^{WCE}	★★	★★★	★	2(WC/CE)
<i>A. foliacea</i>	Af05	nDNA (PEPCK and NaK) + mtDNA (COI)	★ ^{WE}	★	★★★	-	1(WE)
<i>A. antennatus</i>	Aa03	Allozymes (15)	★ ^{WCE}	★★	★	-	1(WCE)
<i>A. antennatus</i>	Aa05	mtDNA (CR)	★★★ ^{WC}	★★	★	-	1(WC)
<i>A. antennatus</i>	Aa06	mtDNA (cytb+16S)	★ ^W	★★★	★	★	2(W)
<i>A. antennatus</i>	Aa07	mtDNA (16S)	★ ^W	★★★	-	-	1(W)
<i>A. antennatus</i>	Aa08	mtDNA (COI+16S)	★ ^{WE}	★★★	★★★	★	2(W/E)
<i>A. antennatus</i>	Aa09	SSRs (8)	★ ^W	★★	-	-	1(W)
<i>A. antennatus</i>	Aa10	AFLP +mtDNA(CR)	★ ^{WC}	★★	★★★	-	1(WC)
<i>A. antennatus</i>	Aa11	nDNA (PEPCK and NaK) + mtDNA (COI)	★ ^{WE}	★	-	-	1(WE)
<i>A. antennatus</i>	Aa12	mtDNA (COI+16S)	★ ^{WC}	★★	★	-	1(WC)
<i>M. merluccius</i>	Mm01	Allozymes (20)	★★★ ^{WCE}	★★★	-	-	1(WCE)
<i>M. merluccius</i>	Mm02	Allozymes (34)	★★★ ^W	★★★	-	-	1(W)
<i>M. merluccius</i>	Mm03	SSRs (6)	★★★ ^{CE}	★★★	★	-	1(CE)
<i>M. merluccius</i>	Mm07	SSRs (5)	★★★ ^{WCE}	★★★	★	★	2(WC/E)
<i>M. merluccius</i>	Mm08	Allozymes (5)	★★★ ^C	★★★	-	-	1(C)
<i>M. merluccius</i>	Mm09	Allozymes (5 loci) + PCR-RFLP (CR)	★★★ ^{WC}	★	★	-	1(WC)
<i>M. merluccius</i>	Mm12	allozymes	★★★ ^{WCE}	★★★	★	★	? cline
<i>M. merluccius</i>	Mm15	SSRs (5) + mtDNA (cytb)	★★★ ^W	na	★	-	1(W)
<i>M. merluccius</i>	Mm18	SNPs (72 outlier)	★★★ ^{WCE}	★★★	★★★	-	1(WCE)
<i>M. merluccius</i>	Mm19	SNPs (381)	★★★ ^{WCE}	★★★	★★★	★★★	4(W/WC/WC/E)
<i>M. merluccius</i>	Mm20	SSRs (5)	★★★ ^W	★★★	★	★	1(W)
<i>M. barbatus</i>	Mb03	RAPD	★★★ ^E	★	-	-	1(E)
<i>M. barbatus</i>	Mb05	Allozymes (14)	★★★ ^{WE}	★★★	★	★	2(WE/E)
<i>M. barbatus</i>	Mb06	Isozymes (25)	★★★ ^{WCE}	★★★	★	-	1(WCE)
<i>M. barbatus</i>	Mb10	RFLP	★★ ^{WCE}	★★	-	-	1(WCE)

		(COI+CR+12s16S)					
<i>M. barbatus</i>	Mb11	SSRs (5)	★★ ^W	★★★	★	★	2(C)
<i>M. barbatus</i>	Mb15	allozymes, RAPDs and mtDNA (RFLPs)	★★ ^E	★★	★★	★★	2(E)
<i>M. barbatus</i>	Mb16	SSRs (10)	★★★ ^W	★★	★	★	4(2W/1WC/E)
<i>M. barbatus</i>	Mb18	SSRs (6)	★★★ ^{WC}	★★★	★	★	2(WC/C)
<i>M. barbatus</i>	Mb20	SSRs (10)	★★ ^W	★★	-	-	1(W)
<i>M. barbatus</i>	Mb24	SSRs (13)	★★★ ^{WCE}	★★★	★	★★	3(WC/C/CE)
<i>N. norvegicus</i>	Nn02	allozymes (17)	★★ ^{CE}	na	-	-	1(CE)
<i>N. norvegicus</i>	Nn03	allozymes (15)	★★★ ^{WCE}	★★	★	★★★	4(2W/1WE/C)
<i>N. norvegicus</i>	Nn06	mtDNA (RFLP)	★★★ ^E	★★★	★	-	1(E)
<i>N. norvegicus</i>	Nn07	allozymes (15)	★★★ ^E	★★★	★	★	3(E)
<i>N. norvegicus</i>	Nn14	mtDNA (CR)	★★★ ^C	★★★	★	★★	1(C)
<i>P. longirostris</i>	Pl05	mtDNA (COI)	★★ ^W	★	-	-	1(W)
<i>P. longirostris</i>	Pl06	AFLP+ mtDNA (CR)	★★★ ^{WCE}	★★	★★★	★★★	4(W/2C/E)

Where:

GDMU = Genetic Dataset MED_Units

Genetic marker = type of molecular marker used;

N GSA = Number of GFCM GSA investigated: ★≤3; ★★4-6; ★★★>6; W, C, E indicates that sampling sites included are from the Western, Central, Eastern Mediterranean, [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/site ★≤20; ★★20-40; ★★★>40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall F_{ST} : ★≤0.05; ★★0.05-0.1; ★★★>0.1)

Genetic structure = ★ low, ★★ medium ★★★ high support (whenever available the score is based on pairwise F_{ST} values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.));

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean;

The following paragraphs summarize all the most relevant information obtained from each species.

Aristaeomorpha foliacea

A few studies (n=5) have investigated the giant red shrimp population genetics within the Mediterranean Sea using different methods (i.e., ISSR, microsatellites, mitochondrial and nuclear genes). Microsatellites data (Cannas et al 2012; Marcias et al., 2010) did not allow identifying significant differentiation among samples from Sardinia and Sicily, suggesting the lack of genetic differentiation among these two areas. Similarly, a substantial genetic homogeneity was identified within the Mediterranean using ISSR and nuclear genes (Fernandez et al 2011a; Fernandez et al 2013b). Using mitochondrial DNA (COI gene sequences, 685 bp), Fernandez et al (2013a) identified the occurrence of genetic differences within the Mediterranean, suggesting that a certain degree of genetic differentiation was present among the Mediterranean local samples. However, when Mediterranean samples were grouped in western and eastern basins, the variation among groups was not statistically significant, even if differences in the frequencies of mitochondrial haplogroups were detected between the Western (Ibiza, Tyrrhenian Sea and Sicily) and the Central-Eastern Mediterranean locations (Ionian and Aegean). The locality of MAZ (Mazara, Strait of Sicily) displayed an intermediate differentiation.

Aristeus antennatus

Several studies (n=9) have investigated the blue and red shrimp population genetics within the Mediterranean Sea using different methods (i.e. allozymes, AFLP, microsatellites, mitochondrial DNA). The vast majority of studies using all kinds of markers (allozymes: Sarda et al. 1998; microsatellites: Cannas et al. 2012, mtDNA sequences: Fernández et al. 2013b, Maggio et al. 2009, Marra et al. 2015; Roldán et al. 2009; Sardà et al. 2010; AFLP: Lo Brutto et al. 2012) did not allow identifying significant differentiation among areas, suggesting a substantial genetic homogeneity within the Mediterranean. On the contrary, Fernandez et al., 2011b using mitochondrial DNA (COI gene sequences, and 16S) indicated the occurrence of genetic differentiation among geographical regions (AO Atlantic Ocean, AS Alboran Sea, WM Western Mediterranean, EM Eastern Mediterranean, and IO Indian Ocean). The outcome of AMOVAs suggested that the AS sample is part of WM. Tests aimed at detecting population substructuring within the WM region (AS included) showed that there was also a small but significant variance among population ($\Phi_{ST} = 0.015$, $p = 0.005$). AMOVA analyses and Bayesian assignment of individuals clearly pointed out the effectiveness of Gibraltar Strait and Sicily Strait constrictions in reducing gene flow and, therefore, producing genetic differentiation between regions.

Merluccius merluccius

Several studies (n=11) have investigated the European hake population genetics within the Mediterranean Sea using different methods (i.e. allozymes, mtDNA-RFLP, mtDNA sequences, microsatellites and SNPs). In general, several studies identified a substantial homogeneity within the Mediterranean but suggested a strong subdivision of Atlantic and Mediterranean hake stocks (Lo Brutto et al., 1998; Roldan et al., 1999; Lundy et al., 1999; Castillo et al., 2004; Levi et al., 2004; Lo Brutto et al., 2004; Pita et al., 2010; Tanner et al., 2014). On the contrary, other studies (Castillo et al., 2004, Cimmaruta et al., 2005) described the occurrence of genetic heterogeneity within the Mediterranean Sea. Similarly, Nielsen et al. (2012) and Milano et al. (2014), using a big number of SNPs with significantly higher resolving power, confirmed the genetic break between Atlantic and Mediterranean populations and described a finer-scale significant genetic population structure. In particular, in the Mediterranean outlier SNPs revealed a strong differentiation among Western, Central and Eastern Mediterranean geographical samples.

Mullus barbatus

Several studies (n= 10) have investigated the red mullet population genetics within the Mediterranean Sea using several methods (i.e., RAPDs, allozymes, PAGE, mtDNA-RFLP, nuclear microsatellite markers). The first genetic studies (based on allozymes, RAPD, PAGE, mtDNA-RFLP), failed to identify differentiation among samples or described feeble genetically different populations within the Aegean Sea (Mamuris et al., 1998 a, c; Arculeo et al., 1999; Mamuris et al., 2001; Apostodolis et al., 2009). Using microsatellites, the Adriatic Sea was found to be differentiated from the other investigated areas (Maggio et al., 2009), whereas red mullet populations along the Mediterranean Spanish coast were genetically homogeneous (Felix-Hackradt et al., 2013; GSA1 and the southern part of GSA6). Finally, Matic-Skoko et al. (2018) identified three different genetic clusters coexisting in the

Adriatic Sea. Contrasting results were obtained for distant locations like the Western and the Eastern Mediterranean; they were found to be non significantly diverse by Matic-Skoko et al. (2018), but differentiated in distinct clusters by Galarza et al. (2009).

Nephrops norvegicus

A few studies (n=5) have investigated the Norway lobster population genetics within the Mediterranean Sea using different methods (i.e. allozymes, mtDNA RFLP, mtDNA CR). Allozyme data (Passamonti et al., 1997; Maltagliati et al., 1998) revealed low or moderate genetic differentiation between geographical regions (Atlantic vs Mediterranean) but no geographical pattern of genetic differentiation, thus genetic variability seems to be randomly distributed among populations. Microsatellites have been isolated but used only in a single study in Iceland (Pampoulie et al., 2011), never in the Mediterranean Sea.

Parapenaeus longirostris

Very few studies (n=2) have investigated the deep-water rose shrimp population genetics within the Mediterranean Sea using mitochondrial DNA markers and AFLP. Garcia-Merchan et al., (2012) analysed the variability of the mtDNA gene COI along the Spanish coasts, and found a substantial homogeneity among locations. Lo Brutto et al. (2013), using mtDNA (CR) and AFLP, identified four clusters (according to the Mediterranean sub basins: Tyrrhenian Vs Strait of Sicily Vs Adriatic Vs Aegean Sea) significantly differentiated. The greatest contribution to the differences among the four Mediterranean sub basins depended on Aegean and Tyrrhenian areas, which represented the most divergent groups.

The full details of the different datasets, data and comments, are reported in the following paragraphs.

Objective

The aim of Task 1.1 is to review and compile all the available genetic information for the 6-target species in the Mediterranean Sea: giant red shrimp (*Aristaeomorpha foliacea*), blue and red shrimp (*Aristeus antennatus*), European hake (*Merluccius merluccius*), red mullet (*Mullus barbatus*), Norway lobster (*Nephrops norvegicus*) and deep water rose shrimp (*Parapenaeus longirostris*).

Therefore, D1.2 contains all the relevant data for the six species along with genetic estimates of population differentiation (indicators/parameters).

This review contributes to identifying stock boundaries, connectivity patterns and population genetic structure of the target species. This information represents the basis for the of sampling design (Task 1.3) and the new genetic analyses performed in Tasks 1.4 and Task 1.5.

Applicability, advantages and disadvantages of different methodologies and techniques, as well as their limitations and strengths are briefly addressed but they will be deeply discussed in the comprehensive methodological review included in another document produced within Task 1.2, that is deliverable D1.3

Methodology

The systematic review is based on peer-review papers, grey literature, and data from national and European research activities. In particular, the main bulk of this review is derived from Deliverable 7 (Cannas *et al.*, 2014) produced during the MAREA Stockmed project (Fiorentino *et al.*, 2014). In MED_UNITS, the review mentioned above has been updated and revised, including genetic information from the papers published in the period 2014 to 2018.

To update the data from published sources, we have followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach (Moher *et al.*, 2009). This approach consists of three steps: (1) systematic article selection using a search engine and precise keywords; (2) article screening; and (3) review of all articles and extraction of information from relevant sources.

DATA SOURCES

We conducted a systematic review of the published papers using a protocol based on the PRISMA statement (Moher *et al.*, 2009, Moher *et al.*, 2015). The electronic bibliographic databases ISI Web and Scopus were searched in the period 3 January 2019 – 15 February 2019. The search was performed using the specific keywords (Table 1), combining the scientific species name or the common species name with the words listed below (A+B and A+C). Additional references were added to the review, when found cited in the publications retrieved in the bibliographic databases.

Table 2 Keywords used in the systematic review

Keyword category	Species 1	Species 2	Species 3	Species 4	Species 5	Species 6
A species name	<i>Aristaeomorpha foliacea</i>	<i>Aristeus antennatus</i>	<i>Merluccius merluccius</i>	<i>Mullus barbatus</i>	<i>Nephrops norvegicus</i>	<i>Parapenaeus longirostris</i>
	Giant red shrimp	Blue and Red shrimp	European hake	Red mullet	Norway lobster	deep water rose shrimp
B genetic	Allozymes	Allozymes	Allozymes	Allozymes	Allozymes	Allozymes
	Microsatellites	Microsatellites	Microsatellites	Microsatellites	Microsatellites	Microsatellites
	Mitochondrial DNA	Mitochondrial DNA	Mitochondrial DNA	Mitochondrial DNA	Mitochondrial DNA	Mitochondrial DNA
	SNPs	SNPs	SNPs	SNPs	SNPs	SNPs
	Genetics	Genetics	Genetics	Genetics	Genetics	Genetics
	Population genetics	Population genetics	Population genetics	Population genetics	Population genetics	Population genetics
	Genetic structure	Genetic structure	Genetic structure	Genetic structure	Genetic structure	Genetic structure
Gene flow	Gene flow	Gene flow	Gene flow	Gene flow	Gene flow	
C generic	Stock identification	Stock identification	Stock identification	Stock identification	Stock identification	Stock identification
	Fishery stocks	Fishery stocks	Fishery stocks	Fishery stocks	Fishery stocks	Fishery stocks
	Population structure	Population structure	Population structure	Population structure	Population structure	Population structure

From the results of the research, all the titles and abstracts found were read, and from these, only articles related to the search criteria were retained, while duplications and items that did not meet the inclusion criteria were removed from the search. All publications/reports that contained information on genetic differentiation/structuring in the investigated species were considered pertinent to this review. Phylogenetics and genetic species identification studies (e.g. barcoding on fresh or canned fishes) were excluded from this review because a priori they have little to do with the objective of this task.

Each dataset (paper/report) was coded as follows: GDMU_AaXX/GDMU_AfXX/ GDMU_MmXX/ GDMU_MbXX/ GDMU_NnXX/ GDMU_PiXX/ (GDMU=Genetic Dataset MEDUNITS; Af = *A. foliacea*, Aa = *A. antennatus*; Mm = *M. merluccius*; Mb = *M. barbatus*; Nn = *N. norvegicus*; Pl = *P. longirostris*; XX = progressive number).

The datasets were further subdivided in three categories:

- **High Priority (HP)**: the ones that contained at least a sample population for the target species in the Mediterranean Sea.
- **Priority (P)**: the ones that contained at least a sample population for the target species in the Black Sea or adjacent Atlantic Ocean.

All datasets have been scrutinized searching for four main type of information:

1 Genetic marker used/ 2 Details of the sampling sites analysed/ 3 Main findings/ 4 Genetic differentiation and stock structuring.

However, only the data from the High Priority datasets are reported in full details in D1.2, while the information from the other datasets (P) is summarized in few paragraphs.

In particular, for the high priority datasets, the identification of genetic differentiation and/or structuring was realized in three steps:

Step one: Scrutiny of the dataset

Step two: Matrix of clustering

Step three: Geo-visualization of clustering

Step one: Since no raw data was available for inspection or re-analysis, the statement of genetic differentiation and/or structuring among populations totally rely on the results from the published papers. However, the description reported was not always complete, accurate or appropriate, in terms of sampling size and design, genetic marker used, and/or data analysis performed. Whenever available, p-values indicating genetic structure were recorded [X^2 tests (Allozymes)/ F_{ST} and/or clustering methods (mtDNA, Microsatellites, SNPs)]. To facilitate the illustration of the main results, the original relevant tables and figures from the literature were also included in D1.2.

Step two: According to the approach used in Stockmed, the genetic data has been transformed into binary matrixes for each GDMU data sets, in order to facilitate the use and integration of genetic data with those from other tasks. Each population sample has been assigned to a specific Cluster (score 1) according to the results in the literature. In some cases, Clusters have been further divided in sub-Clusters to better illustrate more complex patterns of genetic differentiation emerging from the data.

Step three: The genetic variation and structure of GDMU data sets has been illustrated through a Google Earth map on the GFCM Geographical Subareas (GSAs) grid, using the original coordinates (when available) or rough coordinates inferred from sampling maps/locations of the reference literature. The sampling sites have been coloured in order to reflect the estimates of population differentiation provided in each reference literature. The whole process of identifying genetic differentiation and structuring has been realized for each GDMU at least twice, in parallel through the independent work of experts' participant to Task 1.1.

It is worth stressing that the choice of using the software 'Google Earth' instead of other Gis software was largely due to practical reasons (easiness of use and spread of results even for non-specialists) and the very 'low' quality/quantity of the geo-referentiated genetic data that did not require a sophisticated representation.

When multiple high priority datasets are available for a given species, they are all analysed in detail.

A final paragraph for each species summarizes the information provided by the different datasets; to facilitate the comparisons and have a quick overview of the main results, a table has been included for each species summarizing the most relevant features of each dataset:

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic structure (clusters)
XX	★ ^{WCE}	★	★	★

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★ ≤3; ★★ 4-6; ★★★ >6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★ ≤20; ★★ 20-40; ★★★ >40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall F_{ST} : ★ ≤0.05; ★★ 0.05-0.1; ★★★ >0.1)

Genetic structure = ★ low, ★★ medium ★★★ high support (whenever available the score is based on pairwise F_{st} values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

A more detailed discussion on the technical aspects (e.g. the different resolutive power of genetic markers) and/or the impact of different sampling designs will be further addressed within Task 1.2. and Deliverable 1.3.

The final dataset

According to the PRISMA guidelines, a flow diagram (Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, Figure 6) was produced for each species to ensure for the good quality of the systematic review.



PRISMA 2009 Flow Diagram *Aristaeomorpha foliacea*

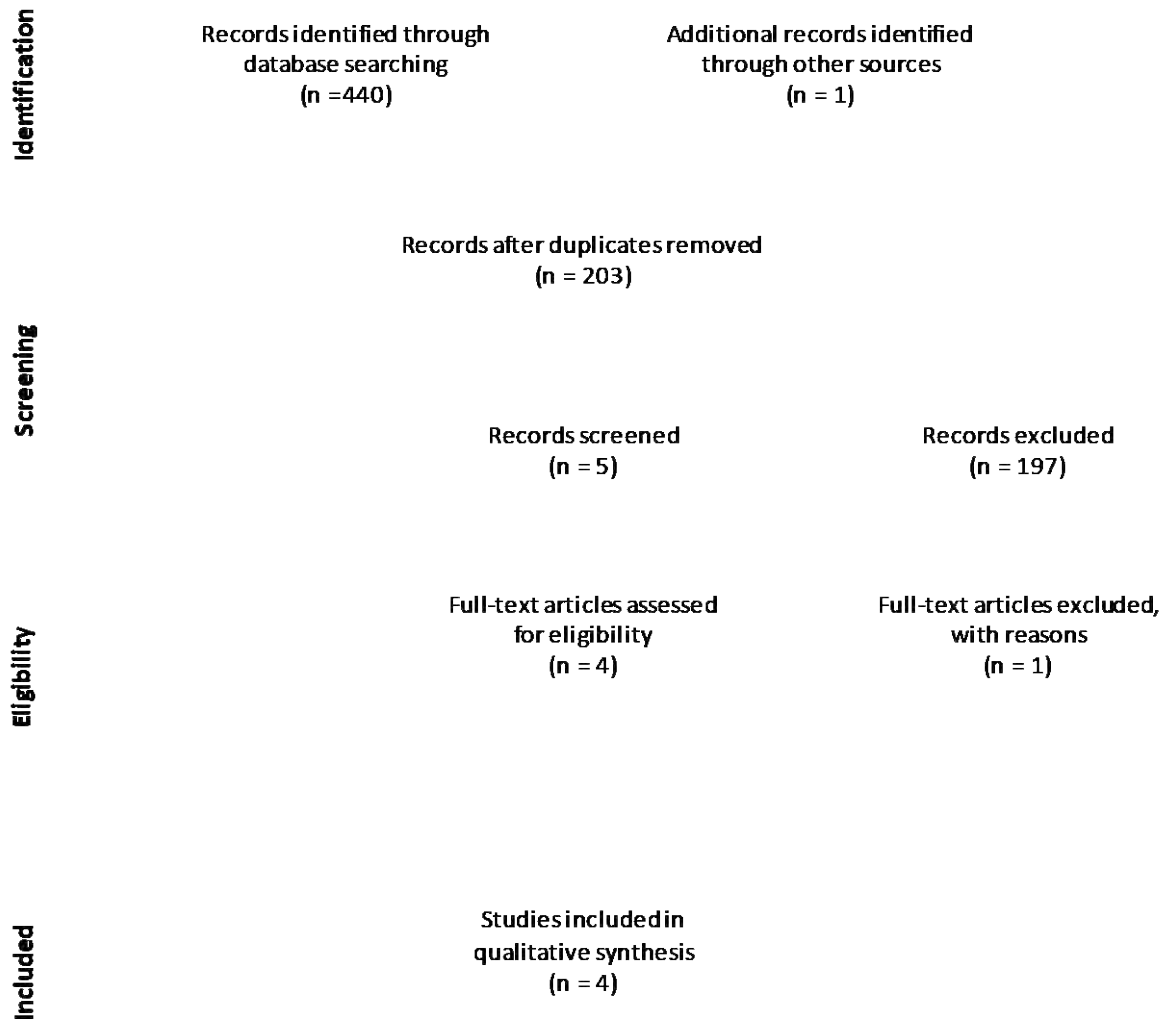


FIGURE 2



PRISMA 2009 Flow Diagram *Aristeus antennatus*

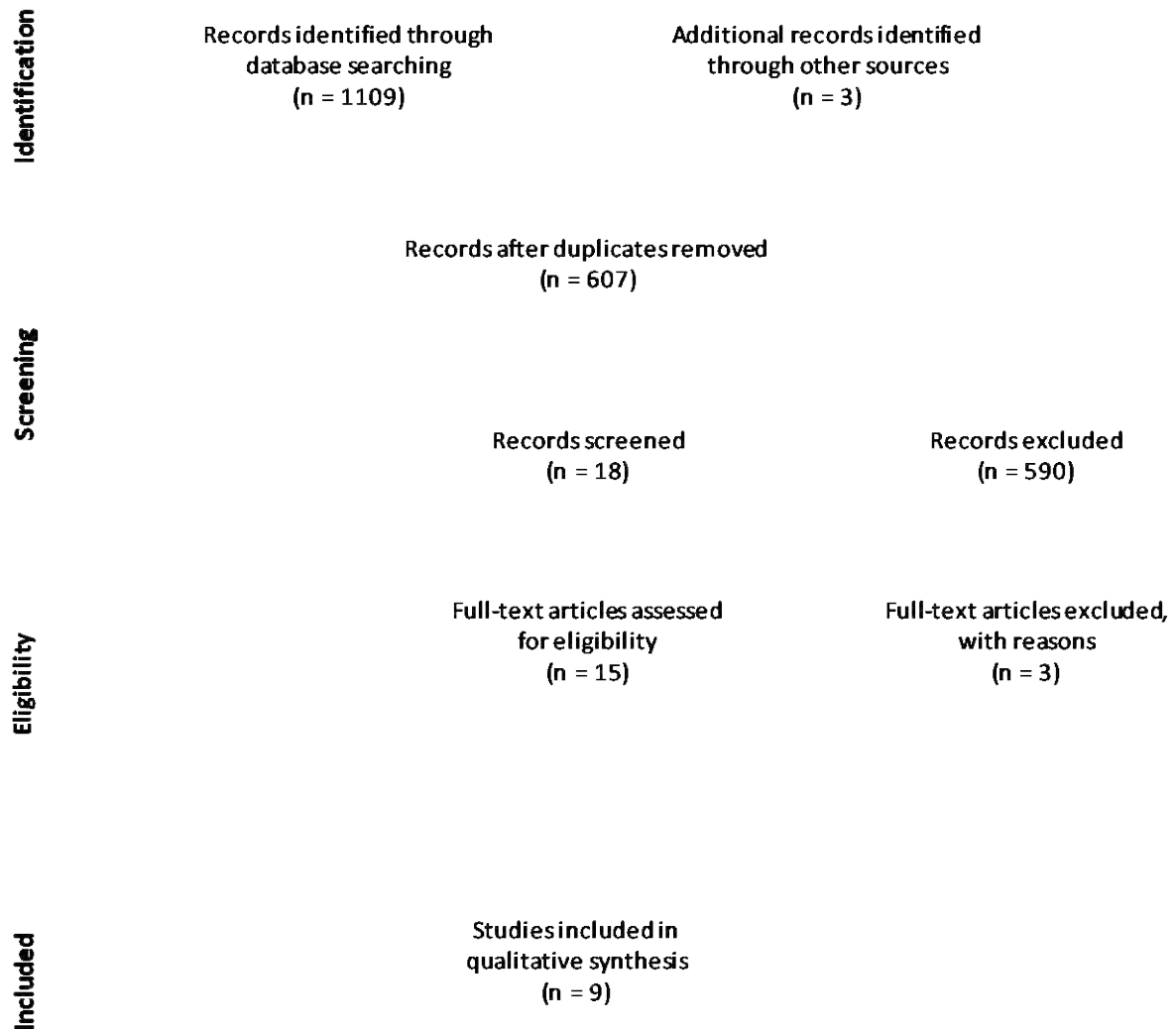


FIGURE 3



PRISMA 2009 Flow Diagram *Merluccius merluccius*

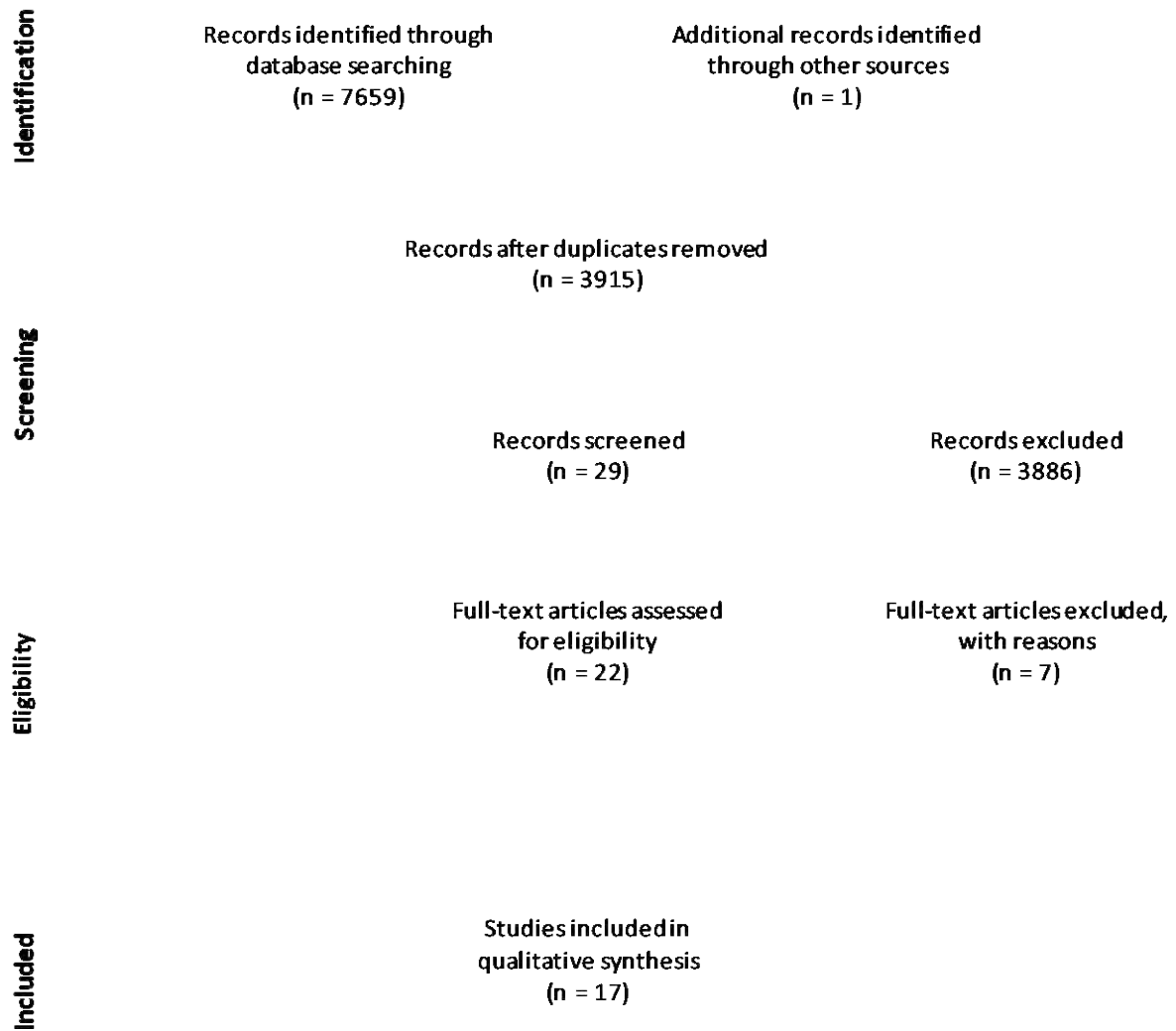


FIGURE 4



PRISMA 2009 Flow Diagram *Mullus barbatus*

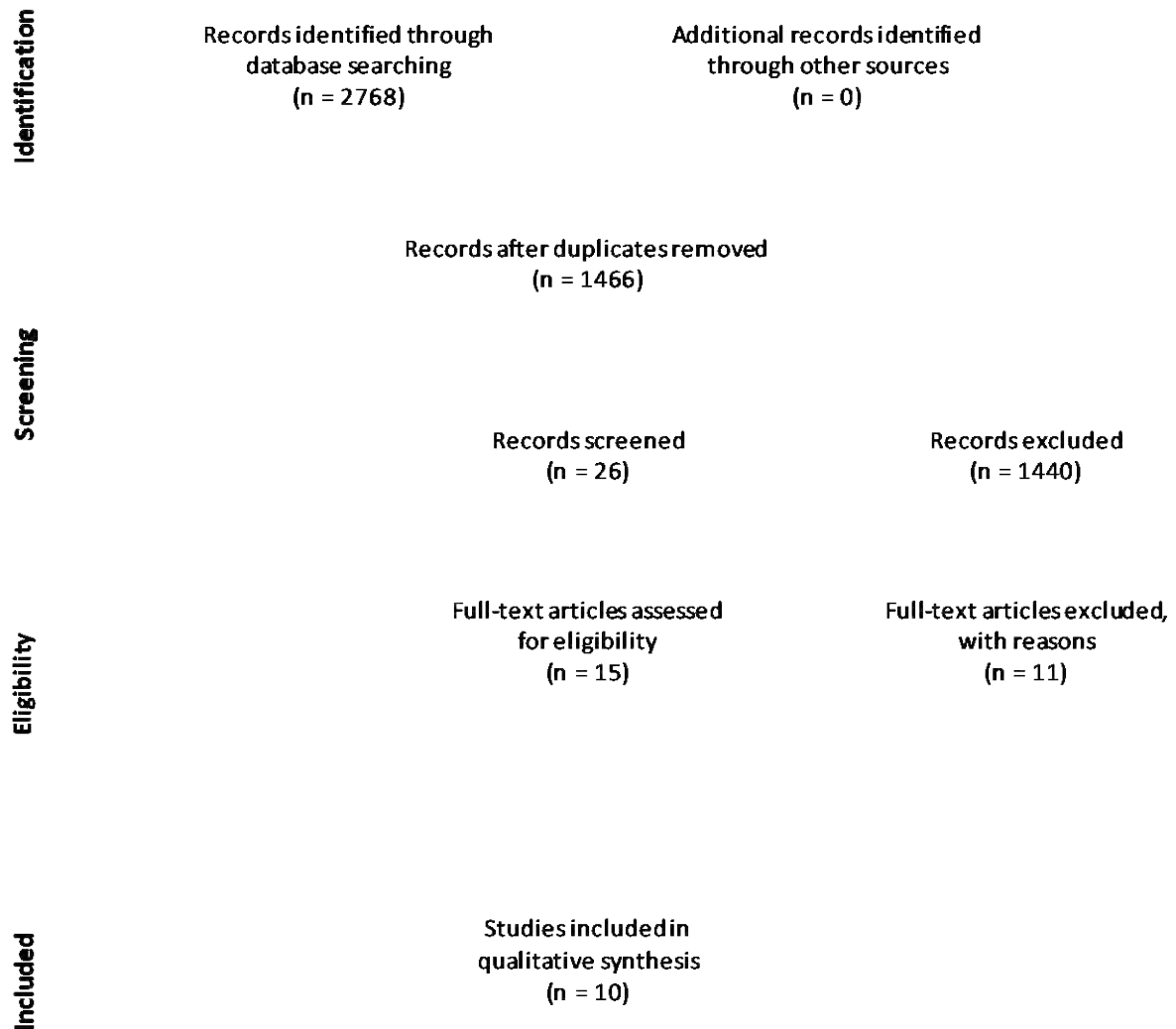


FIGURE 5



PRISMA 2009 Flow Diagram *Nephrops norvegicus*

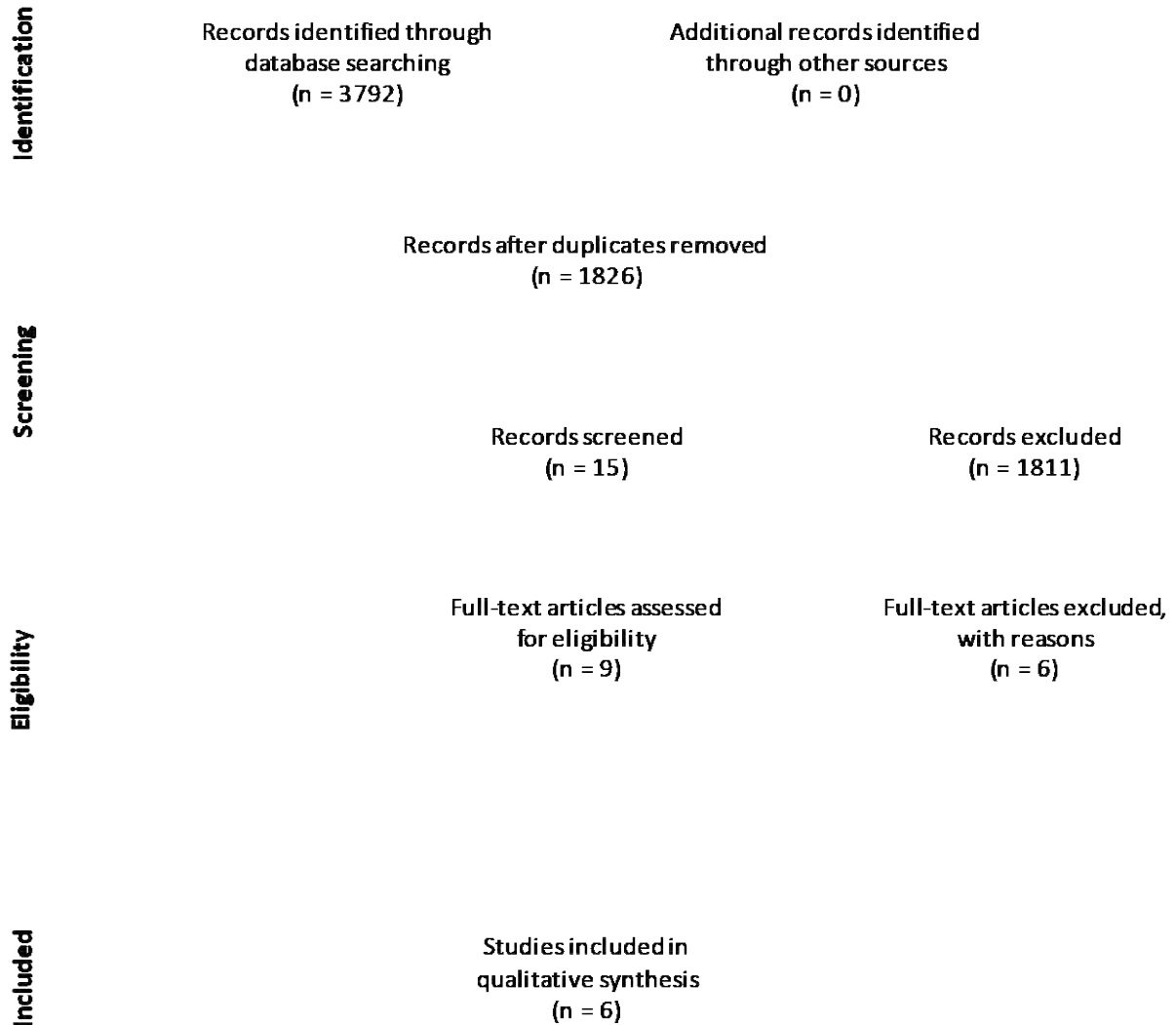


FIGURE 6



PRISMA 2009 Flow Diagram *Parapenaeus longirostris*

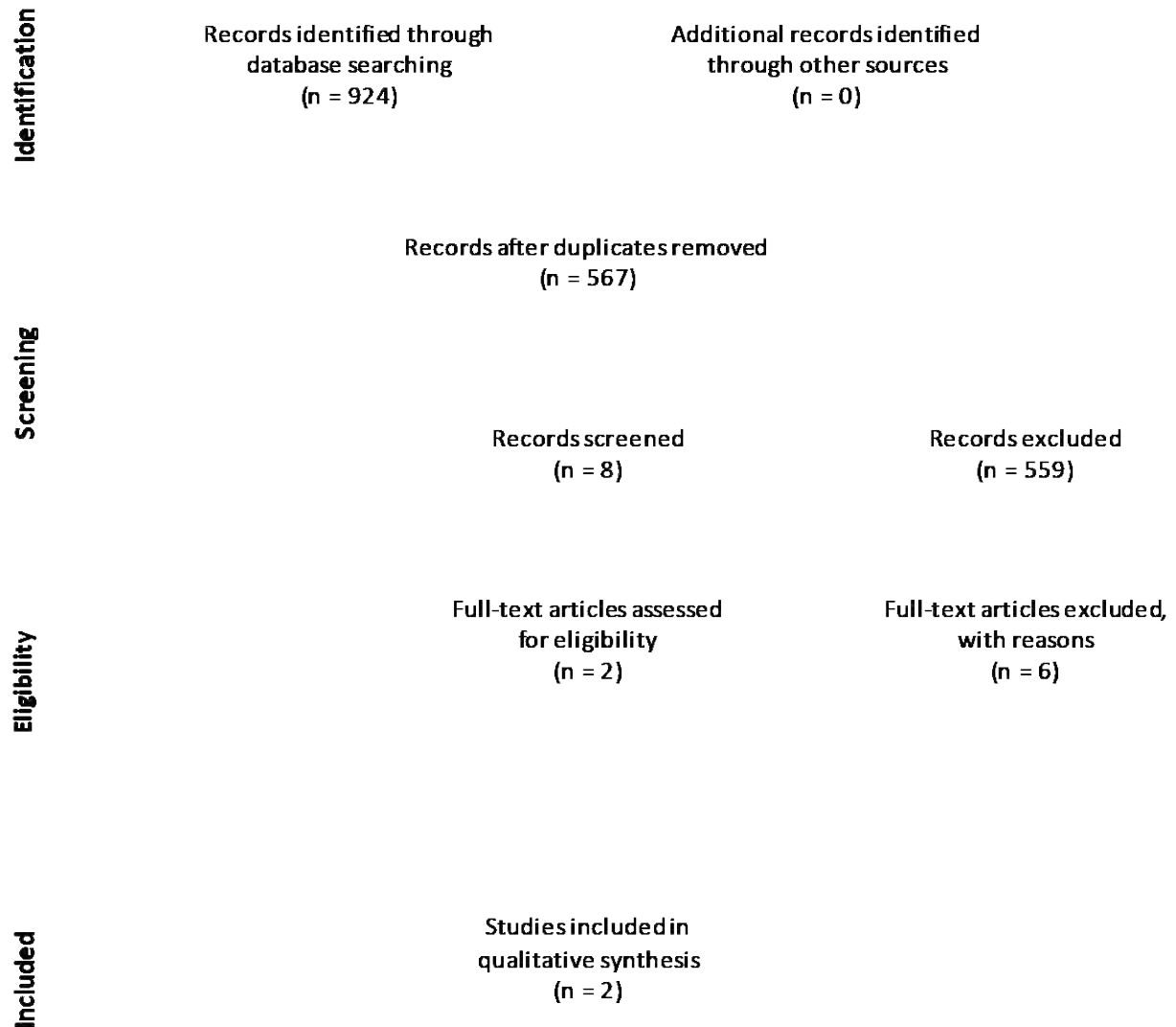


FIGURE 7

In brief, a total of 48 datasets provided useful information on population genetic differentiation for the target species, 41 within the Mediterranean and other 7 from the Black Sea or the adjacent Atlantic Ocean (Table 3 Table 4).

TABLE 3 SUMMARY INFORMATION ON DATASETS SCREENED IN D1.2 HIGH PRIORITY (HP): THE ONES THAT CONTAINED AT LEAST A SAMPLE POPULATION IN THE MEDITERRANEAN, PRIORITY (P): THE ONES THAT CONTAINED AT LEAST A SAMPLE POPULATION OUTSIDE THE MEDITERRANEAN SEA BLACK SEA OR ADJACENT ATLANTIC OCEAN

Species	Pertinent dataset		
	Total	HP	P
<i>Aristaeomorpha foliacea</i>	4	4	0
<i>Aristeus antennatus</i>	9	9	0
<i>Merluccius merluccius</i>	17	11	6
<i>Mullus barbatus</i>	10	10	0
<i>Nephrops norvegicus</i>	6	5	1
<i>Parapenaeus longirostris</i>	2	2	0

TABLE 4 COMPREHENSIVE DATASET OF GENETIC STUDIES. HP HIGH PRIORITY DATASETS (THE ONES THAT CONTAINED AT LEAST A SAMPLE POPULATION FOR THE TARGET SPECIES IN THE MEDITERRANEAN SEA); P PRIORITY DATASETS (THE ONES THAT CONTAINED AT LEAST A SAMPLE POPULATION FOR THE TARGET SPECIES IN THE BLACK SEA OR ADJACENT ATLANTIC OCEAN); R REVIEW, NP NOT PERTINENT; NA: FULL TEXT NOT AVAILABLE

species	code	Full reference	rank	notes
<i>Aristaeomorpha foliacea</i>	GDMU_Af01	Marcias, S., F. Sacco, A. Cau & R. Cannas, 2010. Microsatellite markers for population genetic studies of the giant red shrimp <i>Aristaeomorpha foliacea</i> (Crustacea, Decapoda). Rapp Comm int Mer Médit 39:386.	HP	
<i>Aristaeomorpha foliacea</i>	GDMU_Af02	Fernández, M. V., F. Maltagliati, F. G. Pannacciulli & M. I. Roldán, 2011. Analysis of genetic variability in <i>Aristaeomorpha foliacea</i> (Crustacea, Aristeidae) using DNA-ISSR (Inter Simple Sequence Repeats) markers. Comptes Rendus Biologies 334(10):705-712 doi:https://doi.org/10.1016/j.crv.2011.07.005.	HP	
<i>Aristaeomorpha foliacea</i>	GDMU_Af03	Cannas, R., S. Marcias, F. Sacco, A. Cau & A. M. Deiana, 2012. First isolation and characterization of genomic SSR markers for the giant red shrimp <i>Aristaeomorpha foliacea</i> (Risso, 1827). Genet Mol Res 11(3):2745-8 doi:10.4238/2012.August.17.1.	np	loci isolation; same samples as in GDMU_Af01
<i>Aristaeomorpha foliacea</i>	GDMU_Af04	Fernandez, M. V., S. Heras, F. Maltagliati & M. I. Roldan, 2013a. Deep genetic divergence in giant red shrimp <i>Aristaeomorpha foliacea</i> (Risso, 1827) across a wide distributional range. Journal of Sea Research 76:146-153 doi:10.1016/j.seares.2012.08.004.	HP	
<i>Aristaeomorpha foliacea</i>	GDMU_Af05	Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán, 2013b. Multilocus Comparative Phylogeography of Two Aristeid Shrimps of High Commercial Interest (<i>Aristeus antennatus</i> and <i>Aristaeomorpha foliacea</i>) Reveals Different Responses to Past Environmental Changes. PLOS ONE 8(3):e59033 doi:10.1371/journal.pone.0059033.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa01	Pla, C., Roldán, M. and García-Marín, J. 1995. Biochemical genetics of the pink prawn, <i>Aristeus antennatus</i> Risso, in the western Mediterranean. Rapports Comision International pour l'exploration scientifique de la mer Méditerranée 34:39.	na	no pdf available
<i>Aristeus antennatus</i>	GDMU_Aa02	Marchi A, Cauli G, Greco S, Cau A (1995) Genetic variation in <i>Aristeus antennatus</i> (Crustacea: Aristeidae) in the Mediterranean basin: analysis of 16 enzyme loci. Biologia Marina Mediterranea 2, 495-498	na	no pdf available
<i>Aristeus antennatus</i>	GDMU_Aa03	Sardà, F., C. Bas, M. I. Roldán, C. Pla & J. Lleonart, 1998. Enzymatic and morphometric analyses in mediterranean populations of the rose shrimp, <i>Aristeus antennatus</i> (Risso, 1816). Journal of Experimental Marine Biology and Ecology 221(1):131-144 doi:10.1016/S0022-0981(97)00119-6.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa04	Cannas, R., S. Buccoli, F. Sacco, S. Marcias, S. Salvadori, A. Cau & A. Deiana, 2008. Isolation and characterization of 14 polymorphic microsatellite markers for the blue and red shrimp, <i>Aristeus antennatus</i> (Crustacea, Decapoda). Molecular Ecology Resources 8(6):1420-1422 doi:10.1111/j.1755-0998.2008.02183.x.	np	loci isolation
<i>Aristeus antennatus</i>	GDMU_Aa05	Maggio, T., S. Lo Brutto, R. Cannas, A. M. Deiana & M. Arculeo, 2009. Environmental features of deep-sea habitats linked to the genetic population structure of a crustacean species in the Mediterranean Sea. Marine Ecology 30(3):354-365 doi:10.1111/j.1439-0485.2008.00277.x.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa06	Roldán, M. I., S. Heras, R. Patellani & F. Maltagliati, 2009. Analysis of genetic structure of the red shrimp <i>Aristeus antennatus</i> from the Western Mediterranean employing	HP	

		two mitochondrial regions. <i>Genetica</i> 136(1):1-4 doi:10.1007/s10709-008-9330-2.		
<i>Aristeus antennatus</i>	GDMU_Aa07	Sardà, F., M. I. Roldán, S. Heras & F. Maltagliati, 2010. Influence of the genetic structure of the red and blue shrimp, <i>Aristeus antennatus</i> (Risso, 1816), on the sustainability of a deep-sea population along a depth gradient in the Western Mediterranean. <i>Scientia Marina</i> 74(3):569-575 doi:10.3989/scimar.2010.74n3569.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa08	Fernández, M. V., S. Heras, F. Maltagliati, A. Turco & M. I. Roldán, 2011b. Genetic structure in the blue and red shrimp <i>Aristeus antennatus</i> and the role played by hydrographical and oceanographical barriers. <i>Marine Ecology Progress Series</i> 421:163-171 doi:10.3354/meps08881.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa09	Cannas, R., F. Sacco, M. C. Follesa, A. Sabatini, M. Arculeo, S. Lo Brutto, T. Maggio, A. M. Deiana & A. Cau, 2012. Genetic variability of the blue and red shrimp <i>Aristeus antennatus</i> in the Western Mediterranean Sea inferred by DNA microsatellite loci. <i>Marine Ecology</i> 33(3):350-363 doi:10.1111/j.1439-0485.2011.00504.x.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa10	Lo Brutto, S., T. Maggio, A. M. Deiana, R. Cannas & M. Arculeo, 2012. Further investigations on populations of the deep-water blue and red shrimp <i>Aristeus antennatus</i> (Risso, 1816) (Decapoda, Dendrobranchiata), as inferred from Amplified Fragment Length Polymorphism (AFLP) and mtDNA analyses. <i>Crustaceana</i> 85(11):1393-1408 doi:10.1163/15685403-00003131.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa11	Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán, 2013b. Multilocus Comparative Phylogeography of Two Aristeid Shrimps of High Commercial Interest (<i>Aristeus antennatus</i> and <i>Aristaeomorpha foliacea</i>) Reveals Different Responses to Past Environmental Changes. <i>PLOS ONE</i> 8(3):e59033 doi:10.1371/journal.pone.0059033.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa12	Marra, A., S. Mona, R. M. Sa, G. D'Onghia & P. Maiorano, 2015. Population Genetic History of <i>Aristeus antennatus</i> (Crustacea: Decapoda) in the Western and Central Mediterranean Sea. <i>Plos One</i> 10(3):16 doi:10.1371/journal.pone.0117272.	HP	
<i>Aristeus antennatus</i>	GDMU_Aa13	Heras, S., L. Planella, I. Caldarazzo, M. Vera, J. L. García-Marín & M. I. Roldán, 2016. Development and characterization of novel microsatellite markers by Next Generation Sequencing for the blue and red shrimp <i>Aristeus antennatus</i> . <i>PeerJ</i> 2016(7) doi:10.7717/peerj.2200.	np	loci isolation
<i>Aristeus antennatus</i>	GDMU_Aa14	Taylor, M. L. & C. N. Roterman, 2017. Invertebrate population genetics across Earth's largest habitat: The deep-sea floor. <i>Molecular Ecology</i> 26(19):4872-4896 doi:10.1111/mec.14237.	np	genetic review
<i>Aristeus antennatus</i>	GDMU_Aa15	Carreton, M., J. B. Company, L. Planella, S. Heras, J.-L. García-Marín, M. Agulló, M. Clavel-Henry, G. Rotllant, A. dos Santos & M. I. Roldán, 2019. Morphological identification and molecular confirmation of the deep-sea blue and red shrimp <i>Aristeus antennatus</i> larvae. <i>PeerJ</i> 7:e6063 doi:10.7717/peerj.6063	np	Species identification
<i>Merluccius merluccius</i>	GDMU_Mm01	Lo Brutto, S., M. Arculeo, A. Mauro, M. Scalisi, M. Cammarata & N. Parrinello, 1998. Allozymic variation in Mediterranean hake <i>Merluccius merluccius</i> (Gadidae). <i>Italian Journal of Zoology</i> 65:49-52.	HP	
<i>Merluccius merluccius</i>	GDMU_Mm02	Roldan, M. I., J. L. Garcia-Marín, F. M. Utter & C. Pla, 1998. Population genetic structure of European hake, <i>Merluccius merluccius</i> . <i>Heredity</i> 81:327-334 doi:10.1038/sj.hdy.6883830.	HP	
<i>Merluccius</i>	GDMU_Mm03	Lundy, C. J., P. Moran, C. Rico, R. S. Milner & G. M. Hewitt,	HP	

<i>merluccius</i>		1999. Macrogeographical population differentiation in oceanic environments: a case study of European hake (<i>Merluccius merluccius</i>), a commercially important fish. <i>Molecular Ecology</i> 8(11):1889-1898 doi:10.1046/j.1365-294x.1999.00789.x.		
Merluccius merluccius	GDMU_Mm04	Morán, P., C. Lundy, C. Rico & G. M. Hewitt, 1999. Isolation and characterization of microsatellite loci in European hake, <i>Merluccius merluccius</i> (Merlucidae, Teleostei). <i>Molecular Ecology</i> 8(8):1357-1358 doi:10.1046/j.1365-294X.1999.00701_4.x	NP	loci isolation
Merluccius merluccius	GDMU_Mm05	Lundy, C. J., C. Rico & G. M. Hewitt, 2000. Temporal and spatial genetic variation in spawning grounds of European hake (<i>Merluccius merluccius</i>) in the Bay of Biscay. <i>Molecular Ecology</i> 9(12):2067-2079 doi:10.1046/j.1365-294X.2000.01120.x.	P	
Merluccius merluccius	GDMU_Mm06	Grant, W. S. & R. W. Leslie, 2001. Inter-ocean dispersal is an important mechanism in the zoogeography of hakes (Pisces: <i>Merluccius</i> spp.). <i>Journal of Biogeography</i> 28(6):699-721 doi:10.1046/j.1365-2699.2001.00585.x.	np	phylogenetics
Merluccius merluccius	GDMU_Mm07	Castillo, A. G. F., J. L. Martinez & E. Garcia-Vazquez, 2004. Fine spatial structure of Atlantic hake (<i>Merluccius merluccius</i>) stocks revealed by variation at microsatellite loci. <i>Marine Biotechnology</i> 6(4):299-306 doi:10.1007/s10126-004-3027-z.	HP	
Merluccius merluccius	GDMU_Mm08	Levi, D., B. Patti, P. Rizzo, S. Lo Brutto, N. Parrinello & M. Arculeo, 2004. Genetic and morphometric variations of Mediterranean hake, <i>Merluccius merluccius</i> , in the Strait of Sicily (central Mediterranean): implications for stock assessment of shared resources. <i>Italian Journal of Zoology</i> 71(2):165-170 doi:10.1080/11250000409356568.	HP	
Merluccius merluccius	GDMU_Mm09	Lo Brutto, S., M. Arculeo & N. Parrinello, 2004. Congruence in genetic markers used to describe Mediterranean and Atlantic populations of European hake (<i>Merluccius merluccius</i> L. 1758). <i>Journal of Applied Ichthyology</i> 20(2):81-86 doi:10.1046/j.1439-0426.2003.00514.x.	HP	
Merluccius merluccius	GDMU_Mm10	Utter, F., 2004. Population genetics, conservation and evolution in salmonids and other widely cultured fishes: some perspectives over six decades. <i>Reviews in Fish Biology and Fisheries</i> 14(1):125-144 doi:10.1007/s11160-004-3768-9.	r	REVIEW
Merluccius merluccius	GDMU_Mm11	Castillo, A. G. F., P. Alvarez & E. Garcia-Vazquez, 2005. Population structure of <i>Merluccius merluccius</i> along the Iberian Peninsula coast. <i>ICES Journal of Marine Science</i> 62(8):1699-1704 doi:10.1016/j.icesjms.2005.06.001.	P	
Merluccius merluccius	GDMU_Mm12	Cimmaruta, R., P. Bondanelli & G. Nascetti, 2005. Genetic structure and environmental heterogeneity in the European hake (<i>Merluccius merluccius</i>). <i>Molecular Ecology</i> 14(8):2577-2591 doi:10.1111/j.1365-294X.2005.02595.x.	HP	
Merluccius merluccius	GDMU_Mm13	Martinson, J. T. & R. Ogden, 2009. FishPopTrace-Developing SNP-based population genetic assignment methods to investigate illegal fishing. <i>Forensic Science International: Genetics Supplement Series</i> 2(1):294-296 doi:10.1016/j.fsigss.2009.08.108.	NP	METHODOLOGICAL
Merluccius merluccius	GDMU_Mm14	Reiss, H., G. Hoarau, M. Dickey-Collas & W. J. Wolff, 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. <i>Fish Fish</i> 10(4):361-395 doi:10.1111/j.1467-2979.2008.00324.x	NP	METHODOLOGICAL
Merluccius merluccius	GDMU_Mm15	Pita, A., P. Presa & M. Perez, 2010. GENE FLOW, Multilocus assignment and genetic structuring of the European hake (<i>Merluccius merluccius</i>). <i>Thalassas</i> 26(2):129-133.	HP	
Merluccius	GDMU_Mm16	Milano, I., M. Babbucci, F. Panitz, R. Ogden, R. O. Nielsen,	NP	METHODOLOGICAL

<i>merluccius</i>		M. I. Taylor, S. J. Helyar, G. R. Carvalho, M. Espiñeira, M. Atanassova, F. Tinti, G. E. Maes, T. Patarnello & L. Bargelloni, 2011. Novel tools for conservation genomics: Comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. PLoS ONE 6(11) doi:10.1371/journal.pone.0028008.		
<i>Merluccius merluccius</i>	GDMU_Mm17	Pita, A., M. Perez, S. Cervino & P. Presa, 2011. What can gene flow and recruitment dynamics tell us about connectivity between European hake stocks in the Eastern North Atlantic? Continental Shelf Research 31(5):376-387 doi:10.1016/j.csr.2010.09.010.	P	
<i>Merluccius merluccius</i>	GDMU_Mm18	Nielsen, E. E., A. Cariani, E. Mac Aoidh, G. E. Maes, I. Milano, R. Ogden, M. Taylor, J. Hemmer-Hansen, M. Babbucci, L. Bargelloni, D. Bekkevold, E. Diopere, L. Grenfell, S. Helyar, M. T. Limborg, J. T. Martinsohn, R. McEwing, F. Panitz, T. Patarnello, F. Tinti, J. K. J. Van Houdt, F. A. M. Volckaert, R. S. Waples, G. R. Carvalho, J. E. J. Albin, J. M. V. Baptista, V. Barmintsev, J. M. Bautista, C. Bendixen, J.-P. Berge, D. Blohm, B. Cardazzo, A. Diez, M. Espinera, A. J. Geffen, E. Gonzalez, N. Gonzalez-Lavin, I. Guarniero, M. Jerome, M. Kochzius, G. Krey, O. Mouchel, E. Negrisolo, C. Piccinetti, A. Puyet, S. Rastorguev, J. P. Smith, M. Trentini, V. Verrez-Bagnis, A. Volkov, A. Zanzi & C. FishPopTrace, 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nature Communications 3 doi:10.1038/ncomms1845	HP	
<i>Merluccius merluccius</i>	GDMU_Mm19	Milano, I., M. Babbucci, A. Cariani, M. Atanassova, D. Bekkevold, G. R. Carvalho, M. Espineira, F. Fiorentino, G. Garofalo, A. J. Geffen, J. H. Hansen, S. J. Helyar, E. E. Nielsen, R. Ogden, T. Patarnello, M. Stagioni, F. Consortium, F. Tinti & L. Bargelloni, 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (<i>Merluccius merluccius</i>). Molecular Ecology 23(1):118-135 doi:10.1111/mec.12568	HP	
<i>Merluccius merluccius</i>	GDMU_Mm20	Tanner, S. E., M. Perez, P. Presa, S. R. Thorrold & H. N. Cabral, 2014. Integrating microsatellite DNA markers and otolith geochemistry to assess population structure of European hake (<i>Merluccius merluccius</i>). Estuarine Coastal and Shelf Science 142:68-75 doi:10.1016/j.ecss.2014.03.010	HP	
<i>Merluccius merluccius</i>	GDMU_Mm21	Garcia-Souto, D., T. Troncoso, M. Perez & J. J. Pasantes, 2015. Molecular Cytogenetic Analysis of the European Hake <i>Merluccius merluccius</i> (Merlucciidae, Gadiformes): U1 and U2 snRNA Gene Clusters Map to the Same Location. Plos One 10(12) doi:10.1371/journal.pone.0146150	NP	CHROMOSOMES
<i>Merluccius merluccius</i>	GDMU_Mm22	Mattiucci, S., R. Cimmaruta, P. Cipriani, P. Abaunza, B. Bellisario & G. Nascetti, 2015. Integrating Anisakis spp. parasites data and host genetic structure in the frame of a holistic approach for stock identification of selected Mediterranean Sea fish species. Parasitology 142(1):90-108 doi:10.1017/s0031182014001103	r	REVIEW
<i>Merluccius merluccius</i>	GDMU_Mm23	Casey, J., E. Jardim & J. T. Martinsohn, 2016. The role of genetics in fisheries management under the EU common fisheries policy. Journal of Fish Biology 89(6):2755-2767 doi:10.1111/jfb.13151	r	REVIEW
<i>Merluccius merluccius</i>	GDMU_Mm24	Cuellar-Pinzon, J., P. Presa, S. J. Hawkins & A. Pita, 2016. Genetic markers in marine fisheries: Types, tasks and trends. Fisheries Research 173:194-205 doi:10.1016/j.fishres.2015.10.019.	r	REVIEW
<i>Merluccius merluccius</i>	GDMU_Mm25	Ferrito, V., V. Bertolino & A. M. Pappalardo, 2016. White fish authentication by COI-Bar-RFLP: Toward a common strategy for the rapid identification of species in	np	SPECIES IDENTIFICATION

		convenience seafood. Food Control 70:130-137 doi:10.1016/j.foodcont.2016.05.026		
Merluccius merluccius	GDMU_Mm26	Pita, A., J. Casey, S. J. Hawkins, M. R. Villarreal, M. J. Gutierrez, H. Cabral, F. Carocci, P. Abaunza, S. Pascual & P. Presa, 2016a. Conceptual and practical advances in fish stock delineation Preface. Fisheries Research 173:185-193 doi:10.1016/j.fishres.2015.10.029.	r	REVIEW
Merluccius merluccius	GDMU_Mm27	Pita, A., A. Leal, A. Santafe-Munoz, C. Pineiro & P. Presa, 2016b. Genetic inference of demographic connectivity in the Atlantic European hake metapopulation (<i>Merluccius merluccius</i>) over a spatio-temporal framework. Fisheries Research 179:291-301 doi:10.1016/j.fishres.2016.03.017	P	
Merluccius merluccius	GDMU_Mm28	Pita, A., M. Perez, F. Velasco & P. Presa, 2017. Trends of the genetic effective population size in the Southern stock of the European hake. Fisheries Research 191:108-119 doi:10.1016/j.fishres.2017.02.022	P	
Merluccius merluccius	GDMU_Mm29	Westgaard, J.-I., A. Staby, J. A. Godiksen, A. J. Geffen, A. Svensson, G. Charrier, H. Svedang & C. Andre, 2017. Large and fine scale population structure in European hake (<i>Merluccius merluccius</i>) in the Northeast Atlantic. ICES Journal of Marine Science 74(5):1300-1310 doi:10.1093/icesjms/fsw249	P	
Mullus barbatus	GDMU_Mb01	Cammarata, M., N. Parrinello & M. Arculeo, 1991. Biochemical taxonomic differentiation between <i>Mullus barbatus</i> and <i>Mullus surmuletus</i> (Pisces, Mullidae). Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 99(3):719-722 doi:10.1016/0305-0491(91)90360-p	np	species identification
Mullus barbatus	GDMU_Mb02	Vincent, F., S. Jaunet, F. Galgani, H. Besselink & J. Koeman, 1995. Expression of RAS gene in flounder (<i>Platichthys flesus</i>) and red mullet (<i>Mullus barbatus</i>). Biochemical and Biophysical Research Communications 215(2):659-665 doi:10.1006/bbrc.1995.2515	np	gene identification
Mullus barbatus	GDMU_Mb03	Mamuris, Z., A. P. Apostolidis, A. J. Theodorou & C. Triantaphyllidis, 1998a. Application of random amplified polymorphic DNA (RAPD) markers to evaluate intraspecific genetic variation in red mullet (<i>Mullus barbatus</i>). Marine Biology 132(2):171-178 doi:10.1007/s002270050383.	HP	
Mullus barbatus	GDMU_Mb04	Mamuris, Z., A. P. Apostolidis, A. Triantafyllidis, P. S. Economidis & C. Triantaphyllidis, 1998b. Relationship between morphologic and genetic variation in red mullet (<i>Mullus barbatus</i>) populations from Greek Seas. Folia Zoologica 47(4):295-303.	np	methodological
Mullus barbatus	GDMU_Mb05	Mamuris, Z., A. P. Apostolidis & C. Triantaphyllidis, 1998c. Genetic protein variation in red mullet (<i>Mullus barbatus</i>) and striped red mullet (<i>M. surmuletus</i>) populations from the Mediterranean Sea. Marine Biology 130(3):353-360 doi:10.1007/s002270050255	HP	
Mullus barbatus	GDMU_Mb06	Arculeo, M., S. Lo Brutto, M. Cammarata, M. Scalisi & N. Parrinello, 1999. Genetic variability of the Mediterranean Sea red mullet, <i>Mullus barbatus</i> (Pisces, Mullidae). Russian Journal of Genetics 35(3):292-296	HP	
Mullus barbatus	GDMU_Mb07	Golani, D. & U. Ritte, 1999. Genetic relationship in goatfishes (Mullidae : Perciformes) of the Red Sea and the Mediterranean, with remarks on Suez Canal migrants. Scientia Marina 63(2):129-135 doi:10.3989/scimar.1999.63n2129	np	1 sample
Mullus barbatus	GDMU_Mb08	Mamuris, Z., C. Stamatis, M. Bani & C. Triantaphyllidis, 1999a. Taxonomic relationships between four species of the Mullidae family revealed by three genetic methods: allozymes, random amplified polymorphic DNA and mitochondrial DNA. Journal of Fish Biology 55(3):572-587 doi:10.1006/jfbi.1999.1017	np	phylogenetics

<i>Mullus barbatus</i>	GDMU_Mb09	Apostolidis, A. P., Z. Mamuris & C. Triantaphyllidis, 2001. Phylogenetic relationships among four species of Mullidae (Perciformes) inferred from DNA sequences of mitochondrial cytochrome b and 16S rRNA genes. <i>Biochemical Systematics and Ecology</i> 29(9):901-909 doi:10.1016/s0305-1978(01)00037-0.	np	phylogenetics
<i>Mullus barbatus</i>	GDMU_Mb10	Mamuris, Z., C. Stamatis, K. A. Moutou, A. P. Apostolidis & C. Triantaphyllidis, 2001. RFLP analysis of mitochondrial DNA to evaluate genetic variation in striped red mullet (<i>Mullus surmuletus</i> L.) and red mullet (<i>Mullus barbatus</i> L.) populations. <i>Marine Biotechnology</i> 3(3):264-274 doi:10.1007/s101260000075	HP	
<i>Mullus barbatus</i>	GDMU_Mb11	Garoiá, F., I. Guarniero, C. Piccinetti & F. Tinti, 2004. First microsatellite loci of red mullet (<i>Mullus barbatus</i>) and their application to genetic structure analysis of Adriatic shared stock. <i>Marine Biotechnology</i> 6(5):446-452 doi:10.1007/s10126-004-3045-x	HP	
<i>Mullus barbatus</i>	GDMU_Mb12	Turan, C., 2006. Phylogenetic relationships of Mediterranean Mullidae species (Perciformes) inferred from genetic and morphologic data. <i>Scientia Marina</i> 70(2):311-318 doi:10.3989/scimar.2006.70n2311	np	phylogenetics
<i>Mullus barbatus</i>	GDMU_Mb13	Galarza, J. A., G. F. Turner, E. Macpherson, J. Carreras-Carbonell & C. Rico, 2007. Cross-amplification of 10 new isolated polymorphic microsatellite loci for red mullet (<i>Mullus barbatus</i>) in striped red mullet (<i>Mullus surmuletus</i>). <i>Molecular Ecology Notes</i> 7(2):230-232 doi:10.1111/j.1471-8286.2006.01551.x	np	isolation markers
<i>Mullus barbatus</i>	GDMU_Mb14	Apostolidis, A. P., A. Georgiadis, N. Karaiskou & R. Sandaltzopoulos, 2008. Reliable and rapid discrimination of congeneric species by mtDNA SNP analysis by multiplex PCR: application on three <i>Trachurus</i> and two <i>Mullus</i> fish species as model cases. <i>Hydrobiologia</i> 614(1):401-404 doi:10.1007/s10750-008-9528-4	np	species identification
<i>Mullus barbatus</i>	GDMU_Mb15	Apostolidis, A. P., K. A. Moutou, C. Stamatis & Z. Mamuris, 2009. Genetic structure of three marine fishes from the Gulf of Pagasitikos (Greece) based on allozymes, RAPD, and mtDNA RFLP markers. <i>Biologia</i> 64(5):1005-1010 doi:10.2478/s11756-009-0161-0	HP	
<i>Mullus barbatus</i>	GDMU_Mb16	Galarza, J. A., G. F. Turner, E. Macpherson & C. Rico, 2009. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (<i>Mullus barbatus</i>) and the striped red mullet (<i>Mullus surmuletus</i>). <i>Canadian Journal of Fisheries and Aquatic Sciences</i> 66(9):1478-1490 doi:10.1139/f09-098	HP	
<i>Mullus barbatus</i>	GDMU_Mb17	Keskin, E. & A. Can, 2009. Phylogenetic relationships among four species and a sub-species of Mullidae (Actinopterygii; Perciformes) based on mitochondrial cytochrome B, 12S rRNA and cytochrome oxidase II genes. <i>Biochemical Systematics and Ecology</i> 37(5):653-661 doi:10.1016/j.bse.2009.10.001	np	phylogenetics
<i>Mullus barbatus</i>	GDMU_Mb18	Maggio, T., S. Lo Brutto, F. Garoiá, F. Tinti & M. Arculeo, 2009. Microsatellite analysis of red mullet <i>Mullus barbatus</i> (Perciformes, Mullidae) reveals the isolation of the Adriatic Basin in the Mediterranean Sea. <i>ICES Journal of Marine Science</i> 66(9):1883-1891 doi:10.1093/icesjms/fsp160	HP	
<i>Mullus barbatus</i>	GDMU_Mb19	Vogiatzi, E., R. Hanel, T. Dailianis, J. Lagnel, M. Hassan, A. Magoulas & C. S. Tsigenopoulos, 2012. Description of microsatellite markers in four mullids based on the development and cross-species amplification of 18 new markers in red mullet (<i>Mullus barbatus</i>). <i>Biochemical Systematics and Ecology</i> 44:279-285 doi:10.1016/j.bse.2012.06.006	np	isolation markers
<i>Mullus</i>	GDMU_Mb20	Felix-Hackradt, F. C., C. W. Hackradt, A. Perez-Ruzafa & J.	HP	

barbatus		A. Garcia-Charton, 2013. Discordant patterns of genetic connectivity between two sympatric species, <i>Mullus barbatus</i> (Linnaeus, 1758) and <i>Mullus surmuletus</i> (Linnaeus, 1758), in south-western Mediterranean Sea. Marine Environmental Research 92:23-34 doi:10.1016/j.marenvres.2013.08.008		
Mullus barbatus	GDMU_Mb21	Espineira, M. & J. M. Vieites, 2015. Genetic system for traceability of goatfishes by FINS methodology and authentication of mullets (<i>Mullus barbatus</i> and <i>Mullus surmuletus</i>) by RT-PCR. European Food Research and Technology 240(2):423-429 doi:10.1007/s00217-014-2341-6	np	species identification
Mullus barbatus	GDMU_Mb22	Ahmed, M. I., M. M. Sabrah, R. A. Heneish & M. El-Alwany, 2016. DNA Barcoding Uncover Cryptic Diversity in Goat Fishes (Family: Mullidae) Across the Egyptian Coastal Waters. Pakistan journal of biological sciences : PJBS 19(2):65-70 doi:10.3923/pjbs.2016.65.70	np	barcoding
Mullus barbatus	GDMU_Mb23	Prazdnikov, D. V., 2016. Karyology of <i>Mullus barbatus</i> (Pisces, Perciformes) from the Mediterranean basin. Turkish Journal of Zoology 40(2):279-281 doi:10.3906/zoo-1503-15	np	chromosomes
Mullus barbatus	GDMU_Mb24	Matic-Skoko, S., T. Segvic-Bubic, I. Mandic, D. Izquierdo-Gomez, E. Arneri, P. Carbonara, F. Grati, Z. Ikica, J. Kolutari, N. Milone, P. Sartor, G. Scarcella, A. Tokac & E. Tzanatos, 2018. Evidence of subtle genetic structure in the sympatric species <i>Mullus barbatus</i> and <i>Mullus surmuletus</i> (Linnaeus, 1758) in the Mediterranean Sea. Sci Rep 8:14 doi:10.1038/s41598-017-18503-7	HP	
Nephrops norvegicus	GDMU_Nn01	Deiana, A. M., E. Coluccia, A. Milia & S. Salvadori, 1996. Supernumerary chromosomes in <i>Nephrops norvegicus</i> L (Crustacea, Decapoda). Heredity 76:92-99 doi:10.1038/hdy.1996.12.	np	chromosomes
Nephrops norvegicus	GDMU_Nn02	Passamonti, M., B. Mantovani, V. Scali & C. Froglija, 1997. Allozymic Characterization of Scottish and Aegean Populations of <i>Nephrops norvegicus</i> . Journal of the Marine Biological Association of the United Kingdom 77(3):727-735 doi:10.1017/S0025315400036158.	HP	
Nephrops norvegicus	GDMU_Nn03	Maltagliati, F., L. Camilli, F. Biagi & M. Abbiati, 1998. Genetic structure of Norway lobster, <i>Nephrops norvegicus</i> (L.)(Crustacea: Nephropidae), from the Mediterranean Sea. Scientia Marina 62:91-99.	HP	
Nephrops norvegicus	GDMU_Nn04	Streiff, R., T. Guillemaud, F. Alberto, J. Magalhaes, M. Castro & M. L. Cancela, 2001. Isolation and characterization of microsatellite loci in the Norway lobster (<i>Nephrops norvegicus</i>). Molecular Ecology Notes 1(1-2):71-72 doi:10.1046/j.1471-8278.2001.00029.x.	np	isolation markers
Nephrops norvegicus	GDMU_Nn05	Coluccia, E., R. Cannas, A. Cau, A. M. Deiana & S. Salvadori, 2004. B chromosomes in Crustacea Decapoda. Cytogenetic and Genome Research 106(2-4):215-221 doi:10.1159/000079290.	np	chromosomes
Nephrops norvegicus	GDMU_Nn06	Stamatis, C., A. Triantafyllidis, K. A. Moutou & Z. Mamuris, 2004. Mitochondrial DNA variation in Northeast Atlantic and Mediterranean populations of Norway lobster, <i>Nephrops norvegicus</i> . Molecular Ecology 13(6):1377-1390 doi:10.1111/j.1365-294X.2004.02165.x.	HP	
Nephrops norvegicus	GDMU_Nn07	Stamatis, C., A. Triantafyllidis, K. A. Moutou & Z. Mamuris, 2006. Allozymic variation in Northeast Atlantic and Mediterranean populations of Norway lobster, <i>Nephrops norvegicus</i> . ICES Journal of Marine Science 63(5):875-882 doi:10.1016/j.icesjms.2006.01.006.	HP	
Nephrops norvegicus	GDMU_Nn08	Streiff, R., S. Mira, M. Castro & M. L. Cancela, 2004. Multiple paternity in Norway lobster (<i>Nephrops norvegicus</i> L.) assessed with microsatellite markers. Marine Biotechnology 6(1):60-66 doi:10.1007/s10126-	np	paternity analysis

		003-0015-7.		
<i>Nephrops norvegicus</i>	GDMU_Nn09	Skirnisdottir, S., K. Olafsson, S. Hauksdottir, C. Pampoulie, G. O. Hreggvidsson, G. H. Gunnarsson & S. Hjorleifsdottir, 2010. Isolation and characterisation of eight new microsatellite loci in the Norway lobster, <i>Nephrops norvegicus</i> (Linnaeus, 1758). Molecular Ecology Resources Database, http://tomatobiotrinityedu/manuscripts/10-4/mer-10-0048pdf (accessed 26 October 2010).	np	isolation markers
<i>Nephrops norvegicus</i>	GDMU_Nn10	Pampoulie, C., S. Skirnisdottir, S. Hauksdottir, K. Olafsson, H. Eiriksson, V. Chosson, G. O. Hreggvidsson, G. H. Gunnarsson & S. Hjorleifsdottir, 2011. A pilot genetic study reveals the absence of spatial genetic structure in Norway lobster (<i>Nephrops norvegicus</i>) on fishing grounds in Icelandic waters. ICES Journal of Marine Science 68(1):20-25 doi:10.1093/icesjms/fsq165.	P	
<i>Nephrops norvegicus</i>	GDMU_Nn11	Keskin, E. & H. H. Atar, 2013. DNA barcoding commercially important aquatic invertebrates of Turkey. Mitochondrial DNA 24(4):440-450 doi:10.3109/19401736.2012.762576.	np	barcoding
<i>Nephrops norvegicus</i>	GDMU_Nn12	Gan, H. M., M. H. Tan, H. Y. Gan, Y. P. Lee & C. M. Austin, 2016. The complete mitogenome of the Norway lobster <i>Nephrops norvegicus</i> (Linnaeus, 1758) (Crustacea: Decapoda: Nephropidae). Mitochondrial DNA Part A 27(5):3179-3180 doi:10.3109/19401736.2015.1007325.	np	mitogenomic data
<i>Nephrops norvegicus</i>	GDMU_Nn13	Rotllant, G., N. Tuan Viet, V. Sbragaglia, L. Rahi, K. J. Dudley, D. Hurwood, T. Ventura, J. B. Company, V. Chand, J. Aguzzi & P. B. Mather, 2017. Sex and tissue specific gene expression patterns identified following de novo transcriptomic analysis of the Norway lobster, <i>Nephrops norvegicus</i> . BMC Genomics 18 doi:10.1186/s12864-017-3981-2.	np	gene expression
<i>Nephrops norvegicus</i>	GDMU_Nn14	Gallagher, J., J. A. Finarelli, J. P. Jonasson & J. Carlsson, 2018. Mitochondrial D-loop DNA analyses of Norway Lobster (<i>Nephrops norvegicus</i>) reveals genetic isolation between Atlantic and Mediterranean populations. bioRxiv.	HP	
<i>Nephrops norvegicus</i>	GDMU_Nn15	Rotllant, G., N. Tuan Viet, D. Hurwood, V. Sbragaglia, T. Ventura, J. B. Company, S. Joly, A. Elizur & P. B. Mather, 2018b. Evaluation of genes involved in Norway lobster (<i>Nephrops norvegicus</i>) female sexual maturation using transcriptomic analysis. Hydrobiologia 825(1):137-158 doi:10.1007/s10750-018-3521-3.	np	gene expression
<i>Parapenaeus longirostris</i>	GDMU_PI01	Benzie, J. A. H., 1998. Penaeid genetics and biotechnology. Aquaculture 164(1-4):23-47 doi:10.1016/s0044-8486(98)00175-6.	np	protein characterization
<i>Parapenaeus longirostris</i>	GDMU_PI02	Hisar, S. A., E. Aksakal, O. Hisar, T. Yanik & S. Mol, 2008. Discrimination of penaeid shrimps with PCR-RFLP analysis. Journal of Shellfish Research 27(4):917-920 doi:10.2983/0730-8000(2008)27[917:Dopswp]2.0.Co;2.	np	species discrimination
<i>Parapenaeus longirostris</i>	GDMU_PI03	Pascoal, A., J. Barros-Velazquez, A. Cepeda, J. M. Gallardo & P. Calo-Mata, 2008a. Identification of shrimp species in raw and processed food products by means of a polymerase chain reaction-restriction fragment length polymorphism method targeted to cytochrome b mitochondrial sequences. Electrophoresis 29(15):3220-3228 doi:10.1002/elps.200800132.	np	species discrimination
<i>Parapenaeus longirostris</i>	GDMU_PI04	Ortea, I., B. Canas, P. Calo-Mata, J. Barros-Velazquez & J. M. Gallardo, 2010. Identification of commercial prawn and shrimp species of food interest by native isoelectric focusing. Food Chemistry 121(2):569-574 doi:10.1016/j.foodchem.2009.12.049.	np	species discrimination
<i>Parapenaeus longirostris</i>	GDMU_PI05	Garcia-Merchan, V. H., A. Robainas-Barcia, P. Abello, E. Macpherson, F. Palero, M. Garcia-Rodriguez, L. Gil de Sola & M. Pascual, 2012. Phylogeographic patterns of decapod	HP	

		crustaceans at the Atlantic-Mediterranean transition. <i>Mol Phylogenet Evol</i> 62(2):664-72 doi:10.1016/j.ympev.2011.11.009.		
<i>Parapenaeus longirostris</i>	GDMU_PI06	Lo Brutto, S., T. Maggio & M. Arculeo, 2013. Isolation by distance (IBD) signals in the deep-water rose shrimp <i>Parapenaeus longirostris</i> (Lucas, 1846) (Decapoda, Panaeidae) in the Mediterranean sea. <i>Marine Environmental Research</i> 90:1-8 doi:10.1016/j.marenvres.2013.05.006.	HP	
<i>Parapenaeus longirostris</i>	GDMU_PI07	Bilgin, R., M. A. Utkan, E. Kalkan, S. U. Karhan & M. Bekbolet, 2015. DNA barcoding of twelve shrimp species (Crustacea: Decapoda) from Turkish seas reveals cryptic diversity. <i>Mediterr Mar Sci</i> 16(1):36-45 doi:10.12681/mms.548.	np	species discrimination
<i>Parapenaeus longirostris</i>	GDMU_PI08	Tagliavia, M., A. Nicosia, M. Salamone, G. Biondo, C. D. Bennici, S. Mazzola & A. Cuttitta, 2016. Development of a fast DNA extraction method for sea food and marine species identification. <i>Food Chemistry</i> 203:375-378 doi:10.1016/j.foodchem.2016.02.095.	np	methodological

The analysis: *Aristaeomorpha foliacea* (giant red shrimp)

INFORMATION FROM STUDIES INCLUDING MEDITERRANEAN POPULATION SAMPLES (HP DATASETS)

GDMU_Af01=Af03

DATA FROM:

Marcias, S., F. Sacco, A. Cau & R. Cannas, 2010. **Microsatellite markers for population genetic studies of the giant red shrimp *Aristaeomorpha foliacea* (Crustacea, Decapoda).** *Rapp Comm int Mer Médit* 39:386

GENETIC MARKER USED

microsatellites (6 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

code	Site	latitude	longitude	FAO area, GFCM GSA	No of samples
ASI	Asinara	41° 7.330'N	8° 36.965'E	FAO37:GSA11	20
SIN	Siniscola	40° 35.836'N	9° 53.428'E	FAO37:GSA11	20
CAG	Cagliari	39° 4.231'N	9° 17.067'E	FAO37:GSA11	20
SAN	Sant'Antioco	39° 3.283'N	8° 19.004'E	FAO37:GSA11	20
MES	Messina	38° 21.898'N	15° 27.137'E	FAO37:GSA10	15
MAZ	Mazara	37° 35.245'N	12° 40.200'E	FAO37:GSA16	20
total					115

MAIN FINDINGS (MARCIAS ET AL., 2010)

1. A total of 115 specimens were analysed from 6 populations: 4 samples were collected off the Sardinian coast (Sant'Antioco, Cagliari, Siniscola, Asinara) and two samples off Sicily (Messina and Sicily Channel).
2. *A. foliacea* populations resulted not genetically differentiated as indicated by the low and not significant pair-wise *Fst* values
3. Analysis of molecular variance AMOVA clearly showed that genetic variability was largely due to differences among individuals (99.31%) rather than to differences among populations (0.69%).
4. The absence of population structure was further confirmed by the bayesian clustering method implemented in Structure indicating K=1 as the most probable structure.
5. The PCA performed with Aegenet showed a substantial genetic homogeneity among populations separated by hundreds of kilometres suggesting that western Mediterranean populations could represent a unique panmictic stock (fig 1).
6. Finally, the bottleneck test and interlocus *g*-tests did not find signs of recent bottlenecks (reduction of population size) or growth (increase of population size after the recovering from a demographic collapse), leading to hypothesize that, up to now, the Western Mediterranean giant red shrimps have experienced a sustainable fishing pressure.
7. In summary, microsatellite data revealed a substantial genetic homogeneity and no signs of recent bottlenecks, suggesting the existence of a high gene flow connecting all populations.

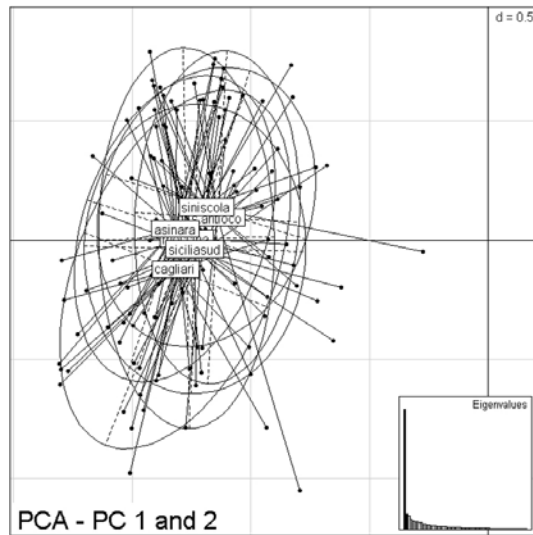


FIGURE 1 FROM MARCIAS *ET AL.*, 2010. TYPOLOGY OF GIANT RED SHRIMP POPULATIONS OBTAINED BY INTERCLASS PCA. EIGENVALUES CORRESPONDING TO THE REPRESENTED COMPONENTS ARE FILLED IN BLACK. POINTS REPRESENT GENOTYPES; SAMPLES ARE LABELLED INSIDE THEIR 95% INERTIA ELLIPSES

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AF01 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Af01	<i>Aristaeomorpha foliacea</i>
Population\cluster	Cluster A
ASI	1
SIN	1
CAG	1
SAN	1
MES	1
MAZ	1

DATA FROM:

Fernández, M. V., F. Maltagliati, F. G. Pannacciulli & M. I. Roldán, 2011a. **Analysis of genetic variability in *Aristaeomorpha foliacea* (Crustacea, Aristeidae) using DNA-ISSR (Inter Simple Sequence Repeats) markers.** *Comptes Rendus Biologies* 334(10):705-712
doi:https://doi.org/10.1016/j.crvi.2011.07.005.

GENETIC MARKER USED

DNA-ISSR (about 150 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

code	Site	latitude	longitude	FAO area, GFCM GSA	No of samples
IBI	Ibiza Channel	39°02'N	02°39'E	FAO37:GSA5	51
TYR	Tyrrhenian Sea	42°28'N	9°43'E	FAO37:GSA8	48
MAZ	Mazara del Vallo	--	--	FAO37:GSA16	40
PPA	Portopalo	--	--	FAO37:GSA16	40
ION	Ionian Sea	37°31'N	21°22'E	FAO37:GSA20	38
AEG	Aegean Sea	37°17'N	22°53'E	FAO37:GSA22	44
MOZ°	Mozambique Channel	25°57'S	34°38'E	FAO51	42
total					303

° not shown in the geovisualization

MAIN FINDINGS (FERNÁNDEZ ET AL., 2011A)

1. A total of 303 adult specimens of *A. foliacea* were sampled from six locations of the Mediterranean Sea and one location in the Western Indian Ocean (Mozambique Channel)
2. Average estimates of genetic diversity did not significantly differ among sampled localities, with a mean value of heterozygosity $H = 0.105 \pm 0.015$.
3. Analysis of molecular variance (AMOVA) allocated > 98% of genetic variability to the within-sample component; however, the F-statistics value associated to the remaining part of variance was significant ($\Phi_{ST} = 0.013$, $P = 0.002$), even after removing MOZ ($\Phi_{ST} = 0.012$, $P < 0.001$).
4. When localities were grouped according to geographical regions (EM, WM, MOZ), no genetic variance was assigned to the among-region level (-0.07%, $\Phi_{ST} = -0.001$, $P = 0.472$). After omitting MOZ from analysis, no genetic variation was either attributed to the among-region level (-0.31%, $\Phi_{ST} = -0.003$, $P = 0.7$) as it was mostly associated to the within-sample component (98.9%, $\Phi_{ST} = 0.011$, $P = 0.007$) and secondly to the within-region component (1.4%, $\Phi_{SC} = 0.014$, $P = 0.006$). Alternative grouping criteria did not provide any different results.
5. The Neighbor-joining tree (Fig. 2) did not show any clustering that could be associated to the geographical position of samples
6. Cluster analyses did not detect geographically or genetically distinct groups. Replicate runs of STRUCTURE yielded consistent results and the uppermost hierarchical structure present in the entire dataset was detected for $K = 2$ by the K statistics. The distribution of the two clusters across individuals did not show any geographical meaning (fig. 3)

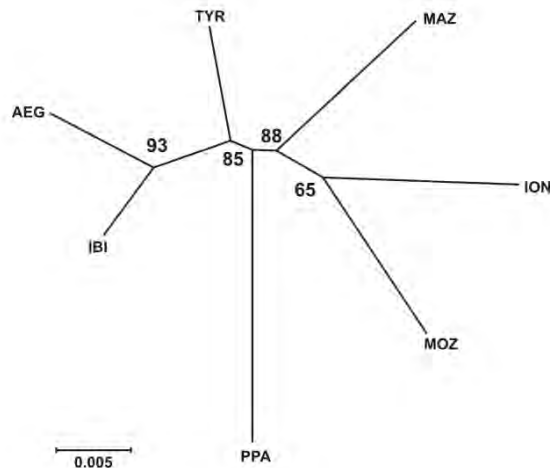


FIGURE 2 FROM FERNANDEZ ET AL., 2011A. NEIGHBOR-JOINING TREE BASED ON CAVALLI-SFORZA AND EDWARDS' CHORD GENETIC DISTANCE ILLUSTRATING THE RELATIONSHIPS AMONG SAMPLED LOCALITIES OF ARISTAEOMORPHA FOLIACEA. BOOTSTRAP VALUES AFTER 1,000 REPLICATES ARE INDICATED ON THE NODES

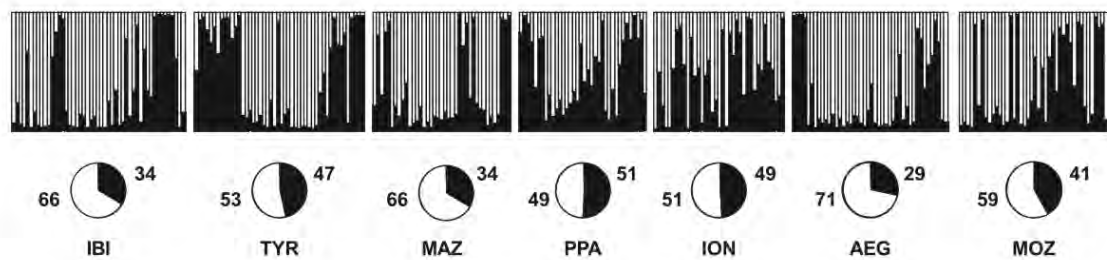


FIGURE 3 FROM FERNANDEZ ET AL., 2011A. BAYESIAN ASSIGNMENT ANALYSIS FOR K=2. EACH BAR REPRESENTS AN INDIVIDUAL OF ARISTAEOMORPHA FOLIACEA AND ITS ASSOCIATED PROBABILITY OF BELONGING TO ONE OF THE TWO GENETIC CLUSTERS DETECTED (BLACK AND WHITE). PIE CHARTS INDICATE PERCENTAGES OF THE HAPLOGROUPS CONTRIBUTING TO EACH SAMPLE.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AF02 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Af02	<i>Aristaeomorpha foliacea</i>
Population\cluster	Cluster A
IBI	0
TYR	0
MAZ	1
PPA	0
ION	1
AEG	1
MOZ ^o	0

^o not shown in the geovisualization

DATA FROM:

Fernandez MV, Heras S, Maltagliati F, Roldan MI. 2013a. **Deep genetic divergence in giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827) across a wide distributional range.** *Journal of Sea Research* 76:146-153.

GENETIC MARKER USED

Mitochondrial DNA (COI gene sequences, 685 bp)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (some in 2008 for the others no information is available) in the following locations:

code	Site	latitude	longitude	FAO area, GFCM GSA	No of samples
IBI	Ibiza Channel	39°02'N	02°39'E	FAO37:GSA5	51
TYR	Tyrrhenian Sea	42°28'N	9°43'E	FAO37:GSA8	45
MAZ	Mazara del Vallo	--	--	FAO37:GSA16	40
PPA	Portopalo	--	--	FAO37:GSA16	37
ION	Ionian Sea	37°31'N	21°22'E	FAO37:GSA20	37
AEG	Aegean Sea	37°17'N	22°53'E	FAO37:GSA22	44
MOZ°	Mozambique Channel	25°57'S	34°38'E	FAO51	42
AUS°	North-western Australia	14°51'S	121°26'E	FAO57	21
total					317

° not shown in the geovisualization

MAIN FINDINGS (FERNANDEZ ET AL., 2013A)

1. Significantly high molecular variance among all localities (85.0% of variance among samples, $\Phi_{ST}=0.850$, $P<0.005$) resulting from genetic differences among the Mediterranean Sea, Mozambique Channel and North Western Australia (92.3% of variance among regions, $\Phi_{CT}=0.923$, $P=0.036$).
2. Within the Mediterranean, although at a lesser extent, molecular variance was also significant (16.9% of variance among samples, $\Phi_{ST}=0.169$, $P<0.005$), suggesting that a degree of genetic differentiation was present among the Mediterranean local samples.
3. However, when Mediterranean samples were grouped in western and eastern basins, the variation among groups was not statistically significant (23.3%, $\Phi_{CT}=0.233$, $P=0.061$).
4. Four haplogroups (HG1-HG4) detected, two restricted to the Mediterranean localities and with minor presence in the Mozambique Channel (HG1, HG2), a third restricted to the Mozambique Channel (HG3) and a fourth restricted to northwestern Australia (HG4). Within the Mediterranean, IBI, TYR and PPA were mostly represented by one of the Mediterranean haplogroups (HG1>75%), whilst ION and AEG were mostly represented by the other Mediterranean haplogroup (HG 2>65%). Notably, the two Mediterranean haplogroups were almost evenly present in MAZ (HG 1=56%, HG 2=44%) (Figure 3B from Fernandez *et al.*, 2013)
5. Pairwise F_{ST} comparisons within the Mediterranean indicated no genetic differentiation between IBI-TYR, IBI-PPA, TYR-PPA and ION-AEG ($F_{ST}=0$ to 0.004) whilst the highest values were detected for the comparisons of ION or AEG against IBI, TYR or PPA ($F_{ST}=0.206$ to 0.382). The locality of MAZ displayed intermediate differentiation values being most similar to ION, followed by PPA, AEG, TYR and IBI ($F_{ST}=0.037$ to 0.149).

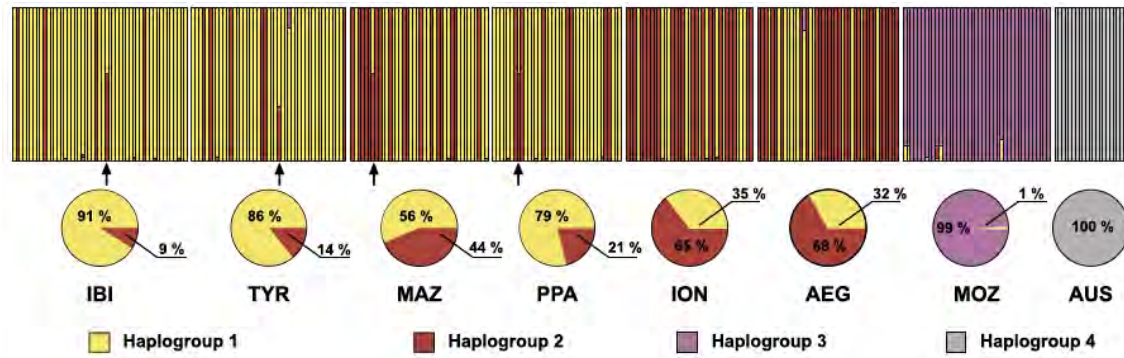
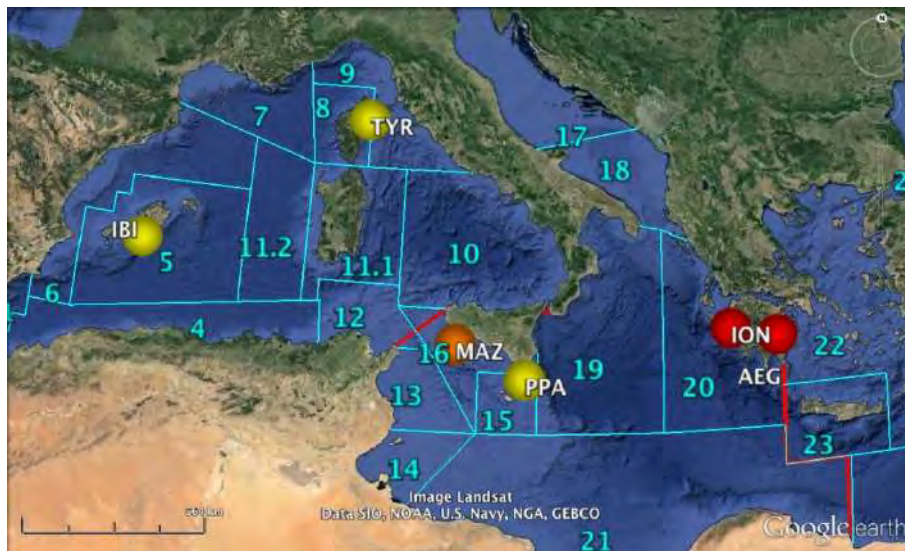


FIGURE 3B FROM FERNANDEZ *ET AL.*, 2013A. BAYESIAN ASSIGNMENT ANALYSIS FOR COI SEQUENCES OF THE EIGHT LOCALITIES SAMPLED. EACH VERTICAL BAR PRESENTS AN INDIVIDUAL AND ITS ASSOCIATED PROBABILITY OF BELONGING TO ONE OF THE 4 HAPLOGROUPS DETECTED ($P=1$). BLACK ARROWS INDICATE INDIVIDUALS WITH UNCERTAIN ASSIGNMENTS ($P<0.05$). PIE CHARTS INDICATE THE PERCENTAGES OF THE HAPLOGROUPS CONTRIBUTING TO EACH LOCAL SAMPLE.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GSMU_AF04 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Af04	<i>Aristaeomorpha foliacea</i>			
Population\cluster	Cluster A	Cluster B	Cluster C	Cluster D
IBI	0	1	0	0
TYR	0	1	0	0
MAZ	1	1	0	0
PPA	0	1	0	0
ION	1	0	0	0
AEG	1	0	0	0
MOZ°	0	0	1	0
AUS°	0	0	0	1

° not shown in the geovisualization

DATA FROM:

Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán, 2013b. **Multilocus comparative phylogeography of two Aristeid shrimps of high commercial interest (*Aristeus antennatus* and *Aristaeomorpha foliacea*) reveals different responses to past environmental changes.** *PLOS ONE* 8(3):e59033 doi:10.1371/journal.pone.0059033.

GENETIC MARKER USED

Two nuclear genes (PEPCK and NaK) and one mitochondrial (COI) gene

DETAILS OF THE SAMPLING SITES ANALYSED

code	Site	latitude	longitude	FAO area, GFCM GSA	No of samples
IBI	Ibiza Channel	39°02'N	02°39'E	FAO37:GSA5	10
AEG	Aegean Sea	37°17'N	22°53'E	FAO37:GSA22	10
MOZ°	Mozambique Channel	25°57'S	34°38'E	FAO51	10
AUS°	North-western Australia	14°51'S	121°26'E	FAO57	10
total					317

° not shown in the geovisualization

MAIN FINDINGS (FERNANDEZ ET AL., 2013B)

1. Maximum Likelihood, Neighbour Joining and Bayesian analyses for the concatenated dataset generated identical tree topologies.
2. Two major phylogroups were detected. One group corresponding to AUS and the second including MED and MOZ, where MED appears monophyletic.
3. The AUS lineage of *A. foliacea* warrants consideration as a distinct species, with consequent implications in systematics and resource management.

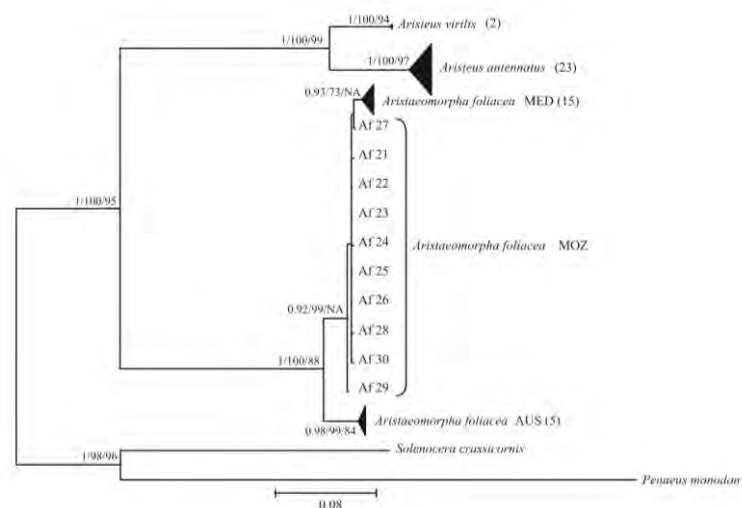


FIGURE 1 FROM FERNANDEZ *ET AL.*, 2013B. BAYESIAN CONDENSED TREE BASED ON CONCATENATED DATASET. THE NUMBERS ON NODES INDICATE POSTERIOR PROBABILITY VALUES FOR BAYESIAN TREE AND BOOTSTRAP VALUES FOR NEIGHBOR-JOINING AND MAXIMUM LIKELIHOOD TREES RESPECTIVELY.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AF05 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Af05	<i>Aristaeomorpha foliacea</i>
Population\cluster	Cluster A
IBI	1
AEG	1
MOZ [°]	1
AUS [°]	1

[°] not shown in the geovisualization

INFORMATION FROM STUDIES INCLUDING POPULATION SAMPLES OUTSIDE THE
MEDITERRANEAN SEA (P DATASETS)

Not data available

COMMENTS AND EVALUATION OF DATA: GIANT RED SHRIMP

The following table summarizes the results of the five population genetic studies performed so far for the giant red shrimp

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic Mediterranean structure	N°of Mediterranean clusters
Af01=Af03	★ ^{WC}	★	-	-	1(WC)
Af02	★ ^{WCE}	★★★	★	-	1(WCE)
Af04	★ ^{WCE}	★★	★★★	★	2(WC/CE)
Af05	★ ^{WE}	★	★★★	-	1(WE)

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★≤3; ★★4-6; ★★★>6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★≤20; ★★20-40; ★★★>40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall FST: ★≤0.05; ★★0.05-0.1; ★★★ >0.1)

Genetic structure = ★low, ★★medium ★★★ high support (whenever available the score is based on pairwise Fst values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean

Comments: two genetic clusters were identified, but the Mazara del Vallo sample showed an intermediate position from these two. These results are not confirmed by other papers of the same authors analysing the same sites with different markers; moreover, they are in contrast with other papers analysing other sites in the same region with different markers, and reporting a substantial genetic homogeneity within the Mediterranean Sea.

Gaps: no data for vast areas of the Mediterranean Sea

Technical notes: no data available for most variable markers (microsatellites, SNPs)

The analysis: *Aristeus antennatus*

INFORMATION FROM STUDIES INCLUDING MEDITERRANEAN POPULATION SAMPLES (HP DATASETS)

GDMU_Aa03

DATA FROM:

Sardà, F., C. Bas, M. I. Roldán, C. Pla & J. Lleonart, 1998. **Enzymatic and morphometric analyses in mediterranean populations of the rose shrimp, *Aristeus antennatus* (Risso, 1816).** *Journal of Experimental Marine Biology and Ecology* 221(1):131-144 doi:10.1016/S0022-0981(97)00119-6.

GENETIC MARKER USED

Allozymes (15 loci, 2 polymorphic within 95%)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
AL	Alicante			FAO37: GSA6	30
PA	Palma			FAO37: GSA5	30
BAd	Barcelona deep >1000m			FAO37: GSA6	30
BA	Barcelona			FAO37: GSA6	30
MAd	Marseille deep >1000m			FAO37: GSA7	30
MA	Marseille			FAO37: GSA7	30
RO	Rome			FAO37: GSA10	24
SI	Mazara			FAO37: GSA16	46
HA	Haifa			FAO37: GSA27	53
LI	Lisbon			FAO34	30
LA	Larache			FAO27	57
total					390

MAIN FINDINGS (SARDÀ ET AL., 1998)

1. Despite the high electrophoretic effort applied to the 27 enzyme systems, only fifteen loci useful in the population survey were identified. Ten of these fifteen loci were monomorphic, and the remaining five loci exhibited some allele variants. Two of these five loci were polymorphic within 95% confidence limits.
2. Heterogeneity was detected at the loci *MDH-1** and *PGM**. When Bonferroni procedure is applied to the five loci, only the *MDH-1** locus remains significant for Fisher's test and *PGM** for the Monte Carlo test. By applying the Bonferroni procedure rather weak evidence exist about genetic differentiation among samples. The measures of genetic differentiation were: D maximum=0.002 and F_{ST} =0.017.
3. The very low values for F_{ST} and genetic distance provide evidence about no population structure among the samples. Comparison of allele frequencies by contingency tables indicated low levels of differentiation among samples.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA03 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa03	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
AL	1
PA	1
BAd	1
BA	1
MAd	1
MA	1
RO	1
SI	1
HA	1
LI	1
LA	1

DATA FROM:

Maggio, T., S. Lo Brutto, R. Cannas, A. M. Deiana & M. Arculeo, 2009. **Environmental features of deep-sea habitats linked to the genetic population structure of a crustacean species in the Mediterranean Sea.** *Marine Ecology* 30(3):354-365 doi:10.1111/j.1439-0485.2008.00277.x.

GENETIC MARKER USED

mtDNA (control region, 369 bp)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected in 2006/2007 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
CA	Cataluna	--	--	FAO37: GSA6	14
SR	Sanremo	--	--	FAO37: GSA9	28
SM	Santa Margherita Ligure	--	--	FAO37: GSA9	26
SI	Siniscola	--	--	FAO37: GSA11	8
SA	Sant'Antioco	--	--	FAO37: GSA11	22
TE	Terrasini	--	--	FAO37: GSA10	29
SV	San Vito lo Capo	--	--	FAO37: GSA10	29
SS	Stretto di Sicilia	--	--	FAO37: GSA16	19
total					175

MAIN FINDINGS (MAGGIO ET AL., 2009)

1. AMOVA showed a high level of genetic variation, more within than between populations, and a low but significant F_{ST} value was recovered (overall $\Phi_{ST} = 0.017$; $P < 0.05$). Clustering the samples from the Algero-Provencal sub-basin (CA, SM, SR and SA), Tyrrhenian sub-basin (TE and SI) and Straits of Sicily (SV and SS) no genetic differentiation was observed.
2. None of the pairwise comparisons was significant, except in two cases, SI versus TE and SI versus SR, where the higher F_{ST} values probably resulted from the small SI sample size.
3. Minimum spanning network did not separate any haplotype group and haplotype distribution does not mirror the geographic origin of the samples. The absence of population substructuring was also observed with a principal coordinate analysis PCA, which uses an individual-by-individual comparison. These results revealed extensive gene flow among populations.
4. Information on demographic history based on mismatch analysis revealed an unstable population, showing an alternate pattern of growth and decline.
5. The results indicated that in the western and central Mediterranean basins *A. antennatus* is a large panmictic population with a fluctuating abundance.

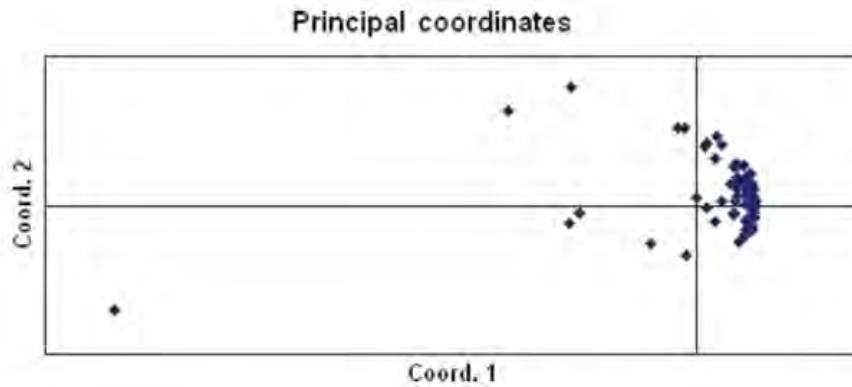
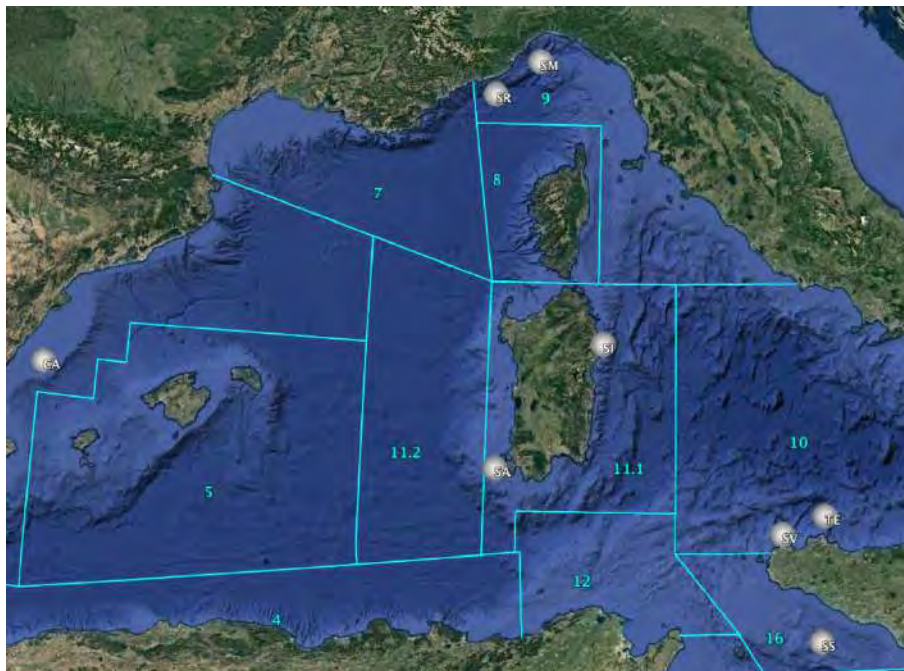


FIGURE 3 FROM MAGGIO *ET AL.*, 2009. PRINCIPAL COORDINATES ANALYSIS (PCA) VIA DISTANCE MATRIX WITH DATA STANDARDIZATION.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA05 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa05	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
CA	1
SR	1
SM	1
SI	1
SA	1
TE	1
SV	1
SS	1

DATA FROM:

Roldán MI, Heras S, Patellani R, Maltagliati F: **Analysis of genetic structure of the red shrimp *Aristeus antennatus* from the Western Mediterranean employing two mitochondrial regions.** *Genetica* 2009, 136: 1-4

GENETIC MARKER USED

Mitochondrial DNA [cytochrome c oxidase subunit I (514 bp); ribosomal 16S subunit (547 bp)]

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
PLM	Palamós	41°48'N	3°26'E	FAO37: GSA6	59
GEN	Genova	44°18'N	8°52'E	FAO37: GSA9	42
PAL	Palermo	38°24'N	12°57'E	FAO37: GSA10	36
total					137

MAIN FINDINGS (ROLDÁN ET AL., 2009)

1. Analysis of molecular variance (AMOVA) partitioned most of variance within samples (96.5%), whereas the remaining portion was attributed to the among-sample component, and the estimate of the fixation index was significantly greater than zero ($\Phi_{ST} = 0.035$, $P = 0.005$).
2. Pairwise values of F_{ST} were significantly different from zero in the comparisons Palamos–Genova ($F_{ST} = 0.053$, $P = 0.005$) and Genova–Palermo ($F_{ST} = 0.035$, $P = 0.033$); conversely, the comparison Palamos–Palermo was not significant ($F_{ST} = 0.002$, $P = 0.303$).
3. Among sample genetic diversity was not significant and no geographical patterns in the distribution of haplotypes were apparent.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA06 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa06	<i>Aristeus antennatus</i>	
Population\cluster	Cluster A	Cluster B
PLM	0	1
GEN	1	0
PAL	0	1

DATA FROM:

Sardà, F., M. I. Roldán, S. Heras & F. Maltagliati, 2010. **Influence of the genetic structure of the red and blue shrimp, *Aristeus antennatus* (Risso, 1816), on the sustainability of a deep-sea population along a depth gradient in the Western Mediterranean.** *Scientia Marina* 74(3):569-575 doi:10.3989/scimar.2010.74n3569.

GENETIC MARKER USED

mitochondrial DNA (16S gene, 547 bp)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected in 2003-2004 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
BA_1	Barcelona 350 m			FAO37: GSA6	45
BA_2	Barcelona 700 m			FAO37: GSA6	206
BA_3	Barcelona 1100 m			FAO37: GSA6	46
BA_4	Barcelona 1500 m			FAO37: GSA6	24
total					321

MAIN FINDINGS (SARDÀ ET AL., 2010)

1. A partial region (547 bp) of the mtDNA 16S gene from 321 individuals collected at four depths in the Catalan Sea off Barcelona was analyzed
2. The AMOVA analysis distributed all the molecular variance in the within-depth components, and the fixation index (F_{ST}) was not significantly different from zero.
3. Replicate runs of BAPS yielded identical results and produced three genetic clusters ($P = 0.563$). The proportions of the three genetic clusters did not differ substantially across the samples from the four different depths.
4. The mismatch distribution results within the samples from each depth as well as in the data set overall were consistent with past population growth. Results suggest population expansion only at the two shallowest depths sampled.
5. The high levels of genetic homogeneity detected among the vertical samples of *A. antennatus* from the Catalan Sea were a reason for rejecting the hypothesis that the deeper-dwelling stocks are isolated from the exploited stocks previously proposed by Sardà et al. (2003).
6. Furthermore, combining these results with those reported in Roldán et al., (2009), a general picture of relative genetic homogeneity emerges in the three spatial dimensions of the species distribution in the western Mediterranean. Larval and adult dispersal is, thus, effective at ensuring gene flow, and as a result significant vertical and geographical genetic structuring were absent.

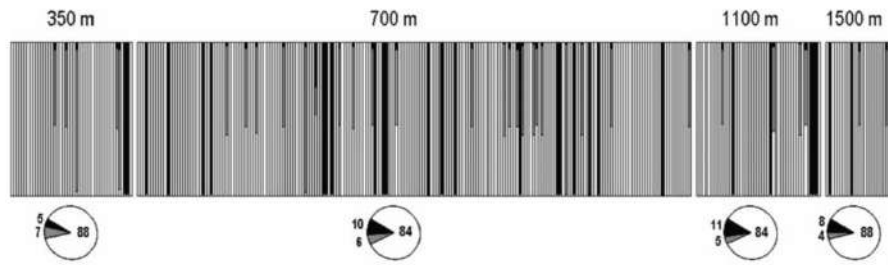
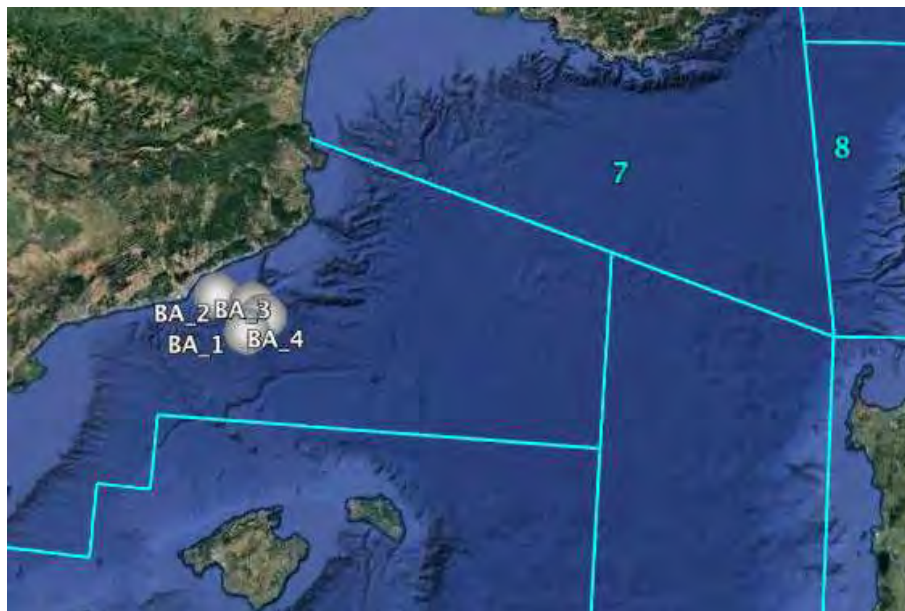


FIGURE 2 FROM SARDÀ *ET AL.*, 2010. RESULTS OF BAYESIAN ASSIGNMENT ANALYSIS. THREE GENETIC CLUSTERS (BLACK, GREY, AND WHITE) WERE IDENTIFIED ($P = 0.563$), THE COLOURS REPRESENTING THE ASSIGNMENT PROBABILITIES OF INDIVIDUAL GENOTYPES TO A CLUSTER. IN THE BAR GRAPHS EACH BAR REPRESENTS AN INDIVIDUAL, WHEREAS OVERALL PROPORTIONS OF THE THREE GENETIC CLUSTERS IN THE DEPTH SAMPLES ARE REPRESENTED BY PIE CHARTS.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA07 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa07	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
BA_1	1
BA_2	1
BA_3	1
BA_4	1

DATA FROM:

Fernández, M. V., S. Heras, F. Maltagliati, A. Turco & M. I. Roldán, 2011b. **Genetic structure in the blue and red shrimp *Aristeus antennatus* and the role played by hydrographical and oceanographical barriers.** *Marine Ecology Progress Series* 421:163-171 doi:10.3354/meps08881.

GENETIC MARKER USED

Mitochondrial DNA (16S rDNA 546 bp, and COI 514 bp)

DETAILS OF THE SAMPLING SITES ANALYSED

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
PO	Faro (AO)	--	--	FAO27	38
AS	Alboran Sea	35° 59' N	03° 05' W	FAO37: GSA2	53
AL	Almeria (WM)	36° 32' N	02° 07' W	FAO37: GSA1	45
SO	Soller (WM)	39° 44' N	02° 09' E	FAO37: GSA5	48
CA	Cabrera (WM)	39° 02' N	02° 38' E	FAO37: GSA5	40
PS	Palamos (WM)	--	--	FAO37: GSA11	59
LI	Gulf of Lion (WM)	42° 35' N	4° 13' E	FAO37: GSA11	51
GE	Genova (WM)	--	--	FAO37: GSA11	44
PA	Palermo (WM)	--	--	FAO37: GSA10	40
IS	Ionian Sea (EM)	37° 37' N	21° 03' E	FAO37: GSA20	40
MO ^o	Mozambique Channel (IO)	17° 32' S	38° 29' E	FAO51	48
total					506

^o not shown in the geovisualization

MAIN FINDINGS (FERNÁNDEZ ET AL., 2011)

1. Results of AMOVA from the unstructured data set showed that 16.4% of genetic heterogeneity was apportioned among samples ($\Phi_{ST} = 0.164$, $p < 0.001$) and 83.6% within samples. These results are indicative of genetic differentiation among samples.
2. AMOVA applied to sequences pooled by geographical regions revealed significant population genetic structure (23.3 % variation among regions, $\Phi_{ST} = 0.236$, $p < 0.001$).
3. The outcome of AMOVAs carried out with these criteria suggested that the AS sample is part of WM.
4. When IO was omitted from the analysis, genetic divergence was lower but still significant ($\Phi_{ST} = 0.233$, $p < 0.001$), with 22.8% of the molecular variance due to differences among regions.
5. The high haplotypic diversity of the Eastern Mediterranean, Atlantic and Indian Ocean samples reflects the occurrence of a number of private haplotypes, which are also responsible for significant genetic divergence between these samples and the Western Mediterranean ones.
6. The Bayesian assignment of individuals (Fig. 2) to haplogroups detected 2 main clusters ($p = 1$). Samples from WM (including AS) were mostly assigned to the first haplogroup (HG1). In contrast, the proportion of specimens belonging to the second haplogroup (HG2) was higher in EM (49 %), AO (63 %) and IO (89%).
7. Tests aimed at detecting population substructuring within the WM region (AS included) highlighted that the largest part of molecular variance occurred within samples (98.5%); but there was also a small but significant variance among samples ($\Phi_{ST} = 0.015$, $p = 0.005$). However, the present results and those of past studies suggest considering WM a genetically homogeneous unit, although the presence and origin of a number of diverse haplotypes remains an open question.

8. The analysis of mismatch distributions, neutrality tests, and star-like patterns present in the network of haplotypes provided consistent inference of past population expansion in all areas, except EM.

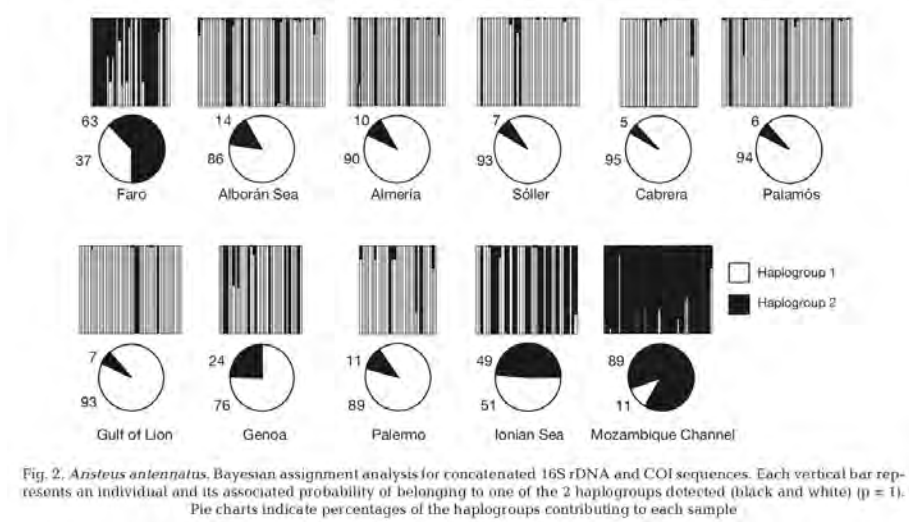


FIGURE 2 FROM FERNÁNDEZ *ET AL.*, 2011B.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA09 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa09	<i>Aristeus antennatus</i>	
Population\cluster	Cluster A	Cluster B
PO	0	1
AS	1	0
AL	1	0
SO	1	0
CA	1	0
PS	1	0
LI	1	0
GE	1	0
PA	1	0
IS	0	1
MO ^o	0	1

^o not shown in the geovisualization

DATA FROM:

Cannas, R., F. Sacco, M. C. Follesa, A. Sabatini, M. Arculeo, S. Lo Brutto, T. Maggio, A. M. Deiana & A. Cau, 2012. **Genetic variability of the blue and red shrimp *Aristeus antennatus* in the Western Mediterranean Sea inferred by DNA microsatellite loci.** *Marine Ecology* 33(3):350-363
doi:10.1111/j.1439-0485.2011.00504.x.

GENETIC MARKER USED

Microsatellites (8 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected in 2006-2008 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
AL	Algeria	--	--	FAO37: GSA4	20
SR	Sanremo	--	--	FAO37: GSA9	29
CG	Cala Gonone	--	--	FAO37: GSA11	20
CA	Cagliari	--	--	FAO37: GSA11	17
VA	Vapore			FAO37: GSA11	16
SA	Sant'Antioco	--	--	FAO37: GSA11	26
PPs	Pesca Profonda sud	--	--	FAO37: GSA11	55
PPn	Pesca Profonda nord	--	--	FAO37: GSA11	24
TE	Terrasini	--	--	FAO37: GSA10	20
total					227

MAIN FINDINGS (CANNAS ET AL., 2012)

1. Genetic variation at eight microsatellite loci was studied in nine populations of the blue and red shrimp *A. antennatus* to quantify the genetic diversity, investigate population genetic structure, and assess whether distinct stocks occur in the Western Mediterranean Sea. Due to a different bathymetric distribution of the sexes and their possible differential dispersal capacity, the hypothesis of instantaneous sex-biased dispersal was also tested.
2. Individuals were collected from both commercial bottom trawling (<800 m depth) and experimental deep-bottom trawling (from 800 to 1600 m).
3. Microsatellites data highlighted a high level of gene flow and no evidence of genetic partitioning. No significant variation was found ($F_{ST} = 0.00673$, $P\text{-value} = 0.067$) even when shrimps from exploited and those from deep-water unexploited grounds were compared.
4. No evidence of reduction or expansion of population size in the recent past was found, as indicated by the bottleneck and interlocus g -tests. The results are consistent with previous studies using mitochondrial gene methods and allozymes, indicating that, for this species, extensive pelagic larval dispersal and adult migration are probably responsible for the genetic homogeneity observed.
5. The hypothesis of sex-biased dispersal was tested; mean values of corrected assignment indices and mean relatedness values were higher for males, suggesting that females are the more widely dispersing sex.

TABLE 4 FROM CANNAS ET AL., 2012.

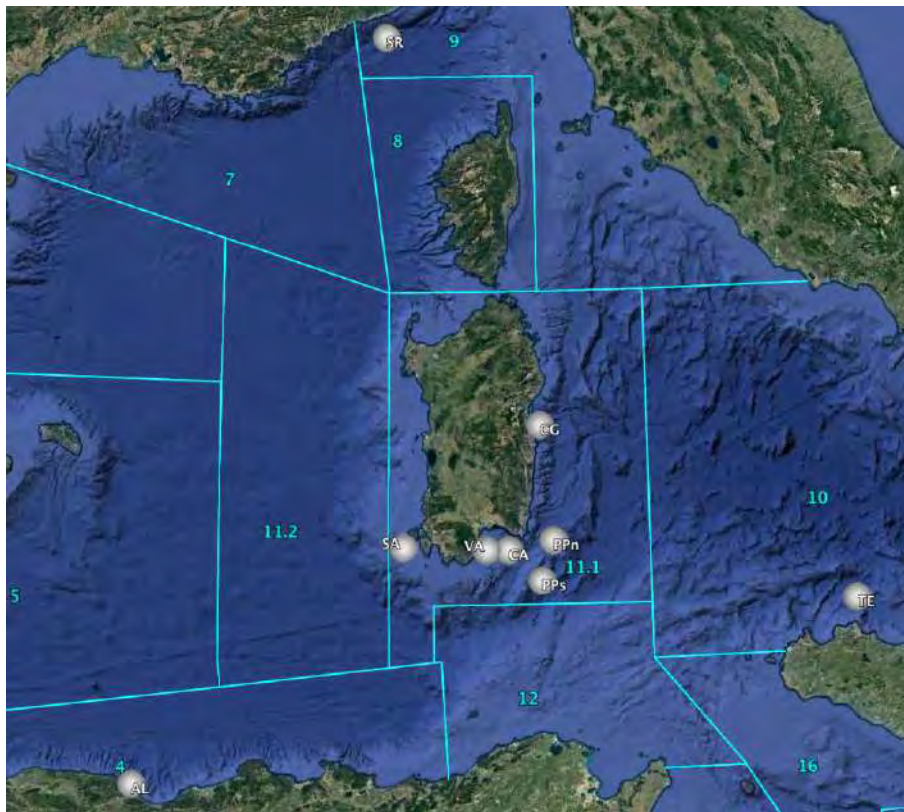
Table 4. Analysis of molecular variance AMOVA.

Source of variation	df	Sum of squares	Variance components	Percentage of variation	Fixation indices
Among populations	8	33.342	0.02118	0.67	$F_{ST} = 0.00673^*$
Within populations	445	1391.065	3.12599	99.33	
Total	453	1424.407	3.14717		

*P-value = 0.06723 ± 0.00076 (1,00,172 permutations).

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA09 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa09	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
AL	1
SR	1
CG	1
CA	1
VA	1
SA	1
PPs	1
PPn	1
TE	1

DATA FROM:

Lo Brutto, S., T. Maggio, A. M. Deiana, R. Cannas & M. Arculeo, 2012. **Further investigations on populations of the deep-water blue and red shrimp *Aristeus antennatus* (Risso, 1816) (Decapoda, Dendrobranchiata), as inferred from Amplified Fragment Length Polymorphism (AFLP) and mtDNA analyses.** *Crustaceana* 85(11):1393-1408 doi:10.1163/15685403-00003131.

GENETIC MARKER USED

AFLP (145 loci) and mtDNA (control region, 369 bp)

DETAILS OF THE SAMPLING SITES ANALYSED

Samples are the same as in GDMU_Aa05 with the addition of the Atlantic sample from Faro:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
CA	Cataluna	--	--	FAO37: GSA6	30
SR	Sanremo	--	--	FAO37: GSA9	30
SM	Santa Margherita Ligure	--	--	FAO37: GSA9	26
SI	Siniscola	--	--	FAO37: GSA11	27
SA	Sant'Antioco	--	--	FAO37: GSA11	30
TE	Terrasini	--	--	FAO37: GSA10	30
SS	Stretto di Sicilia	--	--	FAO37: GSA16	28
PO	Faro	--	--	FAO27	46
total					247

MAIN FINDINGS (LO BRUTTO ET AL., 2012)

1. AMOVA results of AFLP data revealed that the overall genetic variation among-populations was lower (11.81%) than within-populations (88.19%), though the fixation index proved to be significant ($F_{ST} = 0.118$; $p < 0.001$).
2. The genetic variation between the Atlantic and Mediterranean samples was found to be not significant ($F_{CT} = -0.007$; N.S.), indicating that the transition area between the Atlantic Ocean and the Mediterranean Sea does not act as a barrier to gene flow.
3. No differentiation was found also among the three Mediterranean sub-basins as was evident by the non-significant fixation index ($F_{CT} = 0.04$; N.S.)
4. Bayesian analysis conducted with Structure also demonstrated the absence of genetic differentiation between the Atlantic and Mediterranean populations and within the Mediterranean basin (see Fig. 1). The separation into two genetic clusters negated any geographical correspondence with the Atlantic-Mediterranean subdivision and revealed a high degree of membership to the first cluster for the Mediterranean samples.
5. AMOVA results based on mtDNA were in accordance with the AFLP results

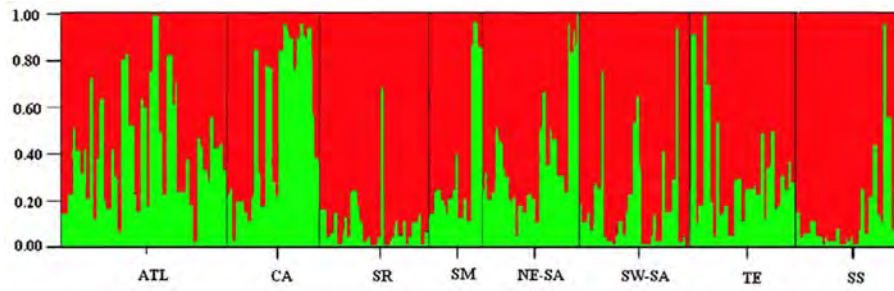


FIGURE 1 FROM LO BRUTTO *ET AL.*, 2012. SUMMARY PLOT OF Q ESTIMATES (PROPORTION OF MEMBERSHIP) FOR K = 2 FROM THE AFLP ANALYSIS.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA10 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa10	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
CA	1
SR	1
SM	1
SI	1
SA	1
TE	1
SS	1
PO	1

DATA FROM:

Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán, 2013b. **Multilocus Comparative Phylogeography of Two Aristeid Shrimps of High Commercial Interest (*Aristeus antennatus* and *Aristaeomorpha foliacea*) Reveals Different Responses to Past Environmental Changes.** *PLOS ONE* 8(3): e59033 doi:10.1371/journal.pone.0059033.

GENETIC MARKER USED

Two nuclear genes (PEPCK and NaK) and one mitochondrial (COI) gene

DETAILS OF THE SAMPLING SITES ANALYSED

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
PO	Faro (AO)	--	--	FAO27	10
AS	Alboran Sea	35° 59' N	03° 05' W	FAO37: GSA2	10
LI	Gulf of Lion (WM)	42° 35' N	4° 13' E	FAO37: GSA11	10
IS	Ionian Sea (EM)	37° 37' N	21° 03' E	FAO37: GSA20	10
MO°	Mozambique Channel (IO)	17° 32' S	38° 29' E		3
total					43

MAIN FINDINGS (FERNÁNDEZ ET AL., 2013B)

1. ML, NJ and BI analyses for the concatenated dataset generated identical tree topologies
2. Within *Aristeus antennatus* no clear associations between geographical distribution and sequences obtained was detected. The lack of a clear geographical pattern in the distribution of genetic diversity in *Aristeus antennatus* was corroborated by the low and non-significant “among samples” component of molecular variance (9.6%, $\Phi_{ST} = 0.096$, $P = 0.079$)

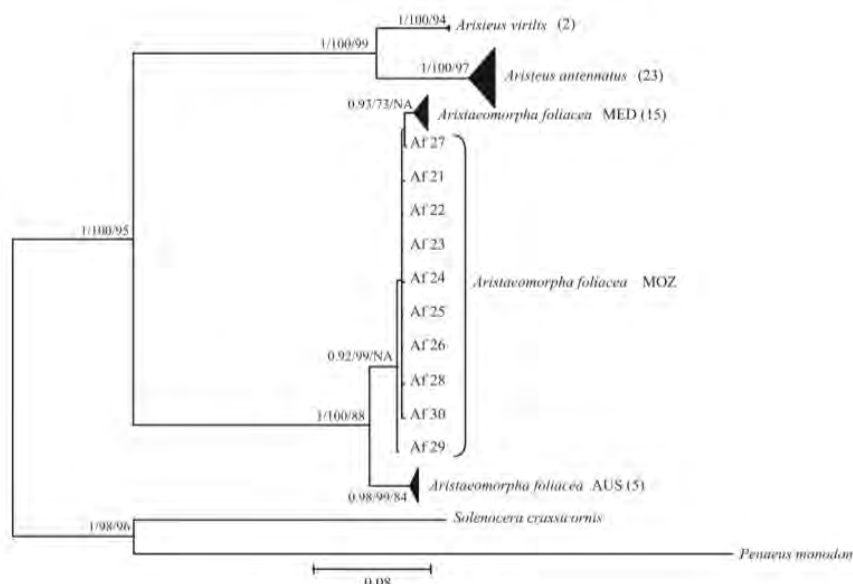


FIGURE 1 FROM FERNANDEZ ET AL., 2013B. BAYESIAN CONDENSED TREE BASED ON CONCATENATED DATASET. THE NUMBERS ON NODES INDICATE POSTERIOR PROBABILITY VALUES FOR BAYESIAN TREE AND BOOTSTRAP VALUES FOR NEIGHBOR-JOINING AND MAXIMUM LIKELIHOOD TREES RESPECTIVELY.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA10 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa11	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
PO	1
AS	1
LI	1
IS	1
MO°	1

DATA FROM:

Marra, A., S. Mona, R. M. Sa, G. D'Onghia & P. Maiorano, 2015. **Population Genetic History of *Aristeus antennatus* (Crustacea: Decapoda) in the Western and Central Mediterranean Sea.** *Plos One* 10(3):16
doi:10.1371/journal.pone.0117272.

GENETIC MARKER USED

Mitochondrial DNA (COI and 16S)

DETAILS OF THE SAMPLING SITES ANALYSED

Samples were collected in 2010–2011 from the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
CSTy1	70	--	--	FAO37: GSA10	43
CSTy2	18	--	--	FAO37: GSA10	39
NWIS1	20	--	--	FAO37: GSA19	36
NWIS2	15	--	--	FAO37: GSA19	48
NWIS3	55	--	--	FAO37: GSA19	44
NWIS4	64	--	--	FAO37: GSA19	24
NWIS5	61	--	--	FAO37: GSA19	38
NWIS6	65	--	--	FAO37: GSA19	8
SAS1	152-98			FAO37: GSA18	24
SAS2	144			FAO37: GSA18	15
total					319

MAIN FINDINGS (MARRA ET AL., 2015)

1. Mitochondrial data highlighted two main results:
2. The genetic variance was mostly partitioned in the within-population component (98.11%), meaning there are no large difference in alleles distribution among populations, even though the fixation index was significantly greater than zero ($\Phi_{ST} = 0.018$, P-value < 0.05).
3. When differentiation was investigated among geographical areas (Tyrrhenian, Ionian and Adriatic basins) or fishing depth ranges (500–600 m, 600–700 m and >700 m), Φ_{CT} was not significant in both cases. This means that neither geography nor depth can explain the pattern of genetic variability in our populations.
4. MDS plot showed that the studied populations are homogeneous, with the non-Mediterranean Sea sample from the Faro station being the only exception.
5. The genetic diversity values are consistent with previous data within the Mediterranean and the absence of barriers to gene flow within the Mediterranean Sea.
6. A constant long-term effective population size in almost all demes but a strong signature of population expansion in the pooled sample about 50,000 years B.P./ago.

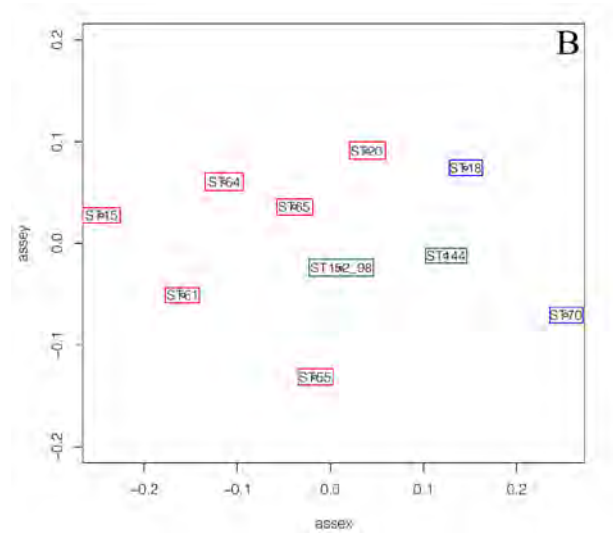


FIGURE A FROM MARRA ET AL., 2015. MDS PLOT BASED ON PAIRWISE F_{ST} VALUES.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_AA12 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Aa12	<i>Aristeus antennatus</i>
Population\cluster	Cluster A
CSTy1	1
CSTy2	1
NWIS1	1
NWIS2	1
NWIS3	1
NWIS4	1
NWIS5	1
NWIS6	1
SAS1	1
SAS2	1
CSTy1	1
CSTy2	1

INFORMATION FROM STUDIES INCLUDING POPULATION SAMPLES OUTSIDE THE MEDITERRANEAN SEA (P DATASETS)

Data not available

COMMENTS AND EVALUATION OF DATA: BLUE AND RED SHRIMP

The following table summarizes the results of the five population genetic studies performed so far for the blue and red shrimp

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic Mediterranean structure	N° of Mediterranean clusters
Aa03	★ ^{WCE}	★★	★	-	1(WCE)
Aa05	★★ ^{WC}	★★	★	-	1(WC)
Aa06	★ ^W	★★★	★	★	2(W)
Aa07	★ ^W	★★★	-	-	1(W)
Aa08	★ ^{WE}	★★★	★★★	★	2(W/E)
Aa09	★ ^W	★★	-	-	1(W)
Aa10	★ ^{WC}	★★	★★★	-	1(WC)
Aa11	★ ^{WE}	★	-	-	1(WE)
Aa12	★ ^{WC}	★★	★	-	1(WC)

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★≤3; ★★4-6; ★★★>6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★≤20; ★★20-40; ★★★>40;

Overall differentiation = ★ low;★★ medium;★★★ high support; (whenever available the score is based on the value of the overall FST: ★≤0.05;★★0.05-0.1;★★★ >0.1)

Genetic structure = ★ low,★★ medium ★★★ high support (whenever available the score is based on pairwise Fst values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean

Comments: two genetic clusters were identified. These results are not confirmed by other papers of the same authors analysing the same sites with different markers; moreover, they are in contrast with other papers analysing other sites in the same region with different markers, and reporting a substantial genetic homogeneity within the Mediterranean Sea.

Gaps: no data for vast areas of the Mediterranean Sea

Technical notes: no data available for most variable markers (microsatellites, SNPs)

The analysis: *Merluccius merluccius*

INFORMATION FROM STUDIES INCLUDING MEDITERRANEAN POPULATION SAMPLES (HP DATASETS)

GDMU_Mm01

DATA FROM:

Lo Brutto, S., M. Arculeo, A. Mauro, M. Scalisi, M. Cammarata & N. Parrinello, 1998. **Allozymic variation in Mediterranean hake *Merluccius merluccius* (Gadidae)**. *Italian Journal of Zoology* 65:49-52.

GENETIC MARKER USED

Allozymes (20 loci, 4 polymorphic)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
C1	Canale di Sicilia 1	--	--	FAO37:GSA16	37
C2	Canale di Sicilia 2	--	--	FAO37:GSA16	46
C3	Canale di Sicilia 3	--	--	FAO37:GSA16	43
C4	Canale di Sicilia 4	--	--	FAO37:GSA16	50
C5	Canale di Sicilia 5	--	--	FAO37:GSA15	73
C6	Canale di Sicilia 6	--	--	FAO37:GSA15	31
C7	Canale di Sicilia 7	--	--	FAO37:GSA15	21
A1	Adriatico 1	--	--	FAO37:GSA18	59
T1	Tirreno 1	--	--	FAO37:GSA9	30
T2	Tirreno 2	--	--	FAO37:GSA11	60
total					450

MAIN FINDINGS (LO BRUTTO ET AL., 1998)

1. FST analysis showed heterogeneity only in one locus (PGI-2*, $F_{ST}=0.024$, $P=0.002$)
2. UPGMA clustering showed that C5 sample was very different from the others; this could be caused by the small sample size
3. If C5 is eliminated there was no significant population subdivision

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM01 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm01	<i>Merluccius merluccius</i>	
Population\cluster	Cluster A	Cluster B
C1	1	0
C2	1	0
C3	1	0
C4	1	0
C5	0	1
C6	1	0
C7	1	0
A1	1	0
T1	1	0
T2	1	0

DATA FROM:

Roldàn, M. I., J. L. Garcia-Marin, F. M. Utter & C. Pla, 1998. **Population genetic structure of European hake, *Merluccius merluccius***. *Heredity* 81:327-334 doi:10.1038/sj.hdy.6883830.

GENETIC MARKER USED

Allozymes (34 loci, 20 polymorphic)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
IR	Irish Sea	--	--	FAO27	125
AR	Arcachon			FAO27	57
CA	Cantabria	--	--	FAO27	84
GAI	Galicia I	--	--	FAO27	61
GAI	Galicia II	--	--	FAO27	46
MA	Malaga	--	--	FAO37:GSA1	97
MU	Murcia	--	--	FAO37:GSA6	149
PE	Peñones	--	--	FAO37:GSA2	101
NA	Napoli			FAO37:GSA10	90
LA	Larache			FAO34	100
total					910

MAIN FINDINGS (ROLDÀN ET AL., 1998)

1. A major subdivision between Atlantic and Mediterranean collections (mean Nei's genetic distance within regions 0.005; between regions 0.014) was particularly evident from allelic differences at the GAPDH-1* and GR-2* loci.
2. Further population subdivision was indicated in the Atlantic: the Larache sample is relatively different from others of the Atlantic and close in genetic distance to the adjacent western Mediterranean collections. Additional substructuring to the north is suggested by the distinction of the Ireland sample which would not correspond to the ICES-proposed northern and southern stocks.
3. Gene flow from adjacent Atlantic to proximal Mediterranean populations was suggested by allele frequencies and relevant oceanographic and geological information.

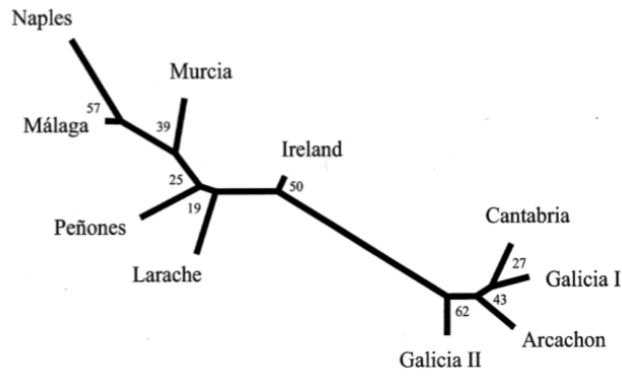
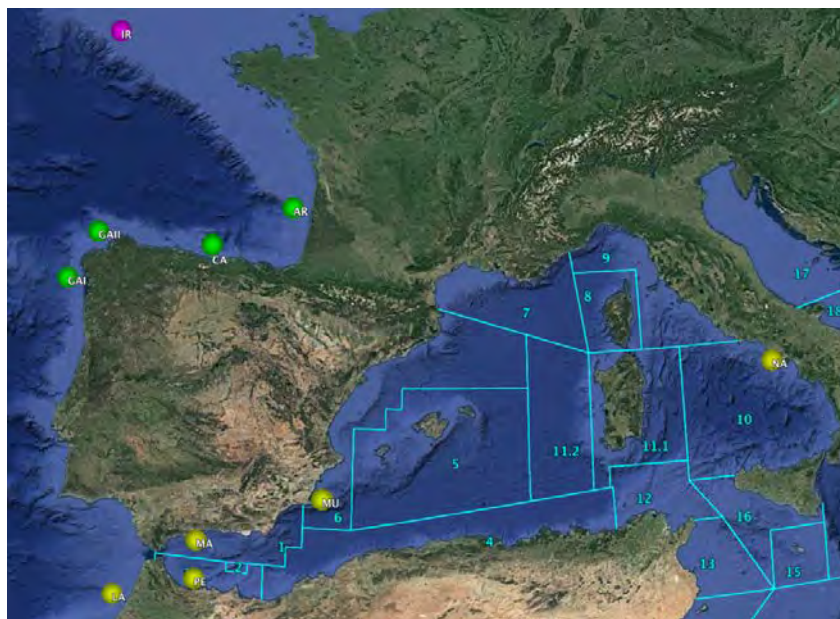


FIGURE 2 FROM ROLDÀN ET AL., 1998. UNROOTED NEIGHBOUR-JOINING TREE CONSTRUCTED FROM THE CAVALLI-SFORZA AND EDWARDS CHORD DISTANCE VALUES. THE NUMBERS REFER TO THE PERCENTAGE OF 1000 BOOTSTRAP REPLICATIONS.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM02 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm02	<i>Merluccius merluccius</i>		
Population\cluster	Cluster A	Cluster B	Cluster C
IR	0	0	1
AR	1	0	0
CA	1	0	0
GAI	1	0	0
GAI	1	0	0
MA	0	1	0
MU	0	1	0
PE	0	1	0
NA	0	1	0
LA	0	1	0

DATA FROM:

Lundy, C. J., P. Moran, C. Rico, R. S. Milner & G. M. Hewitt, 1999. **Macrogeographical population differentiation in oceanic environments: a case study of European hake (*Merluccius merluccius*), a commercially important fish.** *Molecular Ecology* 8(11):1889-1898 doi:10.1046/j.1365-294x.1999.00789.x.

GENETIC MARKER USED

Microsatellites (6 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

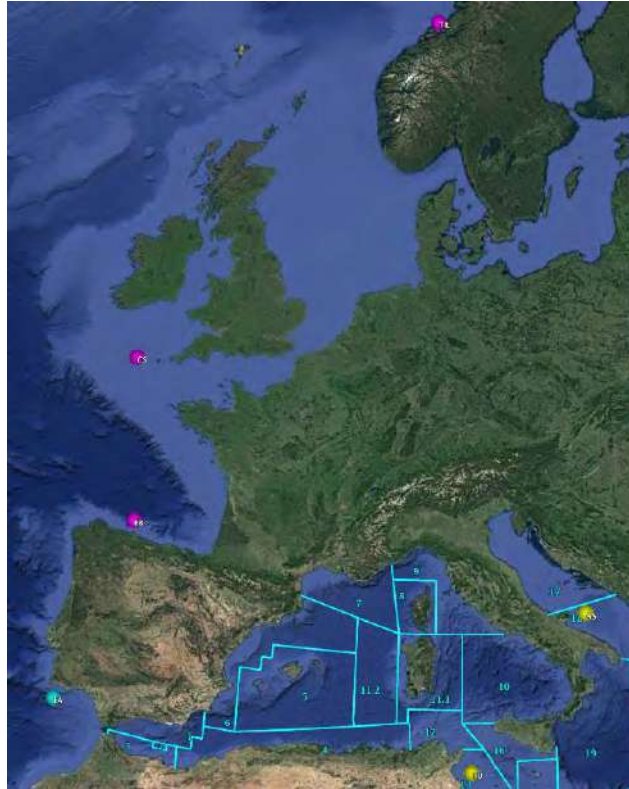
code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
TR	Trondheimsfjord	--	--	FAO27	91
CS	Celtic Sea			FAO27	67
BB	Bay Biscay	--	--	FAO27	100
FA	Faro	--	--	FAO27	64
TU	Tunisia	--	--	FAO37:GSA13	61
AS	Adriatic Sea	--	--	FAO37:GSA18	100
total					483

MAIN FINDINGS (LUNDY ET AL., 1999)

1. Population subdivision was found between Mediterranean and Atlantic samples, $\theta = 0.029$ ($P < 0.001$).
2. No substructuring was found between samples within the Mediterranean Sea, $\theta = 0.003$ and $RST = 0.007$ ($P > 0.05$).
3. The Atlantic population structure appears to be more complex than previously suggested by the placement of stock boundaries by the International Council for the Exploration of the Seas (ICES). Analyses based on microsatellites suggest that differentiation exists between Bay of Biscay and Portuguese samples, $\theta = 0.013$ ($P < 0.001$), $RST = 0.036$ ($P < 0.001$) which are currently managed as one stock. By contrast, fixation indices indicated no differentiation between southern Bay of Biscay samples and Celtic Sea samples, $\theta = 0.003$ ($P = 0.02$), $\varphi_{ST} = 0.007$ ($P = 0.10$) which are managed as separate stocks.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM03 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm03	<i>Merluccius merluccius</i>		
Population\cluster	Cluster A	Cluster B	Cluster C
TR	0	0	1
CS	0	0	1
BB	0	0	1
FA	0	1	0
TU	1	0	0
AS	1	0	0

DATA FROM:

Castillo AG, Martinez JL, Garcia-Vazquez E: **Fine spatial structure of Atlantic hake (*Merluccius merluccius*) stocks revealed by variation at microsatellite loci.** *Marine Biotechnology* 2004, 6(4):299-306.

GENETIC MARKER USED

Microsatellites (5 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2000-2001 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
ICESVI	ICESVI	--	--	FAO27	39
ICESVIIj	ICESVIIj	--	--	FAO27	50
ICESVIIIa,b,d	ICESVIIIa,b,d	--	--	FAO27	60
ICESVIIIc2	ICESVIIIc2	--	--	FAO27	45
ICESVIIIc3	ICESVIIIc3	--	--	FAO27	43
ICESIX	ICESIX	--	--	FAO27	87
MED	MED	--	--	FAO37:GSA6	55
AEG	AEG	--	--	FAO37:GSA22	52
ION	ION	--	--	FAO37:GSA20	73
total					504

MAIN FINDINGS (CASTILLO ET AL., 2004)

1. Significant genetic differentiation was found between samples, suggesting a fine subdivision of Atlantic and Mediterranean hake stocks.
2. All the samples presented a high level of genetic variability
3. Mediterranean samples were less variable than Atlantic samples at these 5 microsatellite loci
4. Genetic differentiation among samples was statistically significant ($P < 0.05$) in most cases.
5. First, Atlantic and Mediterranean hakes clustered in different branches. Second, in each branch (Mediterranean or Atlantic), there were secondary groups of samples.
6. In the Mediterranean, western locations (Spanish coast and Ionian Sea) clustered very close to each other and apart from the Aegean hakes only according the dendrogram based on $(\delta\mu)^2$ genetic distances (Figure 3 from Castillo *et al.*, 2004), while a different pattern was observed in the Neighbor-joining dendrogram based on Nei genetic distances (Figure 2 from Castillo *et al.*, 2004).

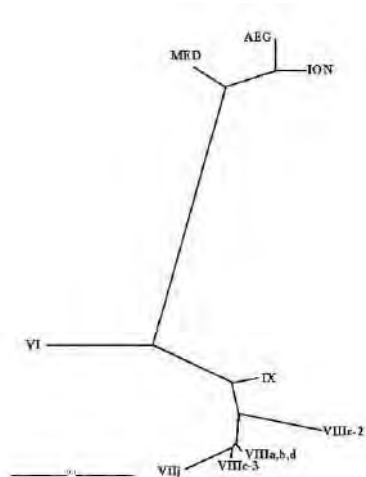


Figure 2. Neighbor-joining dendrogram based on Nei 78 genetic distances between the considered fish samples.

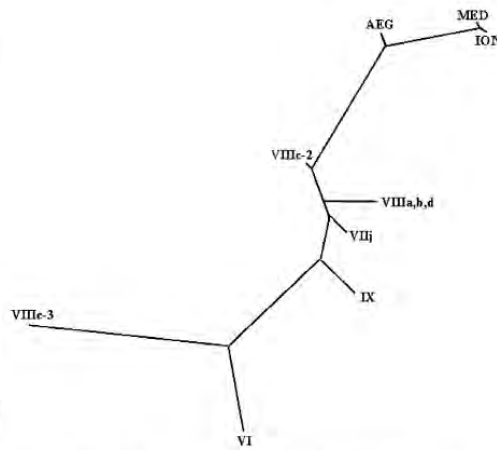


Figure 3. Neighbor-joining distances based on $(\delta\mu)^2$ distances.

FIGURE 2 AND 3 FROM CASTILLO *ET AL.*, 2004

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM07 DATASET

GEOVISUALIZATION



*based on the $(\delta\mu)^2$ genetic distances

MATRIX OF GENETIC CLUSTERING

GDMU_Mm07	<i>Merluccius merluccius</i>				
Population\cluster	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
ICESVI	1	0	0	0	0
ICESVIIj	0	1	0	0	0
ICESVIIIa,b,d	0	1	0	0	0
ICESVIIIc2	0	1	0	0	0
ICESVIIIc3	0	1	0	0	0
ICESIX	0	1	0	0	0
MED	0	0	1	0	0
AEG	0	0	0	0	1
ION	0	0	0	1	0

DATA FROM:

Levi, D., B. Patti, P. Rizzo, S. Lo Brutto, N. Parrinello & M. Arculeo, 2004. Genetic and morphometric variations of Mediterranean hake, *Merluccius merluccius*, in the Strait of Sicily (central Mediterranean): implications for stock assessment of shared resources. *Italian Journal of Zoology* 71(2):165-170 doi:10.1080/11250000409356568.

GENETIC MARKER USED

allozymes (5 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

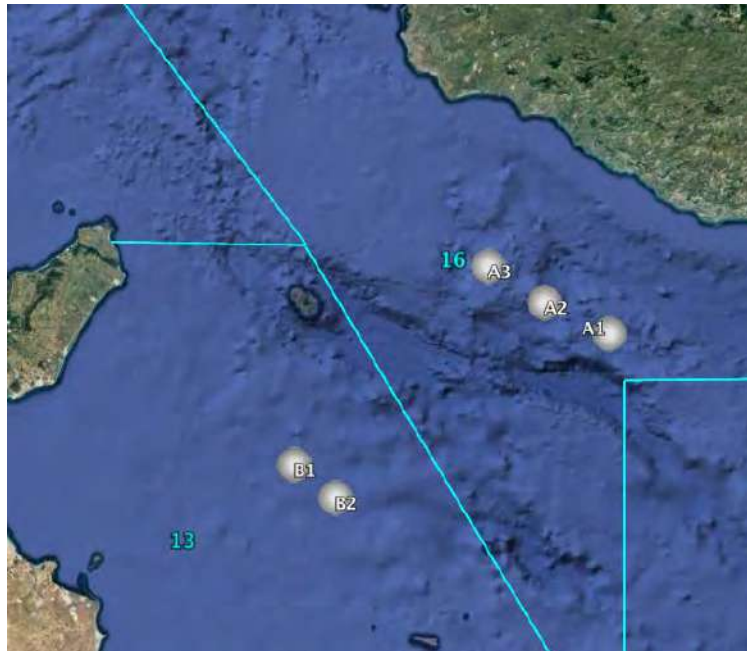
code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
A1	Area A1	--	--	FAO37:16	
A2	Area A1			FAO37:16	
A3	Area A1	--	--	FAO37:16	
B1	Area B1	--	--	FAO37:13	
B2	Area B2	--	--	FAO37:13	
total					270

MAIN FINDINGS (LEVI ET AL., 2004)

1. The population differentiation test performed by θ analysis showed no heterogeneity, except for the single locus PGI-2* locus ($\theta = 0.011$, $P < 0.05$), while the mean value was not significant.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA
AS INFERRED BY THE GDMU_MM08 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm08	<i>Merluccius merluccius</i>
Population\cluster	Cluster A
A1	1
A2	1
A3	1
B1	1
B2	1

DATA FROM:

Lo Brutto, S., M. Arculeo & N. Parrinello. 2004. **Congruence in genetic markers used to describe Mediterranean and Atlantic populations of European hake (*Merluccius merluccius* L. 1758).** *Journal of Applied Ichthyology* 20(2):81-86 doi:10.1046/j.1439-0426.2003.00514.x.

GENETIC MARKER USED

Allozymes (5 loci) + PCR-RFLP (DNA control region)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
TR [°]	Trondheimsfjord	--	--	FAO27	
BA	Balearic Is.	--	--	FAO37:GSA5	
SA	Sardinia	--	--	FAO37:GSA11	
TY1	Tyrrhenian Sea	--	--	FAO37:GSA10	
SS1	Strait Sicily1	--	--	FAO37:GSA16	
SS2	Strait Sicily2	--	--	FAO37:GSA16	
SS3	Strait Sicily3	--	--	FAO37:GSA16	
AD	Adriatic Sea	--	--	FAO37:GSA18	
IS	Israel	--	--	FAO37:GSA27	
total					418

[°] not shown in the geovisualization

MAIN FINDINGS (LO BRUTTO ET AL., 2004)

1. Samples originating from inside the Mediterranean basin appeared genetically homogeneous but the sample originating from the Atlantic was heterogeneous compared with the Mediterranean populations.
2. The pairwise genetic differentiation, obtained by using Fisher's method, and P-value for each population pair across loci described the Norway sample as being significantly different from the others, mainly caused by the allelic distributions of ADH* ($P < 0.005$), PGI-2* ($P = 0$) and PGM* ($P < 0.05$).
3. The analysis showed significant differences with all the samples ($P = 0.006 \pm 0.002$), indicating genetic substructures within the species. Indeed, the UPGMA cluster obtained from nucleotide divergence (Nei and Tajima, 1983), ranging from 0.00003 to 0.0042, showed that distinct Mediterranean and Atlantic populations agreed with the Monte Carlo test. From the viewpoint of spatial heterogeneity, the results obtained from mtDNA agree with those obtained for allozymes.
4. Nuclear and mitochondrial gene analysis showed similar results supporting that the Strait of Gibraltar may be considered as a breakpoint area to gene flow.

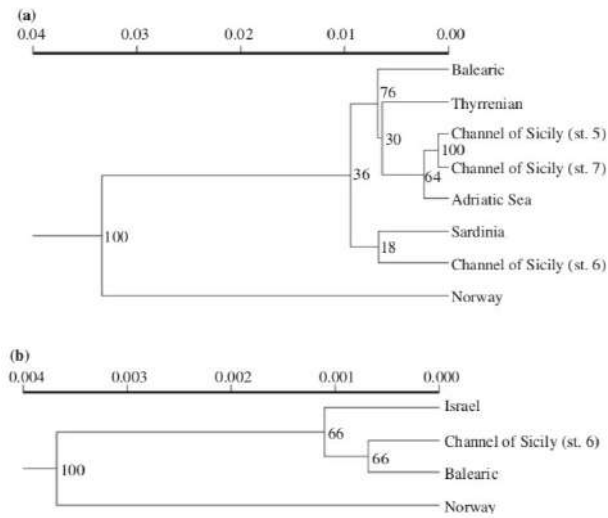
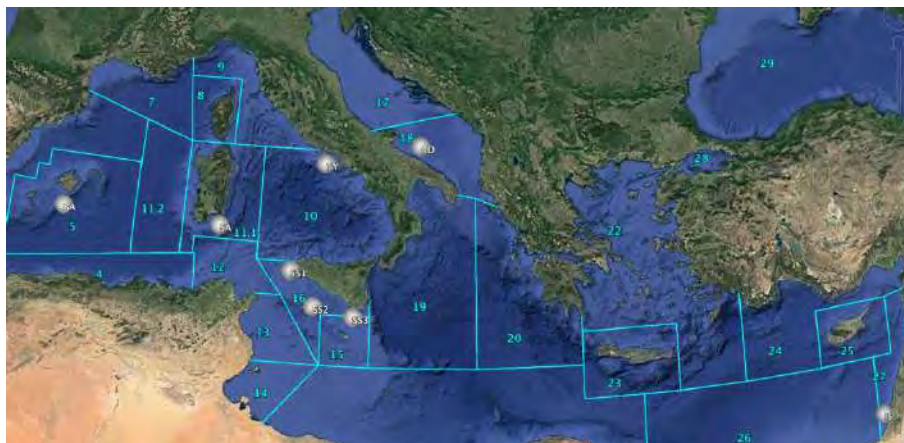


FIGURE 2 FROM LO BRUTTO ET AL., 2004 (A) GENETIC DISTANCE (NEI, 1978) AND (B) NUCLEOTIDE DIVERGENCE (NEI AND TAJIMA, 1981) CLUSTERED BY UPGMA (NUMBERS INDICATE THE PERCENTAGE OF THE 1000 BOOTSTRAP REPLICATES)

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM09 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm09	<i>Merluccius merluccius</i>
Population\cluster	Cluster A
TR	1
BA	1
SA	1
TY1	1
TY2	1
TY3	1
AD	1
IS	1

DATA FROM:

Cimmaruta R, Bondanelli P, Nascetti G: **Genetic structure and environmental heterogeneity in the European hake (*Merluccius merluccius*)**. *Mol Ecol* 2005, **14**(8):2577-2591.

GENETIC MARKER USED

Allozymes: 19 enzyme systems corresponding to 31 putative gene loci, of which 21 loci were polymorphic in all the 15 population samples

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 1998-1999 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
MO	Morocco	--	--	FAO34	96
GA	Galicia	--	--	FAO27	54
BI	Bay of Biscay	--	--	FAO27	60
IR	Ireland	--	--	FAO27	53
AD	Adriatic Sea	--	--	FAO37:GSA17	138
TY	Tyrrhenian Sea	--	--	FAO37:GSA10	136
LI	Ligurian Sea	--	--	FAO37:GSA9	165
SA	Sardinia	--	--	FAO37:GSA11..1	49
BA	Balearic islands	--	--	FAO37:GSA5	44
AL	Alicante	--	--	FAO37:GSA6	44
MA	Malaga	--	--	FAO37:GSA1	60
CY	Cyprus	--	--	FAO37:GSA25	98
CR	Crete	--	--	FAO37:GSA23	110
AE	Aegean Sea	--	--	FAO37:GSA22	96
IO	Ionian Sea	--	--	FAO37:GSA20	103
total					1306

MAIN FINDINGS (CIMMARUTA ET AL., 2005)

1. The European hake populations resulted genetically differentiated into a Mediterranean and an Atlantic group.
2. This partition is due to the allele frequency pattern of a few loci, showing a latitudinal frequency cline (Gapdh and Gpi-2) or a differentiation among Atlantic and Mediterranean samples (Gda and Mpi).
3. The Mediterranean populations indeed have higher levels of genetic variability than those of the Atlantic.
4. A latitudinal cline at loci Gapdh and Gpi-2 within the Mediterranean Sea, with a further steep change across the Almeria-Oran front.
5. The subdivision between Mediterranean and Atlantic samples, these latter joined also by Malaga sample, was confirmed by the NJ tree based on standard distances (Figure 3 from Cimmaruta *et al.*, 2005).

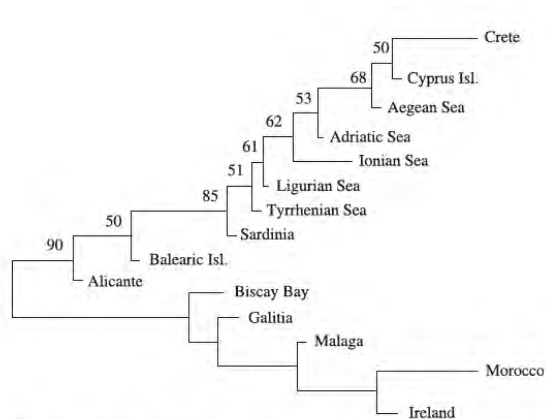
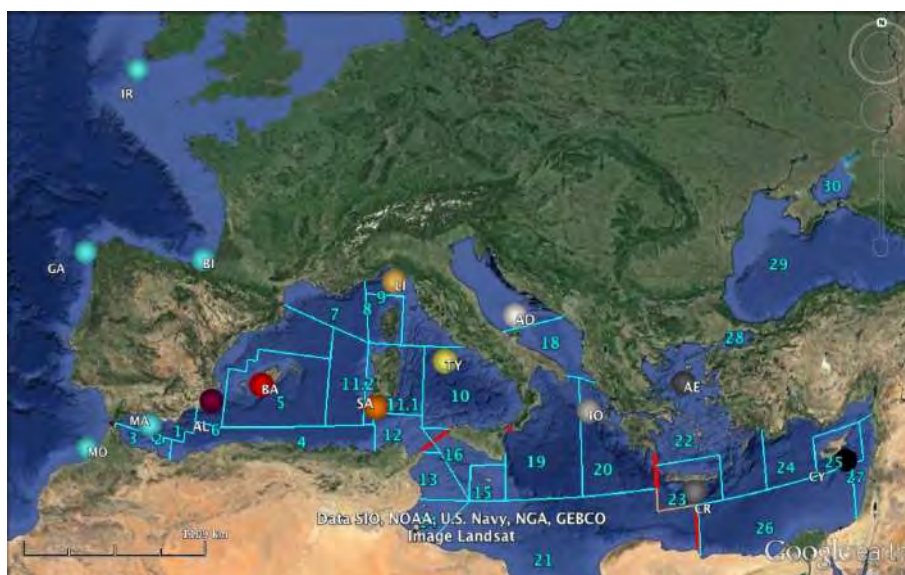


Fig. 3 Neighbour-joining cluster based on standard genetic distances (Nei 1972), showing the genetic relationships between the studied populations of European hake. Bootstrap values over 1000 replicates are shown when $\geq 50\%$.

FIGURE 3 FROM CIMMARUTA *ET AL.*, 2005.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM12 DATASET

GEOVISUALIZATION



*based on the NJ tree in Figure 3

MATRIX OF GENETIC CLUSTERING

GDMU_Mm12	<i>Merluccius merluccius</i>										
Population\cluster	A	B	C	D	E	F	G	H	I	L	M
MO	1	0	0	0	0	0	0	0	0	0	0
GA	1	0	0	0	0	0	0	0	0	0	0
BI	1	0	0	0	0	0	0	0	0	0	0
IR	1	0	0	0	0	0	0	0	0	0	0
AD	0	1	0	0	0	0	0	0	0	0	0
TY	0	0	1	0	0	0	0	0	0	0	0
LI	0	0	0	1	0	0	0	0	0	0	0
SA	0	0	0	0	1	0	0	0	0	0	0
BA	0	0	0	0	0	1	0	0	0	0	0
AL	0	0	0	0	0	0	1	0	0	0	0
MA	1	0	0	0	0	0	0	0	0	0	0
CY	0	0	0	0	0	0	0	1	0	0	0
CR	0	0	0	0	0	0	0	0	1	0	0
AE	0	0	0	0	0	0	0	0	0	1	0
IO	0	0	0	0	0	0	0	0	0	0	1

NB: the binary matrix is not appropriate for describing a clinal distribution; the sites were attributed to different cluster when the bootstrap values was >50.

GDMU_Mm15

DATA FROM:

Pita, A., P. Presa & M. Perez, 2010. GENE FLOW, MULTILOCUS ASSIGNMENT AND GENETIC STRUCTURING OF THE EUROPEAN HAKE (MERLUCCIOUS MERLUCCIOUS). *Thalassas* 26(2):129-133.

GENETIC MARKER USED

Microsatellites (5 loci) + mitochondrial DNA (cytb)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2000 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
1		--	--	FAO27	
2				FAO27	
3		--	--	FAO27	
4		--	--	FAO27	
5		--	--	FAO27	
6				FAO27	
7		--	--	FAO27	
8		--	--	FAO27	
9		--	--	FAO27	
10		--	--	FAO27	
11				FAO27	
12		--	--	FAO27	
13		--	--	FAO27	
14		--	--	FAO27	
15		--	--	FAO27	
16		--	--	FAO27	
17		--	--	FAO27	
18		--	--	FAO27	
19		--	--	FAO27	
20				FAO37:GSA1	
21		--	--	FAO37:GSA1	
22		--	--	FAO37:GSA6	
23		--	--	FAO37:GSA6	
24		--	--	FAO37:GSA6	
25				FAO37:GSA6	
26				FAO37:GSA6	
27				FAO37:GSA6	
total					712

MAIN FINDINGS (PITA ET AL., 2010)

1. We have analysed the molecular variation of five polymorphic microsatellites and a 465 bp fragment from the cytochrome b gene on 27 hake populations to determine the genetic status of this species across European fisheries.

2. While weak genetic differences ($F_{CT} = 0.0092$, $P < 0.01$) exist between the seven major oceanographic regions considered (North Sea, Celtic Sea, Cantabrian Sea, Iberian Atlantic, Iberian Mediterranean, Tyrrhenian Sea and Canarian Sea), the deepest partition resides between Atlantic and Mediterranean populations.
3. However, the probability of the 712 multilocus genotypes scored to be assigned to Atlantic or to Mediterranean basins is fairly 0.5, indicating that these two geographical stocks cannot be reliably identified from each other neither for fishery forensics nor for commercial traceability

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM15 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm15	<i>Merluccius merluccius</i>	
Population\cluster	Cluster A	Cluster B
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	1	0
13	1	0
14	1	0
15	1	0
16	1	0
17	1	0
18	1	0
19	1	0
20	1	0
21	1	1
22	0	1
23	0	1
24	0	1
25	0	1
26	0	1
27	0	1

DATA FROM:

Nielsen, E. E., A. Cariani, E. Mac Aoidh, G. E. Maes, I. Milano, R. Ogden, M. Taylor, J. Hemmer-Hansen, M. Babbucci, L. Bargelloni, D. Bekkevold, E. Diopere, L. Grenfell, S. Helyar, M. T. Limborg, J. T. Martinsohn, R. McEwing, F. Panitz, T. Patarnello, F. Tinti, J. K. J. Van Houdt, F. A. M. Volckaert, R. S. Waples, G. R. Carvalho, J. E. J. Albin, J. M. V. Baptista, V. Barmintsev, J. M. Bautista, C. Bendixen, J.-P. Berge, D. Blohm, B. Cardazzo, A. Diez, M. Espinera, A. J. Geffen, E. Gonzalez, N. Gonzalez-Lavin, I. Guarniero, M. Jerome, M. Kochzius, G. Krey, O. Mouchel, E. Negrisolo, C. Piccinetti, A. Puyet, S. Rastorguev, J. P. Smith, M. Trentini, V. Verrez-Bagnis, A. Volkov, A. Zanzi & C. FishPopTrace, 2012. **Gene-associated markers provide tools for tackling illegal fishing and false eco-certification.** *Nature Communications* 3
doi:10.1038/ncomms1845

GENETIC MARKER USED

395 SNPs, 72 outlier

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2008-2009 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
NS	North Sea	57.996	0.219	FAO27	43
WS	West of Scotland	56.208	-9.212	FAO27	53
SI	Southwest of Ireland	50.426	-9.975	FAO27	58
GC	Galician Coast	43.305	-9.18	FAO27	43
NP	North Portugal	42.123	-9.394	FAO27	46
WA	West Algeria	35.605	-2.123	FAO37:GSA3	41
MA	Malaga	36.634	-4.369	FAO37:GSA1	47
GL	Gulf of Lion	43.35	3.726	FAO37:GSA7	46
NT	North Tyrrhenian	42.515	10.109	FAO37:GSA9	36
SS	South Sardinia	38.997	8.51	FAO37:GSA11.1	44
ST	South Tyrrhenian	40.536	14.77	FAO37:GSA10	48
AB	Adventure Bank	37.333	12.172	FAO37:GSA16	66
MB	Maltese Bank	35.916	14.786	FAO37:GSA15	46
NA	North Adriatic	43.928	13.793	FAO37:GSA17	45
SA	South Adriatic	42.26	16.894	FAO37:GSA18	19
NI	North-western Ionian Sea	38.193	16.366	FAO37:GSA19	33
AE	Aegean Sea	40.344	24.515	FAO37:GSA22	54
TR	Turkish Coast	36.776	30.945	FAO37:GSA24	39
CY	Cyprus	34.52	32.913	FAO37:GSA25	43
total					850

MAIN FINDINGS (NIELSEN ET AL., 2012)

1. 72 of 395 SNPs were outliers. 13 high F_{ST} SNPs (F_{ST} between 0.08 and 0.29) provided 98% (751 of 766 individuals) correct assignment to basin.
2. Fourteen of 15 misassigned individuals originated from western Mediterranean samples (Algerian coast, Malaga) likely to be migrants or the result of admixture with neighbouring

Atlantic populations. One individual sampled in the Atlantic was misassigned to the Mediterranean.

3. Excluding likely migrants from the western Mediterranean, 99% of all individuals were assigned unambiguously to basin of origin. Evaluation of the likelihood of alternative hypotheses of origin showed that 95% of all sampled hake were over 500 times more likely to originate from their basin of sampling than to other basins.

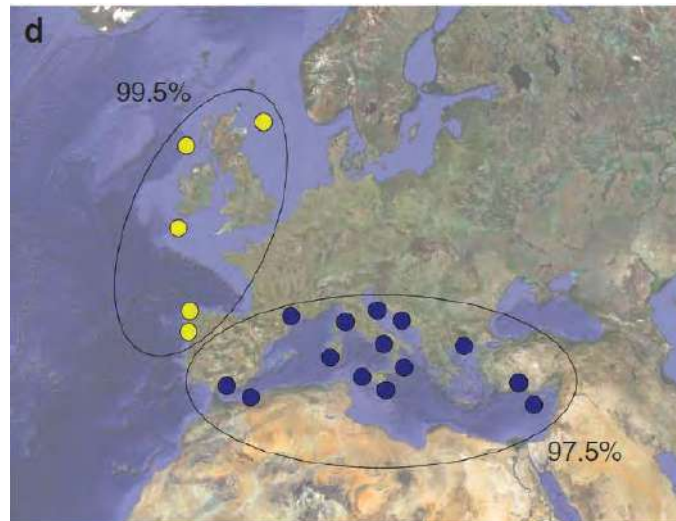


FIGURE 1 FROM NIELSEN *ET AL.*, 2012 SHOWN IS THE PERCENTAGE OF FISH ASSIGNED TO THE SAMPLE/AREA OF ORIGIN AND TO OTHER SAMPLES/

AREAS. MEDITERRANEAN HAKE (BLUE), ATLANTIC HAKE (YELLOW)

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM18 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm18	<i>Merluccius merluccius</i>	
	A	B
Population\cluster		
NS	1	0
WS	1	0
SI	1	0
GC	1	0
NP	1	0
WA	1	0
MA	0	1
GL	0	1
NT	0	1
SS	0	1
ST	0	1
AB	0	1
MB	0	1
NA	0	1
SA	0	1
NI	0	1
AE	0	1
TR	0	1
CY	0	1

DATA FROM:

Milano I, Babbucci M, Cariani A, Atanassova M, Carvalho GR, Espiñeira M, Fiorentino F, Garofalo G, Geffen AJ, Helyar S *et al.* 2014. **Outlier SNP markers reveal fine-scale population genetic structuring across European hake (*Merluccius merluccius*)**. *Molecular Ecology* 23: 118-135

GENETIC MARKER USED

381 SNPs located in transcribed regions

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2008-2009 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
NS	North Sea	57.996	0.219	FAO27	43
WS	West of Scotland	56.208	-9.212	FAO27	53
SI	Southwest of Ireland	50.426	-9.975	FAO27	58
GC	Galician Coast	43.305	-9.18	FAO27	43
NP	North Portugal	42.123	-9.394	FAO27	46
WA	West Algeria	35.605	-2.123	FAO37:GSA3	41
MA	Malaga	36.634	-4.369	FAO37:GSA1	47
GL	Gulf of Lion	43.35	3.726	FAO37:GSA7	46
NT	North Tyrrhenian	42.515	10.109	FAO37:GSA9	36
SS	South Sardinia	38.997	8.51	FAO37:GSA11.1	44
ST	South Tyrrhenian	40.536	14.77	FAO37:GSA10	48
AB	Adventure Bank	37.333	12.172	FAO37:GSA16	66
MB	Maltese Bank	35.916	14.786	FAO37:GSA15	46
NA	North Adriatic	43.928	13.793	FAO37:GSA17	45
SA	South Adriatic	42.26	16.894	FAO37:GSA18	19
NI	North-western Ionian Sea	38.193	16.366	FAO37:GSA19	33
AE	Aegean Sea	40.344	24.515	FAO37:GSA22	54
TR	Turkish Coast	36.776	30.945	FAO37:GSA24	39
CY	Cyprus	34.52	32.913	FAO37:GSA25	43
total					850

MAIN FINDINGS (MILANO ET AL., 2014)

1. Analysis of 850 individuals from 19 locations across the entire distribution range showed evidence for several outlier loci, with significantly higher resolving power.
2. 299 putatively neutral SNPs confirmed the genetic break between basins ($F_{CT}=0.016$) and weak differentiation within basins
3. Outlier loci revealed a dramatic divergence between Atlantic and Mediterranean populations (F_{CT} range 0.275-0.705) and fine-scale significant population structure.
4. In the Atlantic, outlier loci separated North Sea and Northern Portugal populations from all other Atlantic samples (Figure 3B from Milano *et al.*, 2014).
5. In the Mediterranean, outlier loci revealed a strong differentiation among Western, Central and Eastern Mediterranean geographical samples (Figure 3C from Milano *et al.*, 2014).

- Significant correlation of allele frequencies at outlier loci with environmental parameters (seawater temperature and salinity) supported the hypothesis that outlier SNPs might be directly or indirectly involved in the adaptation to local conditions.

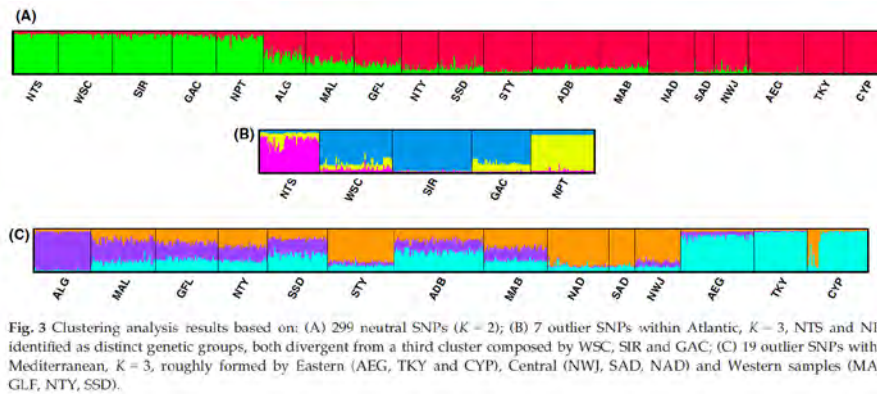
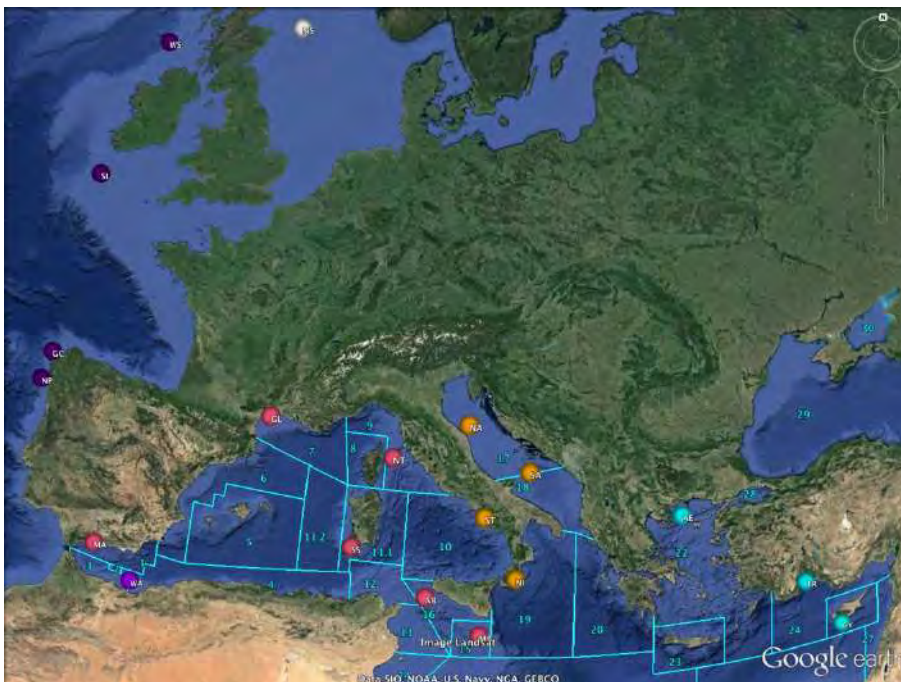


FIGURE 3 FROM MILANO *ET AL.*, 2014

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM19 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm19	<i>Merluccius merluccius</i>					
Population\cluster	A	B	C	E	F	G
NS	1	0	0	0	0	0
WS	0	1	0	0	0	0
SI	0	1	0	0	0	0
GC	0	1	0	0	0	0
NP	0	1	0	0	0	0
WA	0	0	1	0	0	0
MA	0	0	0	1	0	0
GL	0	0	0	1	0	0
NT	0	0	0	1	0	0
SS	0	0	0	1	0	0
ST	0	0	0	0	0	1
AB	0	0	0	1	0	0
MB	0	0	0	1	0	0
NA	0	0	0	0	0	1
SA	0	0	0	0	0	1
NI	0	0	0	0	0	1
AE	0	0	0	0	1	0
TR	0	0	0	0	1	0
CY	0	0	0	0	1	0

DATA FROM:

Tanner, S. E., M. Perez, P. Presa, S. R. Thorrold & H. N. Cabral, 2014. Integrating microsatellite DNA markers and otolith geochemistry to assess population structure of European hake (*Merluccius merluccius*). Estuarine Coastal and Shelf Science 142:68-75 doi:10.1016/j.ecss.2014.03.010.

GENETIC MARKER USED

Microsatellites (5 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
AS	Armorican Shelf	--	--	FAO27	42
CS	Celtic Sea			FAO27	50
GS	Galicia shelf	--	--	FAO27	50
PT	Portugal	--	--	FAO27	50
GC	Gulf Cadiz	--	--	FAO27	50
BI	Balearic Is.	--	--	FAO37:GSA5	50
SA	Sardinia			FAO37:GSA11	47
total					339

MAIN FINDINGS (TANNER ET AL., 2014)

1. Microsatellites provided evidence of a major genetic split in the vicinity of the Strait of Gibraltar, separating the Atlantic and the Mediterranean populations, with the exception of the Gulf of Cádiz.
2. Within the Atlantic Ocean, only the Gulf of Cádiz samples differed significantly from Celtic Sea and Portugal samples. AMOVA analyses showed that most variation among collection locations (2.25%) was due to differences between the Atlantic Ocean and the Mediterranean Sea and to a lesser extent (1.60%) among the two management units currently implemented in the northeast Atlantic and the Mediterranean Sea.
3. The neighbor-joining dendrogram showed a major branching between Atlantic and Mediterranean locations with 98.8% bootstrap support (Fig. 2). Within the branch of the Atlantic locations, samples from the Gulf of Cádiz were positioned apart with 83.2% bootstrap support.
4. Based on the clustering approach performed in STRUCTURE, individuals were assigned to two hypothetical clusters ($K = 2$). The first cluster was composed by individuals collected in the Atlantic Ocean however, only about one third of the individuals collected in the Gulf of Cádiz were allocated to this cluster. The remaining individuals collected in the Gulf of Cádiz were placed in the second cluster with the individuals collected in the Mediterranean Sea (Fig. 4)
5. Coupling genotype and otolith data increased the classification accuracy of individuals to their potential natal origins while providing evidence of movement between the northern and southern stock units in the Atlantic Ocean

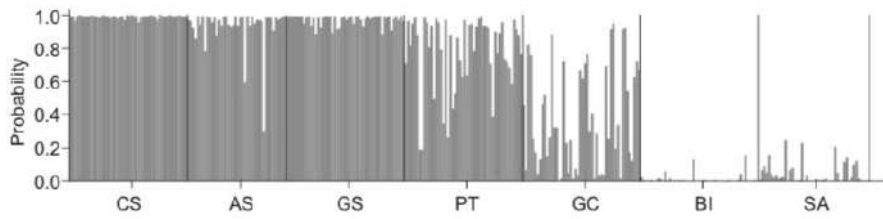
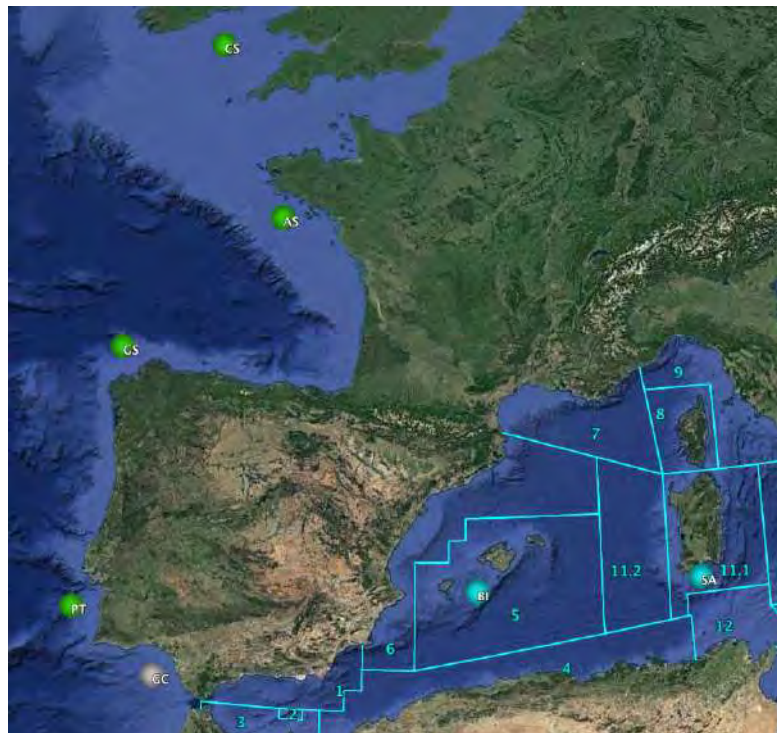


FIGURE 3 FROM TANNER *ET AL.*, 2014

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MM20 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mm20	<i>Merluccius merluccius</i>	
Population\cluster	Cluster A	Cluster B
AS	0	1
CS	0	1
GS*	0	1
PT	0	1
GC	*	*
BI	1	0
SA	1	0

*GC intermediate between the Atlantic and Mediterranean clusters

INFORMATION FROM STUDIES INCLUDING POPULATION SAMPLES OUTSIDE THE MEDITERRANEAN SEA (P DATASETS)

GDMU_Mm05: Polymorphism at five microsatellite loci were screened to determine the genetic variability and the temporal stability of population structure in natural populations of European hake (*Merluccius merluccius*, L.) within the Bay of Biscay. In addition, the control region (900 bp) and two protein coding genes (ATPase, subunits 6 and 8, 842 bp and a partial sequence of the ND1, 800 bp) of the mitochondrial DNA (mtDNA) were sequenced from geographically distant populations from the extremes of the species range. Low levels of population subdivision were found between putative populations within years. Hierarchical analysis of molecular variance (AMOVA) does not confirm the temporal persistence of population structure (Lundy et al., 2000).

GDMU_Mm11: Prespawning hake caught in eight locations around the Iberian Peninsula were genetically analysed. The distribution of variation at five microsatellite loci suggests that the species follows a population model of isolation by distance in this geographical area. Three different areas can be identified: the Mediterranean Sea, the Portuguese coast, and the Cantabric Sea. The lack of differentiation between samples caught in the VIIa,b,d and in the VIIc ICES Areas suggests that, based on genetic information, the boundary between northern and southern stocks of European hake should be reconsidered (Castillo et al., 2005).

GDMU_Mm17: The pattern of gene flow between the two Atlantic stocks of the European hake has been examined for the period 2000–2002. Analyses indicate that a large genetic homogeneity existed among all Atlantic populations in that period, and that a systematic grouping occurred between Porcupine Bank samples and Cantabric ones. This scenario is congruent with an inter-annual gene flow from central grounds of the northern stock (Porcupine and Great Sole) to Iberian grounds inhabited by the southern stock (Pita et al., 2011).

GDMU_Mm27: Using five microsatellite markers from the European hake and an extensive sampling coverage within stocks, we have tested the connectivity pattern between the two managed fish stocks of this species, the Northern stock and the Southern stock, over the last decade. On the spatial domain, the non-significant genetic variance between stocks ($F_{CT} = 0.001$, $k = 1$ gene pools) as opposed to the low but significant divergence within stocks ($F_{SC} = 0.008$) suggests that demographic exchange occurs between stocks. On the temporal domain, the non-significant variation among years ($F_{CT} < 0.001$) suggests that the genetic background within the Atlantic hake metapopulation did not change significantly over the last decade. (Pita et al., 2016b).

GDMU_Mm28: The study aimed at testing both, the temporal stability of the genetic diversity among samples from the Southern stock of European hake and the influence of a high fishing intensity on its N_e . Such goals were addressed with variation of microsatellites and Cytochrome b haplotypes on an 82 sample collection taken in years of deep population depletion (2000–2007) plus a historical anchoring sample from 1976 which predates the industrial exploitation of this fishery. Despite the Southern stock exhibits similar levels of variation at microsatellites and Cytochrome b across the historical window addressed, N_e experienced a 43-fold reduction in parallel to an 80% biomass (N_{ssb}) loss between the historical estimates of 1976 ($N_{ssb} = 20.55 \cdot 10^6$ Mi, $N_e = 12,480$) and their minimums of 2004 ($N_{ssb} = 4.68 \cdot 10^6$ Mi, $N_e = 291$). Neither N_e nor N_e / N_{ssb} matched SSB trends at overharvest in 2000–2007, when a faster N_e reduction was patent. Nevertheless, N_e correlated well with SSB and the N_e / N_{ssb} ratio was maximal at population equilibrium and minimal at population depletion (Pita et al., 2017).

GDMU_Mm29: the authors relieved a population genetic study using neutral and outlier SNP loci assessing population structure in hake in the north-eastern parts of its range in the Atlantic. Hake samples from localities along the west coast of Norway, the Kattegat, the northern North Sea, and one locality in the Bay of Biscay were analysed using 53 SNPs, six of which were outliers potentially influenced by natural selection. Small-scale structure among northern samples was detected, all of

which were also distinct from Bay of Biscay hake, with the exception of a few individuals from the North Sea and the coast of Norway who clustered genetically together with Bay of Biscay hake. These findings suggest that the present management unit of a single northern stock of hake is not biologically correct, and that there is more detail in the fine- scale population structure indicating that independent population dynamics could be expected in response to fishing patterns or changing environmental conditions (Westgaard et al., 2017).

COMMENTS AND EVALUATION OF DATA: EUROPEAN HAKE

The following table summarizes the results of the five population genetic studies performed so far for the European hake

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic structure (clusters)	Number of Mediterranean clusters
Mm01	★★★ ^{WCE}	★★★	-	-	1(WCE)
Mm02	★★★ ^W	★★★	-	-	1(W)
Mm03	★★ ^{CE}	★★★	★	-	1(CE)
Mm08	★★★ ^{WCE}	★★★	★	★	2(WC/E)
Mm09	★★★ ^C	★★★	-	-	1(C)
Mm10	★★★ ^{WC}	★	★	-	1(WC)
Mm13	★★★ ^{WCE}	★★★	★	★	? cline
Mm16	★★★ ^W	na	★	-	1(W)
Mm19	★★★ ^{WCE}	★★★	★★★	-	1(WCE)
Mm20	★★★ ^{WCE}	★★★	★★★	★★★	4(W/WC/WC/E)
Mm21	★★ ^W	★★★	★	★	1(W)

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★≤3; ★★4-6; ★★★>6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★≤20; ★★20-40; ★★★>40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall FST: ★≤0.05; ★★0.05-0.1; ★★★ >0.1)

Genetic structure = ★low, ★★medium ★★★ high support (whenever available the score is based on pairwise Fst values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean

Comments: A few studies reported genetic differentiation within the Mediterranean. Allozyme data showed a latitudinal gradient rather than a clear pattern of clustering. SNP data confirm the genetic structure described a finer-scale significant genetic population structure.

Gaps: No information available from the North African coasts and southeast Med.

Technical notes: Results from Milano et al., 2014 (based on 381 SNPs and a wider sampling coverage) are the most informative, and among the few case-studies based on 'last-generation' markers.

Extensive sampling was realized within the framework of the EU Project FishPopTrace.

The analysis: *Mullus barbatus*

INFORMATION FROM STUDIES INCLUDING MEDITERRANEAN POPULATION SAMPLES (HP DATASETS)

GDMU_Mb03

DATA FROM:

Mamuris Z, Apostolidis AP, Theodorou AJ, Triantaphyllidis C. 1998a. **Application of random amplified polymorphic DNA (RAPD) markers to evaluate intraspecific genetic variation in red mullet (*Mullus barbatus*)**. *Marine Biology*, **132**(2):171-178.

GENETIC MARKER USED

Random amplified polymorphic DNA (RAPD)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
CO	Corfu	--	--	FAO37:GSA20	17
AM	Amvrakikos	--	--	FAO37:GSA20	17
KA	Kavala	--	--	FAO37:GSA22	18
GL	Gulf of Lions	--	--	FAO37:GSA7	18
PL	Platania	--	--	FAO37:GSA22	19
AL	Allonisos	--	--	FAO37:GSA22	20
KY	Kymi	--	--	FAO37:GSA22	19
TH	Thermaikos	--	--	FAO37:GSA22	18
total					146

MAIN FINDINGS (MAMURIS ET AL., 1998A)

- no specific RAPD marker for the discrimination of the populations was detected.
- The data analysis revealed that the genetic diversity among populations is positively related to their geographic distances.
- UPGMA tree based on Nei's genetic distances shows the existence of distinct clades, supported by medium to high bootstrap values (Figure 3 from Mamuris *et al.*, 1998a).

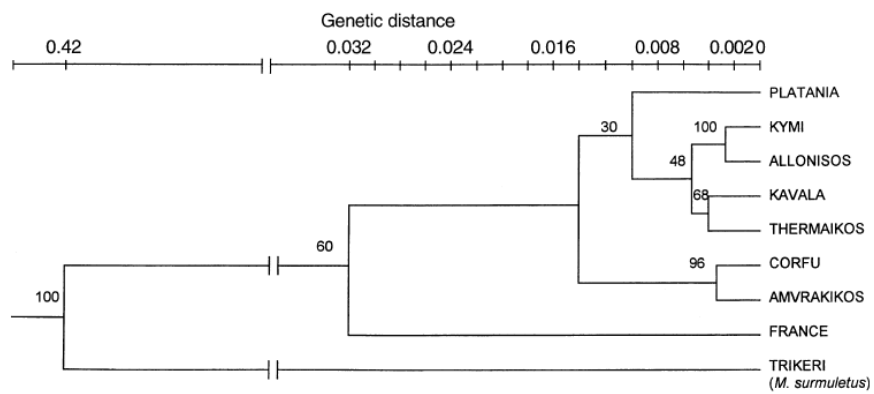
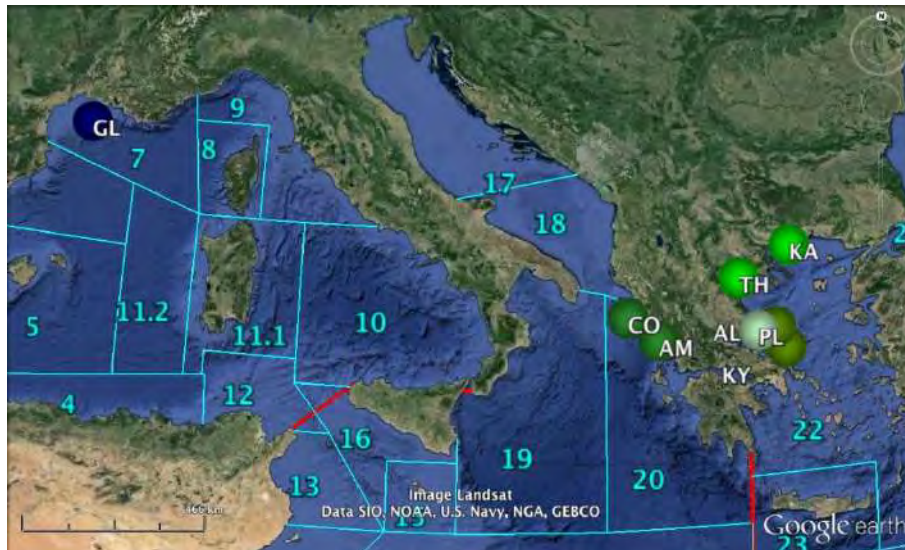


FIGURE 3 FROM MAMURIS *ET AL.*, 1998A *MULLUS BARBATUS*, UPGMA DENDROGRAM BASED ON NEI'S (1978) GENETIC DISTANCE. BOOTSTRAP ESTIMATES (AS A PERCENTAGE) ARE INDICATED ABOVE BRANCHES.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB03 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb03	<i>Mullus barbatus</i>					
Population\cluster	Cluster A	Cluster B	Subcl.b1	Subcl.b2	Subcl.b3	Subcl.b4
CO	0	1	1	0	0	0
AM	0	1	1	0	0	0
KA	0	1	0	0	1	0
GL	1	0	0	0	0	0
PL	0	1	0	1	0	0
AL	0	1	0	0	0	1
KY	0	1	0	0	0	1
TH	0	1	0	0	1	0

DATA FROM:

Mamuris Z, Apostolidis AP, Triantaphyllidis C. 1998c. **Genetic Protein Variation in Red Mullet (*Mullus barbatus*) and Striped Red Mullet (*M. surmuletus*) Populations from the Mediterranean Sea.** *Marine Biology*, 130, no. 3: 353-60.

GENETIC MARKER USED

Allozymes, twenty putative enzyme-coding loci (14 polymorphic)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
CO	Corfu	--	--	FAO37:GSA20	--
AM	Amvrakikos	--	--	FAO37:GSA20	--
KA	Kavala	--	--	FAO37:GSA22	--
GL	Gulf of Lions	--	--	FAO37:GSA7	--
PL	Platania	--	--	FAO37:GSA22	--
AL	Allonisos	--	--	FAO37:GSA22	--
KY	Kymi	--	--	FAO37:GSA22	--
TH	Thermaikos	--	--	FAO37:GSA22	--
total					349

MAIN FINDINGS (MAMURIS ET AL., 1998C)

1. Genetic distance among *M. barbatus* samples was low (maximum Nei's D. 0.012), with the sample from Platania differing most from other *M. barbatus* samples. This is probably due to founder effects existing at this area.
2. The χ^2 contingency analyses indicated that allele frequencies differed significantly ($p < 0.05$) among the eight samples of *Mullus barbatus* for nine of the 14 polymorphic loci after Bonferroni corrections.
3. Mean F_{ST} value for all samples of *M. barbatus* were significant ($F_{ST} = 0.043$; χ^2 test, $p < 0.05$). This would appear to indicate a degree of structuring among the populations of *M. barbatus* sampled, with a low to moderate level of gene flow between them.
4. Genetic population subdivision was also evident in the eastern Mediterranean region ($F_{ST} = 0.041$; χ^2 test, $p < 0.05$). This population subdivision was still evident ($F_{ST} = 0.023$; χ^2 test, $p < 0.05$) even after the omission of the Platania sample, which was the main contributor.
5. Within *M. barbatus* populations, apart from the sample from Platania, the genetic distances were low for any pairwise comparison of samples (mean D. 0.0039). The Platania fish were most genetically distinct (range from 0.009 in Kavala and Thermaikos to 0.019 in France, mean $D = 0.0116$) and this was clearly illustrated in the dendrogram derived from UPGMA cluster analysis (Figure 2 from Mamuris *et al.*, 1998c).

Fig. 2 *Mullus barbatus*, *M. surmuletus*, UPGMA dendrogram based on Nei's (1978) genetic distance, showing the relationship between *M. barbatus* and *M. surmuletus* samples in the Mediterranean Sea

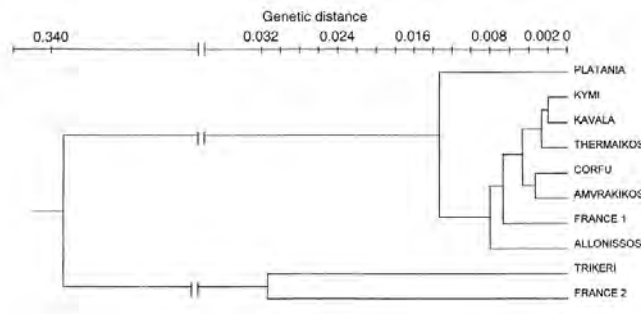
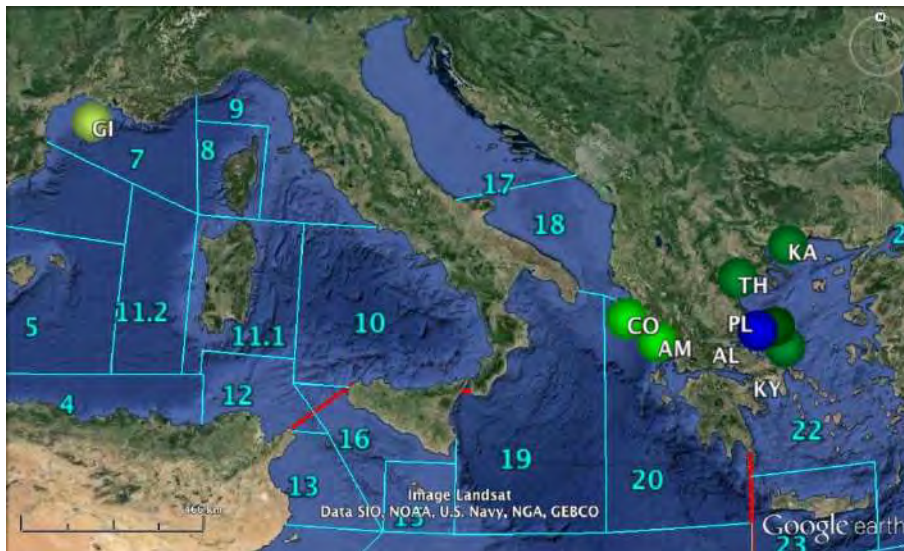


FIGURE 2 FROM MAMURIS ET AL., 1998C. TRIKERI AND FRANCE 2 ARE SAMPLES OF *MULLUS SURMULETUS*.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB05 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb05	<i>Mullus barbatus</i>					
Population\cluster	Cluster A	Cluster B	Subcl.b1	Subcl.b2	Subcl.b3	Subcl.b4
CO	0	1	0	0	1	0
AM	0	1	0	0	1	0
KA	0	1	0	0	0	1
GL	0	1	0	1	0	0
PL	1	0	0	0	0	0
AL	0	1	1	0	0	0
KY	0	1	0	0	0	1
TH	0	1	0	0	0	1

GDMU_Mb06

DATA FROM:

Arculeo, M., S. Lo Brutto, M. Cammarata, M. Scalisi & N. Parrinello, 1999. **Genetic variability of the Mediterranean Sea red mullet, *Mullus barbatus* (Pisces, Mullidae).** *Russian Journal of Genetics* 35(3):292-296

GENETIC MARKER USED

Isozymes (25 loci, 3 polymorphic)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	Number of samples
SE	Sète	--	--	FAO37:GSA7	58
LI	Livorno	--	--	FAO37:GSA9	71
CA	Castellammare	--	--	FAO37:GSA10	65
GE	Gela	--	--	FAO37:GSA16	74
SS	Stretto Sicilia	--	--	FAO37:GSA16	73
CH	Chioggia	--	--	FAO37:GSA17	73
IO	Ionian Sea	--	--	FAO37:GSA20	66
AE	Aegean Sea	--	--	FAO37:GSA22	64
total					533

MAIN FINDINGS (ARCULEO ET AL., 1999)

1. No significant differences in allele frequency variation for each locus and across all samples for all loci were found.
2. F_{ST} values indicated insignificant differences among the sample sites examined. The mean value of F_{ST} equals to 0.009.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB06 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb06	<i>Mullus barbatus</i>
Population\cluster	Cluster A
SE	1
LI	1
CA	1
GE	1
SS	1
CH	1
IO	1
AE	1

DATA FROM:

Mamuris, Z., C. Stamatis, K. A. Moutou, A. P. Apostolidis & C. Triantaphyllidis, 2001. **RFLP analysis of mitochondrial DNA to evaluate genetic variation in striped red mullet (*Mullus surmuletus* L.) and red mullet (*Mullus barbatus* L.) populations.** Marine Biotechnology 3(3):264-274
doi:10.1007/s101260000075.

GENETIC MARKER USED

RFLP (control region, COI, and 12S–16S ribosomal RNA)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	Number of samples
CO	Corfu	--	--	FAO37:GSA20	26
AM	Amvrakikos	--	--	FAO37:GSA20	21
KA	Kavala	--	--	FAO37:GSA22	23
GL	Gulf of Lions	--	--	FAO37:GSA7	27
AL	Alonisos	--	--	FAO37:GSA22	21
RH	Rhodes			FAO37:GSA22	20
total					138

MAIN FINDINGS (MAMURIS ET AL., 2001)

1. On the UPGMA tree, based on the distance matrices of net interpopulation nucleotide divergence, the clustering of *M. barbatus* populations did not seem to follow any particular geographical trend.
2. For *M. barbatus* none of the comparisons in haplotype frequencies among samples showed any statistical significance.
3. For *M. barbatus* there was no apparent correlation of nucleotide divergence with geographical distances (Mantel t test, $P > .05$)

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB10 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb10	<i>Mullus barbatus</i>
Population\cluster	Cluster A
CO	1
AM	1
KA	1
GL	1
AL	1
RH	1

DATA FROM:

Garoia F, Guarniero I, Piccinetti C, Tinti F. 2004. **First microsatellite loci of red mullet (*Mullus barbatus*) and their application to genetic structure analysis of Adriatic shared stock.** Marine Biotechnology, **6**(5):446-452.

GENETIC MARKER USED

Microsatellites (6 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2001 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	Number of samples
AL	Albania	--	--	FAO37:GSA18	53
RI	Rimini	--	--	FAO37:GSA17	58
OR	Ortona	--	--	FAO37:GSA17	47
BA	Bari	--	--	FAO37:GSA18	48
total					206

MAIN FINDINGS (GAROIA ET AL., 2004)

1. Subtle but statistically significant genetic differentiation, indicating that the Adriatic red mullet may group into local, genetically isolated populations (Table 3 from Garoia *et al.*, 2004).
2. No correlation between geographic distance and genetic differentiation was observed

Table 3. Pairwise Estimates of F_{ST} (below diagonal) and Allelic Differentiation (above diagonal) Between Red Mullet Adriatic Samples^a

	RMN	ORT	BAR	ALB
RMN	—	*** (6*, 7*, 26b*, 31*, 39*)	NS	* (7*, 15*)
ORT	0.011 **	—	** (6*, 39*)	*** (6*, 39*)
BAR	0.002	0.002	—	—
ALB	0.007	0.008*	0.001	—

^aSignificance after sequential Bonferroni's correction is denoted as follows: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant. Values in bold were not significant after sequential Bonferroni's correction. The loci that showed significant allelic differentiation for each comparison are given in parentheses.

TABLE 3 FROM GAROIA ET AL., 2004

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB11 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb11	<i>Mullus barbatus</i>	
Population\cluster	Cluster A	Cluster B
AL	1	0
RI	1	0
OR	0	1
BA	1	0

GDMU_Mb15

DATA FROM:

Apostolidis, A. P., K. A. Moutou, C. Stamatis & Z. Mamuris, 2009. **Genetic structure of three marine fishes from the Gulf of Pagasitikos (Greece) based on allozymes, RAPD, and mtDNA RFLP markers.** *Biologia* 64(5):1005-1010 doi:10.2478/s11756-009-0161-0.

GENETIC MARKER USED

allozymes, RAPDs and mtDNA RFLPs

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	Number of samples
PAG1	Pagasitikos1	--	--	FAO37:GSA22	40
PAG2	Pagasitikos2	--	--	FAO37:GSA22	40
PAG3	Pagasitikos3	--	--	FAO37:GSA22	40
TRI	Trikeri			FAO37:GSA22	40
ALO	Alonissos	--	--	FAO37:GSA22	40
total					200

MAIN FINDINGS (APOSTOLIDIS ET AL., 2009)

1. Regardless of the method used, genetic population subdivision was not evident for any species within the Pagasitikos Gulf, suggesting homogeneity within the Gulf. The three populations within Pagasitikos represent thus a panmictic stock. However, there were evidences of genetic population subdivision between localities from inside and outside of the Pagasitikos Gulf.

TABLE 3 FROM APOSTOLIDIS ET AL., 2009

Table 3. N_eM and F_{ST} values among various groupings of populations in each species.

	N_eM Allozymes	N_eM RAPD	F_{ST} Allozymes	F_{ST} RAPD	N_{ST} mtDNA
<i>M. barbatus</i>			0.047*	0.057*	0.50
Within Pagasitikos	9.4	8.3			
Between Pagasitikos - Aegean Sea	7.2	7.1			
Outside Pagasitikos	8.7	8.2			

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB15 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb15	<i>Mullus barbatus</i>	
Population\cluster	Cluster A	Cluster B
PAG1	1	0
PAG2	1	0
PAG3	1	0
TRI	0	1
ALO	0	1

DATA FROM:

Galarza JA, Turner GF, Macpherson E, Rico C. 2009. **Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*).** *Can J Fish Aquat Sci*, **66**(9):1478-1490.

GENETIC MARKER USED

Microsatellites (10 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2003-2005 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	Number of samples
CO*	Conil	--	--	FAO27	30
CA*	Canary Islands	--	--	FAO34	48
CG	Cabo de Gata	--	--	FAO37:GSA1	48
BL	Blanes	--	--	FAO37:GSA6	37
PO*	Porticello	--	--	FAO37:GSA10	15
SI*	Siracusa	--	--	FAO37:GSA19	15
IT	Italy	--	--	FAO37:GSA9	48
GR	Greece	--	--	FAO37:GSA20	48
TR	Turkey	--	--	FAO37:GSA24	48
total					337

*CO; CA, PO and SI not shown in the geovisualization (not included in some analyses);

MAIN FINDINGS (GALARZA ET AL., 2009)

1. Low genetic heterogeneity within the Mediterranean Sea and a substantial gene flow reduction between the Atlantic Ocean and Mediterranean Sea.
2. Single discrete populations exist within the sampling area.
3. For red mullet differentiation was observed between all population pairwise comparisons except for Italian and Greek samples.
4. MDS analysis revealed a complex pattern, with high genetic similarity between Italian and Greek populations, Turkey being the most distinct, while CG and BL clustered separately (Figure 2 from Galarza *et al.*, 2009).

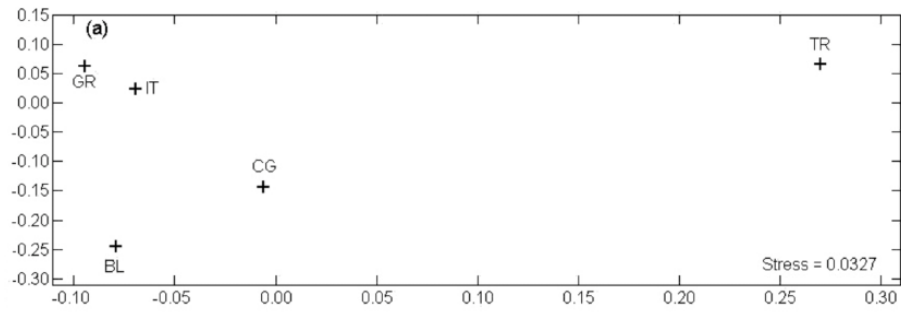
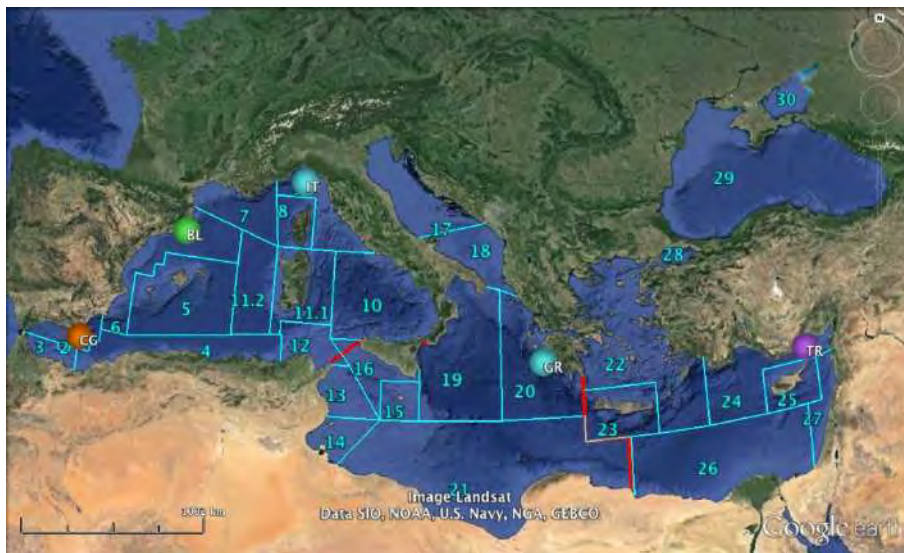


FIGURE 2 FROM GALARZA *ET AL.*, 2009 MULTIDIMENSIONAL SCALING PLOT OF POPULATION'S GENETIC DISTANCE FOR *M. BARBATUS*

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB16 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb16	<i>Mullus barbatus</i>			
Population\cluster	Cluster A	Cluster B	Cluster C	Cluster D
CG	1	0	0	0
BL	0	1	0	0
IT	0	0	1	0
GR	0	0	1	0
TR	0	0	0	1

DATA FROM:

Maggio T, Lo Brutto S, Garoia F, Tinti F, Arculeo M. 2009. **Microsatellite analysis of red mullet *Mullus barbatus* (Perciformes, Mullidae) reveals the isolation of the Adriatic Basin in the Mediterranean Sea.** *ICES Journal of Marine Science*, **66**(9):1883-1891.

GENETIC MARKER USED

Microsatellites (6 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2002-2003 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
GL1	Sete	43°14'N	04°10'E	FAO37:GSA7	24
TIR1	Genova	44°05'N	08°47'E	FAO37:GSA9	20
TIR2	Alghero	40°28'N	08°08'E	FAO37:GSA11.1	20
TIR3	S.Agata di Militello	38°05'N	14°37'E	FAO37:GSA10	41
TIR4	Porticello	38°03'N	13°35'E	FAO37:GSA10	45
TIR5	Castellamare del Golfo	38°09'N	12°55'E	FAO37:GSA10	59
SS1	Tunisia	37°30'N	08°40'E	FAO37:GSA12	22
SS2	Licata	36°49'N	13°59'E	FAO37:GSA16	70
SS3	Siracusa	36°28'N	15°18'E	FAO37:GSA16	43
IO1	Catania	37°24'N	15°29'E	FAO37:GSA19	51
AD1	Rimini	44°10'N	12°47'E	FAO37:GSA17	58
AD2	Fano	43°54'N	13°09'E	FAO37:GSA17	47
AD3	Bari	41°20'N	17°20'E	FAO37:GSA18	48
AD4	Albania	41°49'N	18°45'E	FAO37:GSA18	33
total					581

MAIN FINDINGS (MAGGIO ET AL., 2009)

1. Mean value of F_{ST} was significant ($F_{ST}=0.003$; $p < 0.001$), indicating genetic differentiation among the samples analysed.
1. This differentiation is primarily attributable to the isolation of the Adriatic samples and partly to a weaker substructuring of the populations in the Gulf of Lions, Tyrrhenian Sea, Strait of Sicily, and Ionian Sea.
2. Bayesian analysis also revealed genetic differentiation among the samples analysed, identifying two genetic clusters.
3. PCA analysis clearly separates Adriatic samples from all the other populations (Figure 2 from Maggio *et al.*, 2009).

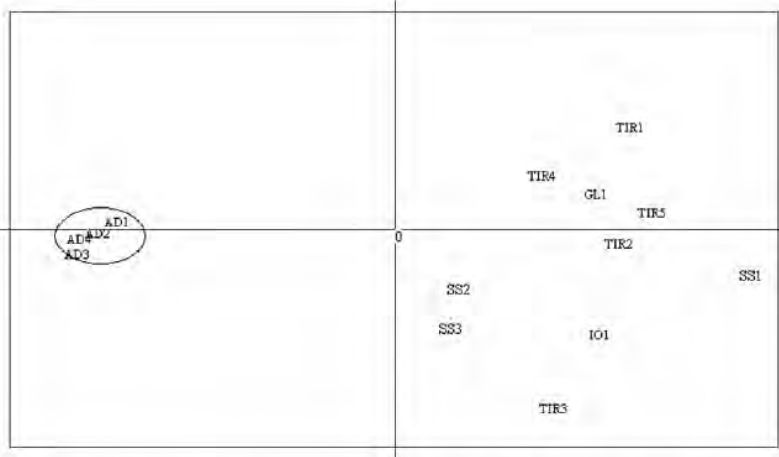


Figure 2. PCA of the genetic distance, calculated for six microsatellite loci among 14 samples of *Mullus barbatus*.

TABLE 2 FROM MAGGIO *ET AL.*, 2009

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB18 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb18	<i>Mullus barbatus</i>	
Population\cluster	Cluster A	Cluster B
GL1	0	1
TIR1	0	1
TIR2	0	1
TIR3	0	1
TIR4	0	1
TIR5	0	1
SS1	0	1
SS2	0	1
SS3	0	1
IO1	0	1
AD1	1	0
AD2	1	0
AD3	1	0
AD4	1	0

DATA FROM:

Felix-Hackradt, F. C., C. W. Hackradt, A. Perez-Ruzafa & J. A. Garcia-Charton, 2013. **Discordant patterns of genetic connectivity between two sympatric species, *Mullus barbatus* (Linnaeus, 1758) and *Mullus surmuletus* (Linnaeus, 1758), in south-western Mediterranean Sea.** *Marine Environmental Research* 92:23-34 doi:10.1016/j.marenvres.2013.08.008

GENETIC MARKER USED

Microsatellites (10 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in 2010-2012 in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	Number of samples
FU	Fuengirola	--	--	FAO37:GSA1	26
AL	Almeria	--	--	FAO37:GSA1	32
GA	Garrucha	--	--	FAO37:GSA1	20
MA	Mazarron	--	--	FAO37:GSA1	21
CA	Cartagena	--	--	FAO37:GSA6	18
VJ	Villajoyosa	--	--	FAO37:GSA6	28
total					145

MAIN FINDINGS (FELIX-HACKRADT ET AL., 2013)

1. 10 microsatellites markers were used on mullets' samples across the Spanish coast to determine the spatial scale of gene flow as well as the origin of post-larvae.
2. Population structure varied from complete homogeneity for *Mullus barbatus*, to high spatial variability in *Mullus surmuletus* samples.
3. No significant F_{ST} population pairwise comparisons was observed for *M. barbatus*

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB20 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb20	<i>Mullus barbatus</i>
Population\cluster	Cluster A
FU	1
AL	1
GA	1
MA	1
CA	1
VJ	1

DATA FROM:

Matic-Skoko, S., T. Segvic-Bubic, I. Mandic, D. Izquierdo-Gomez, E. Arneri, P. Carbonara, F. Grati, Z. Ikica, J. Kolitari, N. Milone, P. Sartor, G. Scarcella, A. Tokac & E. Tzanatos, 2018. **Evidence of subtle genetic structure in the sympatric species *Mullus barbatus* and *Mullus surmuletus* (Linnaeus, 1758) in the Mediterranean Sea.** Sci Rep 8:14 doi:10.1038/s41598-017-18503-7.

GENETIC MARKER USED

Microsatellites (13 loci)

DETAILS OF THE SAMPLING SITES ANALYSED

All samples were collected in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
C_NAS	North Adriatic, Croatia	44.943815°	13.605487°	FAO37:GSA17	48
C_MAS	Middle Adriatic, Croatia	43.568418°	15.769985°	FAO37:GSA17	66
C_SAS	South Adriatic, Croatia	42.615188°	17.779310°	FAO37:GSA17	50
I_NAS	North Adriatic, Italy	45.009044°	13.004869°	FAO37:GSA17	49
I_MAS	Middle Adriatic, Italy	43.750132°	13.757831°	FAO37:GSA17	50
I_SAS	South Adriatic, Italy	41.504365°	16.810572°	FAO37:GSA17	47
I_TS	Tyrrhenian Sea, Italy	41.503912°	11.065175°	FAO37:GSA9	49
MS_AS	South Adriatic, Montenegro	42.266861°	18.618152°	FAO37:GSA18	87
AL_AS	South Adriatic, Albania	41.162552°	19.259209°	FAO37:GSA18	67
GR_IS	Ionian Sea, Greece	38.171978°	22.651481°	FAO37:GSA20	59
TR_AS	Aegean Sea, Turkey	37.947463°	26.720771°	FAO37:GSA22	48
IS_LS	Israel, Levantine Sea	32.335961°	34.363010°	FAO37:GSA27	23
CP_LS	Cyprus, Levantine Sea	34.429656°	32.220743°	FAO37:GSA25	31
SP_BS	Balearic Sea	38.572230°	0.637080°	FAO37:GSA6	46
total					720

MAIN FINDINGS (MATIC-SKOKO ET AL., 2018)

1. At the broad Mediterranean scale, the overall F_{ST} value was 0.027 (95% CI = 0.014 to 0.042) and highly significant ($p < 0.001$), supporting the spatial heterogeneity.
2. Pairwise F_{ST} across all samples of red mullet ranged from -0.030 to 0.084, with 53 of 91 pairwise comparisons at $p < 0.01$ when permuted by Fisher's exact test. The majority of non-significant comparisons were found within populations from the northern and middle Adriatic Sea, including populations from both the western and eastern sides. These populations paired with the samples from the Tyrrhenian Sea, where F_{ST} ranged from 0.001 to 0.005.
3. On the contrary, in the southern Adriatic, populations from the eastern Adriatic coast (CRO_SAS, MN_AS, AL_AS) showed a break in gene flow toward the north and middle Adriatic regions and further to the south, including the Ionian Sea ($0.024 < F_{ST} < 0.049$).
4. No significant pairwise differentiations were observed for populations within the eastern Mediterranean, while reduced gene flow was noted between the western Mediterranean region (Balearic Sea) and Adriatic Sea, but not between the western and eastern Mediterranean.
5. The Bayesian clustering analysis of red mullet revealed three discrete genetic clusters that were supported by the mean likelihood score ($\ln(K)$) and the Delta K method (Fig. 3a). All three clusters were observed within the Adriatic Sea, separating the north and middle Adriatic

- populations (first cluster) from the south-eastern CRO_SAS population (second cluster) and the Montenegro-Albania populations (third cluster).
6. In accordance with FST pairwise results, no gene barrier was observed between the Balearic Sea and eastern Mediterranean, grouping all populations into the third cluster. Interestingly, the population from the Tyrrhenian Sea was assigned into the first cluster, together with the north and middle Adriatic populations.
 7. The DAPC plot clustered groups following the defined structure observed in STRUCTURE (Fig. 4a)

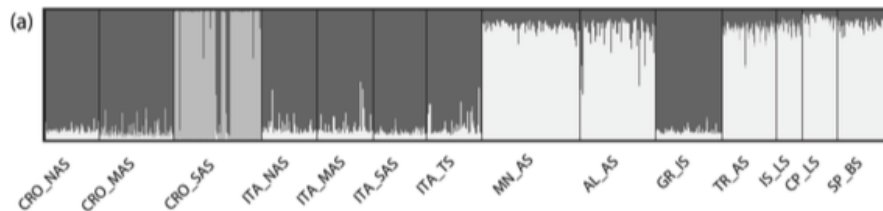


FIGURE 3 FROM MATIC-SKOKO *ET AL.*, 2018. BAYESIAN CLUSTERING OF *MULLUS BARBATUS* ACCORDING TO STRUCTURE ASSIGNMENT SCORES, ASSUMING THREE ($K = 3$) INFERRED CLUSTERS.

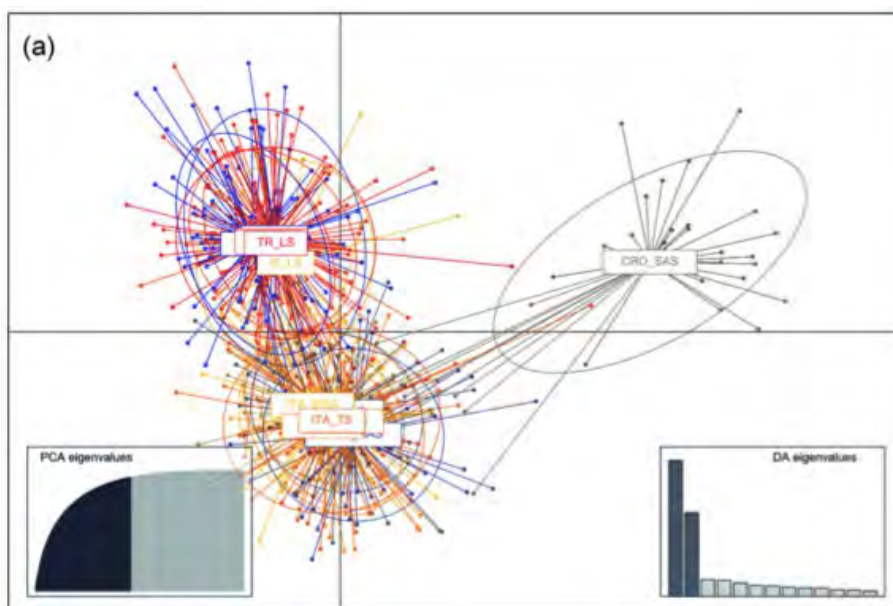


FIGURE 4 FROM MATIC-SKOKO *ET AL.*, 2018. DISCRIMINANT ANALYSIS OF PRINCIPAL COMPONENTS (DAPC) OF *MULLUS BARBATUS*

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_MB24 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Mb24	<i>Mullus barbatus</i>		
Population\cluster	Cluster A	Cluster B	Cluster C
C_NAS	0	1	0
C_MAS	0	1	0
C_SAS	1	0	0
I_NAS	0	1	0
I_MAS	0	1	0
I_SAS	0	1	0
I_TS	0	1	0
MS_AS	0	0	1
AL_AS	0	0	1
GR_IS	0	1	0
TR_AS	0	0	1
IS_LS	0	0	1
CP_LS	0	0	1
SP_BS	0	0	1

INFORMATION FROM STUDIES INCLUDING POPULATION SAMPLES OUTSIDE THE
MEDITERRANEAN SEA (P DATASETS)

Data not available

COMMENTS AND EVALUATION OF DATA: RED MULLET

The following table summarizes the results of the five population genetic studies performed so far for the red mullet

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic structure (clusters)	Number of Mediterranean clusters
Mb03	★★★ ^E	★			1(E)
Mb05	★★★ ^{WE}	★★★	★	★	2(WE/E)
Mb06	★★★ ^{WCE}	★★★	★	-	1(WCE)
Mb10	★★ ^{WCE}	★★	-	-	1(WCE)
Mb11	★★ ^W	★★★	★	★	2(C)
Mb15	★★ ^E	★★	★★	★★	2(E)
Mb16	★★★ ^W	★★★	★	★	4(2W/1WC/E)
Mb18	★★★ ^{WC}	★★★	★	★	2(WC/C)
Mb20	★★ ^W	★★	-	-	1(W)
Mb24	★★★ ^{WCE}	★★★	★	★★	3(WC/C/CE)

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★≤3; ★★4-6; ★★★>6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★≤20; ★★20-40; ★★★>40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall FST: ★≤0.05; ★★0.05-0.1; ★★★>0.1)

Genetic structure = ★ low, ★★ medium ★★★ high support (whenever available the score is based on pairwise Fst values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean

Comments: Old data based on RAPD and allozymes indicate a possible differentiation of Ionian from Aegean and within the Aegean Sea. Microsatellites data identified the coexistence of different clusters in the Adriatic Sea, where the sampling effort was specially concentrated. Contrary to the conspecific *M. surmuletus*, no genetic differentiation on a small scale was found in GSA1 and GSA6.

Gaps: No information available from vast areas (especially in the Western Mediterranean)

Technical notes: impossible to merge or even directly compare data from different sources (different markers or number of loci were used).

The analysis: *Nephrops norvegicus*

INFORMATION FROM STUDIES INCLUDING MEDITERRANEAN POPULATION SAMPLES (HP DATASETS)

GDMU_Nn02

DATA FROM:

Passamonti, M., B. Mantovani, V. Scali & C. Frogliia, 1997. **Allozymic Characterization of Scottish and Aegean Populations of *Nephrops norvegicus***. *Journal of the Marine Biological Association of the United Kingdom* 77(3):727-735 doi:10.1017/S0025315400036158..

GENETIC MARKER USED

Allozymes, (17 loci, 5 polymorphic)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

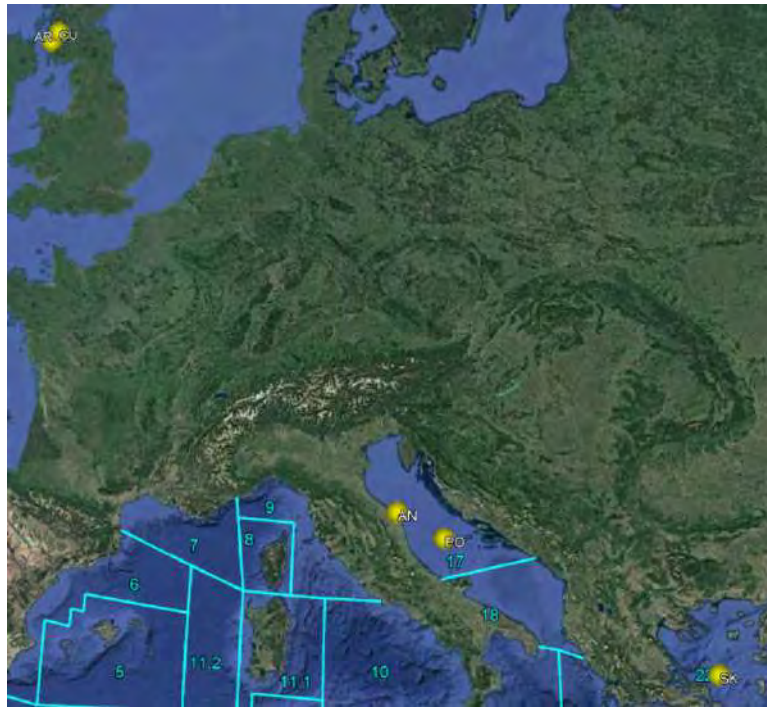
code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
AR	South Arran	--	--	FAO27	
CU	South Cumbrae	--	--	FAO27	
AN	Ancona	--	--	FAO37:GSA17	
PO	Pomo	--	--	FAO37:GSA17	
SK	Skiros	--	--	FAO37:GSA22	
total					

MAIN FINDINGS (PASSAMONTI ET AL., 1997)

1. A highly homogeneous allele pattern has been described.
2. The Nei's genetic distance values turned out to be very low, without any apparent geographical trend.
3. A very low genetic differentiation has been identified among samples.
4. The Pgm locus is the most differentiated, owing both intra- and interpopulation variance of genotype distribution.

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_NN02 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Nn02	<i>Nephrops norvegicus</i>
Population\cluster	Cluster A
AR	1
CU	1
AN	1
PO	1
SK	1

DATA FROM:

Maltagliati F, Camilli L, Biagi F, Abbiati M: **Genetic structure of Norway lobster, *Nephrops norvegicus* (L.) (Crustacea : Nephropidae), from the Mediterranean Sea.** *Scientia Marina* 1998, 62:91-99.

GENETIC MARKER USED

Allozymes, 15 enzymes

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
ADR	Adriatic Sea	43°54'N	13°50'E	FAO37:GSA17	100
AEG	Aegean Sea	38°40'N	23°26'E	FAO37:GSA22	100
ALB	Alboran Sea	35°30'N	04°20'W	FAO37:GSA1	100
BAL	Balearic Islands	39°37'N	02°02'E	FAO37:GSA5	100
CAT	Catalan Sea	41°06'N	02°11'E	FAO37:GSA6	100
FAR	Faro	36°47'N	07°57'W	FAO27	100
GEN	Ligurian Sea	44°04'N	09°26'E	FAO37:GSA9	100
SIC	Sicily Channel	36°30'N	13°30'E	FAO37:GSA16	100
TYR	Tyrrhenian Sea	42°26'N	10°44'E	FAO37:GSA9	100
total					900

MAIN FINDINGS (MALTAGLIATI ET AL., 1998)

1. The values of Nei's genetic distance index showed moderate genetic differentiation between geographical regions and no geographical pattern of genetic differentiation was detected by UPGMA cluster analysis (Figure 2 from Maltagliati *et al.*, 1998).
2. Moreover, no clear clines in allelic frequencies were detected, thus genetic variability seems to be randomly distributed among populations and *Nephrops norvegicus* seems to follow the island model of genetic structure.
3. Genetic distances among the Mediterranean populations of *Nephrops norvegicus* were moderate, but relatively high in all pairwise comparisons including Faro, Genoa and Tyrrhenian samples.
4. Over all loci the mean F_{ST} value was 0.122, this means that 12.2% of the total genetic variation results from differences between populations, with 87.8% coming from within-population variation. Highly significant mean F_{ST} value for the total dataset demonstrated significant differentiation among populations.
5. On the basis of the results of the present study, the species appears fragmented in groups, or islands, which replace a fraction of their residents with individuals migrating at random from a large collection of local populations.

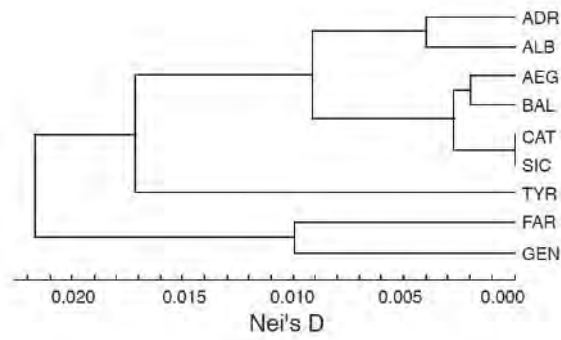
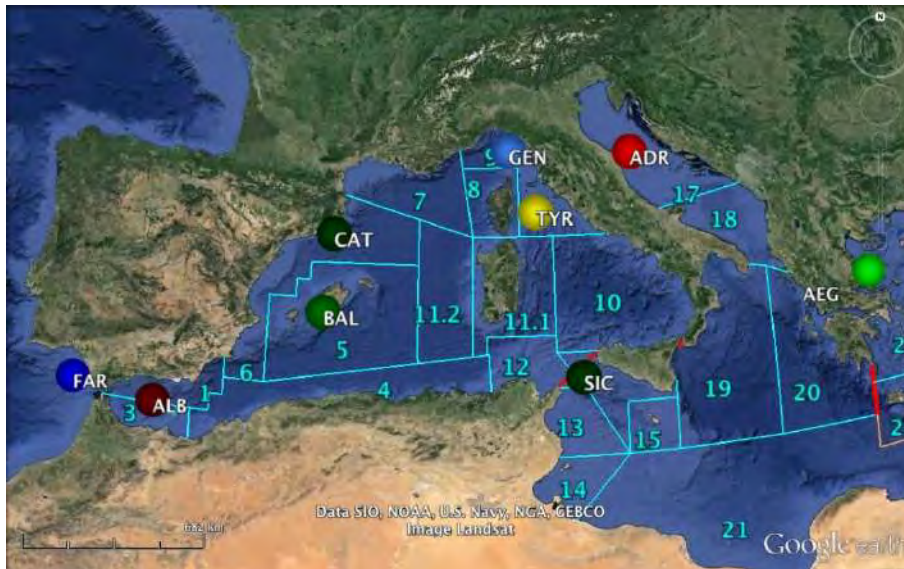


FIG. 2. – UPGMA clustering by Nei's (1978) genetic distances among local populations. Abbreviations for localities are as in Table 1.

FIGURE 2 FROM MALTAGLIATI *ET AL.*, 1998

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_NN03 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Nn03	<i>Nephrops norvegicus</i>										
Population\cluster	A	Subcl.a1	Subcl.a2	B	C	Subcl.c1	Subcl.c2	Subcl.c3	D	Subcl.d1	Subcl.d2
ADR	0	0	0	0	0	0	0	0	1	1	0
AEG	0	0	0	0	1	1	0	0	0	0	0
ALB	0	0	0	0	0	0	0	0	1	0	1
BAL	0	0	0	0	1	0	1	0	0	0	0
CAT	0	0	0	0	1	0	0	1	0	0	0
FAR	1	1	0	0	0	0	0	0	0	0	0
GEN	1	0	1	0	0	0	0	0	0	0	0
SIC	0	0	0	0	1	0	0	1	0	0	0
TYR	0	0	0	1	0	0	0	0	0	0	0

DATA FROM:

Stamatis, C., A. Triantafyllidis, K. A. Moutou & Z. Mamuris, 2004. **Mitochondrial DNA variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus***. *Molecular Ecology* 13(6):1377-1390 doi:10.1111/j.1365-294X.2004.02165.x.

GENETIC MARKER USED

Restriction fragment length polymorphism analysis of two mitochondrial DNA segment

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
PAG	Pagasitikos	--	--	FAO37:GSA22	36
PLA	Platanias	--	--	FAO37:GSA22	27
KER	Keramidi	--	--	FAO37:GSA22	21
EVI	Evia	--	--	FAO37:GSA22	20
ALO	Alonissos	--	--	FAO37:GSA22	30
CHA	Chalkidiki	--	--	FAO37:GSA22	32
POR	Portugal	--	--	FAO27	40
IRI	Irish Sea	--	--	FAO27	82
NS1	North Sea 1	--	--	FAO27	22
NS2	North Sea 2	--	--	FAO27	23
NS3	North Sea 3	--	--	FAO27	21
NS4	North Sea 4	--	--	FAO27	25
total					379

MAIN FINDINGS (STAMATIS ET AL., 2004)

1. Low levels of differentiation were found among the twelve populations from the North Sea, Irish Sea, Portuguese coast and Aegean Sea analysed (FST = 0.018, P < 0.001).
2. Statistically significant differences in haplotype frequencies among all populations (globally) and among the four major basins (Aegean Sea, Portugal, Irish Sea and North Sea) were observed (P < 0.00001 in both cases).
3. The sample that seemed to be the most differentiated, based on pairwise tests, was the Irish Sea sample (11 statistically significant tests out of 11 possible comparisons, even after sequential Bonferroni correction). Even if the Irish Sea sample was omitted, however, global tests remained highly significant.
4. The Irish Sea population was the most differentiated as a result of reduced levels of diversity.
5. There was no significant heterogeneity among samples within the Aegean Sea and North Sea (P = 0.084 and P = 0.061, respectively).
6. There were no signs of an Atlantic–Mediterranean divide or of an isolation-by-distance scheme of differentiation.
7. No evidence was found for a Mediterranean refugium during glaciation periods, separate from the Atlantic, as has been reported for some marine species.

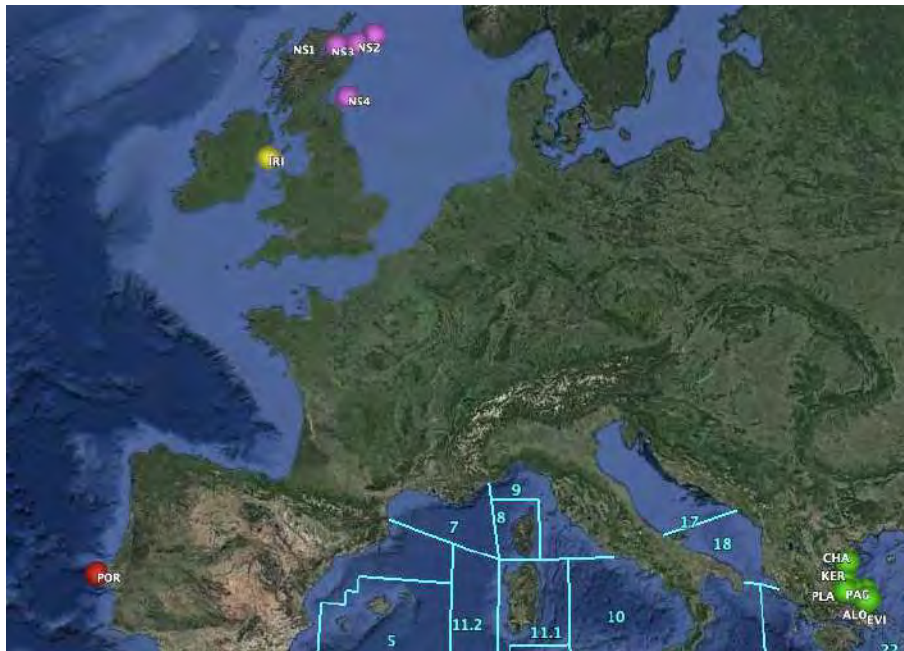
Structure	Source of variation	% Total variance	Fixation indices	P
No structure	Among populations	1.75	$F_{ST} = 0.018$	< 0.001
	Within populations	98.25		
Four major areas	Among regions	1.60	$F_{CT} = 0.016$	0.002
	Among samples/ Within regions	0.51	$F_{SC} = 0.005$	0.093
	Within regions			
	Within samples	97.89	$F_{ST} = 0.021$	< 0.001

The percentage of variation, the probability (P) estimated from permutation tests as well as the F statistics is given at each hierarchical level (Excoffier *et al.* 1992).

TABLE2 FROM STAMATIS ET AL., 2004. RESULTS OF THE HIERARCHICAL ANALYSIS OF MOLECULAR VARIANCE FOR *NEPHROPS NORVEGICUS* USING THE FOUR MAJOR GEOGRAPHICAL AREAS OF SAMPLING AS THE LEVELS OF GROUPING, OR WITHOUT GROUPING

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_NN06 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Nn06	<i>Nephrops norvegicus</i>			
Population\cluster	Cluster A	Cluster B	Cluster C	Cluster D
PAG	1	0	0	0
PLA	1	0	0	0
KER	1	0	0	0
EVI	1	0	0	0
ALO	1	0	0	0
CHA	1	0	0	0
POR	0	0	1	0
IRI	0	1	0	0
NS1	0	0	0	1
NS2	0	0	0	1
NS3	0	0	0	1
NS4	0	0	0	1

DATA FROM:

Stamatis C, Triantafyllidis A, Moutou KA, Mamuris Z: **Allozymic variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus***. *ICES J. Mar. Sci.* 2006, 63:875-882

GENETIC MARKER USED

Allozymes, 15 putative loci

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
PAG	Pagositikos	--	--	FAO37:GSA22	104
PLA	Platanias	--	--	FAO37:GSA22	31
KER	Keramidi	--	--	FAO37:GSA22	22
EVI	Evia	--	--	FAO37:GSA22	20
ALO	Alonissos	--	--	FAO37:GSA22	20
NS1	North Sea 1	--	--	FAO27	30
NS2	North Sea 2	--	--	FAO27	50
NS3	North Sea 3	--	--	FAO27	41
NS4	North Sea 4	--	--	FAO27	48
total					366

MAIN FINDINGS (STAMATIS ET AL., 2006)

1. Results of this study based on genetic distance estimates, F_{ST} analyses, and tests for genetic differentiation revealed a heterogeneous genetic structure within *N. norvegicus*.
2. The analysis of genetic differentiation by F_{ST} statistics gave a significant mean value ($F_{ST}=0.013$, $p<0.001$), indicating a low, but significant, population structuring across the distribution of the species. There were no signs, however, of an Atlantic-Mediterranean division.
3. The F_{ST} value for all *N. norvegicus* populations revealed a heterogeneous genetic structure among all samples (0.013 , $p < 0.001$). This differentiation is mostly due to loci ADA-2 and GLC-1. This agrees with global genetic heterogeneity probability tests of *N. norvegicus* populations, which indicated that there are highly significant differences ($p < 0.001$) in the allele and genotypic frequency distributions only at these two loci.
4. The absence of a clear geographical pattern of genetic differentiation among the populations studied is further evident in the UPGMA dendrogram (Figure 2 from Stamatis *et al.*, 2006), constructed on the basis of Nei's (1978) genetic distances. Population samples cluster in three groupings: samples PAG with PLA, North Sea and KER samples, and EVI and ALO samples. These groupings are not supported by high bootstrap values (all were lower than 65%, out of 1000 iterations).
5. Results of pairwise genetic heterogeneity tests (based on allele frequency distributions) showed that 12 of 37 tests were statistically significant (six after Bonferroni correction), most concerning the comparison of the PAG and ALO population samples to NS2-4 samples. However, there were statistically significant differences even among close geographical samples (i.e. PAG-ALO and NS2-NS3).

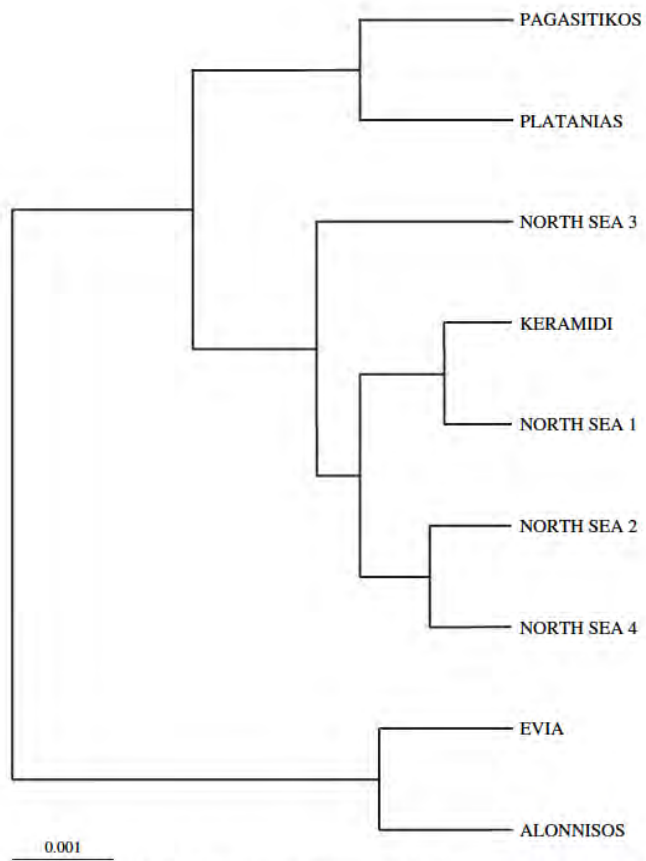
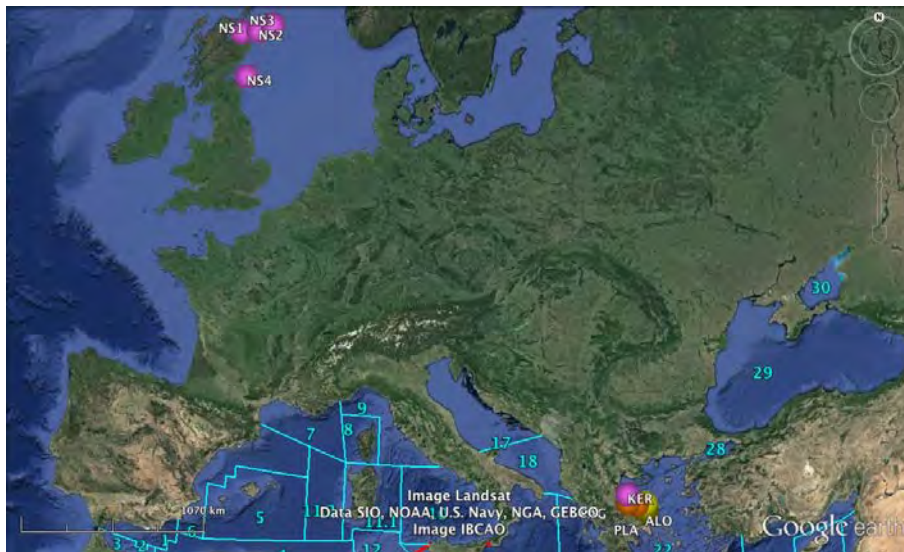


Figure 2. UPGMA clustering of nine *Nephrops norvegicus* populations, based on Nei's (1978) genetic distance.

FIGURE 2 FROM STAMATIS ET AL., 2006

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_NN07 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Nn07	<i>Nephrops norvegicus</i>		
Population\cluster	Cluster A	Cluster B	Cluster C
PAG	1	0	0
PLA	1	0	0
KER	0	1	0
EVI	0	0	1
ALO	0	0	1
NS1	0	1	0
NS2	0	1	0
NS3	0	1	0
NS4	0	1	0

DATA FROM:

Gallagher, J., J. A. Finarelli, J. P. Jonasson & J. Carlsson, 2018. **Mitochondrial D-loop DNA analyses of Norway Lobster (*Nephrops norvegicus*) reveals genetic isolation between Atlantic and Mediterranean populations.** *bioRxiv* doi: <http://dx.doi.org/10.1101/258392>. (not peer reviewed)

GENETIC MARKER USED

Mitochondrial DNA (control region, 375 bp)

DETAILS OF THE SAMPLING SITES ANALYSED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
AN	Ancona	--	--	FAO37:GSA17	30
GC	Cadiz	--	--	FAO27	30
BRE	Breiðamerkurdjúp	--	--	FAO27	30
BOB	Bay of Biscay	--	--	FAO27	30
IR	Irish Sea	--	--	FAO27	30
NS	North Sea	--	--	FAO27	30
NN	Northerb Norway	--	--	FAO27	30
POR	Porcupine	--	--	FAO27	30
SKA	Skagerrak	--	--	FAO27	30
total					270

MAIN FINDINGS (GALLAGHER ET AL., 2018)

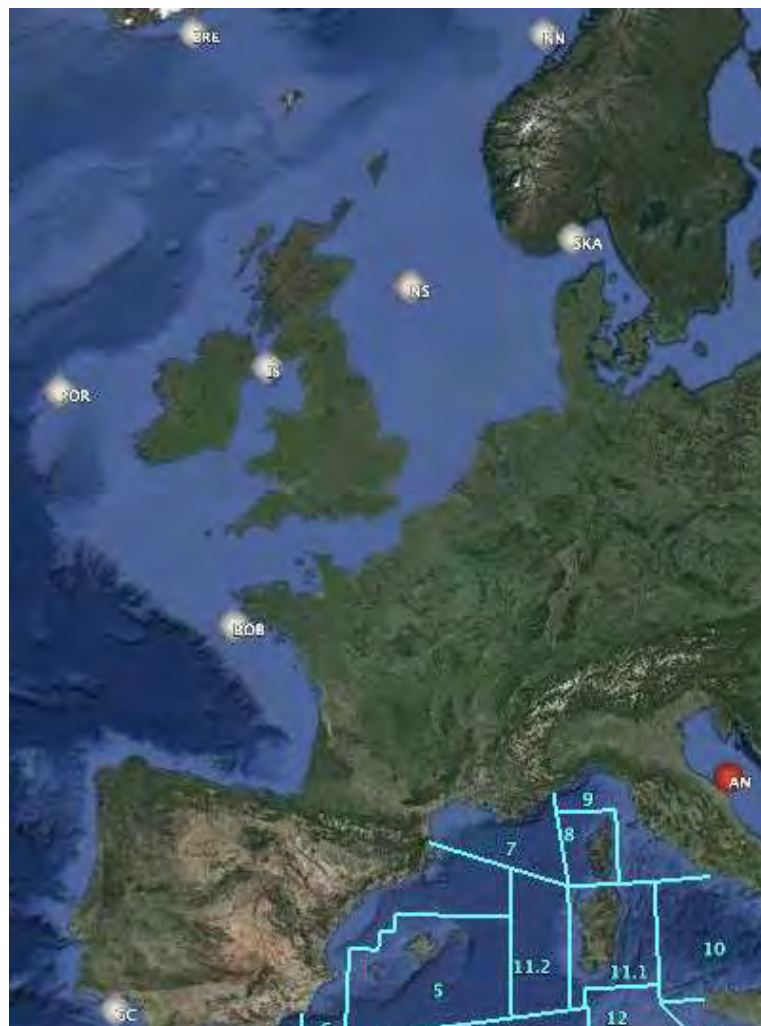
1. The SAMOVA analysis indicated that the best-supported FCT value (FCT = 0.109) was achieved when the samples were clustered into two groups. The first group contained only individuals derived from the Mediterranean (Ancona) while the second group comprised of individuals from all other sampled areas
2. An AMOVA analysis, using the SAMOVA structure and the K80 distance model (Kimura, 1980) revealed significant heterogeneity among the nine samples ($\Phi_{ST} = 0.107$, $P < 0.001$). Within-sample variation accounted for 88.86% of the variance (FCT=0.100, $P < 0.001$)
3. Pairwise Φ_{ST} values revealed population structure between the eastern Mediterranean (Ancona) sample and each of the eight other samples from the North Atlantic. Average Φ_{ST} estimates are at least twice as large between the Mediterranean sample and each Atlantic sample than the estimates among all of the Atlantic samples.

TABLE 4 FROM GALLAGHER *ET AL.*, 2018. PAIRWISE Φ_{ST} ESTIMATES (BELOW DIAGONAL) FOR D-LOOP MTDNA DATA AMONG *N. NORVEGICUS* SAMPLES. P-VALUES (UPPER DIAGONAL) IN BOLD WERE SIGNIFICANT AFTER SEQUENTIAL BONFERRONI CORRECTION.

Population	Anacona	Bay of Biscay	Breiðamer- kurdjúp	Gulf of Cadiz	Irish Sea	Northern Norway	North Sea	Porcupine	Skagerrak
Anacona		0.019	<0.001	0.021	0.024	0.031	0.046	0.026	0.016
Bay of Biscay	0.111		1.336	1.134	0.270	0.870	1.062	0.273	0.717
Breiðamerkurdjúp	0.152	0.046		0.572	0.615	0.627	0.200	0.060	0.670
Gulf of Cadiz	0.144	0.002	0.019		1.196	1.434	1.356	0.140	0.704
Irish Sea	0.143	0.043	-0.009	0.005		0.403	0.333	0.240	0.604
Northern Norway	0.072	0.018	0.011	-0.006	-0.001		1.248	1.554	0.748
North Sea	0.087	-0.007	0.043	-0.015	0.027	-0.010		0.358	0.625
Porcupine	0.057	0.049	0.061	0.051	0.037	-0.010	0.015		0.129
Skagerrak	0.098	-0.014	0.002	-0.010	-0.008	-0.010	-0.009	0.020	

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_NN14 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_Nn14	<i>Nephrops norvegicus</i>	
Population\cluster	Cluster A	Cluster B
AN	1	0
GC	0	1
BRE	0	1
BOB	0	1
IR	0	1
NS	0	1
NN	0	1
POR	0	1
SKA	0	1

INFORMATION FROM STUDIES INCLUDING POPULATION SAMPLES OUTSIDE THE MEDITERRANEAN SEA (P DATASETS)

GDMU_Nn10: Stock structure of Norway lobster off southern Iceland was investigated using 12 microsatellite loci. The overall genetic estimates did not reveal significant F_{ST} ($F_{ST} = -0.0003$, $p > 0.05$). This genetic pattern was reflected in the pairwise F_{ST} comparisons of samples, because none of the comparisons were significant after the Bonferroni correction. No genetic method detected significant genetic differentiation among the locations sampled, even among Icelandic samples and an out-group from Scotland (Pampoulie *et al.*, 2011).

COMMENTS AND EVALUATION OF DATA: NORWAY LOBSTER

The following table summarizes the results of the five population genetic studies performed so far for the Norway lobster

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic structure (clusters)	Number of Mediterranean clusters
Nn02	★★ ^{CE}	na	-	-	1(CE)
Nn03	★★★★ ^{WCE}	★★	★	★★★	4(2W/1WE/C)
Nn06	★★★ ^E	★★★★	★	-	1(E)
Nn07	★★★ ^E	★★★★	★	★	3(E)
Nn14	★★★ ^C	★★★★	★	★★	1(C)

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★≤3; ★★4-6; ★★★>6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★≤20; ★★20-40; ★★★>40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall F_{ST} : ★≤0.05; ★★0.05-0.1; ★★★ >0.1)

Genetic structure = ★ low, ★★ medium ★★★ high support (whenever available the score is based on pairwise F_{ST} values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean

Comments: significant population differentiation across the distribution of the species but without a clear geographical pattern

Gaps: updated data needed for the whole Mediterranean

Technical notes: only allozymes data, old data (sampling and technique used);

The analysis: *Parapenaeus longirostris*

INFORMATION FROM STUDIES INCLUDING MEDITERRANEAN POPULATION SAMPLES (HP DATASETS)

GDMU_P105

DATA FROM:

Garcia-Merchan, V. H., A. Robainas-Barcia, P. Abello, E. Macpherson, F. Palero, M. Garcia-Rodriguez, L. Gil de Sola & M. Pascual, 2012. **Phylogeographic patterns of decapod crustaceans at the Atlantic-Mediterranean transition.** *Mol Phylogenet Evol* 62(2):664-72 doi:10.1016/j.ympev.2011.11.009.

GENETIC MARKER USED

Mitochondrial DNA (COI gene, 561 bp)

DETAILS OF THE SAMPLING SITES ANALYZED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
CAD	Cadiz	--	--	FAO27	22
MAL	Malaga	--	--	FAO37:GSA1	21
ALI	Allcante	--	--	FAO37:GSA6	24
VAL	Valencia	--	--	FAO37:GSA6	13
TAR	Tarragona	--	--	FAO37:GSA6	21
total					101

MAIN FINDINGS (GARCIA-MERCHAN ET AL., 2012)

1. A fragment of the COI gene was analyzed in seven decapod crustacean species from five families and with different bathymetric distributions. Individuals were sampled along the Atlantic–Mediterranean transition area in order to test the effect of three putative barriers to gene flow: Strait of Gibraltar, Almeria–Oran Front and Ibiza Channel.
2. *P. longirostris* samples were characterized by a very low nucleotide and haplotype diversities. Global genetic differentiation within the species was not significant.
3. The Almeria–Oran front, previously defined as a barrier in numerous marine organisms, showed no effect on the genetic structure of the studied species.
4. This result could be due to sampling limitations or could be related to the characteristics of the molecular marker used (e.g. low diversity found in *Parapenaeus*).

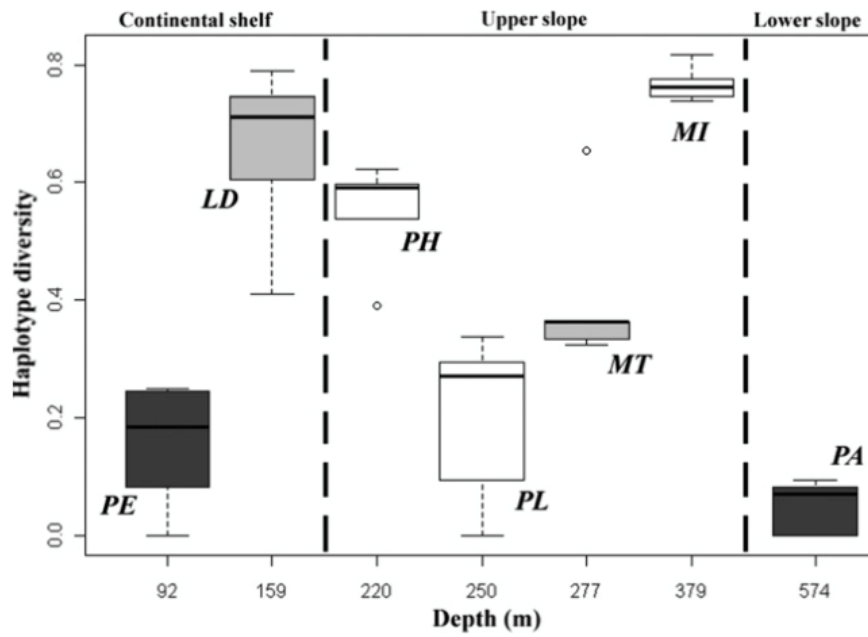


FIGURE 2 FROM GARCIA-MERCHAN *ET AL.*, 2012. PE: *PAGURUS EXCAVATUS*, LD: *LIOCARCINUS DEPURATOR*, PH: *PLESIONIKA HETEROCARPUS*, PL: *PARAPENAEUS LONGIROSTRIS*, MT: *MACROPIPIUS TUBERCULATUS*, MI: *MUNIDA INTERMEDIA* AND PA: *PAGURUS ALATUS*

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_PL05 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_PI05	<i>Parapenaeus longirostris</i>
Population\cluster	Cluster A
CAD	1
MAL	1
ALI	1
VAL	1
TAR	1

DATA FROM:

Lo Brutto S, Maggio T, Arculeo M: **Isolation By Distance (IBD) signals in the deep-water rose shrimp *Parapenaeus longirostris* (Lucas, 1846) (Decapoda, Panaeidae) in the Mediterranean Sea.** Marine Environmental Research 2013, **90**: 1-8.

GENETIC MARKER USED

Mitochondrial DNA (control region 307 bp) and AFLP (143 polymorphic loci)

DETAILS OF THE SAMPLING SITES ANALYZED

The samples were collected (no information is available on the date) in the following locations:

code	site	latitude	longitude	FAO area, GFCM GSA	No of samples
TYR1	Porto S. Stefano	--	--	FAO37:GSA9	40
TYR2	Terrasini	--	--	FAO37:GSA10	22
TYR3	Porticello	--	--	FAO37:GSA10	22
SS1	Scoglitti	--	--	FAO37:GSA16	30
SS2	Porto Palo	--	--	FAO37:GSA16	30
AD1	Fano	--	--	FAO37:GSA17	24
AEG1	Creta	--	--	FAO37:GSA23	24
AEG2	Kavala	--	--	FAO37:GSA22	48
total					240

MAIN FINDINGS (LO BRUTTO ET AL., 2013)

1. 'Isolation by Distance' pattern from central to eastern Mediterranean Sea, within the Mediterranean Sea, which includes geographical areas with some degree of isolation
2. Four clusters are significantly differentiated (according to the Mediterranean sub basins: TYR+SS + AD + AEG) ($\Phi_{CTAFLP} = 0.118$; $p < 0.001$; $\Phi_{CTmDNA} = 0.046$; $p < 0.001$) (Table 2 from Lo Brutto *et al.*, 2013)
3. The greatest contribution to the differences among the four Mediterranean sub basins depended on AEG and TYR areas, which represented the most divergent groups.
4. Deep separation of AEG from the remaining populations' highlights a 'Levantine isolation' that characterizes the eastern Mediterranean basin.

Table 2
 AMOVA conducted on samples of *Parapeneus longirostris* with AFLP (analyses 1_{AFLP}, 2_{AFLP}, 3_{AFLP} and 4_{AFLP}) and mtDNA (analyses 1_{mtDNA}, 2_{mtDNA}, 3_{mtDNA} and 4_{mtDNA}) data, clustering in four different ways. Analysis 1: all together; Analysis 2: grouping Tyrrhenian Sea, Strait of Sicily, Adriatic Sea, Aegean Sea, (TYR + SS + AD + AEG); Analysis 3: grouping TYR + SS + AD, excluding Aegean samples; Analysis 4: three groups Strait of Sicily (SS), Adriatic sea (AD), Aegean sea (AEG) excluding Tyrrhenian sea; * indicates $p < 0.05$ and ** $p < 0.001$.

	Sample clustering	Variation source	Degrees of freedom	Variation (%)	Fixation indices
1 _{AFLP}	All together	Among-populations	7	18.5	ϕ_{ST} 0.185**
		Within-populations	125	81.5	
2 _{AFLP}	Four groups: TYR + SS + AD + AEG	Among-groups	3	11.78	ϕ_{CT} 0.118**
		Among-populations within groups	4	8.69	ϕ_{SC} 0.098**
		Within-populations	125	79.53	ϕ_{ST} 0.204**
3 _{AFLP}	Three groups: TYR + SS + AD	Among-groups	2	-0.08	ϕ_{CT} -0.001
		Among-populations within groups	3	14.02	ϕ_{SC} 0.140*
		Within-populations	67	86.06	ϕ_{ST} 0.139*
4 _{AFLP}	Three groups: SS + AD + AEG	Among-groups	2	7.25	ϕ_{CT} 0.090
		Among-populations within groups	2	20.76	ϕ_{SC} 0.207*
		Within-populations	57	71.99	ϕ_{ST} 0.280*
1 _{mtDNA}	All together	Among-populations	7	5.10	ϕ_{ST} 0.051**
		Within-populations	99	94.90	
2 _{mtDNA}	Four groups: TYR + SS + AD + AEG	Among-groups	3	4.12	ϕ_{CT} 0.046*
		Among-populations within groups	4	1.76	ϕ_{SC} 0.022*
		Within-populations	100	94.12	ϕ_{ST} 0.058**
3 _{mtDNA}	Three groups: TYR + SS + AD	Among-groups	2	-0.51	ϕ_{CT} -0.005
		Among-populations within groups	3	0.99	ϕ_{SC} 0.009
		Within-populations	86	99.52	ϕ_{ST} 0.005
4 _{mtDNA}	Three groups: SS + AD + AEG	Among-groups	2	0.42	ϕ_{CT} 0.004
		Among-populations within groups	2	0.95	ϕ_{SC} 0.009
		Within-populations	57	98.63	ϕ_{ST} 0.013*

TABLE 2 FROM LO BRUTTO ET AL., 2013

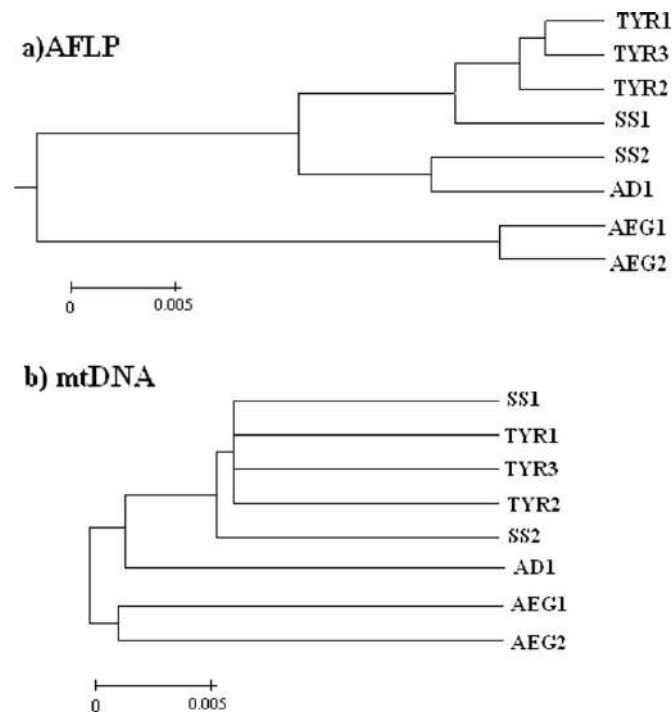


Figure 2 from Lo Brutto et al., 2013. Fig. 2. UPGMA dendrograms based on Nei's genetic distance for AFLP (a) and on Reynolds distance for mtDNA (b). The scale bars represent the genetic distance.

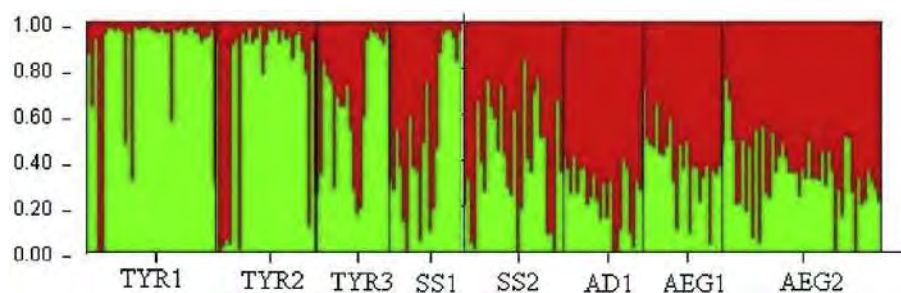
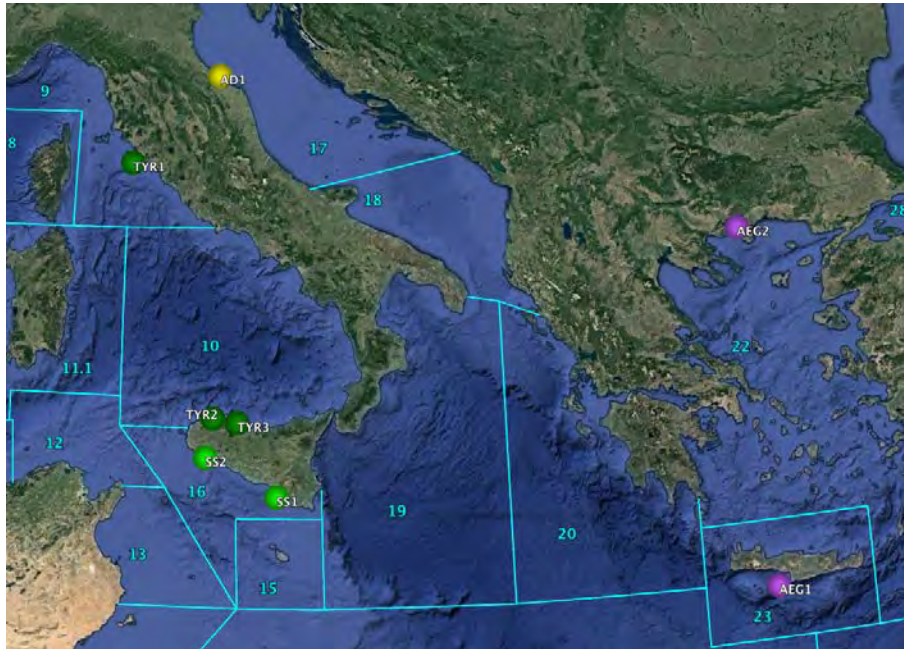


Figure 4 from Lo Brutto et al., 2013. Bayesian assignment analysis for K = 2. Each bar represents an individual of *Parapenaeus longirostris* and its associated probability of belonging to one of the two genetic clusters detected (green and red).

GENETIC DIFFERENTIATION AND STOCK STRUCTURING WITHIN THE MEDITERRANEAN SEA AS INFERRED BY THE GDMU_PL06 DATASET

GEOVISUALIZATION



MATRIX OF GENETIC CLUSTERING

GDMU_PI06	<i>Parapenaeus longirostris</i>			
Population\cluster	Cluster A	Cluster B	Cluster C	Cluster D
TYR1	0	1	0	0
TYR2	0	1	0	0
TYR3	0	1	0	0
SS1	0	0	1	0
SS2	0	0	1	0
AD1	0	0	0	1
AEG1	1	0	0	0
AEG2	1	0	0	0

INFORMATION FROM STUDIES INCLUDING POPULATION SAMPLES OUTSIDE THE
MEDITERRANEAN SEA (P DATASETS)

Data not available

COMMENTS AND EVALUATION OF DATA: DEEP-WATER ROSE SHRIMP

The following table summarizes the results of the five population genetic studies performed so far for the deep-water rose shrimp

GDMU	N GSA	Mean N i/s	Overall differentiation	Genetic structure (clusters)	Number of Mediterranean clusters
PI05	★★ ^W	★	-	-	1(W)
PI06	★★★ ^{WCE}	★★	★★★	★★★	4(W/2C/E)

Where:

GDMU = Genetic Dataset MED_Units

N GSA = Number of GFCM GSAs investigated: ★ ≤3; ★★ 4-6; ★★★ >6; W, C, E indicates that sampling sites are from the Western, Central, Eastern Mediterranean [FAO Subareas 37.1, 37.2 and 37.3, respectively];

Mean N i/s = Mean number of individuals/sites ★ ≤20; ★★ 20-40; ★★★ >40;

Overall differentiation = ★ low; ★★ medium; ★★★ high support; (whenever available the score is based on the value of the overall FST: ★ ≤0.05; ★★ 0.05-0.1; ★★★ >0.1)

Genetic structure = ★ low, ★★ medium ★★★ high support (whenever available the score is based on pairwise Fst values, AMOVA grouping, other metrics or analyses (eg PCA, dendrograms etc.))

N° of Mediterranean clusters = number of differentiated genetic clusters identified within the Mediterranean Sea; between parentheses the location of clusters in W, C, E Mediterranean

Comments: very few data existing to date for the species

Gaps: no data available for the Ionian Sea and most part of the Western Mediterranean (apart from North and South Tyrrhenian Sea)

Technical notes: consistent results with two markers (mtDNA and AFLP)

Conclusions

A total of 48 datasets provided useful information on population genetic differentiation for the target species. The fish species have a higher number of studies compared to the crustaceans, with *M. merluccius* being the most extensively studied species and *P. longirostris* the less investigated one (Figure 8).

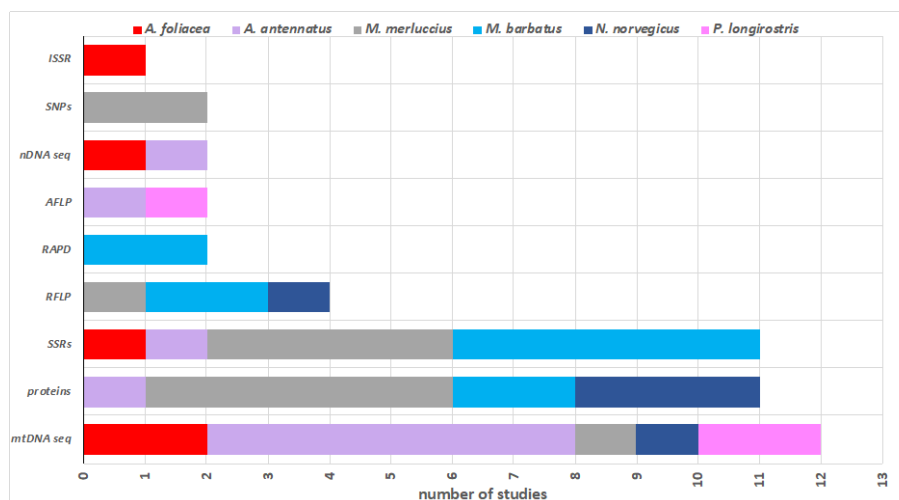


FIGURE 8

However, these studies were based on different numbers and type of markers, different sampling efforts lead sometimes to contrasting results and the impossibility to merge and compare the outputs.

From the re-analysis of the available literature, the following points emerge:

As concerns the sampling:

- For most of the species, the geographical coverage of genetic data is very poor and sporadic; not enough sampling and/or not enough loci have been employed in most of the studies.
- Only very few cases of broad and systematic sampling design are available (e.g. *M. merluccius*), usually within the framework of International project funded by the EU.

As concerns the methodology:

- The old and less resolutive markers (i.e., proteins, RAPD, mtDNA sequencing, nDNA sequencing etc.) have been used in all the 6-target species, often with controversial results. Protein-based analyses (allozymes, isozymes etc.), RFLP, AFLP, Sanger-sequencing of mitochondrial and nuclear DNA are nowadays less and less used in population genetic analyses, in favour of the most performing markers. Sequencing data (particularly mtDNA) can still be useful for reconstructing past evolutionary/demographic events rather than recent scenarios.
- Microsatellite markers (SSR) have been used in four out of the six-target species, both in multiple studies per species (*M. merluccius* and *M. barbatus*) or in a single study (*A. foliacea* and *A. antennatus*). However, limited number of loci are available, their resolutive power, especially in the red shrimp species, proved to be quite low. They are still widely used in population genetic analyses, but only when an adequate number of loci is analyzed, the results can be reliable and resolutive.
- Among the 6-target species of the project, the most modern or 'last-generation' markers (SNPs) have been applied on the European hake only, allowing to detect fine-scale population genetic structuring. They are the most informative markers but also the most demanding in terms of costs and methodological procedures, thus their inherent constraints could limit the number of specimens that are analysed and the spatial/temporal coverage of the study unless

an adequate budget is allocated. When SNPs are available, usually in large numbers, the resolute power of genetic data in identifying differentiation and structuring increases markedly. They should be the first-choice marker for stock delimitation and connectivity studies since information can be transferred and easily used across different labs.

Glossary and acronyms

Admixture (genetic admixture): the process that occurs when individuals from two or more previously separated populations begin interbreeding (the formation of novel genetic combinations through hybridization of genetically distinct groups). Admixture refers also to ancestry from more than one population.

AFLP (Amplified fragment length polymorphism): A highly sensitive and reproducible method for detecting polymorphisms in DNA. Following restriction enzyme digestion of DNA, a subset of DNA fragments is selected for PCR amplification. The amplified fragments are separated and visualized on denaturing polyacrylamide gels, either through autoradiography or fluorescence methodologies, or via automated capillary sequencing instruments.

Alleles: Alternate forms or varieties of a gene. The alleles for a trait occupy the same locus or position on homologous chromosomes and thus govern the same trait. However, because they are different, their action may result in different expressions of that trait.

Allele frequency: The percentage of one allele relative to all alleles at a locus in a population. Also called gene frequency.

Allozyme: any one of a number of different structural forms of the same enzyme coded for by a different allele.

AMOVA (Analysis of Molecular Variance): Analogous to ANOVA in statistics, it is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. It can be defined as a method to partition the total genetic variation in a species into components within and among populations or groups at different level of hierarchical subdivision. AMOVA analysis allows to calculate F-statistics, useful to estimate the proportion of genetic variability found among populations (F_{ST}), among populations within groups (F_{SC}) and among groups (F_{CT}). In particular, for mitochondrial sequence data Φ -statistics (Φ_{ST} , Φ_{CT} , and Φ_{SC} , F-analogues) are obtained. θ -statistic is the unbiased estimator of F_{st} that corrects for error associated with incomplete sampling of a populations (Weir and Cockerham 1984).

Assignment test: A method for quantifying dispersal by assigning each individual to the population in which its genotype has the highest probability of occurring, and from which is therefore most likely originated.

Bayesian: A method of statistical inference that uses prior information to test the likelihood that various parameters can explain a particular data set.

Bayesian assignment analysis: it uses individual genotypes to assign individuals to populations or clusters using a Bayesian approach.

Bonferroni correction: A method used to counteract the problem of multiple comparisons in statistics. The Bonferroni correction is an adjustment made to P values when several dependent or independent statistical tests are being performed simultaneously on a single data set. To perform a Bonferroni correction, the critical P value (α) is divided by the number of comparisons being made. The Bonferroni correction is used to reduce the chances of obtaining false-positive results (type I errors) when multiple pair wise tests are performed on a single set of data.

bp: Abbreviation of 'base pairs' (nucleotides)

branch (of a tree): see evolutionary tree

Clade: A clade is a group of organisms that includes an ancestor and all descendants of that ancestor in an evolutionary tree.

Cline: A geographical gradient in one or more genetic characteristics within a species or other taxon, especially between different populations.

Dendrogram: A branching tree-like diagram representing a hierarchy of categories based on degree of similarity or number of shared genetic characteristics.

Discrete populations (isolated populations): Discrete populations (from the Latin discretus, meaning to separate) are separated and individually distinct populations according to their genetic features.

Distance methods: Methods of phylogenetic reconstruction that are based on the evolutionary (genetic) distances between all pairs of taxa.

Evolutionary tree (phylogeny): It represents the evolutionary relationships among a set of organisms or groups of organisms (Taxa). Evolutionary trees depict clades (see definition). It is a branching diagram or "tree" showing the inferred evolutionary relationships among various biological species or other entities—their phylogeny—based upon similarities and differences in their genetic characteristics. The tips of the tree are sometimes called leaves and are basically where the actual taxa are. The branches of the tree are the lines that connect the leaves and represent the pathway that traces the proposed evolutionary history of the lineages. The points at which branches separate are called nodes and represent hypothetical ancestors of the descendant clades of that node. The root of the tree is the common ancestor of all taxa.

Founder effect: The theory that the founders of a new colony carry only a fraction of the total genetic diversity/variation of the source population.

F_{st}: (overall and pairwise) the most common measurement used to describe the genetic differentiation of populations.

Genes: Unit of inheritance usually occurring at specific locations, or loci, on a chromosome. Physically, a gene is a sequence of DNA bases that specify the order of amino acids in an entire protein or, in some cases, a portion of a protein. A gene may be made up of hundreds of thousands of DNA bases and can have multiple different forms, or alleles, which are defined by different sequences of DNA. Genes are responsible for the hereditary traits in plants and animals.

Gene flow: The transfer of genes from one population to another, usually as a result of migration. The loss or addition of individuals can easily change the gene pool frequencies of both the recipient and donor populations-- that is, they can evolve.

Genetic diversity: The amount of genetic variation that is contained within a population or species.

Genetic distance: Various statistics for measuring the 'genetic distance' between subgroups or populations. Major distance measures include Nei's distance (Nei M, *American Naturalist* 1972, 106:283-292; Nei M, *Genetics*, 1978, 89:583-590), Reynold's distance (Reynolds et al., *Genetics*, 1983, 105(3):767-679).

Genetic divergence: The evolutionary change in allele frequencies between reproductively isolated populations

Genetic markers: Any trait used as a marker of genetic variation within and among individuals and taxa. Traits used include phenotypic traits (e.g. number of dorsal rays), protein products (e.g. allozymes), and segments of the DNA. One might use a particular genetic marker as a diagnostic trait, as a tool for management, as an aid to systematic analyses, or in a huge variety of ways in basic evolutionary biology. Different genetic markers (e.g., microsatellites, mtDNA, allozymes, RAPD's) have different scopes (fine-grained vs. coarse-grained analyses), and different advantages and disadvantages (e.g., specificity, cost, ease of analytical interpretation of the resulting data).

Genetic (population) structuring: Occurrence of multiple populations, subdivided in some way, instead of a single population.

Genetic variability: The tendency of individual genetic characteristics in a population to vary from one another.

Genotype: The genetic makeup or the complement of alleles present in a particular individual's genome that will give rise to the individual's phenotype.

Genotype frequency: The percentage of individuals in a population that possess a specific genotype.

Haplogroup: It is a group of similar haplotypes that share a common ancestor having the same single nucleotide polymorphism (SNP) mutation in all haplotypes.

Haplotype: A particular form of a gene or DNA sequence.

Heterozygous: A genotype consisting of two different alleles of a gene for a particular trait (e.g. Aa). Individuals who are heterozygous for a trait are called heterozygotes. (Antonym: homozygous).

Heterozygosity: An individual- or population-level parameter referred to the proportion of loci expected to be heterozygous in an individual (ranging from 0.0 to 1.0).

Hitchhiking (genetic): A process whereby a gene with neutral value achieves a high value, or even fixation, within a population because it is closely linked to a gene that is being selected for.

Intron: Part of a gene's DNA sequence that does not code for amino acids and is spliced out of mRNA before transcription occurs.

Intron length polymorphism: DNA polymorphism results from difference in length of intron sequences.

Island model (of migration): A model of migration in which a population is subdivided into a series of demes, of size N , that randomly exchange migrants at a given rate, m .

Isolation-by-distance (IBD): The case where genetic differentiation is greater the further individuals (or populations) are from each other because gene flow decreases as geographic distance increases.

Locus (pl. loci): A location of DNA at a particular place on a particular chromosome - often used for a 'gene' in the broad sense, meaning a stretch of DNA being analyzed for variability (e.g., a microsatellite locus).

MDS (multidimensional scaling): A statistical graphing technique used to represent genetic distances between samples in two or three dimensions, and thereby visualizing similarities and differences between different groups or samples.

ME (Minimum Evolution method): Clustering method for the creation of phylogenetic trees.

Microsatellite: A stretch of DNA that consists of a short tandem sequence up to generally five base pairs that is repeated multiple times. Also known as VNTRs, SSRs or STRs.

Mitochondrial DNA (mtDNA): The haploid DNA found in the mitochondrion. In general, it is transmitted maternally and provides both opportunities and problems for phylogenetic analysis. Sequencing of mtDNA is a widely used technique in systematics.

Mitotype: A mitochondrial haplotype.

ML (maximum likelihood): An analytical method that provides an explanation for a particular data set by maximizing the probability of observed data under an explicit model (REF)

Monomorphic: A locus that has only one allele in a population and is therefore lacking in genetic variation

Neighbor joining: Clustering method for the creation of phylogenetic trees.

Nei's genetic distance: see genetic distance

Neutral allele: An allele that is not under selection because it does not affect fitness.

Nuclear marker: A marker located on a chromosomes in the cell nucleus of eukaryotes.

Nucleotide diversity: A measure of genetic diversity that quantifies the mean sequence divergence among several haplotypes by factoring in both the frequencies and the pairwise divergences of different sequences.

Outlier loci: loci that may be under selection (or linked to loci under selection) that are detected because they fall outside the range of expected variation for a given summary statistics (eg extremely low or high F_{ST} compared to most 'neutral' loci in a sample).

Panmictic: A population that is randomly mating.

Panmixia: Absence of any differentiation among subpopulations mainly due to high levels of gene flow, creating effectively one single large population with no sub-structure.

PCA (principal component analysis): A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

PCoA (principal coordinates analysis): It is a method to explore and to visualize similarities or dissimilarities of data. It starts with dissimilarity matrix (= distance matrix) and assigns for each item a location in a low-dimensional space. Ordination technique similar to PCA, but PCA is used for similarities and PCoA for dissimilarities.

Polymerase chain reaction (PCR): Technique for amplifying nucleic acids in a thermal cycler. It involves the use of a set of primers (forward and reverse) that start off the reaction and finally yields many orders of magnitude more DNA of the target sequence than one started with. The resulting amplified DNA can then be visualized with stains or radioactive labeling, or sized with fluorescent markers in a sequencer.

Polymorphic: A locus that has multiple alleles in a population.

Polymorphism: The presence of more than one allele at a locus. Generally defined as having the most common allele at a frequency less than 95% or 99%.

Population: A group of conspecific individuals that are found in a more or less well defined geographical region and exhibit reproductive continuity from generation to generation.

Population genetics: The scientific field in genetics that studies allele and genotype frequencies in populations.

RAPD (Randomly Amplified Polymorphic DNA): A method of analysis where PCR amplification using two copies of an arbitrary oligonucleotide primer is used to create a multilocus fingerprint (i.e. band profile).

Restriction enzymes: Bacterially-derived enzymes that digest DNA sequences at specific locations; each restriction enzyme recognizes and cuts a particular sequence of 4-8 nucleotide

Reynold's distance: see genetic distances

RFLP (Restriction Fragment Length Polymorphism): Dominant molecular marker that generates multiple fragments of DNA by digesting an entire genome or a pre-selected stretch of DNA with one or more restriction enzymes. The resulting banding pattern will vary depending on the underlying DNA sequence, because these will determine the number of restriction sites that are found in each individual.

SAMOVA: A genetic program that implements an approach to define groups of populations that are geographically homogeneous and maximally differentiated from each other.

Sequencing: Molecular techniques for deducing the nucleotide composition of the DNA.

SNP (Single Nucleotide Polymorphism): A variation between two sequences of DNA that is caused by a single nucleotide substitution. It can be used as a marker to assess genetic variation within and among populations. Usually only two alleles exist for a SNP in a population.

SSRs: see microsatellite.

Subpopulations: groups within a population delineated by reduced levels of gene flow with other groups.

Transcribed regions: portions of the gene that are transcribed (process by which messenger RNA is synthesized from a DNA template resulting in the transfer of genetic information from the DNA molecule to the messenger RNA).

Transcription: The first step in gene expression, in which a messenger RNA molecule complementary to particular gene encoded in DNA is synthesized by enzymes called RNA polymerases. To produce a functional protein, transcription is followed by translation.

Unrooted distance tree: Unrooted trees illustrate the relatedness of the leaf nodes without making assumptions about ancestry. They do not require the ancestral root (node corresponding to the (usually imputed) most recent common ancestor of all the entities at the leaves of the tree) to be known or inferred. An unrooted tree specifies the relationships among species and does not define the evolutionary path.

UPGMA: Unweighted pair group method with arithmetic averages.

Variance: The average squared deviation that a set of measurements show from the arithmetic mean.

INDEX

EXECUTIVE SUMMARY	1
1. INTRODUCTION	1
2. PROTOCOL OF BIOLOGICAL SAMPLING CORDINATION	1
2.1 PREPARATORY PHASE	2
2.2 EXECUTIVE PHASE	4

EXECUTIVE SUMMARY

This Deliverable formalizes the Sampling Procedure set to coordinate all the phases of the samplings (biological material for genetic analyses and otoliths) of the Workpackages WP1 and WP2, from the selection of species/areas/sample sizes, to the collection of samples and finally to the shipping of samples for the analyses foreseen by the project.

The samplings are coordinated by Paolo Sartor (CIBM Livorno) and will be concluded within the month 12, December 2019, in order to have enough time for carrying out laboratory analyses, processing data and assessing results. The Sampling Procedure includes the creation of a Genetic Hub (responsible Alessia Cariani, CoNISM University of Bologna) for the biological samples for the genetic analyses and an Otolith Hub (responsible Pierluigi Carbonara, COISPA Bari) for the exchange of otoliths.

1. INTRODUCTION

Two of the main objectives of the MED_UNITS Specific Contract are the genetic characterization of the target species of the study (European hake, red mullet, deep water rose shrimp, red shrimp, blue and red shrimp and Norway lobster), as well as the otolith shape and microchemistry analyses of the European hake and red mullet.

Two specific Workpackages, WP1 and WP2, have been specifically dedicated to accomplish these two aspects. Samples (biological material for genetic analyses and otoliths) will be collected all around the Mediterranean GSAs, both from EU (partners of the Project) and non EU Countries.

In order to guarantee a successful and a standardized collection and shipping of samples, two specific Protocols have been set, one for the collection of biologic material for genetic analysis (Deliverable D 1.1), another for the collection of otoliths for shape and microchemistry analysis (Milestone M 2.1).

Moreover, to ensure the accomplishment of the sampling protocols and to guarantee the collection of samples according to the timetable of the project, avoiding possible risks and sampling failures, a specific Sampling Procedure has been outlined (Milestone M 0.2). The Sampling Procedure has been shared with MARE and EASME during the kick off meeting and it is included in the Inception Report of the MED-UNITS Project (Deliverable D 0.1)

The Sampling Procedure, as well as the Sampling protocols will be presented and discussed with all the Partners during the first plenary meeting of the project (Rome, March 28th 2019) and transmitted on time to non-EU Countries willing to participate in the study.

2. PROTOCOL OF BIOLOGICAL SAMPLING CORDINATION

The samplings (biological material for genetic analyses and otoliths) for WP1 and WP2 are coordinated by **Paolo Sartor** (CIBM) and will be concluded within the month 12 of the Project, December 2019, in order to have enough time for carrying out laboratory analyses, processing data and assessing results. In Table 2.1 the target species and the type of biological sampling are shown.

Tab. 2.1. MED-UNITS. Target species of the study and type of biological sampling.

European hake (<i>Merluccius merluccius</i>)	biological material for genetic analyses and for otolith analyses (shape and microchemistry)
---	--

Red mullet (<i>Mullus barbatus</i>)	biological material for genetic analyses and for otolith analyses (shape and microchemistry)
Deep water rose shrimp (<i>Parapenaeus longirostris</i>)	biological material for genetic analyses
Red shrimp (<i>Aristaeomorpha foliacea</i>)	biological material for genetic analyses
Blue and red shrimp (<i>Aristeus antennatus</i>)	biological material for genetic analyses
Norway lobster (<i>Nephrops norvegicus</i>)	biological material for genetic analyses

The sampling and laboratory activities planned for the genetic characterization of the target species (WP1) of the study include two steps (Figure 2.1): i. Pilot Study. ii. Full Study, based on the evaluation of the results of the Pilot Study (the sampling plan and the methodology will be, possibly, refined).

Also the activities planned for the study of otoliths (WP2) include two steps (Fig. 2.1): i. Shape analysis of the otoliths. ii. Microchemistry analysis of the otoliths.

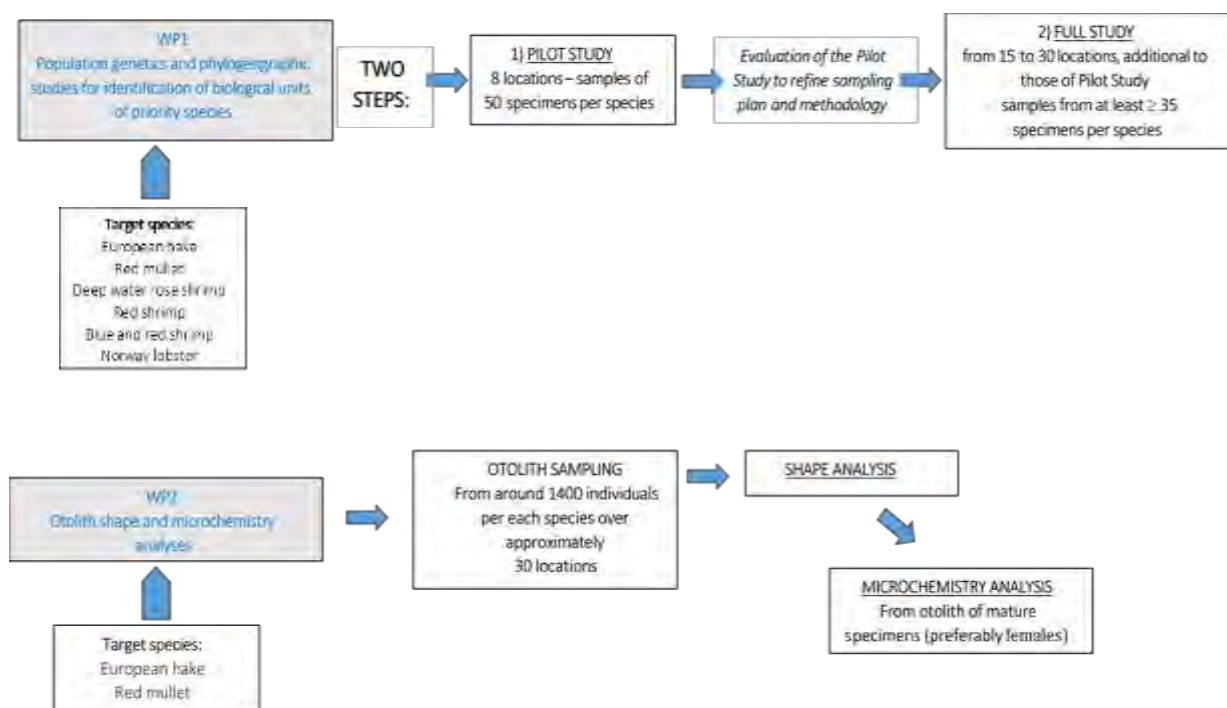


Fig. 2.1. Summary of the activities of WP1 and WP2 concerning genetic and otolith analyses.

2.1 PREPARATORY PHASE

In order to guarantee a successful and a standardized collection and shipping of samples, two specific Protocols have been set, one for the collection of biologic material for genetic analysis (Deliverable D 1.1), another for the collection of otoliths for shape and microchemistry analysis (Milestone M 2.1).

To ensure the accomplishment of the sampling protocols and to guarantee the collection of samples according to the timetable of the project (Table 2.2), avoiding possible risks and sampling failures, a specific Sampling Procedure has been outlined (Milestone M 0.2).

Tab. 2.2. Timetable of the activities of WP1 and WP2 concerning genetic and otolith analyses.

WP1 - POPULATION GENETICS AND PHYLOGEOGRAPHIC STUDIES FOR IDENTIFICATION OF BIOLOGICAL UNITS OF PRIORITY SPECIES	Months	WP2 - OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES
---	---------------	--

Definition of sampling protocols (D1.1) Definition of sampling sites for Pilot study (M1.2)	February 2019	Definition of sampling protocols (M2.1)
	March 2019	Analysis of literature for otolith shape and microchemistry (D.2.1)
Completion of samplings for Pilot Study (M0.7)	May 2019	
	July 2019	Protocol on shape analysis (M.2.3)
	September 2019	Completion of otolith samplings (M2.4)
Definition of sampling sites for Full study (M1.4) Report with the results of the pilot study (D1.4)	October 2019	
Completion of samplings for Full study (M0.11)	December 2019	
	February 2020	Protocol for otolith microchemistry analysis (M.2.5)
	March 2020	Completion of otolith shape analysis (D2.4)
	July 2020	Completion of otolith microchemistry analysis (D2.5)
Completion of the Full study (D1.5)	September 2020	

To organize the workflow for the collection of the genetic and otoliths samples from the Partners of the Projects and to deliver them to the Laboratories for the subsequent analyses, two HUBs have been created:

- **HUB for genetic samplings:** in charge to CoNISMa, University of Bologna, responsible **A. Cariani**;
- **HUB for otolith samplings:** in charge to COISPA of Bari, responsible **P. Carbonara**).

The sampling activities are defined during the first two months of the Contract, according to the following steps:

- Identification of the sampling locations for the Genetic Pilot Study and the Otolith Study; the samples collected in these locations will be used for both purposes (responsibles **R. Cannas/C. Tsigenopoulos** and **K. Mahè/I. Catalan**, Chairs of WP1 and WP2).
- Identification of the Partners/Institutes appointed for the collection of the samples for the Genetic Pilot Study and the Otolith Study, according to the sampling locations selected (responsible **P. Sartor**)¹.
- The same steps as above are followed to select the sampling locations and the Partners/Institutes appointed for the Genetic Full Study (according to the outcomes of the Genetic Pilot Study).
- The whole Mediterranean is the area covered by the biological sampling; however, the selection of the sampling locations will take into account the evidences of previous studies, as well as the geographical distribution of each target species (e.g. the presence/absence in each GSA).
- Sharing detailed sampling protocols for genetic and otolith sampling which include specific instruction on sampling collection, storage and shipping of samples as well as a standardised codification system to label the samples. This is coordinated between WP1 and WP2 (responsibles P. Sartor in cooperation with R. Cannas/C. Tsigenopoulos, K. Mahè/I. Catalan).
- Sharing of the the sampling protocols with project Partners by e-mail, web based tools and during the first Plenary Project Meeting (March 2019).

¹ The selection of the Partners/Institutes appointed for the sampling is made depending on the areas where each Partner has more working experience and logistic facilities for sampling. Knowledge, professional contacts and relationships among the Partners of MED-UNITS can be used to facilitate the communication, as well as the shipping of the samples.

- Testing the key aspects of conservation of biological samplings for genetics in the first three months of the Project (more details are provided in the protocol of Genetic Study, Deliverable D 1.1).

2.2 EXECUTIVE PHASE

Samplings are carried out according to the following procedure:

1. Following the information provided by the coordinator of the sampling activities in cooperation with the coordinators of WP1 and WP2, the responsible of the HUBs for genetic and otolith sampling contact the selected Partners, providing them instructions for sampling, storage, codification, and shipping of material (according to the sampling protocols of WP1 and WP2). In the case of the genetic samples, sampling tools already prepared (e.g. 2 ml tubes with O-ring screw cap; boxes with Racks with 80 apertures) are provided to each Partner. For otolith sampling eppendorf and labels are provided.
2. A courier specialized in the expedition of biological materials is used for shipping.
3. The selected Partners send the samples for genetic and otolith together with a sheet containing the basic biological data of each specimen sampled.
4. Samples, once received by the HUBs for genetic and otolith samplings, are sent to the Laboratories in charge for the analyses.

The biological samples are collected taking advantage of the activities of DCF (Data Collection Framework, EU Reg. 1004/2017) in force in all the EU Mediterranean Countries (e.g. **biological samplings of commercial fishery** and **MEDITS experimental trawl surveys**). The former mainly supplies the samples for the Genetic Pilot Study and partially for the otolith study, the latter is the main source of samples for the Full Genetic Study and for the Otolith Study.

Samples will be collected during the MEDITS survey for the Full Genetic Study on all the possible relevant sites, because the results from the genetic Pilot study will be not yet available.

In case of unforeseen problems in the carrying out of the survey in some areas the sampling will be anyhow accomplished by using the cooperation with the fishing industry and the colleagues of the specific areas.

As regards European hake and red mullet, it is suggested to use the same specimens for the genetic studies also for the otolith study (shape and microchemistry analysis).

The Partners are invited to start the sampling activity as soon as they receive the indications for sampling by the Coordinators. This will allow having enough time to collect samples, in particular in those areas where gathering some specimens is more difficult (e.g. adult females of red mullet, needed for the otolith microchemistry study).

As concerns the non-EU Countries, a liaison managed by DG MARE with GFCM, and in particular with the FAO Regional Projects (CopeMed, MedSudMed, AdriaMed, EastMed), is established in order to support the collection of samples also from these Countries. Samples are collected taking advantage of Fishery monitoring under DCRF (Data Collection Regulation Framework).

In order to ensure the accomplishment of the sampling protocols, to guarantee the collection of samples according to the timetable of the project, the sampling activities are constantly monitored by the sampling coordinator **P. Sartor**².

² During all the sampling phase the responsible of the sampling protocols (P. Sartor), the responsible of the sampling HUBs of WP1 and WP2 (A. Cariani and P. Carbonara, respectively), as well the Chairs of WP1 and WP2 (R. Cannas/C. Tsigenopoulos and K. Mahè/I. Catalan, respectively) will work in close contact, monitoring the progress of the sampling activities and regularly reporting to the MED-UNIT coordinator, M. T. Spedicato.

Each Partner/Institute appointed for the sampling indicates a **reference person** for any communication with the sampling coordinator.

- Specific deadlines are set for each Partner, to deliver the samples.
- Periodic communications (by email or similar ways) are realised between the sampling coordinator and the reference persons of each Partner to monitor the progress of the sampling;
- Possible problems (e.g. insufficient samples size, delay in sample delivery) are tackled with appropriate corrective measures (e.g. sampling from different sources).

The phases of the Procedure coordinating the samplings for WP1 and WP2 are summarised in Figure 2.2.

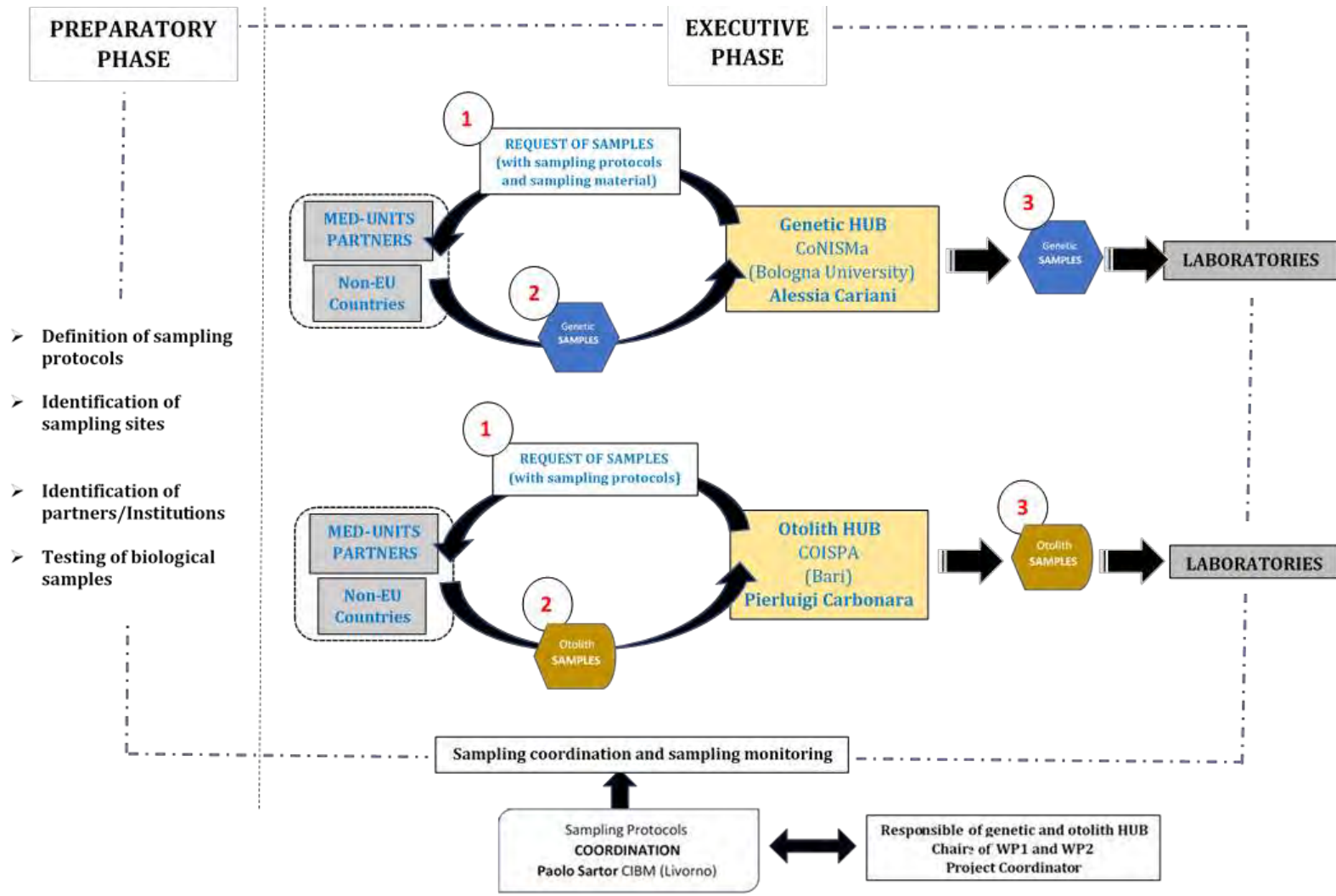


Fig 2.2. Phases of the Coordination of the samplings for WP1 and WP2 (above: Genetic Study; below: Otolith Study).

Below, the main aspects of the sampling procedure for WP1 (Genetic Analyses) and WP2 (Otolith Analyses) are summarised.

CoNISMa-UNIBO (A. Cariani) is in charge of the Genetic HUB for:

1. ensuring the effective and timely delivery of the needed sampling tools to the partners in charge of the tissue collection (specific partners identified in the Sampling Procedure set in this task);
2. verifying bilaterally with partners and other scientists belonging to non partner Institutes and willing to collaborate the use of appropriate materials for sampling and storing biological tissues, if direct shipping of sampling materials is not possible for law restrictions (i.e. non-EU Countries) or for time constrains;
3. establishing a service contract with a courier specialized in the expedition of biological materials;
4. gathering all tissue samples collected by partners identified in the Sampling Procedure set in this task, at specific time points during the project, using a courier specialized in the expedition of biological materials;
5. assessing each tissue samples, in terms of its quantity, quality, presence of the required labelling and reporting to Sampling Coordinator (CIBM) and WP1 Chairs (CoNISMa-UNICA and HCMR) of any discrepancies;
6. guarantee the proper storage of tissues samples in cooling facilities (freezers) and if needed, the replacement of storing solution (non-denatured analytical grade Ethanol 96%);
7. re-distribution of tissues samples to the pertinent laboratories for the subsequent genetic analyses.

A testing phase has been foreseen in the first three months of the project to verify the key aspects of conservation of biological samplings. For the genetic analysis a test will be conducted to identify the optimal tissue type to be sampled for each species. White muscle is the highly preferred tissue type, but fin clip (for fish species) as well as pereopods and pleiopods (for crustacean species) will be assessed in terms of DNA extraction outcomes (quantity and quality yield).

For the otolith sampling HUB, the coordination is in charge to COISPA (P. Carbonara). The activity concerns the following aspects:

1. organizing the collection of samples ensuring that all partners and other scientists belonging to non partner Institutes and willing to collaborate to the project will use appropriate materials for sampling and storing otoliths, following the same procedure in term of cleaning, conservation methods and size of fish (depending on the species, mature/adult females for red mullet and juvenile females for European hake);
2. verifying bilaterally with partners and other scientists belonging to non partner Institutes and willing to collaborate to the project, the preferred way to exchange the sampled materials;
3. closely follow the timing of the samples' collection;
4. coordinating the shipping activity with the laboratories in charge of the otolith analyses; the otolith HUB will re-distribute the samples between the involved laboratories (Otolith Shape analysis at IFREMER, Microchemistry at IMEDEA and Bologna University-CoNISMa-UNIBO) following the scheme of Figure 2.3.

The same specimens of hake and red mullet used for the genetic studies will be used for the otolith study (shape and microchemistry analysis). A subsample of the otolith used for the shape analyses will be used for microchemistry (not less the 25 left side otolith of mature/adult females for red mullet and juvenile females for European hake)

The collection of the otolith samples will start in conjunction with the collection of the genetic samples for the Genetic Pilot Study.

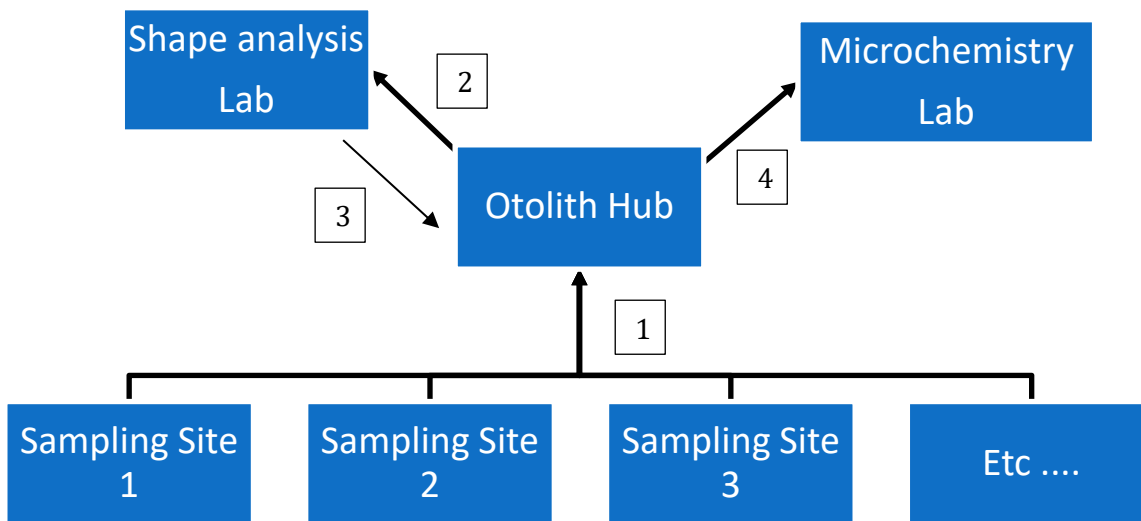


Fig. 2.3. The otolith samples will be sent to the Otolith Hub (1); once all samples will be received, they will be sent to the Shape analysis Lab (2). Here, after the acquisition of the images, the otoliths will be re-sent to the Hub (3) and a sub sample of the otoliths (left otolith only) will be delivered to the Microchemistry Lab (4).

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

Deliverable 1.3 – Report with the review bioinformatic methods and tools

LEAD: LORENZO ZANE (CONISMA) & COSTAS TSIGENOPOULOS (HCMR)

PARTICIPANTS: TEREZA MANOUSAKI (HCMR), RITA CANNAS, ALESSIA CARIANI, ILARIA ANNA MARIA MARINO (CONISMA)

Date: April 2nd, 2019

Contents

Executive Summary	3
Objective	4
Methodology	4
The final dataset	4
Bioinformatic methods and tools	6
Bioinformatic challenges and progress	7
Overall evaluation and proposed pipeline	9
References	11
Appendix 1	14
Appendix 2	33

Executive Summary

Deliverable 1.3 focuses on bioinformatics methods and tools used for the analysis and interpretation of genetic/genomic data in the context of stock assessment and identification of populations of marine organisms. Through a systematic literature review, we aim to identify bioinformatic methods used to infer stock boundaries, connectivity patterns and population genetic structure of the target species. Each method/tool is carefully examined for its peculiarities, appropriateness, robustness and accuracy in stock identification, considering, inter alia, the necessary equipment and expertise, the molecular markers needed, the costs, the annual/seasonal replication requirements, the effectiveness, the associated risks, the transferability, the user-friendliness of the methods. This information represents the background for the new genetic analyses performed in Tasks 1.4 and Task 1.5.

Our work builds up on a recently published review (Cuellar-Pinzon et al. 2016), which focused on identifying the kind and trend of use of genetic markers in genetic fisheries and extensively considered all the papers published on the topic in the period 2004- 2014; the authors identified a switch toward NGS (Next Generation Sequencing) derived markers in fishery research starting from 2011, accompanied by a reduction in the use of “classical” genetic markers such as microsatellites and mitochondrial DNA-based markers.

Here, we use a similar search strategy to identify studies published from 2015 to early 2019 using genetic approaches to study marine populations and stocks and, for each study, we record the kind of marker used, the taxonomic group and the data analysis methods used. Once the available bioinformatic methods and tools have been identified, we examined them, highlighting their peculiarities, appropriateness, robustness and accuracy in stock identification. To this end, we refer to published literature, whenever available, and to our own expertise. Our effort confirms the trend toward the use of NGS methods, but highlight the fact that microsatellites and mtDNA are still the most commonly used markers in genetic fisheries studies. With regard to data analysis methodologies, these seem quite consolidated for microsatellites and mitochondrial DNA-based markers, but still require refinements for NGS data, for which critical improvements became recently available.

Based on NGS needs and solutions, we propose a pipeline for the analysis of the six focal species of MEDUNITS that will be highly reproducible, taking into account the computational cost of analysing the dataset.

Objective

Deliverable 1.3 reviews the bioinformatics methods and tools used for the analysis and interpretation of genetic/genomic data in the context of stock assessment and the identification of populations of marine organisms. Scientific literature databases, such as PubMed, ISI Web of Knowledge, Scopus, as well as grey literature from past and ongoing projects, and reports of working group (i.e., ICES, GFCM or STECF) are exhaustively scrutinized. Each method/tool is carefully examined for its peculiarities, appropriateness, robustness and accuracy in stock identification considering, inter alia, the necessary equipment and expertise, the molecular markers needed, the costs, the annual/seasonal replication requirements, the effectiveness, the associated risks, the transferability, the user-friendliness of the methods.

The review contributes to identify bioinformatic methods needed to infer stock boundaries, connectivity patterns and population genetic structure of the target species. This information represents the background for the new genetic analyses performed in Tasks 1.4 and Task 1.5.

Applicability, advantages and disadvantages of different methodologies and techniques, as well as their limitations and strengths are discussed. Our final aim is to identify consolidated methods and to highlight critical gaps in data analysis due to the specific challenges imposed by NGS techniques.

Methodology

This systematic review is based on peer-review papers, grey literature, and data from national and European research activities. In particular, we extend results obtained in a recently published review (Cuellar-Pinzon et al. 2016), which focused on identifying the kind and trend of use of genetic markers in genetic fisheries and extensively considered all the papers published on the topic in the period 2004- 2014, and identified a switch toward NGS (Next Generation Sequencing) derived markers in fishery research starting from 2011 accompanied by a reduction in the use of “classical” genetic markers such as microsatellites and mitochondrial DNA-based markers. Here, we use a similar search strategy to identify studies published from 2015 to early 2019 using genetic approaches to study marine populations and stocks. Specifically, for the selected period, we searched in PubMed, ISI Web of Knowledge and Scopus, all the papers containing in their title, abstract, and keywords the words “marine” or “fish” or “fisheries”, and the words “stock” and “population genetic structure”. We merged the results, we integrated the records with relevant grey literature and reports of working group, and we scrutinized all the identified studies for the bioinformatics methods and tools used, also taking trace of the kind of marker used, the taxonomic group and relevant ancillary information.

Once the available bioinformatic methods and tools have been identified, we examine them individually, highlighting their peculiarities, the appropriateness, robustness and accuracy in stock identification. To this end, we refer to published literature, whenever available, and to our own research experience, especially for evaluating aspects such as equipment needs, expertise, molecular markers (number), costs, annual/seasonal replication requirements, effectiveness, associated risks, transferability, and user-friendliness.

The final dataset

The search identified 323 studies (Appendix 1), 250 of which, after inspection, provided useful information on the available bioinformatic methods and tools for stock assessment and identification of populations of marine organisms (Figure 1, Appendix 2).

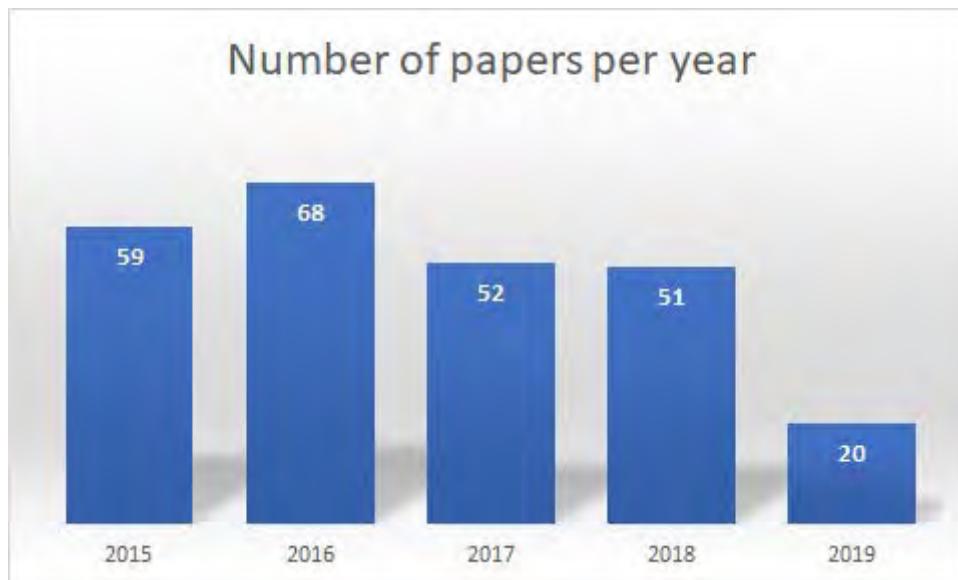


FIGURE 1 NUMBER OF RELEVANT STUDIES IDENTIFIED PER YEAR (2015-2019)

Most of the papers focused on fish species (Figure 2), followed by molluscs and crustaceans, with few studies focusing on mammals (cetaceans) and reptiles (turtles).

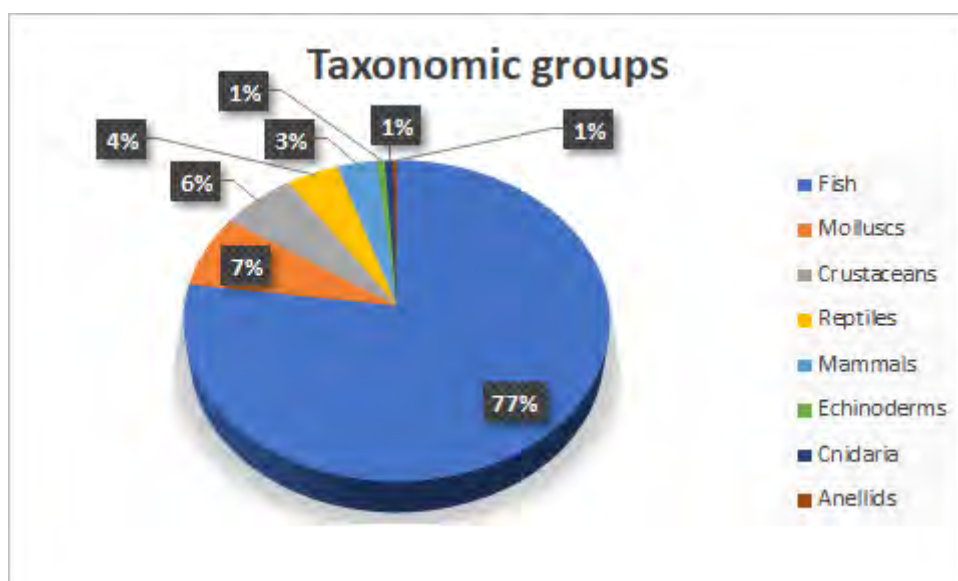


FIGURE 2 DISTRIBUTION OF THE TAXONOMIC GROUPS REPRESENTED IN THE FINAL DATASET (2015-2019)

With regard to the markers used (Figure 3), though 13% of the papers used SNPs, most of them were still based on microsatellites (47%) and, noticeably, on mtDNA markers (37%); the few other studies with different methodologies included allozymes, AFLP, RAPD, EPIC-PCR and nuclear DNA sequencing. Importantly, the percentage of studies using SNPs increases to 37.5% when considering only the studies published so far in 2019. In only 12% of the papers, more than one type of markers was used; in most of the cases, mtDNA and microsatellites were used together. These results indicate that the switch toward the use of NGS sequencing techniques envisaged by Cuellar-Pinzon and colleagues (2016) is ongoing and, while “traditional” markers are still the choice in many studies dealing with genetic analysis of fisheries, a clear trend exists. This puts a strong emphasis on the available bioinformatic tools and the open challenges for the analysis of SNPs data.

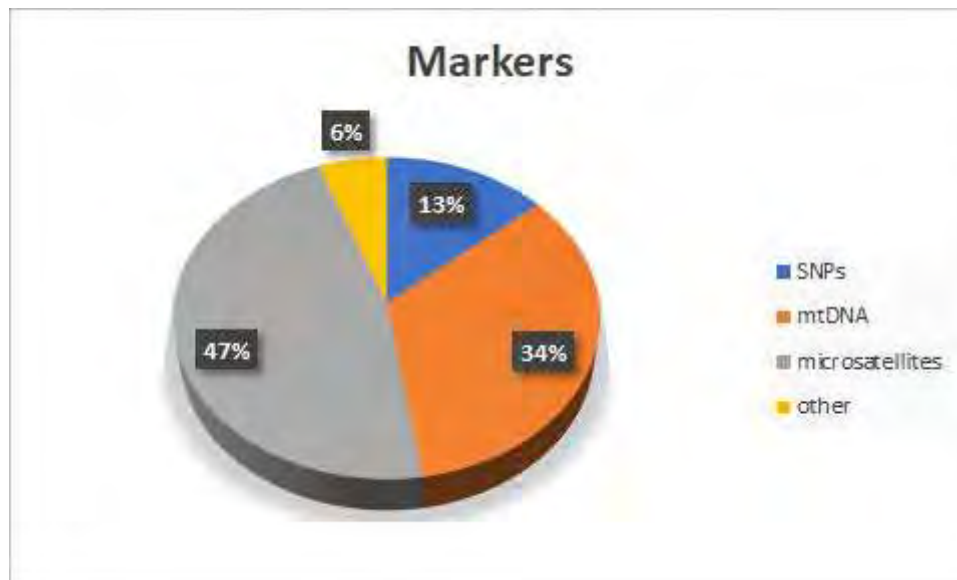


FIGURE 3. MARKERS USED IN THE FINAL DATASET (2015-2019)

Bioinformatic methods and tools

Most of the papers were characterized by the use of different bioinformatic methods, clearly depending on the markers used.

With regard to mtDNA markers, the data analysis strategy is clearly delineated, and relies mainly on AMOVA or pairwise F_{st} , in 82% of the cases performed with Arlequin (Excoffier and Lischer 2010); this is typically accompanied by a graphical representation of the relationships between population samples for which several alternatives emerge, including PCoA using Genalex (Peakall and Smouse 2006, 2012), and haplotypes clustering using TCS (Clement et al. 2000) or Network (Fluxus Technology Limited 2010). Among specific methods for stock analysis, the R package mixstock (Bolker 2012) seems to have gained some popularity.

With regard to the use of microsatellites, though many more alternatives exist for data analysis, still a preliminary comparison between population samples is performed in nearly all the cases. Here, in addition to AMOVA and F_{st} with Arlequin, other well-known software such as Genalex (Peakall and Smouse 2006, 2012), Fstat (Goudet 1995), Genetix (Belkiri et al. 1996-2004) and Genepop are widely used, this latter also to perform exact test of population differentiation. Clustering between individuals follows, in the majority of the cases performed using STRUCTURE (Pritchard et al 2000) and related packages such as STRUCTURE HARVESTER (Earl 2012), CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004). The use of STRUCTURE is often accompanied (but rarely substituted by) other, less model dependent, representations of relationships between individuals such as PCoA in Genalex, FCA in Genetix or DAPC in ADEGENET (Jombart 2008). Several studies also take advantage of the high polymorphism of microsatellites to infer migration patterns among population samples, a task that is often performed using GeneClass (Piry et al 2004), Bayesass (Wilson and Rannala 2003) or Migrate (Beerli and Felsenstein 1999). Finally, though most of the microsatellite studies are still based on few loci, outlier detection is often performed, in this case using Bayescan (Foll and Gaggiotti 2008; Foll 2012) and Lositan (Antao et al 2008).

With regard to the use of SNPs, the situation seems to more more variegated, likely due to the challenges imposed by much bigger dataset. Preliminary comparisons between populations samples is still often performed, in this case also with recent software/packages such as NGSdist (Vieira et al

2016) or *diveRsity* (Keenan et al 2013), but noticeably several studies skip this passage putting more emphasis on individual clustering. In this case, though STRUCTURE is still used, it is often replaced by DAPC, PCoA and MDS, even performed with specialized software such as NGStools (Fumagalli et al 2013, 2014) and Vegan (Oksanen et al 2017). As expected for genome-wide markers, the use of outlier detection test becomes very common, with still Bayescan and Lositan as preferred software. Finally, probably due to the complexity of SNPs dataset, inference of migration patterns or assignment test seems rather uncommon.

Bioinformatic challenges and progress

Next generation sequencing data can be considered 'big data', and as such the task of analysing them poses several computational challenges, that are gradually being dealt by software developers and bioinformaticians. A stark example of how software development evolves to deal with the needs required for analyzing modern sequence data is the RAD data analysis software STACKS (Catchen et al. 2011). STACKS represent the most commonly used pipeline for analysing RAD data since 2011 when it was first launched (cited > 1000 times). Recently, STACKS has been upgraded to accommodate multiple improvements that increase the analysis speed and accuracy. Further it accommodates all different flavours of RAD sequencing (e.g. ddRAD). In the previous versions (v1.x), the forward and reverse reads, that were part of the same DNA fragment were considered by the software as independent loci leading to calling SNPs that are actually linked. In the upgraded version (v2.x), STACKS defines the RAD loci and builds the catalogue by incorporating both reads to one locus and then calls SNPs for each locus. This has increased dramatically the accuracy of the downstream analyses for population genomics analyses, where linked loci need to be eliminated. It also produces outputs that can be directly used by other software.

Further, the parameter selection for analysing the data within STACKS could lead to dramatically different outcomes. The three main parameters that play key role in the analysis are:

- M: defines the distance allowed between stacks within the *ustacks* component
- m: defines the minimum depth coverage per stack within the *ustacks* component
- n: defines the distance allowed between catalogue loci within the *cstacks* component

To provide insights on how to select the right parameters out of the multivariate parameter space two methods have been developed and are proposed by the STACKS developers. The first proposes to iterate among the parameter values and select the set of parameters that maximizes the number of loci found in 80% of the studied individuals (Paris et al. 2017). The second uses technical replicates to identify the parameter values that minimise the error in building loci (Mastretta-Yanes et al. 2014). Downstream data filtering can be applied to ensure a high quality dataset, such as the filters implemented within the R package *stackr* (Gosselin & Bernatchez, 2016 <https://github.com/thierrygosselin/stackr>) and *grur* (Gosselin, 2017; <https://github.com/thierrygosselin/grur>).

Apart from the pipeline building and parameter selection, an inherent issue coupled with "big data" analysis, such as the NGS data, is the computational cost of the analysis. To deal with the computational needs that are increasing in an unprecedented rate, new approaches are being developed that involve the use of high-performance computing and process parallelization. For instance, one of the most commonly used pieces of software in population genetics analyses, STRUCTURE (Prichard et al. 2000), has been designed much earlier from the launch of modern sequencing technologies. STRUCTURE takes a Bayesian approach to estimate global ancestry by

sampling from the posterior distribution over global ancestry parameters using a Gibbs sampler that appropriately accounts for the conditional independence relationships between latent variables and model parameters. However, even well-designed sampling schemes need to generate a large number of posterior samples to resolve convergence and mixing issues and yield accurate estimates of ancestry proportions, greatly increasing the time complexity of inference for large genotype data sets. Due to the computational cost, new software that are more appropriate for large datasets, such as those coming from genome-wide SNP datasets, have been developed. These include fastSTRUCTURE (Raj et al. 2014) and fineSTRUCTURE (Lawson et al. 2012). fastSTRUCTURE represents a more flexible prior distribution over a subset of hidden parameters in the model and demonstrate that estimation of these hyperparameters using an empirical Bayesian framework improves the accuracy of global ancestry estimates when the underlying population structure is more difficult to resolve. They describe a scheme to accelerate computation of the optimal variational distributions and describe a set of scores to help evaluate the accuracy of the results and to help compare models involving different numbers of populations. The variational algorithms used in fastSTRUCTURE are almost two orders of magnitude faster than STRUCTURE. On the other hand, fineSTRUCTURE uses information about the relative positions of these mutations in the genome exploiting correlated variation patterns, at sets of closely positioned markers. Markers on the same chromosome are inherited together unless separated by recombination. Finally, three pieces of software have been developed that parallelise the use of STRUCTURE, enabling a faster implementation of the analysis, ParallelStructure (Besnier & Glover 2013), Structure_threader (Pina-Martins et al. 2017) and parastructure (Lagnel & Manousaki 2015; <https://github.com/jacqueslagnel/ParaStructure>).

Population genetic analyses can be now performed on R packages that have been built allowing the integration of different types of analysis in pipelines and the flexibility to use different algorithms on the same dataset. This not only enriches the possibilities that data analysts have to conduct solid analyses, but it also makes the data analyses process **repeatable** and thus easier to understand and **re-use**. Such software include ape (Paradis et al., 2004) and Adegenet for conducting genome-wide SNP analysis. Multiple other software have been built more specialized in analyzing di-allelic markers such as SNP data for conducting typical population genetic analysis.

The particular functions that are needed for a standard population genetic analysis include: genotypic and allelic frequencies, Hardy–Weinberg equilibrium, FST, analysis of molecular variance, haplotype network, mismatch distribution, Tajima's D and R2 tests for population stability, nucleotide diversity (π), the population parameter θ ($=4N\mu$), the site frequency spectrum (Paradis 2010). One of the first packages released for tackling such analyse in R is pegas (Paradis 2010).

Recently the package Hardy Weinberg has been released, which tests for Hardy-Weinberg equilibrium on bi-allelic markers such as SNP data (Graffelman 2015; <https://cran.r-project.org/web/packages/HardyWeinberg/>). Apart from targeting SNP data it also includes options for power computation, which is the main target of modern bioinformatic software development.

Computational cost is the main issue when analyzing genome-wide data or SNP data from a large number of individuals. Several other R packages have been developed that conduct population genomic analyses using multiple cores enabling the analyses of 'big data'. Such examples are PopGenome (Pfeifer et al. 2014), a software built to conduct population genomics analysis incorporating strategies that parallelise the computation on multiple cores when available and Poppr (Kamvar et al. 2014), a tool for conducting genome-wide population genetics analyses incorporating the factor of clonality found in multiple organisms.

Overall evaluation and proposed pipeline

The different bioinformatic methods and tools, described in the previous paragraphs, have been examined individually based on published literature and our own research experience. The overall evaluation is reported in Table 1; the following aspects have been considered: equipment needs, expertise, molecular markers (number), costs, annual/seasonal replication requirements, effectiveness, associated risks, transferability, and user-friendliness.

Table 1 - Overall evaluation of methods (blue lines) and relative tools, the evaluation is based on the literature data and researchers' expertise

Method/tool	Equipment (software)§	expertise	N° markers	costs	replication requirements	effectiveness	risks	transferability	user-friendliness
mtDNA	*	*	*/**	*	*	*	*	****	****
AMOVA	Arlequin	*	*	*	*	*	*	****	***
pairwise FST	Arlequin	*	*	*	*	*	*	****	***
PCoA	Genalex	*	*	*	*	*	*	****	***
haplotype clustering	TCS/Network	*	**	*	*	*	*	**	**
genetic stock analysis	mixstock in R	**	**	*	*	*	*	**	**
microsatellites	**	**	**/**	**	**	**	**/**	*	**
AMOVA	Arlequin	*	**	**	*	**	**	****	***
pairwise FST	Arlequin	*	**	**	*	**	**	****	***
PCoA	Genalex, Adegenet	*	**	**	*	**	**	**	***
genetic stock analysis	mixstock in R	**	***	**	**	**	***	**	**
population differentiation	Genalex, Fstat, Genetix, Genepop	*	**	**	*	**	**	***	***
clustering analysis	Structure, Clumpp, Distruct	***	***	**	**	**	***	**	***
DAPC	Adegenet	*	***	**	**	**	***	**	***
migration patterns	GeneClass, Bayesass, Migrate	***	***	**	**	**	***	**	**
outlier detection	Lositan, Bayescan	*	***	**	**	*	*	**	**
SNPs	****	****	***/**	****	**	**/**	**/**	**/**	**/**
SNP identification and genotyping	STACKS	****	****	****	**	***	***	**	*
population differentiation	NGSdist, diveRsity	**	***	***	*	**	**	***	***
clustering analysis	Structure, fastSTRUCTURE, fineSTRUCTURE, ParallelStructure, Structure_threader, parastructure	***	****	****	**	***	***	**	*
DAPC	Adegenet	*	***	***	*	***	**	**	***
PCoA	Adegenet, NGStools, Vegan	*	***	***	*	**	**	**	***
MDS	NGStools, Vegan	*	***	***	*	**	**	**	***
outlier detection	Lositan, Bayescan	*	****	***	*	**	**	**	**
genome-wide analyses	PopGenome	***	****	****	**	***	***	**	*

§ = see main text for references.

* = low; ** = medium; *** = high; **** = very high.

With regard to mtDNA markers, the overall need for equipment and expertise is low. This low requirement extends to the number of markers (here, more appropriately to the total sequence length considered, since the entire mitochondrial genome is a single marker by definition), with literature analysis showing that in most of the cases the published mtDNA papers are based on a limited sequence length, typically 1 or 2 mtDNA regions sequenced for a total of 500-1500 bp); this result can be justified by the relatively small increase of information that can be obtained using longer sequences, respect to the laboratory effort needed to obtain them. Overall, costs and risks associated with mtDNA analyses are low and their transferability, and user-friendliness is high. However, these markers provide only a very coarse grain picture of differentiation, resulting in little additional information obtained in most of the cases trough temporal replications, but also in a low effectiveness in detecting subtly differentiated stocks. These considerations extend to most of the methods and software used for the analysis of mtDNA markers: in this case there are no special costs due to particular requests of computational power and most of the programs are quite easy and friendly to handle.

With regard to microsatellite markers, the need in terms of expertise and equipment become more stringent, specially in the case when a high number of markers is advisable for specific analyses such as genetic stock analysis, individual clustering or detection of migration patterns. The costs and the risks associated with this kind of markers can be considered from medium to high, depending on the fact that primers for amplification are already available and tested or that de novo isolation and set up of experimental conditions is needed. Microsatellite generally provide an excellent picture of population differentiation and, provided that a high number of properly optimized loci is used, they have a very high power to detect differences, meaning also that temporal replication of sampling is desirable. Despite this, many of the microsatellites studies reviewed are still based on a small number of loci (Appendix 2) and only about 20% of them use more than 15 loci. This result can be explained by the increased effort needed to isolate and standardize a high number of loci and also to the increasing costs of the analysis which still limit high resolution analysis to special cases. As a further drawback of microsatellite use, their transferability and user-friendliness can be considered poor, specially for the complexity of genotype scoring in sub-optimal cases. With regard to methods and software used for the analysis of microsatellites, these do not imply particular costs in terms of computational power and expertise, except for some analysis linked to individual clustering and inference of migration patterns.

With regard to SNPs markers, the need in terms of expertise and equipment become more important, particularly when considering the first step of pre-processing raw data required by many RAD protocols. Conversely, the number of markers obtained is generally high and, importantly, in many cases it can be scaled up by simple protocol optimization (i.e use of different combination of restriction enzymes) or increasing sequencing effort. Cost and risks of these approaches can be considered high, specially when applied to species for which no pilot analyses were performed. On the other hand, they provide the better power to detect subtle differences and even to investigate putative effect of selection; again this high power makes highly desirable the temporal replication of sampling. The complexity of data analysis implies an investment in computational resources, both for raw data pre-processing than for downstream analysis. Many of the software used are not particularly user friendly and in many of the cases the complexity of the starting dataset and the continuous update of the programs used for the analysis determines some difficulties in transferability and reproducibility of

results. For this reason, considering also the time requested by many of the analyses, it is important to set up a clear strategy for data analysis, specially in the case of research efforts involving different laboratories for which comparative analysis of results is of interest.

Following the aforementioned evaluation, needs and solutions, we aim to build a pipeline for the analysis of the six focal species of MEDUNITS that will be highly reproducible incorporating multiple of the reviewed software (Figure 4) taking into account the computational cost of analysing the dataset.

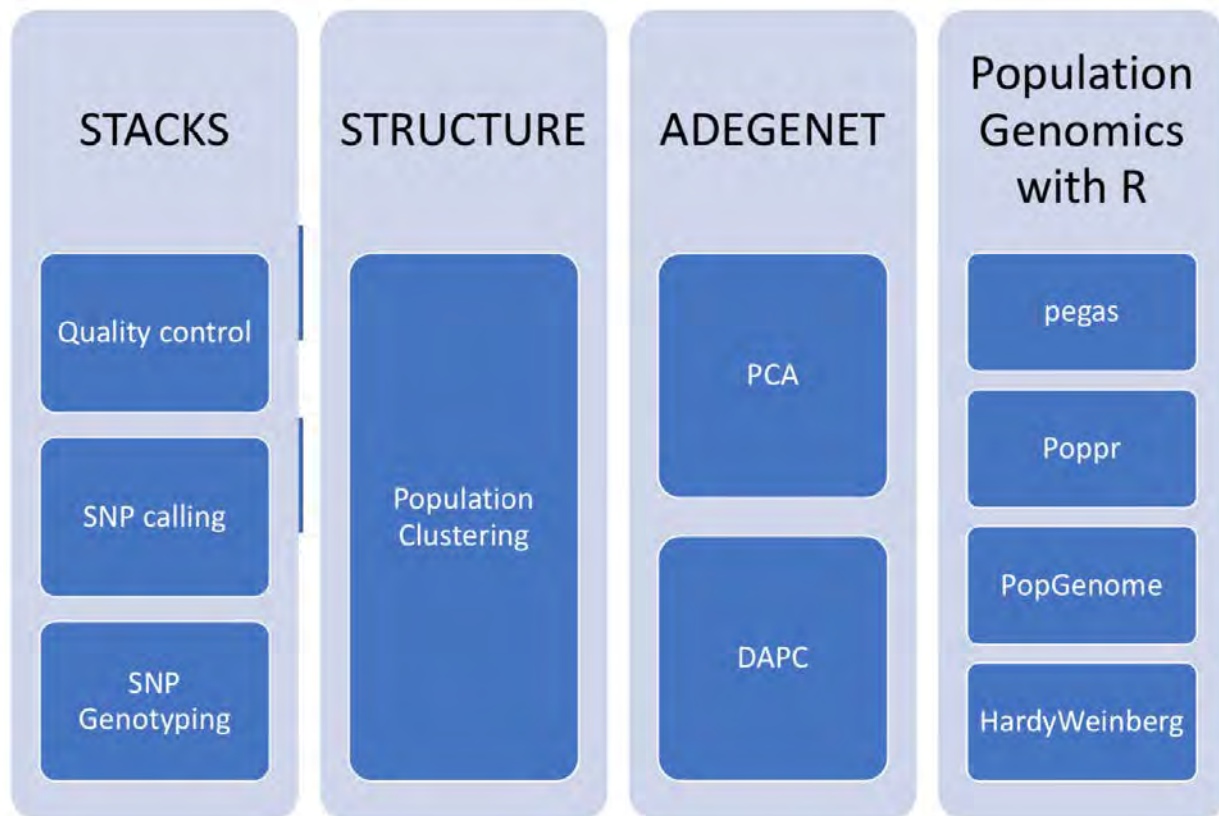


FIGURE 4. THE DESIGNED PIPELINE COMPONENTS FOR ANALYSING THE RAW DATA AND ASSESSING THE POPULATION STRUCTURE OF MEDUNITS SPECIES.

References

- Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008). LOSITAN: a workbench to detect molecular adaptation based on a F_{st} -outlier method. *BMC Bioinformatics*, 9: 323.
- Beerli P, Felsenstein J (1999). Maximum-likelihood estimation of migration rates and effective populations numbers in two populations using a coalescent approach. *Genetics*, 152: 763–773.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Besnier F, Glover KA (2013). ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One*, 8: e70651.

- Bolker BM (2012). Mixed stock analysis in R: getting started with the Mixedstock package. Available online at: <http://ftp.cs.pu.edu.tw/network/CRAN/web/packages/mixstock/vignettes/mixstock.pdf>.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)*, 1: 171-182. doi: 10.1534/g3.111.000240. Epub 2011 Aug 1.
- Clement M, Posada D, Crandall K (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9: 1657-1660.
- Cuéllar-Pinzón J, Presa P, Hawkins SJ, Pita A (2016). Genetic markers in marine fisheries: Types, tasks and trends. *Fisheries Research*, 173: 194-205.
- Earl D A (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4: 359-361. 10.1007/s12686-011-9548-7.
- Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10: 564-567.
- Fluxus Technology Limited (2010). Network version 4.6.1.4: A software for population genetics data analysis, fluxus-engineering.com.
- Foll M (2012). BayeScan v2.1 user manual. *Ecology*, 20: 1450-1462.
- Foll M, Gaggiotti O (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective *Genetics*, 180: 977-993, 10.1534/genetics.108.092221.
- Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, Nielsen R (2013). Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, 195: 979-992. <https://doi.org/10.1534/genetics.113.154740>.
- Fumagalli M, Vieira F, Linderoth T, Nielsen R (2014). ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, 30: 1486-1487. <https://doi.org/10.1093/bioinformatics/btu041>.
- Gosselin T, Bernatchez L (2016). stackr: GBS/RAD data exploration, manipulation and visualization using R. R package version... Retrieved from <https://github.com/thierrygosselinstackr>. doi.zenodo.
- Gosselin T (2017). thierrygosselin/grur: grur package first release (Version v.0.0.1). Zenodo. <http://doi.org/10.5281/zenodo.496176>
- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, 86: 485-486.
- Graffelman J (2015). Exploring Diallelic Genetic Markers: The HardyWeinberg Package. *Journal of Statistical Software*, 64: 1-23.
- Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure *Bioinformatics*, 23: 1801-1806.
- Jombart T (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24: 1403-1405.
- Kamvar ZN, Tabima JF, Grünwald NJ (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2: e281.

- Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013). *diveRsity*: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4: 782–788.
- Lawson DJ, Hellenthal G, Myers S, Falush D (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, 8: e1002453.
- Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15: 28-41.
- Oksanen J, Blanchet GF, Friendly M, Kindt R, Legendre P, McGinn D et al. (2017). *vegan*: community ecology package (Version 2.4-3). <https://CRAN.R-project.org/package=vegan>.
- Paradis E, Claude J, Strimmer K (2004). *APE: Analyses of Phylogenetics and Evolution in R language*. *Bioinformatics*, 20: 289-290.
- Paradis E (2010). *pegas*: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26: 419-420.
- Paris JR, Stevens JR, Catchen JM (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, 8: 1360-1373.
- Peakall R, Smouse PE (2006). *GENALEX 6: genetic analysis in Excel*. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6: 288-295.
- Peakall R, Smouse PE (2012) *GenAlEx 6.5: genetic analysis in Excel*. Population genetic software for teaching and research-an update. *Bioinformatics*, 28: 2537-2539.
- Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ (2014). *PopGenome*: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31: 1929-1936.
- Pina-Martins F, Silva DN, Fino J, Paulo OS (2017). *Structure_threader*: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. *Molecular Ecology Resources*, 17: e268-e274.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004). *GENECLASS2*: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, 95: 536–539.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.
- Raj A, Stephens M, Pritchard JK (2014). *fastSTRUCTURE*: variational inference of population structure in large SNP data sets. *Genetics*, 197: 573-589.
- Rosenberg NA (2004). *Distruct*: a program for the graphical display of population structure *Molecular Ecology Notes*, 4: 137-138.
- Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M (2016). Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biological Journal of the Linnean Society*, 117: 139-149. <https://doi.org/10.1111/bij.12511>
- Wilson GA, Rannala B (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163: 1177-1191.

Appendix 1

This appendix reports the full list of the 323 papers scrutinized in this deliverable, in alphabetic order.

1. Abou-Gabal A.A.A. Abbas E.M. Mohamed H.M. El-Baramawi N. Khaled A.A. El Deeb S.I. (2018). Molecular identification of grey mullet species in the mediterranean sea of egypt. *Egyptian Journal of Aquatic Biology and Fisheries* <http://doi.org/10.21608/ejabf.2018.13263>
2. Adamson E.A.S. Hurwood D.A. (2015). Molecular ecology and stock identification. *Freshwater Fisheries Ecology* <http://doi.org/10.1002/9781118394380.ch58>
3. Aguila R.D. Perez S.K.L. Catacutan B.J.N. Lopez G.V. Barut N.C. Santos M.D. (2015). Distinct yellowfin tuna (*Thunnus albacares*) stocks detected in western and central pacific ocean (WCPO) using DNA microsatellites. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0138292>
4. Al-Breiki R.D. Kjeldsen S.R. Afzal H. Al Hinai M.S. Zenger K.R. Jerry D.R. Al-Abri M.A. Delghandi M. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (*Panulirus homarus*) along the Omani coastline. *BMC Genomics* <http://doi.org/10.1186/s12864-018-5044-8>
5. Amano Y. Shiao J.-C. Ishimura T. Yokouchi K. Shirai K. (2015). Otolith geochemical analysis for stock discrimination and migratory ecology of tunas. *Biology and Ecology of Bluefin Tuna* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045144461&partnerID=40&md5=3daf0926c3fb7e5a9c27d107cb426e1b>
6. Anderson J.D. O'Leary S.J. Cooper P.T. (2019). Population Structure of Atlantic Croakers from the Gulf of Mexico: Evaluating a Single-Stock Hypothesis Using a Genomic Approach. *Marine and Coastal Fisheries* <http://doi.org/10.1002/mcf2.10055>
7. Anderson J.D. Torres A.D. (2016). Genetic Variability and Population Structure of Gulf Menhaden Compared with Yellowfin Menhaden. *Marine and Coastal Fisheries* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84987811341&doi=10.1080%2f19425120.2016.1180334&partnerID=40&md5=d7a316e874f6eecd4405c79437346c37>
8. André C. Svedäng H. Knutsen H. Dahle G. Jonsson P. Ring A.-K. Sköld M. Jorde P.E. (2016). Population structure in Atlantic cod in the eastern North Sea-Skagerrak-Kattegat: Early life stage dispersal and adult migration Ecology. *BMC Research Notes* <http://doi.org/10.1186/s13104-016-1878-9>
9. Antoni L. Saillant E. (2017). Spatial connectivity in an adult-sedentary reef fish with extended pelagic larval phase. *Molecular Ecology* <http://doi.org/10.1111/mec.14263>
10. Antoniou A. Kasapidis P. Kotoulas G. Mylonas C.C. Magoulas A. (2017). Genetic diversity of Atlantic Bluefin tuna in the Mediterranean Sea: Insights from genome-wide SNPs and microsatellites. *Journal of Biological Research (Greece)* <http://doi.org/10.1186/s40709-017-0062-2>
11. Bae S.E. Kim H. Choi S.-G. Kim J.-K. (2018). Evidence of shallow mitochondrial divergence in the slender armorhead, *Pentaceros wheeleri* (Pisces, Pentacerotidae) from the Emperor Seamount Chain. *Mitochondrial DNA Part A: DNA Mapping, Sequencing and Analysis* <http://doi.org/10.1080/24701394.2018.1424842>
12. Baetscher D.S. Hasselman D.J. Reid K. Palkovacs E.P. Garza J.C. (2017). Discovery and characterization of single nucleotide polymorphisms in two anadromous alosine fishes of conservation concern. *Ecology and Evolution* <http://doi.org/10.1002/ece3.3215>
13. Baeza J.A. Behringer D.C. (2017). Integrative taxonomy of the ornamental 'peppermint' shrimp public market and population genetics of *Lysmata boggessi*. *PeerJ* <http://doi.org/10.7717/peerj.3786>
14. Bailleul D. Mackenzie A. Sacchi O. Poisson F. Bierne N. Arnaud-Haond S. (2018). Large-scale genetic panmixia in the blue shark (*Prionace glauca*): A single worldwide population. *Evolutionary Applications* <http://doi.org/10.1111/eva.12591>
15. Balazik M.T. Farrae D.J. Darden T.L. Garman G.C. (2017). Genetic differentiation of spring-spawning and fall-spawning male Atlantic sturgeon in the James River. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0179661>
16. Barton D.P. Taillebois L. Taylor J. Crook D.A. Saunders T. Hearnden M. Greig A. Welch D.J. Newman S.J. Travers M.J. Saunders R.J. Errity C. Maher S. Dudgeon C. Ovenden J. (2018). Stock structure of *Lethrinus laticaudis* (Lethrinidae) across northern Australia determined using genetics. *Marine and Freshwater Research* <http://doi.org/10.1071/MF17087>

17. Bayona-Vásquez N.J. Díaz-Jaimes P. Uribe-Alcocer M. (2015). Isolation and characterization of microsatellite loci in the common dolphinfish *Coryphaena hippurus* (Perciformes: Coryphaenidae) from next generation sequencing and cross amplification in pompano dolphinfish *Coryphaena equiselis*. Conservation Genetics Resources <http://doi.org/10.1007/s12686-014-0372-8>
18. Bekkevold D. Helyar S.J. Limborg M.T. Nielsen E.E. Hemmer-Hansen J. Clausen L.A.W. Carvalho G.R. (2015). Gene-associated markers can assign origin in a weakly structured fish. ICES Journal of Marine Science <http://doi.org/10.1093/icesjms/fsu247>
19. Bell I. Jensen M.P. (2018). Multinational genetic connectivity identified in western Pacific hawksbill turtles. Wildlife Research <http://doi.org/10.1071/WR17089>
20. Ben Khadher S. Agnèse J.-F. Milla S. Teletchea F. Fontaine P. (2015). Patterns of genetic structure of Eurasian perch (*Perca fluviatilis* L.) in Lake Geneva at the end of the spawning season. Journal of Great Lakes Research <http://doi.org/10.1016/j.jglr.2015.04.006>
21. Bennett R.H. Reid K. Gouws G. Bloomer P. Cowley P.D. (2017). Genetic stock structure of white steenbras *Lithognathus lithognathus* (Cuvier). African Journal of Marine Science <http://doi.org/10.2989/1814232X.2017.1303398>
22. Bertram A. Dias P.J. Lukehurst S. Kennington W.J. Fairclough D. Norriss J. Jackson G. (2015). Isolation and characterisation of 16 polymorphic microsatellite loci for bight redfish. Australian Journal of Zoology <http://doi.org/10.1071/ZO15026>
23. Blower D.C. Corley S. Hereward J.P. Riginos C.R. Ovenden J.R. (2015). Characterisation and cross-amplification of 21 novel microsatellite loci for the dusky shark. Conservation Genetics Resources <http://doi.org/10.1007/s12686-015-0499-2>
24. Bohling J. Haffray P. Berrebi P. (2016). Genetic diversity and population structure of domestic brown trout (*Salmo trutta*) in France. Aquaculture <http://doi.org/10.1016/j.aquaculture.2016.04.013>
25. Bozano M.S. Mariani S. Barratt C.D. Sacchi C. Boufana B. Coscia I. (2015). Spatio-temporal variability in the population structure in northeast atlantic stocks of horse mackerel (*trachurus trachurus*). Biology and Environment <http://doi.org/10.3318/BIOE.2015.20>
26. Buchholz-Sørensen M. Vella A. (2016). Population Structure, genetic diversity, effective population size, Demographic History and Regional Connectivity Patterns of the Endangered Dusky Grouper. PLoS one <http://doi.org/10.1371/journal.pone.0159864>
27. Campbell N.R. Kamphaus C. Murdoch K. Narum S.R. (2017). Patterns of genomic variation in Coho salmon following reintroduction to the interior Columbia River. Ecology and Evolution <http://doi.org/10.1002/ece3.3492>
28. Candy J.R. Campbell N.R. Grinnell M.H. Beacham T.D. Larson W.A. Narum S.R. (2015). Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*). Molecular Ecology Resources <http://doi.org/10.1111/1755-0998.12400>
29. Carr-Harris C. Gottesfeld A.S. Moore J.W. (2015). Juvenile Salmon usage of the Skeena River estuary. PLoS ONE <http://doi.org/10.1371/journal.pone.0118988>
30. Catanese G. Montes I. Iriondo M. Estonba A. Iudicone D. Procaccini G. (2016). High resolution SNPs selection in *Engraulis encrasicolus* through Taqman OpenArray. Fisheries Research <http://doi.org/10.1016/j.fishres.2016.01.014>
31. Cauwelier E. Stewart D.C. Millar C.P. Gilbey J. Middlemas S.J. (2018). Across rather than between river genetic structure in Atlantic salmon *Salmo salar* in north-east Scotland. Journal of Fish Biology <http://doi.org/10.1111/jfb.13542>
32. Chabot C.L. (2015). Microsatellite loci confirm a lack of population connectivity among globally distributed populations of the tope shark *Galeorhinus galeus* (Triakidae). Journal of Fish Biology <http://doi.org/10.1111/jfb.12727>
33. Chahdi Ouazzani K. Benazzou T. Charouki N. Bonhomme F. Chlaida M. (2017). Genetic differentiation of European anchovy (*Engraulis encrasicolus*) along the Moroccan coast reveals a phylogeographic break around the 25th parallel North. Marine Biology Research <http://doi.org/10.1080/17451000.2016.1257811>
34. Chan J. Li W. Hu X. Liu Y. Xu Q. (2016). Genetic diversity and population structure analysis of Qinghai-Tibetan plateau schizothoracine fish (*Gymnocypris dobula*) based on mtDNA D-loop sequences. Biochemical Systematics and Ecology <http://doi.org/10.1016/j.bse.2016.09.004>
35. Chandler L.M. Walters L.J. Sharp W.C. Hoffman E.A. (2017). Genetic structure of natural & broodstock populations of the long-spined sea urchin. Bulletin of Marine Science <http://doi.org/10.5343/bms.216.1101>

36. Chang C. Han H.-Z. Wang T.-T. Ma H.-T. Liu Y. Wang F. Zhang M.-L. Wang L.-Y. Jiang H.-B. (2017). Microsatellite Marker Isolation and Genetic Structure Analysis for Five Populations of *Urechis unicinctus* [刺螵 (*Urechis unicinctus*) 微卫星标记及5个地理种群结构分析]. *Oceanologia et Limnologia Sinica* <http://doi.org/10.11693/hyhz20161100238>
37. Chebib J. Renaut S. Bernatchez L. Rogers S.M. (2016). Genetic structure and within-generation genome scan analysis of fisheries-induced evolution in a Lake Whitefish (*Coregonus clupeaformis*) population. *Conservation Genetics* <http://doi.org/10.1007/s10592-015-0797-y>
38. Chen H. Chang C.-H. Sun C.-L. Shao K.-T. Yeh S.-Z. Di Nardo G. (2016). Population structure of Blue marlin. *Zoological Studies* <http://doi.org/10.6620/ZS.2016.55-33>
39. Cheng J. Yanagimoto T. Song N. Gao T.-X. (2015). Population genetic structure of chub mackerel *Scomber japonicus* in the Northwestern Pacific inferred from microsatellite analysis. *Molecular Biology Reports* <http://doi.org/10.1007/s11033-014-3777-2>
40. Chou C.-E. Liao T.-Y. Chang H.-W. Chang S.-K. (2015). Population structure of *hirundichthys oxycephalus* in the northwestern pacific inferred from mitochondrial cytochrome oxidase I gene. *Zoological Studies* <http://doi.org/10.1186/s40555-014-0085-4>
41. Clemens B.J. (2015). A Survey of Steelhead Age and Iteroparity Rates from a Volunteer Angler Program in the Willamette River Basin. *North American Journal of Fisheries Management* <http://doi.org/10.1080/02755947.2015.1079572>
42. Coimbra M.R.M. Lima A.P.S. Oliveira K.K.C. Severi W. (2017). Microsatellite assessment of the genetic diversity in indigenous populations of curimba (*Prochilodus argenteus*) in the São Francisco river (Brazil). *Conservation Genetics* <http://doi.org/10.1007/s10592-017-0947-5>
43. Converse P.E. Kuchta S.R. Hauswaldt J.S. Roosenburg W.M. (2017). Turtle soup. *PloS one* <http://doi.org/10.1371/journal.pone.0181898>
44. Cuéllar-Pinzón J. Presa P. Hawkins S.J. Pita A. (2016). Genetic markers in marine fisheries: Types. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.10.019>
45. Cunningham C.J. Branch T.A. Dann T.H. Smith M. Seeb J.E. Seeb L.W. Hilborn R. (2018). A general model for salmon run reconstruction that accounts for interception and differences in availability to harvest. *Canadian Journal of Fisheries and Aquatic Sciences* <http://doi.org/10.1139/cjfas-2016-0360>
46. Cypriano-Souza A.L. Engel M.H. Caballero S. Olavarría C. Flórez-González L. Capella J. Steel D. Sremba A. Aguayo A. Thiele D. Baker C.S. Bonatto S.L. (2017). Genetic differentiation between humpback whales (*Megaptera novaeangliae*) from Atlantic and Pacific breeding grounds of South America. *Marine Mammal Science* <http://doi.org/10.1111/mms.12378>
47. Da Silva Ferrette B.L. Mendonça F.F. Coelho R. De Oliveira P.G.V. Hazin F.H.V. Romanov E.V. Oliveira C. Santos M.N. Foresti F. (2015). High connectivity of the crocodile shark between the Atlantic and southwest Indian oceans: Highlights for conservation. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0117549>
48. Da Silva R. Sampaio I. Schneider H. Gomes G. (2016). Lack of spatial subdivision for the snapper *lutjanus purpureus* (lutjanidae-perciformes) from southwest atlantic based on multi-locus analyses. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0161617>
49. Da Silva R. Veneza I. Sampaio I. Araripe J. Schneider H. Gomes G. (2015). High levels of genetic connectivity among populations of yellowtail snapper. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0122173>
50. Dahle G. Quintela M. Johansen T. Westgaard J.-I. Besnier F. Aglen A. Jørstad K.E. Glover K.A. (2018). Analysis of coastal cod (*Gadus morhua* L.) sampled on spawning sites reveals a genetic gradient throughout Norway's coastline. *BMC Genetics* <http://doi.org/10.1186/s12863-018-0625-8>
51. Damasceno J.S. Siccha-Ramirez R. Morales M.J.A. Oliveira C. Torres R.A. Costa E.N. Silva-Oliveira G.C. Vallinoto M. Machado L.F. Tosta V.C. Farro A.P.C. Hostim-Silva M. (2015). Mitochondrial DNA evidences reflect an incipient population structure in Atlantic goliath grouper (*epinephelus itajara*). *Scientia Marina* <http://doi.org/10.3989/scimar.04203.20A>
52. Darden T.L. Robinson J.D. Strand A.E. Denson M.R. (2017). Forecasting the Genetic Impacts of Net Pen Failures on Gulf of Mexico Cobia Populations Using Individual-based Model Simulations. *Journal of the World Aquaculture Society* <http://doi.org/10.1111/jwas.12333>
53. Daugherty D.J. Pangle K.L. Karel W. Baker F. Robertson C.R. Buckmeier D.L. Smith N.G. Boyd N. (2017). Population structure of alligator gar in a Gulf coast river: Insights from otolith microchemistry and genetic analyses. *North American Journal of Fisheries Management* <http://doi.org/10.1080/02755947.2016.1269029>

54. De Keyzer E.L.R. De Corte Z. Van Steenberge M. Raeymaekers J.A.M. Calboli F.C.F. Kmentová N. N'Sibula Mulimbwa T. Virgilio M. Vangestel C. Mulungula P.M. Volckaert F.A.M. Vanhove M.P.M. (2019). First genomic study on Lake Tanganyika sprat *Stolothrissa tanganicae*: A lack of population structure calls for integrated management of this important fisheries target species. *BMC Evolutionary Biology* <http://doi.org/10.1186/s12862-018-1325-8>
55. De Luca D. Catanese G. Procaccini G. Fiorito G. (2016). *Octopus vulgaris* (Cuvier). *PLoS ONE* <http://doi.org/10.1371/journal.pone.0149496>
56. Dembkowski D.J. Isermann D.A. Hogler S.R. Larson W.A. Turnquist K.N. (2018). Stock structure, dynamics, demographics, and movements of walleyes spawning in four tributaries to Green Bay. *Journal of Great Lakes Research* <http://doi.org/10.1016/j.jglr.2018.07.002>
57. Dercole F. Della Rossa F. (2017). A deterministic eco-genetic model for the short-term evolution of exploited fish stocks. *Ecological Modelling* <http://doi.org/10.1016/j.ecolmodel.2016.10.016>
58. Di Prinzio C.Y. Rossi C.R. Ciancio J. Garza J.C. Casaux R. (2015). Disentangling the contributions of ocean ranching and net-pen aquaculture in the successful establishment of Chinook salmon in a Patagonian basin. *Environmental Biology of Fishes* <http://doi.org/10.1007/s10641-015-0418-0>
59. Divya P.R. Gopalakrishnan A. Basheer V.S. Swaminathan R. Mohitha C. Joy L. Kumar R. Manoj P. Jena J.K. (2015). Mitochondrial ATPase 6/8 genes to infer the population genetic structure of silver pomfret fish *Pampus argenteus* along the Indian waters. *Mitochondrial DNA* <http://doi.org/10.3109/19401736.2013.879655>
60. Divya P.R. Vineesh N. Kathirvelpandian A. Mohitha C. Basheer V.S. Gopalakrishnan A. Jena J.K. (2016). Identification and characterisation of microsatellite markers in narrow barred Spanish mackerel *Scomberomorus commerson* (Lacepede). *Indian Journal of Fisheries* <http://doi.org/10.21077/ijf.2016.63.4.54028-10>
61. Domínguez-López M. Díaz-Jaimes P. Uribe-Alcocer M. Quiñonez-Velázquez C. (2015). Post-glacial population expansion of the Monterey Spanish mackerel *Scomberomorus concolor* in the Gulf of California. *Journal of Fish Biology* <http://doi.org/10.1111/jfb.12580>
62. Drinan D.P. Gruenthal K.M. Canino M.F. Lowry D. Fisher M.C. Hauser L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evolutionary Applications* <http://doi.org/10.1111/eva.12639>
63. Du Pontavice H. Randon M. Lehuta S. Vermard Y. Savina-Rolland M. (2018). Investigating spatial heterogeneity of von Bertalanffy growth parameters to inform the stock structuration of common sole, in the Eastern English Channel. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2018.05.009>
64. Duncan M. James N. Fennessy S.T. Mutombene R.J. Mwale M. (2015). Genetic structure and consequences of stock exploitation of *Chrysoblephus puniceus*. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2014.10.019>
65. Dutton P.H. LeRoux R.A. LaCasella E.L. Seminoff J.A. Eguchi T. Dutton D.L. (2019). Genetic analysis and satellite tracking reveal origin of the green turtles in San Diego Bay. *Marine Biology* <http://doi.org/10.1007/s00227-018-3446-4>
66. Eikeset A.M. Dunlop E.S. Heino M. Storvik G. Stenseth N.C. Dieckmann U. (2016). Roles of density-dependent growth and life history evolution in accounting for fisheries-induced trait changes. *Proceedings of the National Academy of Sciences of the United States of America* <http://doi.org/10.1073/pnas.1525749113>
67. Ellis C.D. Hodgson D.J. Daniels C.L. Boothroyd D.P. Bannister R.C.A. Griffiths A.G.F. (2015). European lobster stocking requires comprehensive impact assessment to determine fishery benefits. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsu196>
68. Ellis C.D. Hodgson D.J. Daniels C.L. Collins M. Griffiths A.G.F. (2017). Population genetic structure in European lobsters: Implications for connectivity. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps11957>
69. Elvira B. Leal S. Doadrio I. Almodóvar A. (2015). Current occurrence of the Atlantic sturgeon *Acipenser oxyrinchus* in northern Spain: A new prospect for sturgeon conservation in western Europe. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0145728>
70. Erlandsson J. Östman Ö. Florin A.-B. Pekcan-Hekim Z. (2017). Spatial structure of body size of European flounder (*Platichthys flesus* L.) in the Baltic Sea. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2017.01.001>

71. Escobar R. Luna-Acosta A. Caballero S. (2019). DNA barcoding, fisheries and communities: What do we have? Science and local knowledge to improve resource management in partnership with communities in the Colombian Caribbean. *Marine Policy* <http://doi.org/10.1016/j.marpol.2018.08.025>
72. Fainburg L.A. Fernández Iriarte P.J. (2018). Current Knowledge on Commercially Important Fish Stocks of the Buenos Aires Coastal Ecosystem. *Reviews in Fisheries Science and Aquaculture* <http://doi.org/10.1080/23308249.2017.1374343>
73. Farivar S. Jalil-Piran Z. Zarei F. Hosseinzadeh Sahafi H. (2017). Intraspecific phylogeography of the Japanese threadfin bream. *Iranian Journal of Fisheries Sciences* <http://doi.org/>
74. Farrell E.D. Carlsson J.E.L. Carlsson J. (2016). Next gen pop gen: Implementing a high-throughput approach to population genetics in boarfish (*Capros aper*). *Royal Society Open Science* <http://doi.org/10.1098/rsos.160651>
75. Fernández G. Cichero D. Patel A. Martínez V. (2015). Genetic structure of Chilean populations of *Seriola lalandi* for the diversification of the national aquaculture in the north of Chile [Estructura genética de poblaciones chilenas de *Seriola lalandi* para la diversificación de la acuicultura nacional en el norte de Chile]. *Latin American Journal of Aquatic Research* <http://doi.org/10.3856/vol43-issue2-fulltext-14>
76. Fernández-Pérez J. Froufe E. Nantón A. Gaspar M.B. Méndez J. (2017). Genetic diversity and population genetic analysis of *Donax vittatus* (Mollusca: Bivalvia) and phylogeny of the genus with mitochondrial and nuclear markers. *Estuarine, Coastal and Shelf Science* <http://doi.org/10.1016/j.ecss.2017.08.032>
77. Fopp-Bayat D. Kaczmarczyk D. Szczepkowski M. (2015). Genetic characteristics of Polish whitefish (*Coregonus lavaretus maraena*) broodstocks - Recommendations for the conservation management. *Czech Journal of Animal Science* <http://doi.org/10.17221/8131-CJAS>
78. Gabián M. Morán P. (2019). A genetic tool for evaluating male-mediated stock introgression in Atlantic salmon. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.3008>
79. Gama-Maia D.J. Torres R.A. (2016). Fine-scale genetic structuring. *Journal of the Marine Biological Association of the United Kingdom* <http://doi.org/10.1017/S0025315415002052>
80. García-Celdrán M. Cutáková Z. Ramis G. Estévez A. Manchado M. Navarro A. María-Dolores E. Peñalver J. Sánchez J.A. Armero E. (2016). Estimates of heritabilities and genetic correlations of skeletal deformities and uninflated swimbladder in a reared gilthead sea bream (*Sparus aurata* L.) juvenile population sourced from three broodstocks along the Spanish coasts. *Aquaculture* <http://doi.org/10.1016/j.aquaculture.2016.08.004>
81. Gargano F. Garofalo G. Fiorentino F. (2017). Exploring connectivity between spawning and nursery areas of *Mullus barbatus* (L. 1758) in the Mediterranean through a dispersal model. *Fisheries Oceanography* <http://doi.org/10.1111/fog.12210>
82. Gaspari S. Holcer D. Mackelworth P. Fortuna C. Frantzis A. Genov T. Vighi M. Natali C. Rako N. Banchi E. Chelazzi G. Ciofi C. (2015). Population genetic structure of common bottlenose dolphins (*Tursiops truncatus*) in the Adriatic Sea and contiguous regions: Implications for international conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.2415>
83. Geladakis G. Nikolioudakis N. Koumoundouros G. Somarakis S. (2018). Morphometric discrimination of pelagic fish stocks challenged by variation in body condition. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsx186>
84. Glynn F. Houghton J.D.R. Bastian T. Doyle T.K. Fuentes V. Lilley M.K.S. Provan J. (2016). High-resolution genetic analysis reveals extensive gene flow within the jellyfish *Pelagia noctiluca* (Scyphozoa) in the North Atlantic and Mediterranean Sea. *Biological Journal of the Linnean Society* <http://doi.org/10.1111/bij.12654>
85. Gonçalves Da Silva A. Appleyard S.A. Upston J. (2015). Establishing the evolutionary compatibility of potential sources of colonizers for overfished stocks: A population genomics approach. *Molecular Ecology* <http://doi.org/10.1111/mec.13046>
86. Gonzalez E.B. Aritaki M. Knutsen H. Taniguchi N. (2015). Effects of large-scale releases on the genetic structure of red sea bream (*Pagrus major*). *PLoS ONE* <http://doi.org/10.1371/journal.pone.0125743>
87. González-Wevar C. Salinas P. Hüne M. Segovia N. Vargas-Chacoff L. Oda E. Poulin E. (2015). Contrasting Genetic Structure and Diversity of *Galaxias maculatus* (Jenyns). *Journal of Heredity* <http://doi.org/10.1093/jhered/esv005>
88. Gopalakrishnan A. Vineesh N. Ismail S. Menon M. Akhilesh K.V. Jeena N.S. Paulton M.P. Vijayagopal P. (2018). Mitochondrial signatures revealed panmixia in *Lutjanus argentimaculatus* (Forsskål 1775). *Journal of Genetics* <http://doi.org/10.1007/s12041-018-0899-7>

89. Grahl-Nielsen O. Glover K.A. (2016). Fatty acids in fish scales. *Marine Biology* <http://doi.org/10.1007/s00227-010-1430-8>
90. Gubili C. Macleod K. Perry W. Hanel P. Batzakas I. Farrell E.D. Lynghammar A. Mancusi C. Mariani S. Menezes G.M. Neat F. Scarcella G. Griffiths A.M. (2016). Connectivity in the deep: Phylogeography of the velvet belly lanternshark. *Deep-Sea Research Part I: Oceanographic Research Papers* <http://doi.org/10.1016/j.dsr.2016.07.002>
91. Guiñez R. Pita A. Pérez M. Briones C. Navarrete S.A. Toro J. Presa P. (2016). Present-day connectivity of historical stocks of the ecosystem engineer *Perumytilus purpuratus* along 4500 km of the Chilean Coast. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.02.013>
92. Hargrove J.S. Murie D.J. Parkyn D.C. Saarinen E.V. Austin J.D. (2018). Mixing rates in weakly differentiated stocks of greater amberjack (*Seriola dumerili*) in the Gulf of Mexico. *Genetica* <http://doi.org/10.1007/s10709-018-0031-1>
93. Harris L.N. Moore J.-S. Bajno R. Tallman R.F. (2016). Genetic stock structure of anadromous arctic char in Canada's central arctic: Potential implications for the management of Canada's largest arctic char commercial fishery. *North American Journal of Fisheries Management* <http://doi.org/10.1080/02755947.2016.1227399>
94. Hasselman D.J. Anderson E.C. Argo E.E. David Bethoney N. Gephard S.R. Post D.M. Schondelmeier B.P. Schultz T.F. Willis T.V. Palkovacs E.P. (2016). Genetic stock composition of marine bycatch reveals disproportional impacts on depleted river herring genetic stocks. *Canadian Journal of Fisheries and Aquatic Sciences* <http://doi.org/10.1139/cjfas-2015-0402>
95. Haxton T. Nienhuis S. Punt K. Baker T. (2015). Assessing Walleye Movement among Reaches of a Large. *North American Journal of Fisheries Management* <http://doi.org/10.1080/02755947.2015.1012278>
96. Hedeholm R.B. Jacobsen R.B. Nielsen E.E. (2016). Learning from 'apparent consensus' in TAC disputes: Exploring knowledge overlaps in LEK and genetic categorization of Atlantic cod. *Marine Policy* <http://doi.org/10.1016/j.marpol.2016.04.020>
97. Henriques R. Nielsen E.S. Durholtz D. Japp D. von der Heyden S. (2017). Genetic population sub-structuring of kingklip (*Genypterus capensis* – Ophidiidae). *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.11.007>
98. Henriques R. von der Heyden S. Lipinski M.R. du Toit N. Kainge P. Bloomer P. Matthee C.A. (2016). Spatio-temporal genetic structure and the effects of long-term fishing in two partially sympatric offshore demersal fishes. *Molecular Ecology* <http://doi.org/10.1111/mec.13890>
99. Hollenbeck C.M. Portnoy D.S. Saillant E. Gold J.R. (2015). Population structure of red snapper (*Lutjanus campechanus*) in U.S. waters of the western Atlantic Ocean and the northeastern Gulf of Mexico. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.06.020>
100. Hughes J.M. Stewart J. Gillanders B.M. Collins D. Suthers I.M. (2016). Relationship between otolith chemistry and age in a widespread pelagic teleost *Arripis trutta*: Influence of adult movements on stock structure and implications for management. *Marine and Freshwater Research* <http://doi.org/10.1071/MF14247>
101. Hurtado-Alarcón J.C. Campos Campos N.H. Bermúdez Tobón A. Márquez E.J. (2018). Phylogeographic patterns in *Maguimithrax spinosissimus* (Decapoda: Mithracidae) from Colombian Caribbean. *New Zealand Journal of Marine and Freshwater Research* <http://doi.org/10.1080/00288330.2017.1353528>
102. Isari S. Pearman J.K. Casas L. Michell C.T. Curdia J. Berumen M.L. Irigoien X. (2017). Exploring the larval fish community of the central Red Sea with an integrated morphological and molecular approach. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0182503>
103. Ismail S. Vineesh N. Mohitha C. Vijayagopal P. Gopalakrishnan A. (2018). Development and validation of microsatellite markers in a protandrous fish species *Eleutheronema tetradactylum* (Shaw). *Indian Journal of Fisheries* <http://doi.org/10.21077/ijf.2018.65.3.81674-06>
104. Izzo C. Huveneers C. Drew M. Bradshaw C.J.A. Donnellan S.C. Gillanders B.M. (2016). Vertebral chemistry demonstrates movement and population structure of bronze whaler. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps11840>
105. Izzo C. Ward T.M. Ivey A.R. Suthers I.M. Stewart J. Sexton S.C. Gillanders B.M. (2017). Integrated approach to determining stock structure: implications for fisheries management of sardine. *Reviews in Fish Biology and Fisheries* <http://doi.org/10.1007/s11160-017-9468-z>
106. Jardine S.L. Sanchirico J.N. (2015). Fishermen. *Journal of Environmental Economics and Management* <http://doi.org/10.1016/j.jeem.2015.06.004>

107. Järv L. Raid T. Pärnoja M. Soome A. (2016). Developments in Estonian coastal perch fishery in the Baltic Sea during the recent two decades (1994-2015). Proceedings of 3rd International Conference on Maritime Technology and Engineering <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016992596&partnerID=40&md5=0a42bf676838b8c5207f81c2dcfda144>
108. Jasonowicz A.J. Goetz F.W. Goetz G.W. Nichols K.M. (2017). Love the one you're with: Genomic evidence of panmixia in the sablefish (*Anoplopoma fimbria*). Canadian Journal of Fisheries and Aquatic Sciences <http://doi.org/10.1139/cjfas-2016-0012>
109. Jeena N.S. Gopalakrishnan A. Kizhakudan J.K. Radhakrishnan E.V. Kumar R. Asokan P.K. (2016). Population genetic structure of the shovel-nosed lobster *Thenus unimaculatus* (Decapoda). Hydrobiologia <http://doi.org/10.1007/s10750-015-2458-z>
110. Jemaa S. Bacha M. Khalaf G. Amara R. (2015). Evidence for population complexity of the European anchovy (*Engraulis encrasicolus*) along its distributional range. Fisheries Research <http://doi.org/10.1016/j.fishres.2015.04.004>
111. Jemaa S. Bacha M. Khalaf G. Dessailly D. Rabhi K. Amara R. (2015). What can otolith shape analysis tell us about population structure of the European sardine. Journal of Sea Research <http://doi.org/10.1016/j.seares.2014.11.002>
112. Jiang L. Chen Y. Zhang J. Zhu A. Wu C. Liu L. Lin Z. Dong Y. (2015). Population structure of large yellow croaker (*Larimichthys crocea*) revealed by single nucleotide polymorphisms. Biochemical Systematics and Ecology <http://doi.org/10.1016/j.bse.2015.09.025>
113. Jolanta K. Maciej K. Remigiusz P. Prüller K. Sławomir K. (2017). Development of the method for identification of selected populations of torpedo scad. Food Chemistry <http://doi.org/10.1016/j.foodchem.2016.11.070>
114. Jónsdóttir Ó.D.B. Schregel J. Hagen S.B. Tobiassen C. Aarnes S.G. Imsland A.K.D. (2018). Population genetic structure of lumpfish along the Norwegian coast: aquaculture implications. Aquaculture International <http://doi.org/10.1007/s10499-017-0194-2>
115. Jorge P.H. Mastrochirico-Filho V.A. Hata M.E. Mendes N.J. Ariede R.B. de Freitas M.V. Vera M. Porto-Foresti F. Hashimoto D.T. (2018). Genetic characterization of the fish *Piaractus brachipomus* by microsatellites derived from transcriptome sequencing. Frontiers in Genetics <http://doi.org/10.3389/fgene.2018.00046>
116. Joy L. Mohitha C. Divya P.R. Gopalakrishnan A. Basheer V.S. Jena J.K. (2016). Weak genetic differentiation in cobia. Mitochondrial DNA <http://doi.org/10.3109/19401736.2015.1053083>
117. Kan C. Lichter J. Douhovnikoff V. (2017). Genetic Population Structure and Accuracy of Morphological Assessment in *Alosa aestivalis* (Blueback Herring) and *A. pseudoharengus* (Alewife). Northeastern Naturalist <http://doi.org/10.1656/045.024.0407>
118. Kaplan D.M. Cuif M. Fauvelot C. Vigliola L. Nguyen-Huu T. Tiavouane J. Lett C. (2017). Uncertainty in empirical estimates of marine larval connectivity. ICES Journal of Marine Science <http://doi.org/10.1093/icesjms/fsw182>
119. Katsumura T. Oda S. Mitani H. Oota H. (2019). Medaka population genome structure and demographic history described via genotyping-by-sequencing. G3: Genes, Genomes, Genetics <http://doi.org/10.1534/g3.118.200779>
120. Kenchington E.L. Nakashima B.S. Taggart C.T. Hamilton L.C. (2015). Genetic structure of capelin (*Mallotus villosus*) in the northwest Atlantic Ocean. PLoS ONE <http://doi.org/10.1371/journal.pone.0122315>
121. Kennington W.J. Keron P.W. Harvey E.S. Wakefield C.B. Williams A.J. Halafihi T. Newman S.J. (2017). High intra-ocean. Fisheries Research <http://doi.org/10.1016/j.fishres.2017.04.015>
122. Kenthao A. Wangsomnuk P.P. Jearranaiprepame P. (2018). Genetic variations and population structure in three populations of beardless barb. Mitochondrial DNA Part A: DNA Mapping, Sequencing and Analysis <http://doi.org/10.1080/24701394.2016.1242581>
123. Khan M.A. Nazir A. (2019). Stock delineation of the long-whiskered catfish, *Sperata aor* (Hamilton 1822), from River Ganga by using morphometrics. Marine and Freshwater Research <http://doi.org/10.1071/MF17306>
124. Khrustaleva A.M. Klovach N.V. Seeb J.E. (2017). Genetic variability and population structure of sockeye salmon from the Asian Coast of Pacific Ocean. Russian Journal of Genetics <http://doi.org/10.1134/S1022795417100052>

125. Kijewska A. Kalamarż-Kubiak H. Arciszewski B. Guellard T. Petereit C. Wenne R. (2016). Adaptation to salinity in Atlantic cod from different regions of the Baltic Sea. *Journal of Experimental Marine Biology and Ecology* <http://doi.org/10.1016/j.jembe.2016.02.003>
126. Kikko T. Okamoto H. Ujiiie M. Usuki T. Nemoto M. Saegusa J. Ishizaki D. Fujioka Y. Kai Y. Nakayama K. (2016). Genetic evaluation of hatchery stocks of Honmoroko *Gnathopogon caerulescens* by mitochondrial DNA sequence for stock enhancement. *Fisheries Science* <http://doi.org/10.1007/s12562-015-0958-4>
127. King J.R. Wetklo M. Supernault J. Taguchi M. Yokawa K. Sosa-Nishizaki O. Withler R.E. (2015). Genetic analysis of stock structure of blue shark (*Prionace glauca*) in the north Pacific ocean. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.06.029>
128. Kiper I.E. Bloomer P. Borsa P. Hoareau T.B. (2018). Characterization of genome-wide microsatellite markers in rabbitfishes. *Molecular Biology Reports* <http://doi.org/10.1007/s11033-017-4136-x>
129. Knutsen H. Jorde P.E. Gonzalez E.B. Eigaard O.R. Pereyra R.T. Sannæs H. Dahl M. André C. Søvik G. (2015). Does population genetic structure support present management regulations of the northern shrimp (*Pandalus borealis*) in Skagerrak and the North Sea?. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsu204>
130. Komoroske L.M. Jensen M.P. Stewart K.R. Shamblin B.M. Dutton P.H. (2017). Advances in the application of genetics in marine turtle biology and conservation. *Frontiers in Marine Science* <http://doi.org/10.3389/fmars.2017.00156>
131. Komoroske L.M. Miller M.R. O'Rourke S.M. Stewart K.R. Jensen M.P. Dutton P.H. (2019). A versatile Rapture (RAD-Capture) platform for genotyping marine turtles. *Molecular Ecology Resources* <http://doi.org/10.1111/1755-0998.12980>
132. Kumar G. Kocour M. (2015). Population genetic structure of tunas inferred from molecular markers: A review. *Reviews in Fisheries Science and Aquaculture* <http://doi.org/10.1080/23308249.2015.1024826>
133. Kumari P. Pavan-Kumar A. Kumar G. Alam A. Parhi J. Gireesh-Babu P. Chaudhari A. Krishna G. (2017). Genetic diversity and demographic history of the giant river catfish *Sperata seenghala* inferred from mitochondrial DNA markers. *Mitochondrial DNA Part A: DNA Mapping, Sequencing and Analysis* <http://doi.org/10.1080/24701394.2016.1209195>
134. Labastida-Estrada E. Machkour-M'Rabet S. Díaz-Jaimes P. Cedeño-Vázquez J.R. Hénaut Y. (2019). Genetic structure, origin, and connectivity between nesting and foraging areas of hawksbill turtles of the Yucatan Peninsula: A study for conservation and management. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.2999>
135. Lacerda A.L.F. Kersanach R. Cortinhas M.C.S. Prata P.F.S. Dumont L.F.C. Proietti M.C. Maggioni R. D'Incao F. (2016). High connectivity among blue crab (*Callinectes sapidus*) populations in the Western South Atlantic. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0153124>
136. Lait L.A. Marshall H.D. Carr S.M. (2018). Phylogeographic mitogenomics of Atlantic cod *Gadus morhua*: Variation in and among trans-Atlantic. *Ecology and Evolution* <http://doi.org/10.1002/ece3.3873>
137. Lal M.M. Southgate P.C. Jerry D.R. Bosserelle C. Zenger K.R. (2016). A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: The case of the Fijian black-lip pearl Oyster *Pinctada Margaritifera*. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0161390>
138. Lal M.M. Southgate P.C. Jerry D.R. Zenger K.R. (2016). Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate. *Marine Genomics* <http://doi.org/10.1016/j.margen.2015.10.010>
139. Lallias D. Boudry P. Batista F.M. Beaumont A. King J.W. Turner J.R. Lapègue S. (2015). Invasion genetics of the Pacific oyster *Crassostrea gigas* in the British Isles inferred from microsatellite and mitochondrial markers. *Biological Invasions* <http://doi.org/10.1007/s10530-015-0896-1>
140. Landi P. Hui C. Dieckmann U. (2015). Fisheries-induced disruptive selection. *Journal of Theoretical Biology* <http://doi.org/10.1016/j.jtbi.2014.10.017>
141. Landínez-García R.M. Márquez E.J. (2016). Development and characterization of 24 polymorphic microsatellite loci for the freshwater fish *Ichthyoelephas longirostris* (Characiformes: Prochilodontidae). *PeerJ* <http://doi.org/10.7717/peerj.2419>

142. Lane H.S. Symonds J.E. Ritchie P.A. (2016). The phylogeography and population genetics of *Polyprion oxygeneios* based on mitochondrial DNA sequences and microsatellite DNA markers. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.08.009>
143. Langangen Ø. Färber L. Stige L.C. Diekert F.K. Barth J.M.I. Matschiner M. Berg P.R. Star B. Stenseth N.C. Jentoft S. Durant J.M. (2019). Ticket to spawn: Combining economic and genetic data to evaluate the effect of climate and demographic structure on spawning distribution in Atlantic cod. *Global Change Biology* <http://doi.org/10.1111/gcb.14474>
144. Larson S. Farrer D. Lowry D. Ebert D.A. (2015). Preliminary observations of population genetics and relatedness of the broadnose sevengill shark. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0129278>
145. Ledoux J.-B. Antunes A. Haguenaer A. Pratlong M. Costantini F. Abbiati M. Aurelle D. (2016). Molecular forensics into the sea: How molecular markers can help to struggle against poaching and illegal trade in precious corals?. *The Cnidaria* http://doi.org/10.1007/978-3-319-31305-4_45
146. Lehnert S.J. DiBacco C. Van Wyngaarden M. Jeffery N.W. Ben Lowen J. Sylvester E.V.A. Wringe B.F. Stanley R.R.E. Hamilton L.C. Bradbury I.R. (2019). Fine-scale temperature-associated genetic structure between inshore and offshore populations of sea scallop (*Placopecten magellanicus*). *Heredity* <http://doi.org/10.1038/s41437-018-0087-9>
147. Leidig J.M. Shervette V.R. McDonough C.J. Darden T.L. (2015). Genetic Population Structure of Black Drum in U.S. Waters. *North American Journal of Fisheries Management* <http://doi.org/10.1080/02755947.2015.1017123>
148. Leslie M.S. Archer F.I. Morin P.A. (2018). Mitogenomic differentiation in spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S. attenuata*) from the eastern tropical Pacific Ocean. *Marine Mammal Science* <http://doi.org/10.1111/mms.12545>
149. Leslie M.S. Morin P.A. (2016). Using genome-wide SNPs to detect structure in high-diversity and low-divergence populations of severely impacted eastern tropical pacific spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S. attenuata*). *Frontiers in Marine Science* <http://doi.org/10.3389/fmars.2016.00253>
150. Lester R.J.G. Moore B.R. (2015). Parasites as valuable stock markers for fisheries in Australasia. *Parasitology* <http://doi.org/10.1017/S003118201400016X>
151. Levsen A. Cipriani P. Mattiucci S. Gay M. Hastie L.C. MacKenzie K. Pierce G.J. Svanevik C.S. Højgaard D.P. Nascetti G. González A.F. Pascual S. (2018). Anisakis species composition and infection characteristics in Atlantic mackerel. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2017.07.030>
152. Li W. Chen X. Xu Q. Zhu J. Dai X. Xu L. (2015). Genetic Population Structure of *Thunnus albacares* in the Central Pacific Ocean Based on mtDNA COI Gene Sequences. *Biochemical Genetics* <http://doi.org/10.1007/s10528-015-9666-0>
153. Li Y. Lin L. Song N. Zhang Y. Gao T. (2018). Population genetics of *Pampus echinogaster* along the Pacific coastline of China: Insights from the mitochondrial DNA control region and microsatellite molecular markers. *Marine and Freshwater Research* <http://doi.org/10.1071/MF17218>
154. Liu B.-J. Zhang B.-D. Xue D.-X. Gao T.-X. Liu J.-X. (2016). Population structure and adaptive divergence in a high gene flow marine fish: The small yellow croaker (*Larimichthys polyactis*). *PLoS ONE* <http://doi.org/10.1371/journal.pone.0154020>
155. Liu Y. Yang R. Liu Y. Si F. (2017). Use of Microsatellite DNA Profiling to Identify Japanese Flounder. *Journal of the World Aquaculture Society* <http://doi.org/10.1111/jwas.12330>
156. Losee J.P. Seamons T.R. Jauquet J. (2017). Migration patterns of anadromous Cutthroat Trout in South Puget Sound: A fisheries management perspective. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.12.006>
157. Louati M. Kohlmann K. Ben Hassine O.K. Kersten P. Poulet N. Bahri-Sfar L. (2016). Genetic characterization of introduced Tunisian and French populations of pike-perch (*Sander lucioperca*) by species-specific microsatellites and mitochondrial haplotypes. *Czech Journal of Animal Science* <http://doi.org/10.17221/8847-CJAS>
158. Loughnan S.R. Beheregaray L.B. Robinson N.A. (2015). Microsatellite marker development based on next-generation sequencing for the smooth marron (*Cherax cainii*). *BMC Research Notes* <http://doi.org/10.1186/s13104-015-1345-z>
159. Lowther-Thieleking J.L. Archer F.I. Lang A.R. Weller D.W. (2015). Genetic differentiation among coastal and offshore common bottlenose dolphins. *Marine Mammal Science* <http://doi.org/10.1111/mms.12135>

160. Lu C.-P. Smith B.L. Hinton M.G. Alvarado Bremer J.R. (2016). Bayesian analyses of Pacific swordfish (*Xiphias gladius* L.) genetic differentiation using multilocus single nucleotide polymorphism (SNP) data. *Journal of Experimental Marine Biology and Ecology* <http://doi.org/10.1016/j.jembe.2016.03.010>
161. Luo W. Qu H. Li J. Wang X. Lin Q. (2015). A novel method for the identification of seahorses (genus *Hippocampus*) using cross-species amplifiable microsatellites. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.07.017>
162. Luo W. Wang W.-M. Wan S.-M. Lin Q. Gao Z.-X. (2017). Assessment of parental contribution to fast- and slow-growth progenies in the blunt snout bream (*Megalobrama amblycephala*) based on parentage assignment. *Aquaculture* <http://doi.org/10.1016/j.aquaculture.2016.07.003>
163. Machado V.N. Willis S.C. Teixeira A.S. Hrbek T. Farias I.P. (2017). Population genetic structure of the Amazonian black flannelmouth characin (*Characiformes*). *Environmental Biology of Fishes* <http://doi.org/10.1007/s10641-016-0547-0>
164. Maggi C. González-Wangüemert M. (2015). Genetic differentiation among *Parastichopus regalis* populations in the Western Mediterranean Sea: Potential effects from its fishery and current connectivity. *Mediterranean Marine Science* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84951969835&partnerID=40&md5=303d2fc3dc4d26015c98c7a5208a9390>
165. Mamoozadeh N.R. McDowell J.R. Rooker J.R. Graves J.E. (2018). Genetic evaluation of population structure in white marlin (*Kajikia albida*): The importance of statistical power. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsx047>
166. Mandal S. Jena J.K. Singh R.K. Mohindra V. Lakra W.S. Deshmukhe G. Pathak A. Lal K.K. (2016). De novo development and characterization of polymorphic microsatellite markers in a schilbid catfish. *Molecular Biology Reports* <http://doi.org/10.1007/s11033-016-3941-y>
167. Marengo M. Baudouin M. Viret A. Laporte M. Berrebi P. Vignon M. Marchand B. Durieux E.D.H. (2017). Combining microsatellite. *Journal of Sea Research* <http://doi.org/10.1016/j.seares.2017.07.003>
168. Maroso F. Casanova A. do Prado F.D. Bouza C. Pardo B.G. Blanco A. Hermida M. Fernández C. Vera M. Martínez P. (2018). Species identification of two closely exploited flatfish, turbot (*Scophthalmus maximus*) and brill (*Scophthalmus rhombus*), using a ddRADseq genomic approach. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.2932>
169. Maroso F. Franch R. Dalla Rovere G. Arculeo M. Bargelloni L. (2016). RAD SNP markers as a tool for conservation of dolphinfish *Coryphaena hippurus* in the Mediterranean Sea: Identification of subtle genetic structure and assessment of populations sex-ratios. *Marine Genomics* <http://doi.org/10.1016/j.margen.2016.07.003>
170. Matsuzawa Y. Kamezaki N. Ishihara T. Omuta K. Takeshita H. Goto K. Arata T. Honda H. Kameda K. Kashima Y. Kayo M. Kawazu I. Kodama J. Kumazawa Y. Kuroyanagi K. Mizobuchi K. Mizuno K. Oki K. Watanabe K.K. Yamamoto A. Yamashita Y. Yamato T. Hamabata T. Ishizaki A. Dutton P.H. (2016). Fine-scale genetic population structure of loggerhead turtles in the Northwest Pacific. *Endangered Species Research* <http://doi.org/10.3354/esr00724>
171. Mattiucci S. Cimmaruta R. Cipriani P. Abaunza P. Bellisario B. Nascetti G. (2015). Integrating *Anisakis* spp. parasites data and host genetic structure in the frame of a holistic approach for stock identification of selected Mediterranean Sea fish species. *Parasitology* <http://doi.org/10.1017/S0031182014001103>
172. Mattiucci S. Giulietti L. Paoletti M. Cipriani P. Gay M. Levsen A. Klapper R. Karl H. Bao M. Pierce G.J. Nascetti G. (2018). Population genetic structure of the parasite *Anisakis simplex* (s. s.) collected in *Clupea harengus* L. from North East Atlantic fishing grounds. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2017.08.002>
173. McBride M.C. Hasselman D.J. Willis T.V. Palkovacs E.P. Bentzen P. (2015). Influence of stocking history on the population genetic structure of anadromous alewife (*Alosa pseudoharengus*) in Maine rivers. *Conservation Genetics* <http://doi.org/10.1007/s10592-015-0733-1>
174. McDowell J.R. Brightman H.L. (2018). High level of genetic connectivity in a deep-water reef fish *Caulolatilus microps*. *Journal of Fish Biology* <http://doi.org/10.1111/jfb.13779>
175. McKeown N.J. Arkhipkin A.I. Shaw P.W. (2015). Integrating genetic and otolith microchemistry data to understand population structure in the Patagonian Hoki (*Macruronus magellanicus*). *Fisheries Research* <http://doi.org/10.1016/j.fishres.2014.10.004>

176. McKeown N.J. Arkhipkin A.I. Shaw P.W. (2017). Regional genetic population structure and fine scale genetic cohesion in the Southern blue whiting *Micromesistius australis*. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.09.006>
177. McKeown N.J. Hauser L. Shaw P.W. (2017). Microsatellite genotyping of brown crab cancer *pagurus* reveals fine scale selection and 'non-chaotic' genetic patchiness within a high gene flow system. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps12044>
178. McKeown N.J. Watson H.V. Coscia I. Wootton E. Ironside J.E. (2018). Genetic variation in Irish Sea brown crab (*Cancer pagurus* L.): Implications for local and regional management. *Journal of the Marine Biological Association of the United Kingdom* <http://doi.org/10.1017/S0025315418000632>
179. McKinney G.J. Seeb J.E. Seeb L.W. (2017). Managing mixed-stock fisheries: Genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences* <http://doi.org/10.1139/cjfas-2016-0443>
180. McWilliam R.A. Adams K.R. Minchinton T.E. Ayre D.J. (2016). Characterization of 11 polymorphic microsatellite markers for black drummer (*Girella elevata*) developed using 454 next-generation sequencing. *Journal of Heredity* <http://doi.org/10.1093/jhered/esw066>
181. Mejri M. Trojette M. Allaya H. Ben Faleh A. Jmil I. Chalh A. Quignard J.-P. Trabelsi M. (2018). Use of otolith shape to differentiate two lagoon populations of *pagellus erythrinus* (Actinopterygii: Perciformes: Sparidae) in Tunisian waters. *Acta Ichthyologica et Piscatoria* <http://doi.org/10.3750/AIEP/02376>
182. Midway S.R. Wagner T. Arnott S.A. Biondo P. Martinez-Andrade F. Wadsworth T.F. (2015). Spatial and temporal variability in growth of southern flounder (*Paralichthys lethostigma*). *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.03.009>
183. Miller A.D. van Rooyen A. Rašić G. Ierodionou D.A. Gorfine H.K. Day R. Wong C. Hoffmann A.A. Weeks A.R. (2016). Contrasting patterns of population connectivity between regions in a commercially important mollusc *Haliotis rubra*: integrating population genetics. *Molecular ecology* <http://doi.org/10.1111/mec.13734>
184. Miller L.M. Schreiner D.R. Blankenheim J.E. Ward M.C. Quinlan H.R. Moore S. (2016). Effects of restrictive harvest regulations on rehabilitation of coaster brook trout in Minnesota's portion of Lake Superior. *Journal of Great Lakes Research* <http://doi.org/10.1016/j.jglr.2016.05.006>
185. Mims M.C. Day C.C. Burkhart J.J. Fuller M.R. Hinkle J. Bearlin A. Dunham J.B. DeHaan P.W. Holden Z.A. Landguth E.E. (2019). Simulating demography, genetics, and spatially explicit processes to inform reintroduction of a threatened char. *Ecosphere* <http://doi.org/10.1002/ecs2.2589>
186. Mkare T.K. Groeneveld J.C. Teske P.R. Matthee C.A. (2017). Comparative genetic structure in two high-dispersal prawn species from the south-west Indian Ocean. *African Journal of Marine Science* <http://doi.org/10.2989/1814232X.2017.1402089>
187. Mohanty P. Sahoo L. Pillai B.R. Jayasankar P. Das P. (2016). Genetic divergence in Indian populations of *M. rosenbergii* using microsatellite markers. *Aquaculture Research* <http://doi.org/10.1111/are.12508>
188. Momigliano P. Jokinen H. Calboli F. Aro E. Merilä J. (2019). Cryptic temporal changes in stock composition explain the decline of a flounder (*Platichthys* spp.) assemblage. *Evolutionary Applications* <http://doi.org/10.1111/eva.12738>
189. Monteiro S.S. Méndez-Fernandez P. Piertney S. Moffat C.F. Ferreira M. Vingada J.V. López A. Brownlow A. Jepson P. Mikkelsen B. Niemeyer M. Carvalho J.C. Pierce G.J. (2015). Long-finned pilot whale population diversity and structure in Atlantic waters assessed through biogeochemical and genetic markers. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps11455>
190. Morgan J.A.T. Sumpton W.D. Jones A.T. Campbell A.B. Stewart J. Hamer P. Ovenden J.R. (2018). Assessment of genetic structure among Australian east coast populations of snapper *Chrysophrys auratus* (Sparidae). *Marine and Freshwater Research* <http://doi.org/10.1071/MF18146>
191. Mutoloki S. Jøssund T.B. Ritchie G. Munang'andu H.M. Evensen Ø. (2016). Infectious pancreatic necrosis virus causing clinical and subclinical infections in atlantic salmon have different genetic fingerprints. *Frontiers in Microbiology* <http://doi.org/10.3389/fmicb.2016.01393>
192. Nantón A. Arias-Pérez A. Freire R. Fernández-Pérez J. Nóvoa S. Méndez J. (2017). Microsatellite variation in *Donax trunculus* from the Iberian Peninsula with particular attention to Galician estuaries (NW Spain). *Estuarine, Coastal and Shelf Science* <http://doi.org/10.1016/j.ecss.2017.08.011>

193. Nathan L.R. Sloss B.L. VanDeHey J.A. Andvik R.T. Claramunt R.M. Hansen S. Sutton T.M. (2016). Temporal stability of lake whitefish genetic stocks in Lake Michigan. *Journal of Great Lakes Research* <http://doi.org/10.1016/j.jglr.2016.01.006>
194. Nayfa M.G. Zenger K.R. (2016). Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). *Marine Genomics* <http://doi.org/10.1016/j.margen.2016.02.005>
195. Nazish N. Abbas K. Abdullah S. Zia M.A. (2018). Microsatellite diversity and population structure of hypophthalmichthys molitrix in hatchery populations of Punjab. *Turkish Journal of Fisheries and Aquatic Sciences* http://doi.org/10.4194/1303-2712-v18_9_10
196. Nesa N.U. Faroque A.A. Sarder M.R.I. Mollah M.F.A. (2018). Assessment of genetic structure of wild populations of mrigal carp, *Cirrhinus cirrhosus* by microsatellite DNA markers. *Aquaculture Research* <http://doi.org/10.1111/are.13875>
197. Neves A. Vieira A.R. Sequeira V. Paiva R.B. Janeiro A.I. Gaspar L.M. Gordo L.S. (2019). Otolith shape and isotopic ratio analyses as a tool to study *Spondyliosoma cantharus* population structure. *Marine Environmental Research* <http://doi.org/10.1016/j.marenvres.2018.11.012>
198. Newman S.J. Wakefield C.B. Williams A.J. O'Malley J.M. Taylor B.M. Nicol S.J. Nichols R.S. Hesp S.A. Hall N.G. Hill N. Ong J.J.L. Andrews A.H. Wellington C.M. Harvey E.S. Mous P. Oyafuso Z.S. Pardee C. Bunce M. DiBattista J.D. Moore B.R. (2017). International workshop on advancing methods to overcome challenges associated with life history and stock assessments of data-poor deep-water snappers and groupers. *Marine Policy* <http://doi.org/10.1016/j.marpol.2017.02.009>
199. Nishizawa H. Joseph J. Chong Y.K. Syed Kadir S.A. Isnain I. Ganyai T.A. Jaaman S. Zhang X. (2018). Comparison of the rookery connectivity and migratory connectivity: insight into movement and colonization of the green turtle (*Chelonia mydas*) in Pacific–Southeast Asia. *Marine Biology* <http://doi.org/10.1007/s00227-018-3328-9>
200. Ogburn M.B. Hasselman D.J. Schultz T.F. Palkovacs E.P. (2017). Genetics and juvenile abundance dynamics show congruent patterns of population structure for depleted river herring populations in the upper Chesapeake Bay. *North American Journal of Fisheries Management* <http://doi.org/10.1080/02755947.2017.1339649>
201. Oosthuizen C.J. Cowley P.D. Kyle S.R. Bloomer P. (2016). High genetic connectivity among estuarine populations of the riverbream *Acanthopagrus vagus* along the southern African coast. *Estuarine* <http://doi.org/10.1016/j.ecss.2016.10.024>
202. Östergren J. Nilsson J. Lundqvist H. Dannewitz J. Palm S. (2016). Genetic baseline for conservation and management of sea trout in the northern Baltic Sea. *Conservation Genetics* <http://doi.org/10.1007/s10592-015-0770-9>
203. Östman Ö. Olsson J. Dannewitz J. Palm S. Florin A.-B. (2017). Inferring spatial structure from population genetics and spatial synchrony in demography of Baltic Sea fishes: implications for management. *Fish and Fisheries* <http://doi.org/10.1111/faf.12182>
204. Otto S.A. Simons S. Stoll J.S. Lawson P. (2016). Making progress on bycatch avoidance in the ocean salmon fishery using a transdisciplinary approach. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsw061>
205. Oudejans M.G. Visser F. Englund A. Rogan E. Ingram S.N. (2015). Evidence for distinct coastal and offshore communities of bottlenose dolphins in the north east Atlantic. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0122668>
206. Ovenden J.R. Berry O. Welch D.J. Buckworth R.C. Dichmont C.M. (2015). Ocean's eleven: A critical evaluation of the role of population. *Fish and Fisheries* <http://doi.org/10.1111/faf.12052>
207. Ovenden J.R. Tillett B.J. Macbeth M. Broderick D. Filardo F. Street R. Tracey S.R. Semmens J. (2016). Stirred but not shaken: Population and recruitment genetics of the scallop (*Pecten fumatus*) in Bass Strait. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsw068>
208. Ozerov M. Vähä J.-P. Wennevik V. Niemelä E. Svenning M.-A. Prusov S. Fernandez R.D. Unneland L. Vasemägi A. Falkegård M. Kalske T. Christiansen B. (2017). Comprehensive microsatellite baseline for genetic stock identification of Atlantic salmon (*Salmo salar* L.) in northernmost Europe. *ICES Journal of Marine Science* <http://doi.org/>
209. Pampoulie C. Slotte A. Óskarsson G.J. Helyar S.J. Jónsson Á. Ólafsdóttir G. Skírnisdóttir S. Libungan L.A. Jacobsen J.A. Joensen H. Nielsen H.H. Sigurdsson S.K. Daníelsdóttir A.K. (2015). Stock structure of Atlantic herring *Clupea harengus* in the Norwegian Sea and adjacent waters. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps11114>

210. Park H.S. Kim C.-G. Kim S. Park Y.-J. Choi H.-J. Xiao Z. Li J. Xiao Y. Lee Y.-H. (2018). Population Genetic Structure of Rock Bream (*Oplegnathus fasciatus* Temminck & Schlegel. *Ocean Science Journal* <http://doi.org/10.1007/s12601-018-0009-z>
211. Parsons K.M. Everett M. Dahlheim M. Park L. (2018). Water, water everywhere: Environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science* <http://doi.org/10.1098/rsos.180537>
212. Patrício A.R. Formia A. Barbosa C. Broderick A.C. Bruford M. Carreras C. Catry P. Ciofi C. Regalla A. Godley B.J. (2017). Dispersal of green turtles from Africa's largest rookery assessed through genetic markers. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps12078>
213. Pazmiño D.A. Maes G.E. Simpfendorfer C.A. Salinas-de-León P. van Herwerden L. (2017). Genome-wide SNPs reveal low effective population size within confined management units of the highly vagile Galapagos shark (*Carcharhinus galapagensis*). *Conservation Genetics* <http://doi.org/10.1007/s10592-017-0967-1>
214. Pearse D.E. Garza J.C. (2015). You can't unscramble an egg: Population genetic structure of *oncorhynchus mykiss* in the California central valley inferred from combined microsatellite and single nucleotide polymorphism data. *San Francisco Estuary and Watershed Science* <https://doi.org/10.15447/sfews.2015v13iss4art3>
215. Pecoraro C. Zudaire I. Bodin N. Murua H. Taconet P. Díaz-Jaimes P. Cariani A. Tinti F. Chassot E. (2017). Putting all the pieces together: integrating current knowledge of the biology. *Reviews in Fish Biology and Fisheries* <http://doi.org/10.1007/s1160-016-9460-z>
216. Perina A. Mari-Mena N. Torrecilla Z. González-Tizón A.M. González-Castellano I. González-Ortegón E. Martínez-Lage A. (2019). Assessment of genetic diversity and population structure of the common littoral shrimp *Palaemon serratus* (Pennant, 1777) by microsatellites: Towards a sustainable management. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.3011>
217. Persat H. Mattersdorfer K. Charlat S. Schenekar T. Weiss S. (2016). Genetic integrity of the European grayling (*Thymallus thymallus*) populations within the Vienne River drainage basin after five decades of stockings. *Cybiurn* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84971474328&partnerID=40&md5=298490270c68361ed2fa51ca003f7b83>
218. Pil M.W. Baggio R.A. Tschá M.K. Marteleto F.M. Orélis-Ribeiro R. Patella L. Chammas M. Ostrensky A. Boeger W.A. (2018). The influence of paleoclimate on the distribution of genetic variability and demography of fishes in a large and highly fragmented neotropical river. *Hydrobiologia* <http://doi.org/10.1007/s10750-017-3285-1>
219. Pita A. Casey J. Hawkins S.J. Villarreal M.R. Gutiérrez M.-J. Cabral H. Carocci F. Abaunza P. Pascual S. Presa P. (2016). Conceptual and practical advances in fish stock delineation. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.10.029>
220. Pita A. Leal A. Santafé-Muñoz A. Piñeiro C. Presa P. (2016). Genetic inference of demographic connectivity in the Atlantic European hake metapopulation (*Merluccius merluccius*) over a spatio-temporal framework. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.03.017>
221. Plough L.V. (2017). Population Genomic Analysis of the Blue Crab *Callinectes sapidus* Using Genotyping-By-Sequencing. *Journal of Shellfish Research* <http://doi.org/10.2983/035.036.0128>
222. Pobedintseva M.A. Makunin A.I. Kichigin I.G. Kulemzina A.I. Serdyukova N.A. Romanenko S.A. Vorobieva N.V. Interesova E.A. Korentovich M.A. Zaytsev V.F. Mischenko A.V. Zadelenov V.A. Yurchenko A.A. Sherbakov D.Y. Graphodatsky A.S. Trifonov V.A. (2019). Population genetic structure and phylogeography of sterlet (*Acipenser ruthenus*, *Acipenseridae*) in the Ob and Yenisei river basins. *Mitochondrial DNA Part A: DNA Mapping, Sequencing and Analysis* <http://doi.org/10.1080/24701394.2018.1467409>
223. Prado F.D. Vera M. Hermida M. Blanco A. Bouza C. Maes G.E. Volckaert F.A.M. Martínez P. (2018). Tracing the genetic impact of farmed turbot *Scophthalmus maximus* on wild populations. *Aquaculture Environment Interactions* <http://doi.org/10.3354/AEI00282>
224. Pukk L. Gross R. Vetemaa M. Vasemägi A. (2016). Genetic discrimination of brackish and freshwater populations of Eurasian perch (*Perca fluviatilis* L.) in the Baltic Sea drainage: Implications for fish forensics. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.05.027>
225. Qi P. Qin J. Xie C. (2015). Determination of genetic diversity of wild and cultured topmouth culter (*Culter alburnus*) inhabiting China using mitochondrial DNA and microsatellites. *Biochemical Systematics and Ecology* <http://doi.org/10.1016/j.bse.2015.06.023>

226. Radhakrishnan D.P. Nedumpally V. Kathirvelpandian A. Valaparambil Saidmuhammed B. Gopalakrishnan A. (2018). Population structure of Spanish mackerel *Scomberomorus commerson* (Lacepede 1800) in the Northern Indian Ocean determined using microatellite markers. *Aquatic Living Resources* <http://doi.org/10.1051/alr/2018011>
227. Rahimi P. Rezvani Gilkoliae S. Ghavam Mostafavi P. Jamili Sh. Rahnema M. (2016). Population genetic structure of the white sardine. *Iranian Journal of Fisheries Sciences* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84992660470&partnerID=40&md5=cd8270125f463a00b43047d4bd297a20>
228. Ravago-Gotanco R. Kim K.M. (2019). Regional genetic structure of sandfish *Holothuria* (*Metriatyala*) *scabra* populations across the Philippine archipelago. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2018.09.021>
229. Reid K. Palkovacs E.P. Hasselman D.J. Baetscher D. Kibele J. Gahagan B. Bentzen P. McBride M.C. Garza J.C. (2018). Comprehensive evaluation of genetic population structure for anadromous river herring with single nucleotide polymorphism data. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2018.04.014>
230. Renán X. Montero-Muñoz J. Garza-Pérez J.R. Brulé T. (2016). Age and Stock Analysis Using Otolith Shape in Gags from the Southern Gulf of Mexico. *Transactions of the American Fisheries Society* <http://doi.org/10.1080/00028487.2016.1217928>
231. Revaldaves E. Renesto E. Gold J.R. (2016). Genetic variation of *Prochilodus lineatus* (Valenciennes). *Genetics and Molecular Research* <http://doi.org/10.4238/gmr.15048829>
232. Riekkola L. Zerbini A.N. Andrews O. Andrews-Goff V. Baker C.S. Chandler D. Childerhouse S. Clapham P. Dodémont R. Donnelly D. Friedlaender A. Gallego R. Garrigue C. Ivashchenko Y. Jarman S. Lindsay R. Pallin L. Robbins J. Steel D. Tremlett J. Vindenes S. Constantine R. (2018). Application of a multi-disciplinary approach to reveal population structure and Southern Ocean feeding grounds of humpback whales. *Ecological Indicators* <http://doi.org/10.1016/j.ecolind.2018.02.030>
233. Rodrigues A.D.S. Brandão J.H.S.G. Bitencourt J.D.A. Jucá-Chagas R. Sampaio I. Schneider H. Affonso P.R.A.D.M. (2016). Molecular identification and traceability of illegal trading in lignobrycon myersi (Teleostei: Characiformes). *Scientific World Journal* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84988944375&doi=10.1155%2f2016%2f9382613&partnerID=40&md5=49d0c416a868c0d6c7a342059b1f30ba>
234. Sabatini L. Bullo M. Cariani A. Celić I. Ferrari A. Guarniero I. Leoni S. Marčeta B. Marcone A. Polidori P. Raicevich S. Tinti F. Vrgoč N. Scarcella G. (2018). Good practices for common sole assessment in the Adriatic Sea: Genetic and morphological differentiation of *Solea solea* (Linnaeus, 1758) from *S. aegyptiaca* (Chabanaud, 1927) and stock identification. *Journal of Sea Research* <http://doi.org/10.1016/j.seares.2018.04.004>
235. Sahoo L. Mohanty M. Meher P.K. Murmu K. Sundaray J.K. Das P. (2019). Population structure and genetic diversity of hatchery stocks as revealed by combined mtDNA fragment sequences in Indian major carp *Catla catla*. *Mitochondrial DNA Part A: DNA Mapping* <http://doi.org/10.1080/24701394.2018.1484120>
236. Saint-Pé K. Blanchet S. Tissot L. Poulet N. Plasseraud O. Loot G. Veyssièrre C. Prunier J.G. (2018). Genetic admixture between captive-bred and wild individuals affects patterns of dispersal in a brown trout (*Salmo trutta*) population. *Conservation Genetics* <http://doi.org/10.1007/s10592-018-1095-2>
237. Saitoh K. Suzuki N. Ozaki M. Ishii K. Sado T. Morosawa T. Tsunagawa T. Tsuchiya M. (2017). Natural habitats uncovered?-Genetic structure of known and newly found localities of the endangered bitterling *Pseudorhodeus tanago* (Cyprinidae). *Nature Conservation* <http://doi.org/10.3897/natureconservation.17.10939>
238. Samani N.K. Esa Y. Amin S.M.N. Ikhsan N.F.M. (2016). Phylogenetics and population genetics of *Plotosus canius* (Siluriformes: Plotosidae) from Malaysian coastal waters. *PeerJ* <http://doi.org/10.7717/peerj.1930>
239. Santos C.H.A. Santana G.X. Sá Leitão C.S. Paula-Silva M.N. Almeida-Val V.M.F. (2016). Loss of genetic diversity in farmed populations of *Colossoma macropomum* estimated by microsatellites. *Animal Genetics* <http://doi.org/10.1111/age.12422>
240. Sather N.K. Johnson G.E. Teel D.J. Storch A.J. Skalski J.R. Cullinan V.I. (2016). Shallow Tidal Freshwater Habitats of the Columbia River: Spatial and Temporal Variability of Fish Communities and

- Density. *Transactions of the American Fisheries Society*
<http://doi.org/10.1080/00028487.2016.1150878>
241. Satterthwaite W.H. Cordoleani F. O'Farrell M.R. Kormos B. Mohr M.S. (2018). Central Valley spring-run Chinook Salmon and ocean fisheries: Data availability and management possibilities. *San Francisco Estuary and Watershed Science* <http://doi.org/10.15447/sfews.2018v16iss1/art4>
 242. Savoy T. Maceda L. Roy N.K. Peterson D. Wirgin I. (2017). Evidence of natural reproduction of Atlantic sturgeon in the Connecticut River from unlikely sources. *PLoS ONE*
<http://doi.org/10.1371/journal.pone.0175085>
 243. Scribner K. Howell P. Thomas M. Smith K. Hanchin P. Wolgamood M. Whelan G. (2015). Spatial genetic structure of suspected remnant and naturalized populations of muskellunge and evidence for introgression between stocked and native strains. *Journal of Great Lakes Research*
<http://doi.org/10.1016/j.jglr.2015.08.011>
 244. Sebastian W. Sukumaran S. Zacharia P.U. Gopalakrishnan A. (2017). Genetic population structure of Indian oil sardine. *Conservation Genetics* <http://doi.org/10.1007/s10592-017-0946-6>
 245. Seyoum S. Collins A.B. Puchulutegui C. McBride R.S. Tringali M.D. (2015). Genetically determined population structure of hogfish (*Labridae: Lachnolaimus maximus*) in the southeastern United States. *Fishery Bulletin* <http://doi.org/10.7755/FB.113.4.7>
 246. Seyoum S. McBride R.S. Tringali M.D. Villanova V.L. Puchutulegui C. Gray S. Van Bibber N. (2018). Genetic population structure of the spotted seatrout (*Cynoscion nebulosus*): Simultaneous examination of the mtDNA control region and microsatellite marker results. *Bulletin of Marine Science*
<http://doi.org/10.5343/bms.2017.1060>
 247. Shamblin B.M. Dutton P.H. Bjorndal K.A. Bolten A.B. Naro-Maciel E. Santos A.J.B. Bellini C. Baptistotte C. Marcovaldi M.Â. Nairn C.J. (2015). Deeper Mitochondrial Sequencing Reveals Cryptic Diversity and Structure in Brazilian Green Turtle Rookeries. *Chelonian Conservation and Biology*
<http://doi.org/10.2744/CCB-1152.1>
 248. Sharifi M. Sourinejad I. Hosseini S.J. Qasemi S.A. (2015). Application of AFLP molecular marker for genetic analysis of black pomfret *Parastromateus niger* from the Persian Gulf. *Iranian Journal of Fisheries Sciences* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84945956370&partnerID=40&md5=f48e0794d8db66dd8cdf0729ed80439>
 249. Sharpe D.L. Castellote M. Wade P.R. Cornick L.A. (2019). Call types of Bigg's killer whales (*Orcinus orca*) in western Alaska: using vocal dialects to assess population structure. *Bioacoustics*
<http://doi.org/10.1080/09524622.2017.1396562>
 250. Sherman K.D. Dahlgren C.P. Stevens J.R. Tyler C.R. (2016). Integrating population biology into conservation management for endangered Nassau grouper *Epinephelus striatus*. *Marine Ecology Progress Series* <http://doi.org/10.3354/meps11771>
 251. Shokoohmand M. Zolgharnein H. Mashjoor S. Laloi F. Foroughmand A.M. Savari A. (2018). Analysis of genetic diversity of white shrimp (*Metapenaeus Affinis*) from the northwest of the Persian Gulf using microsatellite markers. *Turkish Journal of Fisheries and Aquatic Sciences*
http://doi.org/10.4194/1303-2712-v18_3_04
 252. Singh N.S. Behera B.K. Kunal S.P. Das P. Paria P. Sharma A.P. (2016). Genetic stock structure of *Osteobrama belangeri* (Valenciennes. Mitochondrial DNA
<http://doi.org/10.3109/19401736.2014.883602>
 253. Spaet J.L.Y. Jabado R.W. Henderson A.C. Moore A.B.M. Berumen M.L. (2015). Population genetics of four heavily exploited shark species around the Arabian Peninsula. *Ecology and Evolution*
<http://doi.org/10.1002/ece3.1515>
 254. Spies I. Punt A.E. (2015). The utility of genetics in marine fisheries management: A simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*
<http://doi.org/10.1139/cjfas-2014-0050>
 255. Srivastava S. Kushwaha B. Prakash J. Pandey M. Agarwal S. Kumar R. Nagpure N.S. Singh M. Das P. Joshi C. Jena J.K. (2017). Identification and characterization of SSRs in *Clarias batrachus* and their application in population study. *Fisheries Science* <http://doi.org/10.1007/s12562-017-1066-4>
 256. Su J. Ji W. Zhang Y. Gleeson D.M. Lou Z. Ren J. Wei Y. (2015). Genetic diversity and demographic history of the endangered and endemic fish (*Platypharodon extremus*): implications for stock enhancement in Qinghai Tibetan Plateau. *Environmental Biology of Fishes*
<http://doi.org/10.1007/s10641-014-0310-3>

257. Sukumaran S. Sebastian W. Gopalakrishnan A. (2016). Population genetic structure of Indian oil sardine. *Gene* <http://doi.org/10.1016/j.gene.2015.10.043>
258. Sultana F. Abbas K. Xiaoyun Z. Abdullah S. Qadeer I. Hussnain R. (2015). Microsatellite markers reveal genetic degradation in hatchery stocks of *Labeo rohita*. *Pakistan Journal of Agricultural Sciences* <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84944038073&partnerID=40&md5=52942f771d8ce5f978e7c0f8f534d660>
259. Sungani H. Ngatunga B.P. Genner M.J. (2016). Migratory behaviour shapes spatial genetic structure of cyprinid fishes within the Lake Malawi catchment. *Freshwater Biology* <http://doi.org/10.1111/fwb.12767>
260. Taillebois L. Barton D.P. Crook D.A. Saunders T. Taylor J. Hearnden M. Saunders R.J. Newman S.J. Travers M.J. Welch D.J. Greig A. Dudgeon C. Maher S. Ovenden J.R. (2017). Strong population structure deduced from genetics. *Evolutionary Applications* <http://doi.org/10.1111/eva.12499>
261. Taillebois L. Dudgeon C. Maher S. Crook D.A. Saunders T.M. Barton D.P. Taylor J.A. Welch D.J. Newman S.J. Travers M.J. Saunders R.J. Ovenden J. (2016). Characterization. *PeerJ* <http://doi.org/10.7717/PEERJ.2418>
262. Takács P. Eros T. Specziár A. Sály P. Vítál Z. Ferincz Á. Molnár T. Szabolcsi Z. Bíró P. Csoma E. (2015). Population genetic patterns of threatened European mudminnow (*Umbra krameri* Walbaum). *PLoS ONE* <http://doi.org/10.1371/journal.pone.0138640>
263. Taylor A.T. Long J.M. Schwemm M.R. Brewer S.K. (2018). Hybridization and Genetic Structure of Neosho Smallmouth Bass in the Ozark Highlands. *North American Journal of Fisheries Management* <http://doi.org/10.1002/nafm.10225>
264. Telli M. Gürleyen B. (2017). Genetic and morphological divergences between wild and captive-bred populations of *Salmo trutta abanticus*. *Aquaculture Research* <http://doi.org/10.1111/are.13384>
265. Thorburn J. Jones R. Neat F. Pinto C. Bendall V. Hetherington S. Bailey D.M. Leslie N. Jones C. (2018). Spatial versus temporal structure: Implications of inter-haul variation and relatedness in the North-east Atlantic spurdog *Squalus acanthias*. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.2922>
266. Tian C.-X. He Y.-H. Liang X.-F. Guo W.-J. Sun L.-F. Lv L.-Y. He S. (2017). Population genetics of wild *Siniperca kneri* Garman. *Journal of Applied Ichthyology* <http://doi.org/10.1111/jai.13412>
267. Tikochinski Y. Bradshaw P. Mastrogiacomo A. Broderick A. Daya A. Demetropoulos A. Demetropoulos S. Eliades N.-G. Fuller W. Godley B. Kaska Y. Levy Y. Snape R. Wright L. Carreras C. (2018). Mitochondrial DNA short tandem repeats unveil hidden population structuring and migration routes of an endangered marine turtle. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.2908>
268. Timi J.T. MacKenzie K. (2015). Parasites in fisheries and mariculture. *Parasitology* <http://doi.org/10.1017/S0031182014001188>
269. Tolve L. Casale P. Formia A. Garofalo L. Lazar B. Natali C. Novelletto A. Vallini C. Bužan E. Chelazzi G. Gaspari S. Fortuna C. Kocijan I. Marchiori E. Novarini N. Poppi L. Salvemini P. Ciofi C. (2018). A comprehensive mitochondrial DNA mixed-stock analysis clarifies the composition of loggerhead turtle aggregates in the Adriatic Sea. *Marine Biology* <http://doi.org/10.1007/s00227-018-3325-z>
270. Tomano S. Sanchez G. Ueno K. Ueta Y. Ohara K. Umino T. (2015). Microsatellite DNA variation of oval squid *Sepioteuthis* sp. 2 reveals a single fishery stock on the coastline of mainland Japan. *Fisheries Science* <http://doi.org/10.1007/s12562-015-0905-4>
271. Toomey L. Welsford D. Appleyard S.A. Polanowski A. Faux C. Deagle B.E. Belchier M. Marthick J. Jarman S. (2016). Genetic structure of Patagonian toothfish populations from otolith DNA. *Antarctic Science* <http://doi.org/10.1017/S0954102016000183>
272. Torati L.S. Taggart J.B. Varela E.S. Araripe J. Wehner S. Migaud H. (2019). Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genetics* <http://doi.org/10.1186/s12863-018-0711-y>
273. Trojette M. Ben Faleh A. Fatnassi M. Marsaoui B. Mahouachi N.E.H. Chalh A. Quignard J.-P. Trabelsi M. (2015). Stock discrimination of two insular populations of *diplodus annularis* (Actinopterygii: Perciformes: Sparidae) along the coast of Tunisia by analysis of otolith shape. *Acta Ichthyologica et Piscatoria* <http://doi.org/10.3750/AIP2015.45.4.04>
274. Turner S.M. Limburg K.E. Palkovacs E.P. (2015). Can different combinations of natural tags identify river herring natal origin at different levels of stock structure?. *Canadian Journal of Fisheries and Aquatic Sciences* <http://doi.org/10.1139/cjfas-2014-0403>

275. Vähä J.-P. Erkinaro J. Falkegård M. Orell P. Niemelä E. (2017). Genetic stock identification of Atlantic salmon and its evaluation in a large population complex. *Canadian Journal of Fisheries and Aquatic Sciences* <http://doi.org/10.1139/cjfas-2015-0606>
276. Van Doornik D.M. Hess M.A. Johnson M.A. Teel D.J. Friesen T.A. Myers J.M. (2015). Genetic Population Structure of Willamette River Steelhead and the Influence of Introduced Stocks. *Transactions of the American Fisheries Society* <http://doi.org/10.1080/00028487.2014.982178>
277. Vargas S.M. Jensen M.P. Ho S.Y.W. Mobaraki A. Broderick D. Mortimer J.A. Whiting S.D. Miller J. Prince R.I.T. Bell I.P. Hoenner X. Limpus C.J. Santos F.R. Fitzsimmons N.N. (2016). Phylogeography. *Journal of Heredity* <http://doi.org/10.1093/jhered/esv091>
278. Vargas S.M. Lins L.S.F. Molfetti E. Ho S.Y.W. Monteiro D. Barreto J. Colman L. Vila-Verde L. Baptistotte C. Thomé J.C.A. Santos F.R. (2019). Revisiting the genetic diversity and population structure of the critically endangered leatherback turtles in the South-west Atlantic Ocean: Insights for species conservation. *Journal of the Marine Biological Association of the United Kingdom* <http://doi.org/10.1017/S002531541700193X>
279. Vargas-Caro C. Bustamante C. Bennett M.B. Ovenden J.R. (2017). Towards sustainable fishery management for skates in South America: The genetic population structure of *Zearaja chilensis* and *Dipturus trachyderma* (Chondrichthyes). *PLoS ONE* <http://doi.org/10.1371/journal.pone.0172255>
280. Vasconcellos A.V.D. Lima D. Bonhomme F. Vianna M. Solé-Cava A.M. (2015). Genetic population structure of the commercially most important demersal fish in the Southwest Atlantic: The whitemouth croaker (*Micropogonias furnieri*). *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.03.008>
281. Vella A. Vella N. Karakulak F.S. Oray I. Garcia-Tiscar S. de Stephanis R. (2016). Population genetics of Atlantic bluefin tuna. *Journal of Applied Ichthyology* <http://doi.org/10.1111/jai.13035>
282. Vera M. Martinez P. Bouza C. (2018). Stocking impact. *Aquatic Conservation: Marine and Freshwater Ecosystems* <http://doi.org/10.1002/aqc.2856>
283. Vergara-Chen C. Rodrigues F. González-Wangüemert M. (2015). Population genetics of *Cerastoderma edule* in Ria Formosa (southern Portugal): The challenge of understanding an intraspecific hotspot of genetic diversity. *Journal of the Marine Biological Association of the United Kingdom* <http://doi.org/10.1017/S0025315414001313>
284. Vineesh N. Divya P. Kathirvelpandian A. Mohitha C. Shanis C. Basheer V. Gopalakrishnan A. (2018). Four evolutionarily significant units among narrow-barred Spanish mackerel (*Scomberomorus commerson*) in the Indo-West Pacific region. *Marine Biodiversity* <http://doi.org/10.1007/s12526-017-0714-3>
285. Vineesh N. Kathirvelpandian A. Divya P.R. Mohitha C. Basheer V.S. Gopalakrishnan A. Jena J.K. (2016). Hints for panmixia in *Scomberomorus commerson* in Indian waters revealed by mitochondrial ATPase 6 and 8 genes. *Mitochondrial DNA* <http://doi.org/10.3109/19401736.2015.1053084>
286. Viret A. Tsaparis D. Tsigenopoulos C.S. Berrebi P. Sabatini A. Arculeo M. Fassatoui C. Magoulas A. Marengo M. Morales-Nin B. Caill-Milly N. Durieux E.D.H. (2018). Absence of spatial genetic structure in common dentex (*Dentex dentex* Linnaeus, 1758) in the Mediterranean Sea as evidenced by nuclear and mitochondrial molecular markers. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0203866>
287. Vitorino C.A. Nogueira F. Souza I.L. Araripe J. Venere P.C. (2017). Low genetic diversity and structuring of the arapaima (*Osteoglossiformes*). *Frontiers in Genetics* <http://doi.org/10.3389/fgene.2017.00159>
288. Vollmer N.L. Rosel P.E. (2017). Fine-scale population structure of common bottlenose dolphins (*Tursiops truncatus*) in offshore and coastal waters of the US Gulf of Mexico. *Marine Biology* <http://doi.org/10.1007/s00227-017-3186-x>
289. Wallace E.M. (2015). High intraspecific genetic connectivity in the Indo-Pacific bonefishes: implications for conservation and management. *Environmental Biology of Fishes* <http://doi.org/10.1007/s10641-015-0416-2>
290. Wang L. Liu S. Zhuang Z. Lin H. Meng Z. (2015). Mixed-stock analysis of small yellow croaker *Larimichthys polyactis* providing implications for stock conservation and management. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2014.06.006>
291. Wang L. Wan Z.Y. Lim H.S. Yue G.H. (2015). Genetic heterogeneity and local adaptation of Asian seabass across Indonesian Archipelago revealed with gene-associated SNP markers. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.06.012>

292. Wang Q. Wang X. Xie Z. Li Y. Xiao L. Peng C. Zhang H. Li S. Zhang Y. Lin H. (2016). Microsatellite analysis of the genetic relationships between wild and cultivated giant grouper in the South China Sea. *Journal of Genetics* <http://doi.org/10.1007/s12041-016-0647-9>
293. Wang V.H. McCartney M.A. Scharf F.S. (2015). Population genetic structure of southern flounder inferred from multilocus DNA profiles. *Marine and Coastal Fisheries* <http://doi.org/10.1080/19425120.2015.1037473>
294. Was A. Bernaś R. (2016). Long-term and seasonal genetic differentiation in wild and enhanced stocks of sea trout (*Salmo trutta m. trutta* L.) from the Vistula River. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2015.11.006>
295. Watson H.V. McKeown N.J. Coscia I. Wootton E. Ironside J.E. (2016). Population genetic structure of the European lobster (*Homarus gammarus*) in the Irish Sea and implications for the effectiveness of the first British marine protected area. *Fisheries Research* <http://doi.org/10.1016/j.fishres.2016.06.015>
296. Welch D.J. Newman S.J. Buckworth R.C. Ovenden J.R. Broderick D. Lester R.J.G. Gribble N.A. Ballagh A.C. Charters R.A. Stapley J. Street R. Garrett R.N. Begg G.A. (2015). Integrating different approaches in the definition of biological stocks: A northern Australian multi-jurisdictional fisheries example using grey mackerel. *Marine Policy* <http://doi.org/10.1016/j.marpol.2015.01.010>
297. Westgaard J.-I. Saha A. Kent M. Hansen H.H. Knutsen H. Hauser L. Cadrin S.X. Albert O.T. Johansen T. (2017). Genetic population structure in greenland halibut (*Reinhardtius hippoglossoides*) and its relevance to fishery management. *Canadian Journal of Fisheries and Aquatic Sciences* <http://doi.org/10.1139/cjfas-2015-0430>
298. Westgaard J.-I. Staby A. Aanestad Godiksen J. Geffen A.J. Svensson A. Charrier G. Svedäng H. André C. (2017). Large and fine scale population structure in European hake (*Merluccius merluccius*) in the Northeast Atlantic. *ICES Journal of Marine Science* <http://doi.org/10.1093/icesjms/fsw249>
299. Williams R.C. Jackson B.C. Duvaux L. Dawson D.A. Burke T. Sinclair W. (2015). The genetic structure of *Nautilus pompilius* populations surrounding Australia and the Philippines. *Molecular Ecology* <http://doi.org/10.1111/mec.13255>
300. Williams S.M. Bennett M.B. Pepperell J.G. Morgan J.A.T. Ovenden J.R. (2016). Spatial genetic subdivision among populations of the highly migratory black marlin *Istiompax indica* within the central Indo-Pacific. *Marine and Freshwater Research* <http://doi.org/10.1071/MF14370>
301. Willis S.C. Winemiller K.O. Montaña C.G. Macrander J. Reiss P. Farias I.P. Ortí G. (2015). Population genetics of the speckled peacock bass (*Cichla temensis*). *Conservation Genetics* <http://doi.org/10.1007/s10592-015-0744-y>
302. Wilson C.C. Liskauskas A.P. Wozney K.M. (2016). Pronounced Genetic Structure and Site Fidelity among Native Muskellunge Populations in Lake Huron and Georgian Bay. *Transactions of the American Fisheries Society* <http://doi.org/10.1080/00028487.2016.1209556>
303. Winans G.A. Gayeski N. Timmins-Schiffman E. (2015). All dam-affected trout populations are not alike: fine scale geographic variability in resident rainbow trout in Icicle Creek. *Conservation Genetics* <http://doi.org/10.1007/s10592-014-0659-z>
304. Wisenden B.D. Stumbo A.D. McEwen D.C. McIntire K. Scheierl J. Aasand J. North H. Gilbertson J. Grant D. Joseph F. Mammenga E. Walsh R. Brisch E. (2016). Population-specific co-evolution of offspring anti-predator competence and parental brood defence in Nicaraguan convict cichlids. *Environmental Biology of Fishes* <http://doi.org/10.1007/s10641-016-0476-y>
305. Xiao Y. Guan S. Liu Q. Liu H. Yu D. Ma D. Xu S. Liu J. Dai M. Xiao Z. Li J. (2018). Genetic characteristics of broodstock and offspring of the seven-band grouper (*Hyporthodus septemfasciatus*) using fluorescent-AFLP markers. *Journal of the Marine Biological Association of the United Kingdom* <http://doi.org/10.1017/S0025315416001260>
306. Xiao Y. Ma D. Xu S. Liu Q. Wang Y. Xiao Z. Li J. (2016). Significant population genetic structure detected in the rock bream *Oplegnathus fasciatus* (Temminck & Schlegel). *Chinese Journal of Oceanology and Limnology* <http://doi.org/10.1007/s00343-016-4404-y>
307. Xu D. Lou B. Zhou W. Chen R. Zhan W. Liu F. (2017). Genetic diversity and population differentiation in the yellow drum *Nibea albiflora* along the coast of the China Sea. *Marine Biology Research* <http://doi.org/10.1080/17451000.2016.1274033>
308. Xu L. Li Q. Xu C. Yu H. Kong L. (2019). Genetic diversity and effective population size in successive mass selected generations of black shell strain Pacific oyster (*Crassostrea gigas*) based on microsatellites and mtDNA data. *Aquaculture* <http://doi.org/10.1016/j.aquaculture.2018.10.007>

309. Xu R. Bo Q. Zheng X. (2018). A divergent lineage among *Octopus minor* (Sasaki. Marine Biology Research <http://doi.org/10.1080/17451000.2018.1427866>
310. Xue D.-X. Yang Q.-L. Li Y.-L. Zong S.-B. Gao T.-X. Liu J.-X. (2019). Comprehensive assessment of population genetic structure of the overexploited Japanese grenadier anchovy (*Coilia nasus*): Implications for fisheries management and conservation. Fisheries Research <http://doi.org/10.1016/j.fishres.2019.01.012>
311. Yan R.-J. Zhang G.-R. Guo X.-Z. Ji W. Chen K.-C. Zou G.-W. Wei K.-J. Gardner J.P.A. (2018). Genetic diversity and population structure of the northern snakehead (*Channa argus* Channidae: Teleostei) in central China: implications for conservation and management. Conservation Genetics <http://doi.org/10.1007/s10592-017-1023-x>
312. Yanagimoto T. Sakai T. Ochi Y. Ebina Y. Fujino T. (2015). Population structure of *alfonsin Beryx splendens* inferred from nucleotide sequences of the mitochondrial control region. Nippon Suisan Gakkaishi (Japanese Edition) <http://doi.org/10.2331/suisan.81.958>
313. Yang J.M. Sun G.H. Zheng X.D. Ren L.H. Wang W.J. Li G.R. Sun B.C. (2015). Genetic differentiation of *Octopus minor* (Mollusca. Genetics and Molecular Research <http://doi.org/10.4238/2015.December.1.13>
314. Yang J.P. Li X.H. Li Y.F. Li J. Shuai F.M. Zhu S.L. (2016). Complete mitochondrial genome of *Ochetobius elongatus* (Cyprinidae. Mitochondrial DNA <http://doi.org/10.3109/19401736.2015.1089542>
315. Yang M. Tian C. Liang X.-F. Zheng H. Zhao C. Zhu K. (2015). Genetic structure and diversity in natural and stocked populations of the mandarin fish (*Siniperca chuatsi*) in China. Genetics and Molecular Research <http://doi.org/10.4238/2015.May.18.5>
316. Ye Y.Y. Senanan W. Li J.J. Cai H.W. Wu C.W. (2016). Genetic diversity and population structure of the Asian green mussel *Perna viridis* in South China Sea based on microsatellite markers. Biochemical Systematics and Ecology <http://doi.org/10.1016/j.bse.2016.07.010>
317. Yellapu B. Jeffs A. Battaglione S. Lavery S.D. (2017). Population subdivision in the tropical spiny lobster *Panulirus ornatus* throughout its Indo-West Pacific distribution. ICES Journal of Marine Science <http://doi.org/10.1093/icesjms/fsw184>
318. Yu Y. Sun P. Cui H. Sheng H. Zhao F. Tang Y. Chen Z. (2015). Effects of trawl selectivity and genetic parameters on fish body length under long-term trawling. Journal of Ocean University of China <http://doi.org/10.1007/s11802-015-2885-5>
319. Zhang D. Wang L. Guo H. Ma Z. Jiang S. (2016). The complete mitochondrial genome of snubnose pompano *Trachinotus blochii* (Teleostei. Mitochondrial DNA <http://doi.org/10.3109/19401736.2014.898291>
320. Zhang Y. Yang F. Wang Z. You Q. Lou B. Xu D. Chen R. Zhan W. Liu F. (2017). Mitochondrial DNA variation and population genetic structure in the small yellow croaker at the coast of Yellow Sea and East China Sea. Biochemical Systematics and Ecology <http://doi.org/10.1016/j.bse.2017.03.003>
321. Zhao H. Fuller A. Thongda W. Mohammed H. Abernathy J. Beck B. Peatman E. (2019). SNP panel development for genetic management of wild and domesticated white bass (*Morone chrysops*). Animal Genetics <http://doi.org/10.1111/age.12747>
322. Zhao J. Hsu K.-C. Luo J.-Z. Wang C.-H. Chan B.-P. Li J. Kuo P.-H. Lin H.-D. (2018). Genetic diversity and population history of *Tanichthys albonubes* (Teleostei: Cyprinidae): Implications for conservation. Aquatic Conservation: Marine and Freshwater Ecosystems <http://doi.org/10.1002/aqc.2840>
323. Zheng W. Zou L. Han Z. (2015). Genetic analysis of the populations of Japanese anchovy *Engraulis japonicus* from the Yellow Sea and East China Sea based on mitochondrial cytochrome b sequence. Biochemical Systematics and Ecology <http://doi.org/10.1016/j.bse.2014.12.007>

Appendix 2

This appendix reports the relevant papers considered. For each paper is reported: Doi, organism scientific name, taxonomic group, and marker used. In addition, if appropriate, the software used for Fst/AMOVA/Similar analysis, for clustering, to estimate migration or mixed stock analysis, for ABC, and for outliers detection is reported together with relevant notes. In the case of papers using different kind of markers, the record is duplicated.

- <http://doi.org/10.1016/j.ecolind.2018.02.030>; Megaptera novaeangliae; Mammal; microsatellites; Cervus; Notes: 15 loci (but 11 loci were used for analyses)
- <http://doi.org/10.1016/j.ecolind.2018.02.030>; Megaptera novaeangliae; Mammal; mtDNA; Arlequin (Fst); SPAM 3.7; Notes: 470 bp CR
- http://doi.org/10.4194/1303-2712-v18_3_04; Metapenaeus affinis; Crustacean; microsatellites; Genepop (Fst, Rst); Genalex (AMOVA); MEGA; Notes: 5 loci
- http://doi.org/10.4194/1303-2712-v18_9_10; Hypophthalmichthys molitrix; Fish; microsatellites; Arlequin (AMOVA); Notes: 16 loci
- <http://doi.org/10.5343/bms.2017.1060>; Cynoscion nebulosus; Fish; mtDNA; Arlequin(AMOVA); PCA; Notes: 335 bp CR
- <http://doi.org/10.5343/bms.2017.1060>; Cynoscion nebulosus; Fish; microsatellites; Arlequin(AMOVA); FCA, PCA (on Da calculated with poptree), STRUCTURE, STRUCTURE HARVESTER, CLUMPP; Notes: 38 loci
- <http://doi.org/10.1016/j.ecss.2017.08.032>; Donax vittatus; Molluscs; mtDNA; Arlequin(AMOVA); Notes: 550 bp COI, 400 bp COI, 447 bp 16S
- <http://doi.org/10.1016/j.ecss.2017.08.032>; Donax vittatus; Molluscs; nDNA; Arlequin(AMOVA); Notes: 350 bp H3, 1700 bp 18S, 700 bp 28S
- <http://doi.org/>; Salmo salar; Fish; microsatellites; Genodive(Fst); STRUCTURE(parallelstructure); ONCOR, CBAYES; GSI(ONCOR,CBAYES) Notes: 31 loci
- <http://doi.org/10.11693/hyhz20161100238>; URECHIS UNICINCTUS; Anellida; microsatellites; Fst; Notes: 22 loci
- structure with estimated effective migration surfaces. Nat Genet. 2016;48:94–100."); Notes: 16 loci
- structure with estimated effective migration surfaces. Nat Genet. 2016;48:94–100."); Notes: 441(187) loci ddRAD
- <http://doi.org/10.7717/peerj.3786>; Lysmata boggessi; Crustacean; mtDNA; Arlequin(AMOVA); Notes: 554 bp 16S
- <http://doi.org/>; Nemipterus japonicus; Fish; mtDNA; Arlequin(AMOVA), BARRIER; Notes: 665 bp COI
- <http://doi.org/10.1016/j.gene.2015.10.043>; Sardinella longiceps; Fish; mtDNA; Arlequin (Phist, AMOVA); Arlequin (mismatch distribution), DnaSP (Tajima and Fu), XLStat (IBD) Notes: 758 bp CR, 576 bp COI
- <http://doi.org/10.21077/ijf.2016.63.4.54028-10>; Scomberomorus commerson; Fish; microsatellites; GenAlEx (AMOVA); Micro-checker Notes: 10 loci
- <http://doi.org/10.6620/ZS.2016.55-33>; Makaira nigricans; Fish; mtDNA; Arlequin(AMOVA), SAMOVA; Notes: 1140 bp cyt b, 905 bp CR
- <http://doi.org/10.7717/peerj.1930>; Plotosus canius ; Fish; mtDNA; Arlequin(AMOVA); Notes: 655 bp COI
- <http://doi.org/10.7717/peerj.1930>; Plotosus canius ; Fish; microsatellites; Arlequin(AMOVA); STRUCTURE; GeneClass; bottleneck Notes: 5 loci
- <http://doi.org/10.17221/8847-CJAS>; Sander lucioperca; Fish; microsatellites; Genetix(Fst); STRUCTURE; GeneClass; Notes: 9 loci
- <https://doi.org/10.15447/sfews.2015v13iss4art3>; Oncorhynchus mykiss; Fish; Microsats & SNPs; HP-RARE, GENETIX, GENEPOP ; STRUCTURE 2.2, GSI_SIM, PHYLIP; Notes: Separate analyses for SNPs are not specified

<http://doi.org/10.7755/FB.113.4.7>; *Lachnolaimus maximus*; Fish; Microsats; GENEPOP 4.3, GENETIX, ML-NULLFREQ, FSTAT 2.9.3.2, NeESTIMATOR 2; GENALEX, POPTREE2, ARLEQUIN 3.5.1.3, STRUCTURE 2.3.4; STRUCTURE HARVESTER, CLUMPP 1.1.2 Notes: 24 msats

<http://doi.org/10.1002/mcf2.10055>; *Micropogonias undulatus*; Fish; SNPs ; Arlequin 3.5 (AMOVA, Fst); STRUCTURE, DAPC; FDIST, Bayescan; Notes: 3682 SNPs, AMOVA locus by locus, qualitative recovery of Ks (no structureharvester)

<http://doi.org/10.1002/mcf2.10055>; *Micropogonias undulatus*; Fish; mtDNA; Arlequin 3.5 (AMOVA, PhiST); Haplotype cluster membership; Notes: 424 bp of control region

<http://doi.org/10.1002/aqc.2999>; *Eretmochelys imbricata*; Reptiles; mtDNA; Arlequin 3.5 (Fst, SAMOVA); mixstock R package; Notes: 740 bp D-loop

<http://doi.org/10.1002/aqc.3008>; *Salmo salar*; Fish; nuclear sequencing; Newtorks; Notes: SDY (main male sex determining gene, on Y chromosome), first intron, 200 bp, no pop genetics

<http://doi.org/10.1002/aqc.3011>; *Palaemon serratus*; Crustacean; microsatellites; Arlequin 3.5 (AMOVA), GENEPOP (IBD); STRUCTURE, STRUCTURE HARVESTER, CLUMPP 1.1.2, STRUCTURE PLOT ; FREENA Notes: 17 loci

<http://doi.org/10.1002/nafm.10225>; *Micropterus dolomieu velox*; Fish; microsatellites; GenALEX, FSTAT, Arlequin (Fst, AMOVA), MMOD (Dst); Structure, DAPC, CLUMPP; NewHybrids, COANCESTRY Notes: 7 loci

<http://doi.org/10.1002/aqc.2922>; *Squalus acanthias*; Fish; microsatellites; Diversity, Genepop, Arlequin, Genetix; STRUCTURE, GenAlex, MLRelate; FSTAT (sex-biased dispersal); Notes: 7 loci

<http://doi.org/10.1002/aqc.2922>; *Squalus acanthias*; Fish; mtDNA; MEGA 6, DnaSP, Arlequin; PhyML, HAPVIEW, Arlequin (Fu and Tajima tests) Notes: 828 bp CR

<http://doi.org/10.1002/aqc.2932>; *Scophthalmus maximus* and *Scophthalmus rhombus*; Fish; SNPs; Notes: ddRAD, 83 diagnostic SNPs for both species, but NO GENETICS IN

<http://doi.org/10.1002/aqc.2908>; *Chelonia mydas*; Reptiles; mtDNA; Arlequin 3.5 (FST and exact tests of population differentiation); Genealex 6.5 (PCoA); Bayes (Bayesian MSA) Notes: four AT short tandem repeats (STRs) in the mtDNA D-loop (control region)

<http://doi.org/10.1002/aqc.2840>; *Tanichthys albonubes*; Fish; mtDNA; DnaSP (Fst, Gst, Nst); STRUCTURE; DIYABC; DAMBE, MEGA, MrBayes, BEAST for phylogenetic analysis Notes: 2032 bp dLoop and cyt b

<http://doi.org/10.1002/aqc.2840>; *Tanichthys albonubes*; Fish; nDNA; STRUCTURE; DIYABC; DAMBE, MEGA, MrBayes, BEAST for phylogenetic analysis Notes: 2241 bp ENC1 and RAG1

<http://doi.org/10.1002/aqc.2840>; *Tanichthys albonubes*; Fish; microsatellites; Fstat, Genepop, Populations (Nei's Da distance); STRUCTURE; DIYABC; DAMBE, MEGA, MrBayes, BEAST for phylogenetic analysis Notes: 13 loci

<http://doi.org/10.1002/aqc.2856>; *Salmo trutta*; Fish; microsatellites; Fstat, Genepop, NeEstimator, Arlequin (AMOVA); STRUCTURE, Adegnet; Micro-checker Notes: 10 loci

<http://doi.org/10.1002/aqc.2856>; *Salmo trutta*; Fish; RFLP; STRUCTURE; Notes: LDH-C*

<http://doi.org/10.1002/aqc.2415>; *Tursiops truncatus*; Mammal; Microsats; GENALEX, GENEPOP 4.1, FSTAT 2.9.3.2, ; STRUCTURE 2.3.3; BAYESASS, MIGRATE 3.4.2; STRUCTURE-HARVESTER

<http://doi.org/10.1002/aqc.2415>; *Tursiops truncatus*; Mammal; mtDNA (control region); ARLEQUIN 3.5.2

<http://doi.org/10.1007/s00227-018-3446-4>; *Chelonia mydas*; Reptiles; mtDNA; Geneious Pro; BAYES; Satallite telemetry Notes: 770 bp CR

<http://doi.org/10.1007/s12526-017-0714-3>; *Scomberomorus commerson*; Fish; mtDNA; DnaSP, Arlequin (Fst, AMOVA); Arlequin (mismatch distribution, Tajima's and Fu's tests, Harpending raggedness index); MEGA (sequence divergence), PopART (haplotype network) Notes: CR

<http://doi.org/10.1007/s10592-018-1095-2>; *Salmo trutta*; Fish; microsatellites; MicroChecker, FSTAT, Genetix, Genepop, Adegnet (Fst); STRUCTURE, STRUCTURE HARVESTER, CLUMPP, DISTRICT; GeneClass 2, predict() R function; Lositan; Notes: 17 loci

<http://doi.org/10.1007/s12601-018-0009-z>; *Oplegnathus fasciatus*; Fish; mtDNA; Arlequin (Fst, AMOVA); MrBayes, TCS, Network, Arlequin (Fu and Tajima tests, mismatch distribution Notes: 642 bp COI

<http://doi.org/10.1007/s10592-017-1023-x>; *Channa argus*; Fish; microsatellites; GenALEX, Genepop, PopGene, Cervus, Populations (Da), Arlequin (AMOVA), NeEstimator; GenALEX (PCoA), STRUCTURE, STRUCTURE

HARVESTER, CLUMPAK, Adegnet (DAPC), AWclust 3.1, Alleles In Space, Barrier 2.2, SAMOVA; GeneClass 2.0, BayesAss 3.0; Lositan; BELS, PopTree, Bottleneck, Micro-checker Notes: 10 loci

<http://doi.org/10.1007/s00227-018-3325-z>; *Caretta caretta*; Reptiles; mtDNA; Arlequin (Phist; GenAlEx (PCoA); MixStock (MSA); PopART Notes: 755 bp CR

<http://doi.org/10.1007/s00227-018-3328-9>; *Chelonia mydas*; Reptiles; mtDNA; Arlequin (Fst), Geosphere (IBD); SAMOVA, GenAlEx (PCoA), BAYES; Notes: 770 bp CR

<http://doi.org/10.1007/s12041-018-0899-7>; *Lutjanus argentimaculatus*; Fish; mtDNA; DnaSP, Arlequin (Fst, AMOVA); Arlequin (mismatch distribution, Fu and Tajima tests), PopART Notes: 842 bp ATPase 6/8, 1105 bp cyt b

<http://doi.org/10.1007/s11033-017-4136-x>; *Siganus sutor*; Fish; microsatellites; Genetix, Arlequin (AMOVA); Bayescan, Lositan; Notes: 19 loci, cross amplification in 12 cogenetic

<http://doi.org/10.1007/s10499-017-0194-2>; *Cyclopterus lumpus*; Fish; microsatellites; Fst (Arlequin?); STRUCTURE; Notes: 14 loci

<http://doi.org/10.1007/s10750-017-3285-1>; *Leporinus piau*, *Megaleporinus reinhardti*, *Pimelodus maculatus*, *Prochilodus argenteus*, and *Pygocentrus piraya*; Fish; mtDNA; Arlequin (AMOVA); BSP, lamarck Notes: CR, Cytb, COI

<http://doi.org/10.1007/s10750-017-3285-1>; *Pimelodus maculatus*; Fish; microsatellites; Arlequin (AMOVA, Fst, Rst); Bottleneck Notes: 5 loci

<http://doi.org/10.1007/s10592-017-0967-1>; *Carcharhinus galapagensis*; Fish; mtDNA; Arlequin(AMOVA); Notes: 997 bp CR

<http://doi.org/10.1007/s10592-017-0967-1>; *Carcharhinus galapagensis*; Fish; SNPs; Adegnet(Fst); Netview; divmigrate; Lositan; Notes: 8103 loci DArT Seq

<http://doi.org/10.1007/s10592-017-0946-6>; *Sardinella longiceps*; Fish; microsatellites; Arlequin (AMOVA, Fst), Genepop (Rst), Smogd (Jost Dst), IBDWS (mantel test), ZT (mantel test); PCAGEN (PCA), STRUCTURE, STRUCTURE HARVESTER, STRUCTURAMA, BARRIER; Migrate; micro-checker, Powsim, Spagedi, Bottleneck Notes: 12 loci (6 loci used)

<http://doi.org/10.1007/s10592-017-0947-5>; *Prochilodus argenteus*; Fish; microsatellites; Arlequin (AMOVA, Fst); STRUCTURE; micro-checker Notes: 8 loci

<http://doi.org/10.1007/s00227-017-3186-x>; *Tursiops truncatus*; Mammal; mtDNA; Arlequin (AMOVA, Fst); MrBayes, Network Notes: CR

<http://doi.org/10.1007/s00227-017-3186-x>; *Tursiops truncatus*; Mammal; microsatellites; Arlequin (AMOVA, Fst), Genodive(G'st); STRUCTURE, ML-Relate; Adegnet (DAPC), Migrate; bayescan; Notes: 19 loci

<http://doi.org/10.1007/s00227-017-3186-x>; *Tursiops truncatus*; Mammal; SNPs; Arlequin (AMOVA, Fst), Genodive(G'st); STRUCTURE; Migrate; Phase Notes: 39 SNPs

<http://doi.org/10.1007/s10641-016-0547-0>; *Prochilodus nigricans*; Fish; mtDNA; AMOVA, SAMOVA, BARRIER; Notes: 834 bp CR, no access

<http://doi.org/10.1007/s12041-016-0647-9>; *Epinephelus lanceolatus*; Fish; microsatellites; Arlequin (Fst, AMOVA); Genetix (3D-FCA), Structure; Arlequin; Notes: 11 loci

<http://doi.org/10.1007/s00343-016-4404-y>; *Oplegnathus fasciatus*; Fish; AFLP; GenAlEx (Nei's distance), Arlequin (Fst, AMOVA); Structure; Notes: 1264 bands

<http://doi.org/10.1007/s10592-015-0797-y>; *Coregonus clupeaformis*; Fish; microsatellites; FSTAT (thetast), NeEstimator; Structure; Micro-checker, Bottleneck Notes: 20 loci

<http://doi.org/10.1007/s10592-015-0797-y>; *Coregonus clupeaformis*; Fish; SNPs; Lositan; Notes: 51 SNPs

<http://doi.org/10.1007/s12562-015-0958-4>; *Gnathopogon caeruleus*; Fish; mtDNA; Arlequin (Phist, AMOVA); TCS Notes: 417 bp Cyt b

<http://doi.org/10.1007/s10592-015-0770-9>; *Salmo trutta*; Fish; microsatellites; FSTAT (Fst), ISOLDE (IBD), HIERFSTAT (hierarchical Fst); GenAlEx (PCoA); GeneClass, Oncor; Micro-checker, Colony Notes: 10 loci

<http://doi.org/10.1007/s10750-015-2458-z>; *Thenus unimaculatus*; Crustacean; RAPD; PopGene (Gst)

<http://doi.org/10.1007/s10750-015-2458-z>; *Thenus unimaculatus*; Crustacean; mtDNA; Arlequin (Fst, AMOVA); Arlequin (Fu and Tajima test, mismatch distribution), Network Notes: 800 bp COI, 650 bp Cyt b

<http://doi.org/10.1007/s11033-016-3941-y>; *Silonia silondia*; Fish; microsatellites; Genetix (Fst); Notes: 11 loci
<http://doi.org/10.1007/s00227-010-1430-8>; Notes: review
<http://doi.org/10.1007/s10592-015-0744-y>; *Cichla temensis*; Fish; Microsats; GENODIVE 2.0b25, MICROCHECKER; STRUCTURE, ARLEQUIN; MIGRATE-N; DIY-ABC
<http://doi.org/10.1007/s10592-015-0744-y>; *Cichla temensis*; Fish; mtDNA (control region); MAFFT, DNASP, ARLEQUIN; BEAST 1.75, MrMODELEST
<http://doi.org/10.1007/s10592-015-0733-1>; *Alosa pseudoharengus*; Fish; Microsats; MICROCHECKER 2.2.3, GENEPOP 4.0.6, ARLEQUIN 3.1, FSTAT 2.9.3.2; POPTREE2, GENALEX-PCoA, STRUCTURE 2.3.3; COLONY 2; CLUMPP, DISTRUCT, ARCGIS 10.2, VEGAN-R Notes: 14 msat
<http://doi.org/10.1007/s10641-015-0416-2>; *Albula*; Fish; mtDNA (cyt b); BEAST, TRACER; TREEANOTATOR; MIGRATE-N
<http://doi.org/10.1007/s10641-015-0416-2>; *Albula*; Fish; Microsats; GENEPOP 4.2, GENETIX 4.05; STRUCTURE 2.3; STRUCTURE HARVESTER 0.6.93, CLUMPP 1.1.2
<http://doi.org/10.1007/s12562-015-0905-4>; *Sepioteuthis SP.*; molluscs; Microsats; POPULATIONS 1.2.32, SAMOVA 2.0, ARLEQUIN 3.5, SPAGEDI 1.4, BOTTLENECK 1.2.02; STRUCTURE 2.3.3, ; FIGTREE 1.4.2 Notes: 10 msats
<http://doi.org/10.1007/s10530-015-0896-1>; *Crassostrea gigas*; molluscs; mtDNA (MNR); DNASP 5, ARLEQUIN 3.1, SAMOVA 1.0, ; NETWORK 4.6.1; Notes: Major Noncoding Region (MNR)
<http://doi.org/10.1007/s10530-015-0896-1>; *Crassostrea gigas*; molluscs; Microsats; GENEPOP 4.2, FSTAT 2.9.3.2, GENALEX 6.41, DEMETics; POPTREE2, STRUCTURE 2.3.3; STRUCTURE HARVESTER, CLUMPP 1.1.2, DISTRUCT 1.1 Notes: 8 EST-SSRs
<http://doi.org/10.1007/s10641-015-0418-0>; *Oncorhynchus tshawytscha*; Fish; SNPs; GENEPOP 4, FSTAT 2.9.3.2, HP-RARE 1, ; GSI_SIM, ADEGENET; Notes: 96 SNPs
<http://doi.org/10.1007/s10528-015-9666-0>; *Thunnus albacares*; Fish; mtDNA (COI); ARLEQUIN 2.000, DNASP 4.0, MEGA 4.0; TCS 1.21
<http://doi.org/10.1007/s11033-014-3777-2>; *Scomber japonicus*; Fish; Microsats; POPGENE 1.32, FSTAT 2.9.3, GENEPOP 3.4, SPAGEDI, ; GENETIX 4.05, STRUCTURE 2.1; IBDWS
<http://doi.org/10.1007/s10641-014-0310-3>; *Platypharodon extremus*; Fish; mtDNA (Control region & Cyt-b); PAUP 4.0b10, MEGA 5, jMODELTEST, ARLEQUIN 3.0; MRBAYES 3.1.2
<http://doi.org/10.1007/s10592-014-0659-z>; *Oncorhynchus mykiss*; Fish; Microsats; GENEPOP 3.3, GENETIX 4.05, LDNE, ARLEQUIN 3.5, FSTAT 2.9; POPULATIONS, STRUCTURE 2.2
<http://doi.org/10.1017/S002531541700193X>; *Dermochelys coriacea*; Reptiles; mtDNA; Arlequin 3.5 (AMOVA, FST, IBD); mixstock R package; Notes: 695 bp of control region
<http://doi.org/10.1017/S0025315416001260>; *Hyporthodus septemfasciatus*; Fish; AFLP; Arlequin (AMOVA, Fst); STRUCTURE, GENALEX (pca); Notes: 602 bands
<http://doi.org/10.1017/S0025315418000632>; *Cancer pagurus*; Crustacean; microsatellites; Arlequin (Exact test, Fst, AMOVA); STRUCTURE; Fstat(randomization on Fst, Fis, Ar...) Notes: 11 loci
<http://doi.org/10.1017/S0025315415002052>; *Cardisoma guanhumi*; Crustacean; microsatellites; Arlequin(AMOVA); STRUCTURE; lositan; Notes: 9 loci
<http://doi.org/10.1017/S0954102016000183>; *Dissostichus eleginoides*; Fish; mtDNA; AMOVA, Dest; PCA; Notes: 158 and 121 bp CR, 197 bp cytB, 162 COI
<http://doi.org/10.1017/S0954102016000183>; *Dissostichus eleginoides*; Fish; nuclear sequencing; AMOVA, Dest; PCA; Notes: about 200 bp of 4 nuclear genes (LDHA...)
<http://doi.org/10.1017/S0025315414001313>; *Cerastoderma edule*; Molluscs; mtDNA (COI); DNASP 5.10.1, ARLEQUIN 3.5, ; TCS 1.21, CA in R
<http://doi.org/10.1038/s41437-018-0087-9>; *Placopecten magellanicus*; Molluscs; SNPs; Arlequin (Fst); vegan (nMDS), parallelSTRUCTURE, STRUCTURE HARVESTER, CLUMPAK, ADEGENET (SPCA); vegan (redundancy analysis), randomFOREST Notes: 96 loci fluidigm, randomFOREST for correlation of SNPs genotype with environmental parameters

<http://doi.org/10.1051/alr/2018011>; *Scomberomorus commerson*; Fish; microsatellites; Genepop (Fst), Genalex (Fst), Arlequin (AMOVA); STRUCTURE; Notes: 12 loci

<http://doi.org/10.1071/MF18146>; *Chrysophrys auratus*; Fish; microsatellites; Genalex (Dest), Arlequin (AMOVA); STRUCTURE, tree on Dest, DAPC; GeneClass 2.0; IBD Notes: 9 loci

<http://doi.org/10.1071/MF17218>; *Pampus echinogaster*; Fish; mtDNA; Arlequin (AMOVA); BEAST (bayesian skyline plots) Notes: 359 bp CR

<http://doi.org/10.1071/MF17218>; *Pampus echinogaster*; Fish; microsatellites; Arlequin (AMOVA); STRUCTURE, DAPC; BOTTLENECK Notes: 6 loci

<http://doi.org/10.1071/WR17089>; *Eretmochelys imbricata*; Reptile; mtDNA; Bayes(MSA), CHIRXC(rrandomised chi square on Haplotype frequencies); Notes: 800 bp CR

<http://doi.org/10.1071/MF17087>; *Lethrinus laticaudis*; Fish; microsatellites; Arlequin(AMOVA, Fst), GENEPOP(Exact test); STRUCTURE, DAPC; Notes: 10 loci

<http://doi.org/10.1071/MF14370>; *Istiompax indica*; Fish; mtDNA; Arlequin(PhiST); Notes: 896 bp CR

<http://doi.org/10.1071/MF14370>; *Istiompax indica*; Fish; nuclear; Arlequin(FST, AMOVA); STRUCTURE; Notes: 18 loci

<http://doi.org/10.1073/pnas.1525749113>; *gadus morhua*; Fish; Notes: no genetics in

<http://doi.org/10.1080/24701394.2018.1484120>; *Catla catla*; Fish; mtDNA; Arlequin 3.1 (Fst); PCA in DARwin 5, NETWORK; Notes: 306 bp cytb, 710 D-loop

<http://doi.org/10.1080/24701394.2018.1467409>; *Acipenser ruthenus*; Fish; mtDNA; Arlequin (AMOVA); BEAST, DNASP (mismatch distribution) Notes: 626 bp D-loop

<http://doi.org/10.1080/24701394.2018.1424842>; *Pentaceros wheeleri*; Fish; mtDNA; Arlequin (Fst); MEGA7, BEAST, Arlequin (mismatch distribution, Fu's e Tajima test) Notes: 667 bp CR

<http://doi.org/10.1080/17451000.2018.1427866>; *Octopus minor*; Molluscs; mtDNA; Arlequin (Phist); DnaSP (mismatch distribution), Arlequin (Fu's test), MEGA, PopArt, Network, MrBayes Notes: 565 bp COI and 493 bp 16S

<http://doi.org/10.1080/24701394.2016.1242581>; *Cyclocheilichthys apogon*; Fish; mtDNA; Arlequin (AMOVA); MEGA, Network; Notes: 1100 bp of cyt b

<http://doi.org/10.1080/00288330.2017.1353528>; *Maguimithrax spinosissimus*; Crustacean; mtDNA; Arlequin (AMOVA, Fst); MEGA (ML), MrBayes; Notes: 1100 bp CR, 1400 bp COI

<http://doi.org/10.1080/24701394.2016.1209195>; *Sperata seenghala*; Fish; mtDNA; Arlequin(AMOVA); MEGA; DNAsp Notes: 870 bp CR, 1130 cyt b

<http://doi.org/10.1080/02755947.2017.1339649>; *Alosa pseudoharengus*, *Alosa aestivalis*; Fish; microsatellites; GENEPOP(FST), Fstat(Fst), Recodedata(F'st); STRUCTURE, Genetix(FCA); Notes: 15 loci

<http://doi.org/10.1080/17451000.2016.1274033>; *Nibea albiflora*; Fish; microsatellites; Genepop(Fst), Arlequin(AMOVA); STRUCTURE; Notes: 12 loci

<http://doi.org/10.1080/17451000.2016.1257811>; *Engraulis encrasicolus*; Fish; microsatellites; Genetix(Fst), chifish(exact test); Genetix(FCA), STRUCTURE, STRUCTURE HARVESTER, barrier; Notes: 6 loci

<http://doi.org/10.1080/02755947.2016.1269029>; *Atractosteus spatula*; Fish; mtDNA; Arlequin(AMOVA); Notes: 319 bp CR

<http://doi.org/10.1080/02755947.2016.1269029>; *Atractosteus spatula*; Fish; microsatellites; Arlequin(AMOVA), Fstat; STRUCTURE; Notes: 14 loci

<http://doi.org/10.1080/02755947.2016.1227399>; *Salvelinus alpinus*; Fish; microsatellites; Genepop(exact test), Fstat; STRUCTURE,PHYLIP; Notes: 18 loci

<http://doi.org/10.1080/00028487.2016.1209556>; *Esox masquinongy*; Fish; microsatellites; HIERFSTAT, FSTAT, SMOGD; STRUCTURE, STRUCTURE HARVESTER; GeneClass, bayesass; Notes: 20 loci

<http://doi.org/10.1080/00028487.2016.1150878>; *Oncorhynchus tshawytscha*; Fish; Notes: no access to the journal

<http://doi.org/10.1080/00028487.2014.982178>; *Oncorhynchus mykiss*; Fish; Microsats; GENALEX, H-P RARE; PHYLIP, PCoA, STRUCTURE 2.0, CLUMPP; Notes: 15 msat loci

<http://doi.org/10.1080/02755947.2015.1017123>; Pogonias cromis; Fish; Microsats; GENEPOP 4.0.10, CERVUS 3.0.3, MICRO-CHECKER 2.2.3, ARLEQUIN 3.5.1.2, FSTAT 2.9.3.2, LDNe; STRUCTURE 2.3.3, ; STRUCTURE-HARVESTER, CLUMPP 1.1.2, BOTTLENECK 1.2.02

<http://doi.org/10.1080/19425120.2015.1037473>; Paralichthys lethostigma; Fish; AFLPs; ARLEQUIN 3.5.1.2; STRUCTURE; STRUCTURE-HARVESTER, CLUMPP, DISTRUCT

<http://doi.org/10.1080/19425120.2015.1037473>; Paralichthys lethostigma; Fish; mtDNA (control region); ARLEQUIN 3.5.1.2, MEGA 6, DNASP 5.10; NETWORK 4.6.1.2

<http://doi.org/10.1080/02755947.2015.1012278>; Sander vitreus; Fish; Microsats; GENEPOP 4.0, FSTAT, GENALEX, LDNe ; STRUCTURE 2.3; BAYESASS 3.0.1; BOTTLENECK 1.2.02

<http://doi.org/10.1093/icesjms/fsx047>; Kajikia albida; Fish; microsatellites; Arlequin (Fst), Genepop, HP-Rare, GeneClone, DiveRcity; STRUCTURE, STRUCTURE HARVESTER, Adegnet (PCA); Lositan; Micro-checker, Powsim Notes: 24 loci

<http://doi.org/10.1093/icesjms/fsx047>; Kajikia albida; Fish; mtDNA; Arlequin (Fst); Notes: 858 bp CR

<http://doi.org/10.1093/icesjms/fsw249>; Merluccius merluccius; Fish; SNPs; Genepop(Fst), Arlequin(AMOVA); BAPS; Arlequin(assignment); Lositan, bayescan; ecodist(geneticvs environment) Notes: 53 loci

<http://doi.org/10.1093/icesjms/fsw184>; Panulirus ornatus; Crustacean; mtDNA; phiST; Notes: 461 bp CR, no access

<http://doi.org/10.1093/icesjms/fsw068>; Pecten fumatus; Molluscs; microsatellites; Fstat; DAPC; Notes: 11 loci

<http://doi.org/10.1093/jhered/esv091>; Eretmochelys imbricata; Reptiles; mtDNA; Arlequin (Fst, AMOVA, exact test of differentiation, IBD); Network, Arlequin (Fu e Tajima test), DnaSP (R2), BEAST, MrBayes Notes: 766 bp CR

<http://doi.org/10.1093/jhered/esw066>; Girella elevata; Fish; microsatellites; Notes: 11 loci (microsatellites isolation)

<http://doi.org/10.1093/icesjms/fsu204>; Pandalus borealis; Crustacean; Microsats; FSTAT 2.9.3.2, GENEPOP 4.0, MICROCHECKER, ARLEQUIN 3.5, POWSIM; BARRIER 2.2; Notes: 11 msats

<http://doi.org/10.1093/jhered/esv005>; Galaxias maculatus; Fish; mtDNA (control region); DNASP 5.00.07, ARLEQUIN 3.5,

<http://doi.org/10.1098/rsos.180537>; Phocoena phocoena; Mammal; eDNA/mtDNA; R package STRATAG (PhiST and χ^2); IQ-TREE (ML Tree); Notes: mtDNA derived from eDNA

<http://doi.org/10.1098/rsos.160651>; Capros aper; Fish; microsatellites GBS; Arlequin(AMOVA); STRUCTURE, BARRIER; divMigrate; lositan; Notes: 85 loci

<http://doi.org/10.1111/eva.12738>; Platichthys flesus, Platichthys solemdali ; Fish; SNPs ; Notes: Only 5 SNPs, 2 species, ad hoc species assignment test in R based on Toli, Calboli, Shikano, and Merilä (2016)

<http://doi.org/10.1111/1755-0998.12980>; Dermochelys coriacea; Reptiles; SNPs ; NGSdist, ANGSD and realSFS; PCA in NGStools, NGSAdmix; Notes: About 2000 snps on focal species, 11042-1542 candidates in other 5 species by RAD-capture

<http://doi.org/10.1111/age.12747>; Morone chrysops; Fish; SNPs ; Genalex (Fst); MEGA (UPGMA), STRUCTURE, STRUCTURE HARVESTER, CLUMPAK; Notes: 426 filtered SNPs

<http://doi.org/10.1111/are.13875>; Cirrhinus cirrhosus; Fish; microsatellites; POPGENE (Fst, Nm, Nei's genetic distance); Notes: 6 loci

<http://doi.org/10.1111/jfb.13779>; Caulolatilus microps; Fish; microsatellites; GenAlEx (IBD), Genepop (Fis), FSTAT, Arlequin (Fst, AMOVA); Adegnet (PCA), STRUCTURE, STRUCTURE HARVESTER, CLUMPP, DISTRUCT; Notes: 25 loci

<http://doi.org/10.1111/jfb.13779>; Caulolatilus microps; Fish; mtDNA; DnaSP, Arlequin (Fst, AMOVA); Arlequin (Fu's test), PopART (network) Notes: 489 bp CR

<http://doi.org/10.1111/eva.12639>; Gadus macrocephalus; Fish; SNPs; Genepop, NeEstimator, GenAlEx (IBD), diversity (Gst); Adegnet, MLRelate, GeneClass 2, Assigner, STRUCTURE, STRUCTURE HARVESTER; Bayscan, OutFLANK; Notes: 6425 SNP (RAD seq)

<http://doi.org/10.1111/eva.12591>; Prionace glauca; Fish; microsatellites; Genetix, Fstat, NeEstimator; STRUCTURE, STRUCTURE HARVESTER, Adegnet; Powsim, Bottleneck, SimuPop Notes: 9 loci

<http://doi.org/10.1111/eva.12591>; *Prionace glauca*; Fish; mtDNA; DnaSP, Arlequin (Fst, AMOVA, Phist); STRUCTURE, STRUCTURE HARVESTER, Adegnet; Powsim, Network Notes: 758 bp cyt b

<http://doi.org/10.1111/jfb.13542>; *Salmo salar*; Fish; SNPs; Genalex (Fst, Da); BAPS, NJ and MDS on Da; Arlequin; Notes: 4300 SNPs

<http://doi.org/10.1111/mms.12545>; *Stenella longirostris*, *Stenella attenuata*; Mammal; mtDNA; strataG(Fst, PhiST); Notes: mitogenomics

<http://doi.org/10.1111/eva.12499>; *Protonibea diacanthus*; Fish; microsatellites; Arlequin(AMOVA); STRUCTURE, DAPC; Notes: 11 loci

<http://doi.org/10.1111/are.13384>; *Salmo trutta abanticus*; Fish; microsatellites; GENEPOP(FST); STRUCTURE; Notes: 7 loci

<http://doi.org/10.1111/jai.13412>; *Siniperca kneri*; Fish; microsatellites; Arlequin(AMOVA); Notes: 12 loci

<http://doi.org/10.1111/mec.14263>; *Balistes caprisiscus*; Fish; microsatellites; GENEPOP(FST), Fstat(Fst); TESS; MIGRAINE; Notes: 17 loci

<http://doi.org/10.1111/jwas.12330>; *Paralichthys olivaceus*; Fish; microatellites; Arlequin(AMOVA); Notes: 18 loci

<http://doi.org/10.1111/mms.12378>; *Megaptera novaeangliae*; Mammal; mtDNA; Arlequin(Fst); Notes: 465 bp CR

<http://doi.org/10.1111/mms.12378>; *Megaptera novaeangliae*; Mammal; microsatellites; Arlequin(Fst); STRUCTURE, STRUCTURE HARVESTER, DISTRUCT; GENECLASS, bayesass, migrate; Notes: 16 loci

<http://doi.org/10.1111/mec.13890>; *Merluccius paradoxus*, *Merluccius capensis*; Fish; mtDNA; Arlequin(AMOVA); TCS; Notes: 406 bp CR

<http://doi.org/10.1111/mec.13890>; *Merluccius paradoxus*, *Merluccius capensis*; Fish; microsatellites; Adegnet(PCA),STRUCTURE; bayesass; Notes: 10 loci

<http://doi.org/10.1111/mec.13734>; *Haliotis rubra*; Molluscs; microsatellites; genalex(Amova); genetix(FCA), STRUCTURE; SPAGEDI, GENALEX(Autocorrelation analysis) Notes: 15 loci

<http://doi.org/10.1111/mec.13734>; *Haliotis rubra*; Molluscs; SNPs; GENEPOP(Fst); DAPC; bayescan Notes: 1425-1723 loci depending on filtering

<http://doi.org/10.1111/fwb.12767>; *Opsaridium microlepis*, *Opsaridium microcephalum*, *Opsaridium tweddleorum*; Fish; microsatellites; Arlequin (Fst), Genepop (genetic differentiation); MASS (MDS), Structure, Structure harvester; micro-checker, Powsim Notes: 10 loci

<http://doi.org/10.1111/fwb.12767>; *Opsaridium microlepis*, *Opsaridium microcephalum*, *Opsaridium tweddleorum*; Fish; mtDNA; Arlequin (Phist); Hapstar Notes: 700 COI

<http://doi.org/10.1111/age.12422>; *Colossoma macropomum*; Fish; microsatellites; Arlequin (AMOVA, Rst); Molkin, Structure; micro-checker Notes: 13 loci

<http://doi.org/10.1111/jai.13035>; *Thunnus thynnus*; Fish; microsatellites; Arlequin (AMOVA); Structure; micro-checker, HP-Rare, powsim Notes: 13 loci

<http://doi.org/10.1111/are.12508>; *Macrobrachium rosenbergii*; Crustacean; microsatellites; Arlequin (Fst, AMOVA); GenALEX (PCA), Structure, Structure Harvester; Micro-checker Notes: 11 loci

<http://doi.org/10.1111/bij.12654>; *Pelagia noctiluca*; Cnidaria; microsatellites; Arlequin (Fst, AMOVA); Powsim Notes: 8 loci

<http://doi.org/10.1111/bij.12654>; *Pelagia noctiluca*; Cnidaria; mtDNA; Arlequin (Phist, AMOVA); BAPS, GenALEX (PCA); Network, DnaSP (Tajima, Fu and Li's test, mismatch distribution) Notes: 532 bp COI

<http://doi.org/10.1111/mec.13046>; *Hoplostethus atlanticus*; Fish; SNPs; ARLEQUIN 3.5.2, CHOPSTICKS in R, ; STRUCTURE 2.3.4, ADEGENET; BAYESASS; BAYESCAN 2.1, LOSITAN; HIERFSTAT, GDISTANCE in R Notes: 4273 SNPs

<http://doi.org/10.1111/jfb.12727>; *Galeorhinus galeus*; Fish; Microsats; GENEPOP 4.0, GENALEX 6.501, MICRO-CHECKER, FSTAT; STRUCTURE 2.3.3, ; BAYESASS 3.0.3; STRUCTURE-HARVESTER 0.6.94, CLUMPP 1.1.2, DISTRUCT, TRACER 1.5

<http://doi.org/10.1111/jfb.12580>; *Scomberomorus concolor*; Fish; mtDNA (control region); MEGA 6.0, ARLEQUIN, jMODELTEST; BEAST 1.7.1; HAPSTAR 0.5, TRACER 1.5

<http://doi.org/10.1111/mms.12135>; *Tursiops truncatus*; Mammal; mtDNA (control region); ARLEQUIN 3.1, MEGA 3.1. jMODELTEST

<http://doi.org/10.1111/mms.12135>; *Tursiops truncatus*; Mammal; Microsats; GENALEX, GENEPOP 4.0.10; STRUCTURE 2.3.1,

<http://doi.org/10.1111/1755-0998.12400>; *Thaleichthys pacificus*; Fish; SNPs; GENEPOP, HIERFSTAT, ARLEQUIN 3.5.1.2, ; PHYLIP 3.5.7, ADEGENET 3.1.0; LOSITAN; FIGTREE, NeESTIMATOR, ONCOR Notes: RAD

<http://doi.org/10.1111/mec.13255>; *Nautilus pompilius*; molluscs; Microsats; SPAGEDI, GENEPOP, MICRO-CHECKER, CERVUS 3.0, MICROSATELLITE ANALYSER, DEMETICS; STRUCTURE, TESS, GENELAND; ABC TOOLBOX; STRUCTURE HARVESTER 0.6.93, CLUMPP 1.1.2, DISTRUCT 1.1, FASTSIMCOAL, ARLSUMSTAT

<http://doi.org/10.1134/S1022795417100052>; *Oncorhynchus nerka*; Fish; microsatellites; Fst(?); STRUCTURE; Notes: 6 loci

<http://doi.org/10.1134/S1022795417100052>; *Oncorhynchus nerka*; Fish; SNPs; Fst(?); STRUCTURE; Notes: 45 loci fluidigm

<http://doi.org/10.1139/cjfas-2016-0360>; *Oncorhynchus nerka*; Fish; Notes: no genetics in

<http://doi.org/10.1139/cjfas-2015-0430>; *Reinhardtius hippoglossoides*; Fish; SNPs; Arlequin; Genalex(PCA), barrier; lositan,bayescan, arlequin; Notes: 96 loci

<http://doi.org/10.1139/cjfas-2015-0606>; *Salmo salar*; Fish; microsatellites; Fstat; Powermarker(Dc)+Phylip, STRUCTURE; ONCOR,cBAYES (both for MSA); passage(IBD) Notes: 31 loci

<http://doi.org/10.1139/cjfas-2016-0443>; *Oncorhynchus tshawytscha*; Fish; SNPs; Genepop(Fst); GSIsim; Notes: 14494 loci, interesting re-analysis using multiSNPS

<http://doi.org/10.1139/cjfas-2016-0012>; *Anoplopoma fimbria*; Fish; SNPs; Genepop(exact test, Fst), StAMPP(Fst significance); smatPCA, BAPS; LFMM(genotype-environment) Notes: 2661 loci

<http://doi.org/10.1139/cjfas-2015-0402>; *Alosa pseudoharengus*, *Alosa aestivalis*; Fish; microsatellites; micro-checker Notes: 15 loci, genetic stock assessment

<http://doi.org/10.1186/s12863-018-0711-y>; *Arapaima gigas*; Fish; SNPs ; Arlequin (Fst, IBD); STRUCTURE, CLUMPAK, DAPC; Notes: ddRAD, 393 SNPs

<http://doi.org/10.1186/s12862-018-1325-8>; *Stolothrissa tanganicae*; Fish; SNPs ; diveRsity (Fst); ADEGENET (PCoA, DAPC), fineRADstructure; diveRsity v1.9.90, BayeScan v2.1; LEA (individual-based latent fixed mixed model to correlate SNPs frequency to latitude) Notes: RAD 3504 SNPs

<http://doi.org/10.1186/s12862-018-1325-8>; *Stolothrissa tanganicae*; Fish; mtDNA; diveRsity (Fst); PopART 1.7 (Median joining network); Notes: 643 bp COI

<http://doi.org/10.1186/s12864-018-5044-8>; *Panulirus homarus*; Crustacean; SNPs; Genetix, NeEstimator, inbreedR, Arlequin (AMOVA), Adegnet; Adegnet, DAPC, NetVIEW R; Bayescan, Arlequin; Notes: 7988 SNPs

<http://doi.org/10.1186/s12863-018-0625-8>; *Gadus morhua*; Fish; microsatellites; MICROSATELLITE ANALYSER, GenAlEx, Genepop on the web, Arlequin; BARRIER 2.2, STRUCTURE/ParallelStructure; LOSITAN; LDNE (Ne)

<http://doi.org/10.1186/s40709-017-0062-2>; *Thunnus thynnus*; Fish; microsatellites; Genetix(Fst); STRUCTURE, GENELAND, ADEGENET(PCA), ; "EEMS(Petkova D, Novembre J, Stephens M. Visualizing spatial population

<http://doi.org/10.1186/s40709-017-0062-2>; *Thunnus thynnus*; Fish; SNPs; Genetix(Fst); STRUCTURE, GENELAND, ADEGENET(PCA), ; "EEMS(Petkova D, Novembre J, Stephens M. Visualizing spatial population

<http://doi.org/10.1186/s13104-016-1878-9>; *Gadus morhua*; Fish; microsatellites; Genepop (Fst); CMNDScale (MDS); GeneClass; Lositan; Notes: 8 loci

<http://doi.org/10.1186/s13104-015-1345-z>; *Cherax cainii*; Crustacean; Microsats; Notes: Microsat isolation

<http://doi.org/10.1186/s40555-014-0085-4>; *Hirundichthys oxycephalus*; Fish; mtDNA (COI); ARLEQUIN 3.5.1.3, DNASP 5, ; MEGA 5.02, MRBAYES 3.1.2; BEAST; MrMODELEST, TRACER 1.5

<http://doi.org/10.1371/journal.pone.0203866>; *Dentex dentex*; Fish; mtDNA; ARLEQUIN 3.5 (AMOVA and mismatch distribution); Network 5.0.0.0 (Haplotype networks), PHYLIP - platform ATGC (ML Tree)

<http://doi.org/10.1371/journal.pone.0203866>; *Dentex dentex*; Fish; microsatellites; GENETIX (diversity, F statistics, IBD, FCA); STRUCTURE, STRUCTURE HARVESTER

<http://doi.org/10.1371/journal.pone.0179661>; *Acipenser oxyrinchus oxyrinchus*; Fish; microsatellites; GenALEx (Fst, F'st); STRUCTURE, STRUCTURE HARVESTER; Colony Notes: 4 loci?

<http://doi.org/10.1371/journal.pone.0175085>; *Acipenser oxyrinchus*; Fish; mtDNA; Arlequin(PhiST, AMOVA); Notes: 560 bp CR

<http://doi.org/10.1371/journal.pone.0175085>; *Acipenser oxyrinchus*; Fish; microsatellites; Genepop(exact test), Fstat, Genalex(F'st); STRUCTURE, STRUCTURE HARVESTER; ONCOR(MSA); colony Notes: 11 loci

<http://doi.org/10.1371/journal.pone.0172255>; *Zearaja chilensis*, *Dipturus trachyderma*; Fish; mtDNA; Arlequin(PhiST); Notes: 560 bp CR

<http://doi.org/10.1371/journal.pone.0172255>; *Zearaja chilensis*, *Dipturus trachyderma*; Fish; microsatellites; Arlequin(FST); STRUCTURE, DAPC; Notes: 10 loci

<http://doi.org/10.1371/journal.pone.0181898>; *Malaclemys terrapin*; Reptiles; microsatellites; STRUCTURE, DAPC; Migrate, BayesASS; BOTTLENECK Notes: 6 loci

<http://doi.org/10.1371/journal.pone.0161390>; *Pinctada margaritifera*; Molluscs; SNPs; Arlequin(Fst), Genalex(AMOVA); DAPC, PLINK+NETVIEW; bayescan, lositan; Notes: 4123 loci ddRAD

<http://doi.org/10.1371/journal.pone.0161617>; *Lutjanus purpureus*; Fish; mtDNA, nuclear DNA; Arlequin (Fst, AMOVA); Structure, Structure harvester, Structurama; Haploviewer, Arlequin (D e Fs test), BEAST Notes: CR, Cytb, ND4, S7, RPL3, GH5, Myo, Prl, ANT, IGF, Delt8, La1

<http://doi.org/10.1371/journal.pone.0154020>; *Larimichthys polyactis*; Fish; microsatellites; Genepop (Fst), Arlequin (AMOVA), IBD web service (IBD); Structure, GenALEx (PCoA); Lositan; Bottleneck Notes: 15 loci

<http://doi.org/10.1371/journal.pone.0153124>; *Callinectes sapidus*; Crustacean; microsatellites; Arlequin (Fst, AMOVA, IBD); Structure, Structure harvester; Migrate; micro-checker Notes: 9 loci

<http://doi.org/10.1371/journal.pone.0149496>; *Octopus vulgaris*; Molluscs; microsatellites; Arlequin (Fst), DEMetics (Jost's D), XLSTAT (IBD), NeEstimator; ML-Relate, Adegnet (DAPC); GeneClass; Lositan; Micro-checker, GenClone Notes: 13 loci

<http://doi.org/10.1371/journal.pone.0159864>; *Epinephelus marginatus*; Fish; microsatellites; Arlequin(AMOVA); GENELAND; Arlequin; Notes: 14 loci

<http://doi.org/10.1371/journal.pone.0138292>; *Thunnus albacares*; Fish; Microsats; GENEPOP 4, ARLEQUIM 3.5.1.2, GENALEX 6.5; STRUCTURE 2.2; STRUCTURE HARVESTER

<http://doi.org/10.1371/journal.pone.0138640>; *Umbra krameri*; Fish; Microsats; GENEPOP 4.2.2, MICROCHECKER 2.2.3, GENALEX 6.5; STRUCTURE, GENECLASS2; STRUCTURE HARVESTER, CLUMPP 1.1.2 Notes: 9 msats

<http://doi.org/10.1371/journal.pone.0129278>; *Notorynchus cepedianus*; Fish; Microsats; GENEPOP 3.1, MICROCHECKER, GENALEX, BOTTLENECK; MLRELATE, COANCESTRY, COLONY Notes: 7 msats

<http://doi.org/10.1371/journal.pone.0125743>; *Pagrus major*; Fish; Microsats; FSTAT 2.9.3.2, GENEPOP 4.0, MICROCHECKER 2.2.1; PCAGEN 1.2.1, STRUCTURE 2.3.3, DISTRICT 1.1, ARLEQUIN 3.5, LDNE 1.31; IBDWS 3.23, BARRIER 2.2, DIVERSITY R; BAYESCAN; Notes: 15 msats

<http://doi.org/10.1371/journal.pone.0122315>; *Mallotus villosus*; Fish; Microsats; GENEPOP 4.0.10, GENETIX 4.05.2, ADZE 1.0; GENECLASS 2.0, STRUCTURE 2.3.4; SPADE, PRIMER 6.1.5, nMDS, IBDWS

<http://doi.org/10.1371/journal.pone.0122173>; *Ocyurus chrysurus*; Fish; mtDNA (cyt b & control region); ARLEQUIN 3.5.1.3, DNASP 5.10.01, MEGA 6; STRUCTURE 2.3.4, STRUCTURAMA; BEAST; SPLITS TREE 4.12.6, HAPLOVIEWER, PHYLIP 3.6, IBDWS, jMODELTEST, TRACER 1.5

<http://doi.org/10.1371/journal.pone.0122173>; *Ocyurus chrysurus*; Fish; nuclear (3 intragenic regions)

<http://doi.org/10.1371/journal.pone.0118988>; *Oncorhynchus spp.*; Fish; Microsats; CBAYES

<http://doi.org/10.1371/journal.pone.0117549>; *Pseudocarcharias kamoharai*; Fish; mtDNA (control region); ARLEQUIN 3.5.1.3; NETWORK 4.611

<http://doi.org/10.1002/ece3.1515>; *Carcharhinus limbatus*, *Carcharhinus sorrah*, *Rhizoprionodon acutus*, and *Sphyrna lewini*; Fish; mtDNA (control region); jMODELTEST v2.1.4, ARLEQUIN

<http://doi.org/10.1002/ece3.1515>; *Carcharhinus limbatus*, *Carcharhinus sorrah*, *Rhizoprionodon acutus*, and *Sphyrna lewini*; Fish; Microsats; GENEPOP 4.2, ARLEQUIN 3.2, SMOGD 1.2.5; STRUCTURE 2.3.4; STRUCTURE HARVESTER

<http://doi.org/10.1534/g3.118.200779>; *Oryzias latipes*; Fish; SNPs; SNPRelate (PCA); ADMIXTURE 1.23; DYABC Notes: GBS 13177 SNPs

<http://doi.org/10.1656/045.024.0407>; *Alosa aestivalis*, *Alosa pseudoharengus*; Fish; microsatellites; Fstat; STRUCTURE, CLUMPP, NEWHYBRIDS; HYBRIDLAB Notes: 7 loci

<http://doi.org/10.4238/2015.May.18.5>; *Siniperca chuatsi*; Fish; Microsats; MICROCHECKER, GENEPOP 4.0, POPGENE, ARLQUIN 3.11, ; STRUCTURE 2.2; Notes: 12 msats

<http://doi.org/10.4238/2015.December.1.13>; *Octopus minor*; molluscs; AFLPs; POPGENE 3.1, ARLEQUIN 3, MEGA4; Notes: 243 AFLP bands from 5 AFLP primer pairs

<http://doi.org/10.1016/j.fishres.2014.06.006>; *Larimichthys polyactis*; Fish; Microsats; FSTAT2.9.3.2, GENEPOP 4.0; STRUCTURE 2.3.3, ONCOR, BAPS2; Notes: MICRO-CHECKER

<http://doi.org/10.1016/j.fishres.2014.10.004>; *Macruronus magellanicus*; Fish; mtDNA (control region); ARLEQUIN 3.1; NETWORK, POWSIM

<http://doi.org/10.1016/j.fishres.2014.10.004>; *Macruronus magellanicus*; Fish; Microsats; FSTAT 2.9.3.2, GENEPOP 3.3, ; STRUCTURE 2.3.2, ADEGENET; POWSIM, MICRO CHECKER

<http://doi.org/10.1016/j.fishres.2014.10.019>; *Chrysoblephus puniceus*; Fish; mtDNA (control region); ARLEQUIN, ; STRUCTURE 2.3.2, ; MIGRATE 3.2.16, NeESTIMATOR; IBDWS, NETWORK 4.6.10

<http://doi.org/10.1016/j.fishres.2014.10.019>; *Chrysoblephus puniceus*; Fish; Microsats; ARLEQUIN. FSTAT 2.9.3.2, SMOGD 1.2.5

<http://doi.org/10.1016/j.bse.2014.12.007>; *Engraulis japonicus*; Fish; mtDNA (cyt b); ARLEQUIN 3.1, MEGA 6.0,

<http://doi.org/10.1016/j.fishres.2015.03.008>; *Micropogonias furnieri*; Fish; EPIC; Notes: 13 INTRONS, not clear what they did with...

<http://doi.org/10.1016/j.fishres.2015.03.008>; *Micropogonias furnieri*; Fish; Microsats; GENETIX 4.05, ARLEQUIN 3.5, SMOGD; STRUCTURE 2.3.3, CLUMPP 1.1.2, DISTRUCT 1.1, KMEANS; XLSTAT; STRUCTURE HARVESTER Notes: 10 msats

<http://doi.org/10.1016/j.jglr.2015.04.006>; *Perca fluviatilis*; Fish; Microsats; Notes: 12 msats, No Access

<http://doi.org/10.1016/j.fishres.2015.06.012>; *Lates calcarifer*; Fish; SNPs; GENEPOP 4.0, FSTAT 2.9.3.2, ARLEQUIN 3.5, BOTTLENECK 1.2.02, IBD; BAYESCAN 2.01, ARLEQUIN; Notes: 96 SNPs

<http://doi.org/10.1016/j.fishres.2015.06.020>; *Lutjanus campechanus*; Fish; Microsats; GENEPOP 4.7, MICROCHECKER, ARLEQUIN 3.5, FSTAT 2.9.3.2; Notes: 16 msats

<http://doi.org/10.1016/j.fishres.2015.06.020>; *Lutjanus campechanus*; Fish; mtDNA (ND4); jMODELTEST 2.1.1, ARLEQUIN; NETWORK 4.6.11; MIGRATE-N 3.2.16; MSVAR 1.3; LDNe Notes: 16 msats

<http://doi.org/10.1016/j.bse.2015.06.023>; *Culter alburnus*; Fish; mtDNA (control region); ARLEQUIN 3.4; NETWORK 4.1, MEGA 5.0

<http://doi.org/10.1016/j.bse.2015.06.023>; *Culter alburnus*; Fish; Microsats; POPGENE 1.32, MICROCHECKER, FSTAT; Notes: 8 microsats

<http://doi.org/10.1016/j.fishres.2015.06.029>; *Prionace glauca*; Fish; Microsats; FSTAT, ARLEQUIN 3.5.1.3; STRUCTURE 2.3.3, ; MIGRATE; NeESTIMATOR Notes: 14 msats

<http://doi.org/10.1016/j.fishres.2015.08.009>; *Polyprion oxygeneios*; Fish; mtDNA; Arlequin (Phist, IBD); HP-Rare, PhyML Notes: 488 bp CR

<http://doi.org/10.1016/j.fishres.2015.08.009>; *Polyprion oxygeneios*; Fish; microsatellites; Genepop (exact test of differentiation), Arlequin (Fst and Rst, AMOVA, IBD), Smogd (Dst), Populations; Structure, Structure Harvester, GenALEX (PCA); GeneClass; Micro-checker, Powsim Notes: 9 loci

<http://doi.org/10.1016/j.jglr.2015.08.011>; *Esox masquinongy*; Fish; Microsats; Fstat; STRUCTURE, BAPS; Notes: 12 msats,

<http://doi.org/10.1016/j.bse.2015.09.025>; *Larimichthys crocea*; Fish; SNPs; POPGENE 3.2, ARLEQUIN 3.11, PIC-calc, ; Notes: 44 SNPs

<http://doi.org/10.1016/j.margen.2015.10.010>; *Pinctada margaritifera*; Molluscs; SNPs; NeEstimator (Ne), Genetix (Fst); Adegnet (DAPC), PLINK, Cytoscape; Bayescan, Lositan; Notes: 5243 ddRAD SNPs

<http://doi.org/10.1016/j.fishres.2015.11.006>; *Salmo trutta*; Fish; microsatellites; FSTAT (Fst); Colony, Structure, Structure harvester, Clumpp, Distruct, FAMM (MPCA); Bottleneck Notes: 12 loci

<http://doi.org/10.1016/j.fishres.2016.01.014>; *Engraulis encrasicolus*; Fish; SNPs; Populations (Nei's distance), GenAlEx (IBD), Arlequin (Fst, AMOVA); Structure, Structure harvester; GeneClass; Bayescan; Notes: 424 SNPs
<http://doi.org/10.1016/j.margen.2016.02.005>; *Pinctada maxima*; Molluscs; SNPs; Arlequin (Fst), Microsat; Adegnet (DAPC); Bayescan, Lositan; Notes: 1130 SNPs
<http://doi.org/10.1016/j.fishres.2016.02.013>; *Perumytilus purpuratus*; Molluscs; microsatellites; Fstat (Fst), Arlequin (AMOVA), DEMETICS (Dest Jost), ISOLDE (IBD); Geneland, GenAlEx (PCoA); Bayesass; Lositan; micro-checker, Powsim Notes: 5 loci
<http://doi.org/10.1016/j.jembe.2016.03.010>; *Xiphias gladius*; Fish; SNPs; GENEPOP(exact test), Arlequin(AMOVA); STRUCTURE, PCA; lositan; Notes: 20 loci
<http://doi.org/10.1016/j.fishres.2016.03.017>; *Merluccius merluccius*; Fish; microsatellites; Fstat (Fst), Arlequin (AMOVA), DEMETICS (Dest Jost), ISOLDE (IBD); Geneland, GenAlEx (PCoA); Bayesass; micro-checker, Powsim Notes: 5 loci
<http://doi.org/10.1016/j.aquaculture.2016.04.013>; *Salmo trutta*; Fish; microsatellites; Arlequin(FST); STRUCTURE, DAPC; Notes: 12 loci
<http://doi.org/10.1016/j.marpol.2016.04.020>; *Gadus morhua*; Fish; SNPs; GeneClass; Notes: 96 SNPs
<http://doi.org/10.1016/j.jglr.2016.05.006>; Notes: no access to the journal
<http://doi.org/10.1016/j.fishres.2016.05.027>; *Perca fluviatilis*; Fish; microsatellites; GENEPOP(exact test), Fstat, Arlequin(AMOVA); STRUCTURE, PCAGEN, POPULATIONS; GENECLASS, ONCOR; Notes: 16 loci
<http://doi.org/10.1016/j.fishres.2016.06.015>; *Homarus gammarus*; Crustacean; microsatellites; GENEPOP(exact test), Fstat; PCAGEN, STRUCTURE; BAYESASS; LOSITAN; Notes: 12 loci
<http://doi.org/10.1016/j.dsr.2016.07.002>; *Etmopterus spinax*; Fish; mtDNA, nuclear DNA; Arlequin(FST), mmod(Jost D); mdiv; Notes: CR and ITS2
<http://doi.org/10.1016/j.aquaculture.2016.07.003>; *Megalobrama amblycephal*; Fish; microsatellites; Popgene(Fst); Notes: 9 loci
<http://doi.org/10.1016/j.margen.2016.07.003>; *Coryphaena hippurus*; Fish; SNPs; Arlequin (Fst), Genepop (Fst); Structure, structure harvester; Bayescan, Lositan; Notes: 3324 RAD SNPs
<http://doi.org/10.1016/j.bse.2016.07.010>; *Perna viridis*; Molluscs; microsatellites; Genalex, AMOVA; STRUCTURE, MSA+Phylip; Notes: 9 loci
<http://doi.org/10.1016/j.bse.2016.09.004>; *Gymnocypris dobula*; Fish; mtDNA; Arlequin(AMOVA,FST); Notes: 914 bp CR
<http://doi.org/10.1016/j.fishres.2016.09.006>; *Micromesistius australis*; Fish; microsatellites; Fstat, Arlequin(AMOVA), Genepop(exact test); STRUCTURE, genetix(FCA); lositan; Notes: 5 loci
<http://doi.org/10.1016/j.ecolmodel.2016.10.016>; Notes: no genetics in
<http://doi.org/10.1016/j.ecss.2016.10.024>; *Acanthopagrus vagus*; Fish; mtDNA; Arlequin(AMOVA); TCS; BSP, DNAsp Notes: cytb ND2
<http://doi.org/10.1016/j.ecss.2016.10.024>; *Acanthopagrus vagus*; Fish; microsatellites; Arlequin(AMOVA); STRUCTURE,FCA; Notes: 13 loci
<http://doi.org/10.1016/j.fishres.2016.11.007>; *Genypterus capensis*; Fish; mtDNA; Notes: CR, no access
<http://doi.org/10.1016/j.fishres.2016.11.007>; *Genypterus capensis*; Fish; microsatellites; Notes: 10 loci, no access
<http://doi.org/10.1016/j.fishres.2016.12.006>; *Oncorhynchus clarkii clarkii*; Fish; microsatellites; Fstat; ONCOR, geneclass; Notes: 7 loci
<http://doi.org/10.1016/j.fishres.2016.12.006>; *Oncorhynchus clarkii clarkii*; Fish; SNPs; Fstat; ONCOR, geneclass; Notes: 96 SNPs
<http://doi.org/10.1016/j.bse.2017.03.003>; *Larimichthys polyacis*; Fish; mtDNA; Arlequin(AMOVA); Notes: 773 bp of COI
<http://doi.org/10.1016/j.fishres.2017.04.015>; *Pristipomoides zonatus*; Fish; mtDNA; Arlequin(AMOVA); Notes: 598 bp COI
<http://doi.org/10.1016/j.fishres.2017.04.015>; *Pristipomoides zonatus*; Fish; microsatellites; Arlequin(AMOVA),FSTAT; STRUCTURE, STRUCTURE HARVESTER; Notes: 6 loci

<http://doi.org/10.1016/j.seares.2017.07.003>; *Dentex dentex*; Fish; microsatellites; Genetix(Fst); STRUCTURE, DAPC, FLOCK; bayescan; Notes: 8 loci

<http://doi.org/10.1016/j.fishres.2017.07.030>; *Scomber scombrus* and *Anisakis* spp; Fish and Nematods; allozymes; Notes: 3 loci diagnostic for *Anisakis* spp (Adk-2, Pep C-1 and Pep C-2), No GENETICS IN

<http://doi.org/10.1016/j.fishres.2017.08.002>; *Clupea harengus* and *Anisakis simplex*; Fish and Nematods; allozymes; Arlequin (Fst, AMOVA), Arlequin? (IBD); Notes: 3 loci diagnostic for *Anisakis* spp (Adk-2, Pep C-1 and Pep C-2)

<http://doi.org/10.1016/j.fishres.2017.08.002>; *Clupea harengus* and *Anisakis simplex*; Fish and Nematods; mtDNA; Arlequin (Fst, AMOVA), Arlequin? (IBD); Arlequin (Fu and Tajima test, mismatch distribution) Notes: cox2

<http://doi.org/10.1016/j.fishres.2017.08.002>; *Clupea harengus* and *Anisakis simplex*; Fish and Nematods; nDNA; Arlequin (Fst, AMOVA), Arlequin? (IBD); Network Notes: 409 bp EF1 α -1 region

<http://doi.org/10.1016/j.ecss.2017.08.011>; *Donax trunculus*; Molluscs; microsatellites; Arlequin(AMOVA); STRUCTURE, STRUCTURE HARVESTER; Notes: 15 loci

<http://doi.org/10.1016/j.seares.2018.04.004>; *Solea solea* + *S. aegyptiaca*; Fish; mtDNA; ARLEQUIN v.3.5.2.2 (Φ ST); Haploviewer (+ dnaml program of the PHYLIP package)

<http://doi.org/10.1016/j.fishres.2018.04.014>; *Alosa pseudoharengus* and *A. aestivalis*; Fish; SNPs; Genepop, Microsatellite Toolkit, FSTAT, Genetix, Vegan R package (IBD); STRUCTURE, CLUMPP, DISTRUCT, Adegenet; POPTREEW, GSI_SIM (assignment test) Notes: 92 SNPs for *Alosa pseudoharengus* and 95 SNPs for *A. aestivalis*

<http://doi.org/10.1016/j.jglr.2018.07.002>; *Sander vitreus*; Fish; Notes: no access to the journal

<http://doi.org/10.1016/j.marpol.2018.08.025>; *Lutjanus* spp. and *Caranx* spp; Fish; mtDNA; Arlequin (AMOVA, Fst, Phist); Notes: 700 bp COI, barcoding and structure on samples with >5 specimens

<http://doi.org/10.1016/j.fishres.2018.09.021>; *Holothuria (Metriatyla) scabra*; Echinoderms; microsatellites; Arlequin (AMOVA, Fst), Genalex (IBD); Genalex (PCoA), STRUCTURE HARVESTER, CLUMPP, DISTRUCT; divMIGRATE; Lositan; Notes: 11 loci

<http://doi.org/10.1016/j.aquaculture.2018.10.007>; *Crassostrea gigas*; Molluscs; mtDNA; Notes: 606 bp COI, only Ne estimation

<http://doi.org/10.1016/j.aquaculture.2018.10.007>; *Crassostrea gigas*; Molluscs; microsatellites; Arlequin (AMOVA); FreeNA Notes: 11 loci, AMOVA between broodstock, F1-F4, wild relatives

<http://doi.org/10.1016/j.fishres.2019.01.012>; *Coilia nasus*; Fish; Microsats; FSTAT 2.9.3, GenALEX 6.5, Arlequin 3.5; STRUCTURE 2.2, adegenet 2.0.1, Poptree2; GeneClass 2.0

<http://doi.org/10.1007/s10709-018-0031-1>; *Seriola dumerili*; Fish; microsatellites; MicroChecker, GenoDive; BAPS, ONCOR; Notes: 11 loci

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84987811341&doi=10.1080%2f19425120.2016.1180334&partnerID=40&md5=d7a316e874f6eecd4405c79437346c37>; *Brevoortia patronus*; Fish; microsatellites; Arlequin(AMOVA); STRUCTURE; Notes: 15 loci

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84988944375&doi=10.1155%2f2016%2f9382613&partnerID=40&md5=49d0c416a868c0d6c7a342059b1f30ba>; *Lignobrycon myersi*; Fish; mtDNA; Notes: barcode, COI

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016992596&partnerID=40&md5=0a42bf676838b8c5207f81c2dcfda144>; Perch; Fish; Notes: review

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84992660470&partnerID=40&md5=cd8270125f463a00b43047d4bd297a20>; *Sardinella albella*; Fish; mtDNA; Arlequin(Fst, exact test); Notes: 500 bp CR

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84971474328&partnerID=40&md5=298490270c68361ed2fa51ca003f7b83>; *Thymallus thymallus*; Fish; mtDNA; Notes: complete CR

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84971474328&partnerID=40&md5=298490270c68361ed2fa51ca003f7b83>; *Thymallus thymallus*; Fish; microsatellites; Fst; STRUCTURE; Notes: 10 loci

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84951969835&partnerID=40&md5=303d2fc3dc4d26015c98c7a5208a9390>; Parastichopus regalis; Echinoderms; mtDNA (COI & 16S); GENEPOP, DNASP; TCS

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84944038073&partnerID=40&md5=52942f771d8ce5f978e7c0f8f534d660>; Labeo rohita; Fish; Microsats; FSTAT 2.9.3.2, GENEPOP 3.3d, ARLEQUIN 2.0; TFPGA, PCAGEN

<http://doi.org/10.2331/suisan.81.958>; Beryx splendens; Fish; mtDNA (control region); DNASP, ARLEQUIN, MEGA 5, ; TCS; Notes: IN JAPANESE, softwares from refs

<http://doi.org/10.1002/ecs2.2589>; Salvelinus confluentus; Fish; microsatellites; gStudio (Dyer 2014), adegenet (Jombart 2008), and poppr (Kamvar et al. 2014) for AMOVA and ANOVA estimation; Notes: Simulated dataset, downstream analysis of simulated data

<http://doi.org/10.2744/CCB-1152.1>; Chelonia mydas; Reptiles; mtDNA (control region); Notes: No info, No access

<http://doi.org/10.2983/035.036.0128>; Callinectes sapidus; Crustacean; SNPs; Notes: 9600 loci GBS, no access

<http://doi.org/10.2989/1814232X.2017.1402089>; Fenneropenaeus indicus and Metapenaeus monoceros; Crustacean; mtDNA; Arlequin(AMOVA); Notes: CR, Cytb, COI

<http://doi.org/10.2989/1814232X.2017.1303398>; Lithognathus lithognathus; Fish; mtDNA; Arlequin, genepop; Notes: 720 bp CR

<http://doi.org/10.2989/1814232X.2017.1303398>; Lithognathus lithognathus; Fish; microsatellites; Arlequin, genepop, smogd; STRUCTURE; Notes: 15 loci

<http://doi.org/10.3109/19401736.2015.1053084>; Scomberomorus commerson; Fish; mtDNA; Arlequin (Fst); Notes: 842 bp ATPse 6-8

<http://doi.org/10.3109/19401736.2014.883602>; Osteobrama belangeri; Fish; mtDNA; Arlequin (Phist, AMOVA); Arlequin (Fu and Tajima, Harpending's raggedness index, SSD) Notes: 842 bp ATPas 6-8

<http://doi.org/10.3109/19401736.2014.898291>; Trachinotus blochii; Fish; mtDNA; Notes: complete mitochondrial genoma, no genetics in

<http://doi.org/10.3109/19401736.2015.1053083>; Rachycentron canadum; Fish; mtDNA; Arlequin(AMOVA, Fst); Notes: 848 bp ATPase 6,8

<http://doi.org/10.3109/19401736.2013.879655>; Pampus argenteus; Fish; mtDNA (ATPase 6/8); PHYLIP ver 3.7a, MEGA 5, DNASP 5.0, ARLEQUIN 3.5

<http://doi.org/10.1002/ece3.3215>; Alosa pseudoharengus, Alosa aestivalis; Fish; SNPs; Genetix(Fst); STRUCTURE, DAPC; GeneClass; Notes: 96 loci per species, ddRAD then FLUIDIGM

<http://doi.org/10.3318/BIOE.2015.20>; Trachurus trachurus; Fish; Microsats; FSTAT, GENEPOP, FreeNA, POWSIM, APE, LDNe 3.1; STRUCTURE; STRUCTURE-HARVESTER Notes: MDS R,

<http://doi.org/10.3354/AEI00282>; Scophthalmus maximus; Fish; SNPs; Arlequin(Fst); STRUCTURE, STRUCTURE HARVESTER, CLUMPP, DAPC; Bayescan, Lositan; BLASTN,BLAST2GO Notes: 755 loci ddRAD

<http://doi.org/10.3354/meps12078>; Chelonia mydas; Reptiles; mtDNA; Arlequin(Phist, AMOVA); genalex(PcoA); Mixstock, winbug(MSA); Notes: 856 bp CR

<http://doi.org/10.3354/meps12044>; Cancer pagurus; Crustacean; microsatellites; Arlequin(AMOVA), Fstat; PcoA, STRUCTURE; lositan; Notes: 8 loci

<http://doi.org/10.3354/meps11957>; Homarus gammarus ; Crustacean; microsatellites; Fstat, Genepop(exact test), Demetics(Jost); STRUCTURE; Notes: 15 loci

<http://doi.org/10.3354/esr00724>; Caretta caretta; Reptiles; mtDNA; Arlequin(AMOVA); Notes: 820 bp CR

<http://doi.org/10.3354/meps11114>; Clupea harengus; Fish; Microsats; POWSIM, GENEPOP; STRUCTURE, MDS in R; LOSITAN, BAYESCAN; STRUCTURE-HARVESTER, DISTRICT

<http://doi.org/10.3389/fgene.2018.00046>; Piaractus brachypomus; Fish; microsatellites; Fstat; STRUCTURE, STRUCTURE HARVESTER; Cervus (null alleles) Notes: 8 loci

<http://doi.org/10.3389/fgene.2017.00159>; Arapaima gigas; Fish; microsatellites; Arlequin(AMOVA); STRUCTURE, STRUCTURE HARVESTER; migrate-n; bottleneck Notes: 7 loci

<http://doi.org/10.3389/fmars.2016.00253>; *Stenella longirostris*, *Stenella attenuata*; Mammal; SNPs; strataG(Fst); DAPC; 4381, 3721 loci GBS

<http://doi.org/10.1002/ece3.3492>; *Oncorhynchus kisutch*; Fish; SNPs; Genalex(PCA); Bayescan, Lositan, linear regression across cohorts; Notes: 5392 loci RAD-seq

<http://doi.org/10.3856/vol43-issue2-fulltext-14>; *Seriola lalandi*; Fish; Microsats; ARLEQUIN 3.11, MICROCHECKER 2.2.3; STRUCTURE; TANDEM 1.09, COANCESTRY

<http://doi.org/10.1002/ece3.3873>; *Gadus morhua*; Fish; mtDNA genomes; Arlequin v3.5.1 (AMOVA, Φ_{ST}), ?? (SAMOVA), GenALEx v6.5 (PCoA); BAPS v6, TCS v1.21; Genepop v4.2 (IBD), MrBayes v3.2, PAUP* v4.10 (NJ Tree), BEAST v2.3, DnaSP v5.10

<http://doi.org/10.3897/natureconservation.17.10939>; *Pseudorhodeus tanago*; Fish; mtDNA; SAMOVA; Notes: 1219 bp cyt b

<http://doi.org/10.3989/scimar.04203.20A>; *Epinephelus itajara*; Fish; mtDNA (control region); MEGA 5.2, DNASP 5, ARLEQUIN 3.5, SAMOVA 2.0, ; BioESTAT 5.3

SPECIFIC CONTRACT NR. 3 **MED_Units** PROJECT

SAMPLING PROTOCOL FOR THE GENETIC SAMPLING

MATERIALS

Provided to each partner

2 ml Tubes with O-ring screw cap

Boxes with Racks with 81 apertures



Not-provided

Tweezers

Surgical scissors

Disposable scalpels

Gloves

Non-denatured **analytical grade** Ethanol 96% or absolute (100%) analytical grade Ethanol is also fine. **DO NOT use absolutely technical grade or denatured (pink) ethanol, which compromises the possibility of extracting suitable DNA.** In case there is no availability of this specific ethanol in your institution, please contact immediately the Genetic Hub for the supply of it at your premises.

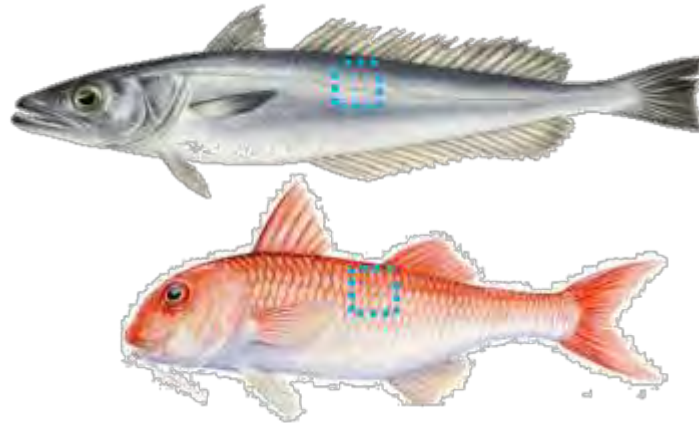
TYPE OF SAMPLING:

For each species 50 individuals per sampling area are required, at least 25 of which should be females.

Fish species

Tissue type: white skeletal muscle on the sides of the fish (highly preferred).

Volume of the muscle sample: 0.5 cm³

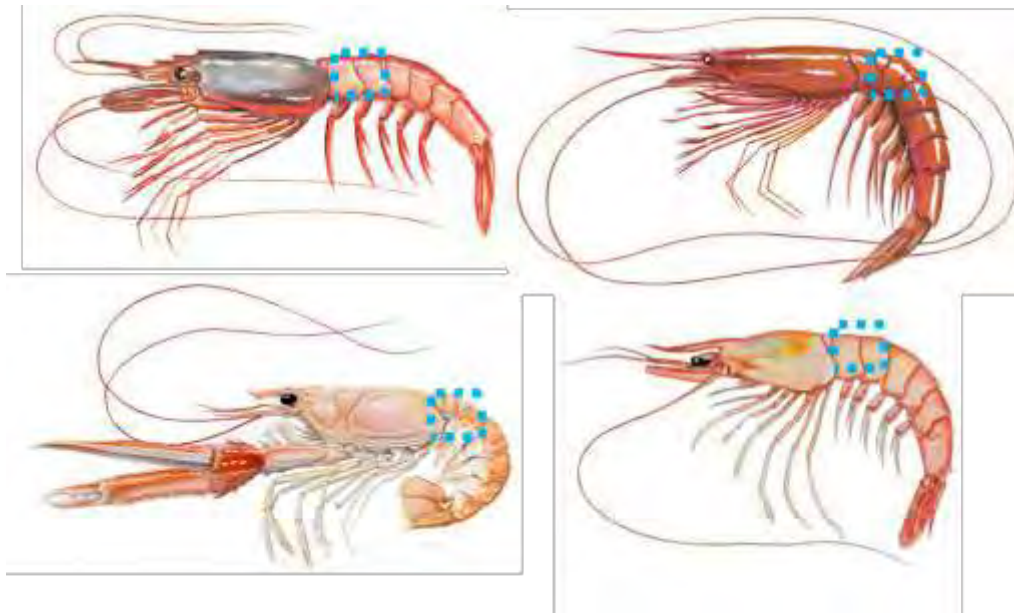


Animals and size of cut (blue rectangle) not drawn to scale

Crustacean species

Tissue type: white abdominal muscle after removal of the carapace.

Volume of the muscle sample: 0.5 cm³



Animals and size of cut (blue rectangle) not drawn to scale

PROCEDURE

Tubes already properly labelled will be provided.

Please USE the tubes provided. The sampling procedure is described at point c (see below).

In case of need to use additional tubes different from the ones provided (highly discouraged),

they are not yet ready to be used (without alcohol and label), so please strictly follow this procedure for the labeling and filling (a+b):

a) Labeling and codes

The correct labeling of the samples is an indispensable step. The specimen to be sampled must be first classified by area, then numbered from 1 to N.

Sample labels must be clearly visible on the tube.

AREA CODE - Species CODE – n° individual

Example of code:

11-Aa-25 (GSA11, *Aristeus antennatus*, individual 25)

When multiple locations are in the same GSA, indicate them with a different lowercase

Example of code with multiple sampling in the same GSA:

06a- Mm-01 (GSA6 location 'a', *Merluccius merluccius*, individual 01)

06b- Mm -03 (GSA6 location 'b', *Merluccius merluccius*, individual 03)

06c- Mm -22 (GSA6 location 'c', *Merluccius merluccius*, individual 22)

Species codes to be used:

Aa: *Aristeus antennatus*

Af: *Aristaeomorpha foliacea*

Mb: *Mullus barbatus*

Mm: *Merluccius merluccius*

Nn: *Nephrops norvegicus*

Pl: *Parapenaeus longirostris*

b) Preparation of the material before sampling

Before sampling:

- fill the tubes with at least 1mL (preferentially 1.5ml) of non-denatured **analytical grade** Ethanol 96%, or absolute (100%). **DO NOT use absolutely technical grade or denatured (pink) ethanol, which compromises the possibility of extracting suitable DNA.** Each microtube must be labeled with Sample ID according to the labeling code reported above (step a).
- label the tubes with pens containing water-resistant ink, write the code and on the side of the vial and use Scotch tape to prevent label erasing due to probable ethanol leak.
- If the tube has a flat cap, please write (if possible) the same code on the cap also.

Please remember that a sample without the code or with a non-readable code cannot be used for subsequent genetic analysis and will probably be discarded.

c) Sampling procedure

IMPORTANT: Sampling of tissue should be carried **out twice from the same individual**. Two tubes will be provided for each individual; one tube **should be immediately shipped** following the shipping instructions (see step d). **The second replicated tube, labelled 'R', should be kept in the laboratory and sent on request** in case of problems with the shipping or the laboratory procedures.

- The operator has to **wear clean gloves**
- **Collect properly identified live/freshly caught adult animals; alternatively, use animals kept in the freezer (at least -20°C), only if they have been frozen live or no more than 1 hour after collection (shrimps)**. The time elapsed from the collection/death of specimens to ethanol fixation is a critical phase; try to minimize this time whenever possible. **Avoid to use shrimps threated with sulphite**. Avoid keeping samples at field temperature for hours before processing them.
- Cut with a sharp scissor and tweezers preferentially **a biopsy (a small portion) 0.5 cm³ muscle sample from each individual**. Larger pieces are not needed and can result in bad DNA quality due to low ethanol/tissue ratio.
- Put the tissue into the ID labeled microtube and fill with ethanol 96%. Please do not get stressed and ensure that the tissue volume is no more than 10-20% of the liquid volume; it is better to have a smaller tissue sample correctly fixed in ethanol than a bigger one which is going to be badly preserved.



IMPORTANT:

Be sure than the tissue volume is no more than 20% of the liquid volume and tightly close the cap!

- To avoid contamination **clean surgical instruments for each sampled animal first with water then with commercial ethanol and dry it with a new paper towel each time**. We need to avoid cross-contamination of tissues from different individuals.
- Make use of the **Excel sampling sheet provided** to collect the available data (location, latitude, longitude, date, depth, length, weight, sex etc.). Additional phenotypic data are very welcome (e.g., reproductive stage, maturity, etc.).
- Insert data for the different species in different sheets, and the tubes in different boxes by species;
- temporarily store the tube containing the tissue in dark and cool place and store as soon as possible at -20°C. If it is not possible to keep it in freezer, make sure that the temperature does not exceed 4° C and store them in a simple fridge.
- 1-3 days after sampling, carefully remove the ethanol from the microtube and replace it **by new ethanol**. The water contained in the sample may have partly diluted the first ethanol used, so this ethanol replacement will ensure a better long-term conservation of DNA. Only in the case this is not possible, please ship the samples promptly (within 1-3 days) to do it for

you at the genetic hub.

ALWAYS USE analytical grade Ethanol 96%, or absolute (100%)

DO NOT use absolutely technical grade or denatured (pink) ethanol, which compromises the possibility of extracting suitable DNA.

If your animal is already frozen:

- Take the sample from each frozen individual and put it directly into the tube with ethanol 96%.
- The ice contained in the sample may dilute the ethanol. Therefore, change the ethanol after 1-3 days, in order to ensure proper sample preservation; alternatively see the remark above.

d) Shipping procedure

- Put the boxes of microtubes with tissue samples in a proper storage box
- Include the data sheets in the box and send the excel file or a scan copy of the datasheet also to my email to alessia.cariani@unibo.it (this will guarantee that we'll have the data file correctly archived even if there is ethanol leak)
- When the sampling is completed for the whole area you are in charge of, and the box is ready please alert alessia.cariani@unibo.it in order to have the box picked up by the courier and sent to the Genetic Hub in Bologna.

For any doubt or technical information on the sampling procedure please contact alessia.cariani@unibo.it

WP1. M1.2 - DEFINITION OF SITES FOR PILOT GENETIC STUDIES

***SPECIFIC CONTRACT NO. 03 IMPLEMENTING FRAMEWORK CONTRACT NO
EASME/EMFF/2016/032***

***STUDY ON ADVANCING FISHERIES ASSESSMENT AND MANAGEMENT ADVICE IN THE
MEDITERRANEAN AND BLACK SEA BY ALIGNING BIOLOGICAL AND MANAGEMENT UNITS OF
PRIORITY SPECIES***

***WP1 - POPULATION GENETICS AND PHYLOGEOGRAPHIC STUDIES FOR IDENTIFICATION OF
BIOLOGICAL UNITS OF PRIORITY SPECIES***

MILESTONE M1.2 V.1

DEFINITION OF SITES FOR PILOT GENETIC STUDIES

COMPILED BY: RITA CANNAS (CONISMA) - UPDATED VERSION APRIL 2019

The present document updates and complements the document included as Annex 7 of the Inception report.

It provides a 'revised' list of preferred locations for the sampling of the six species under investigation within Med_Units (**Table 1**).

The original list of the proposed sites for the collection of tissue samples (Table A7.13) has been updated based on the inputs received during the kick-off meeting with the Contracting Authority in February 2019 and feedbacks from partners during the first plenary meeting of the project in March 2019.

It includes indications for the future samplings to be accomplished for the full study (Task 1.6). The list (**Table 2**) updates Table A7.15. The final list of sites to be included in the full studies will be provided at month 10, after the completion of the pilot studies.

To minimize risks and to compensate the eventual failure of sampling in the non-EU countries (in GSA3, GSA4, GSA12-14, GSA21, GSA24, GSA26 and GSA27, **Table 3, Figure 1**), 48 additional samples have been allocated to European GSAs adjacent to the non-EU GSAs (**Table 4**). These extra 48 samples represent spare samples; they will be used only in case of unavailability of some sampling sites in order to guarantee anyhow, at the end of the sampling phase (month 12), the planned number of sites for the genetic analyses (a total of maximum 228 locations).

The document includes the list of the sampling sites by GSA and species (**Tables 5-21**), along with maps (**Figures 2-18**) with detailed indications of areas (subareas) where the sampling effort has to be performed.

Table 1 – List of the proposed sites for the pilot genetic studies

Pilot studies																
Species	GSA1	GSA3	GSA5	GSA6	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	total
<i>A. foliacea</i>			1		1	1	1	1			1	1	1			8
<i>A. antennatus</i>	1		1	1	1		1	1			1	1 ^a				8
<i>M. merluccius</i>	1	1 ^b					1*	1		1	1		1		1	8
<i>M. barbatus</i>			1		1				2 ^c	2 ^d			1		1	8
<i>N. norvegicus</i>	1		1	1	1		1	1	1				1			8
<i>P. longirostris</i>					1	1		1	1	1	1		1	1		8
total	3	1	4	2	5	2	4	5	4	4	4	2	5	1	2	48

* no otoliths available, because the samples were collected before the sampling procedure for otoliths was defined

^a if not available replace with GSA18

^b ask Copemed

^c 1 north and 1 central Adriatic, Italy

^d 1 Italy and 1 Montenegro

Table 2 – List of the provisional overall sites (pilot + additional sites) for the genetic studies. In color the non-EU GSAs.

Species	GSA1	GSA3	GSA4	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA12	GSA13	GSA14	GSA15	GSA16	GSA17	GSA18	GSA19	GSA20	GSA21	GSA22	GSA23	GSA24	GSA25	GSA26	GSA27	Total
<i>A. foliacea</i>			1	2	3		2	3	3	3	1	1	1	1	3		2	3	2	1	2	1	1	1	1		38
<i>A. antennatus</i>	1	1	1	2	3	2	1	3	2	3	1	1	1	1	2		2	3	2	1	1	1	1	1	1	1	38
<i>M. merluccius</i>	1	1	1	2	3	1	1	2	2	3	1	1	1	1	2	2	2	2	1	1	2	1	1	1	1	1	38
<i>M. barbatus</i>	1	1	1	1	2	2	1	2	2	2	1	1	1	1	2	3	3	2	1	1	2	1	1	1	1	1	38
<i>N. norvegicus</i>	1	1	1	1	3	1	1	2	3	4	1	1	1	1	3	2	2	2	2	1	3	1					38
<i>P. longirostris</i>	1	1	1	1	2	2	1	2	2	2	1	1	1	1	3	2	2	2	2	1	2	1	1	1	1	1	38
Total	5	5	6	9	16	8	7	14	14	17	6	6	6	6	15	9	13	14	10	6	12	6	5	5	5	3	228

Table 3 – List of the sites per species in the non-EU GSAs.

Species	GSA3	GSA4	GSA12	GSA13	GSA14	GSA21	GSA24	GSA26	GSA27	total
<i>A. foliacea</i>		1	1	1	1	1	1	1		7
<i>A. antennatus</i>	1	1	1	1	1	1	1	1		8
<i>M. merluccius</i>	1	1	1	1	1	1	1	1	1	9
<i>M. barbatus</i>	1	1	1	1	1	1	1	1	1	9
<i>N. norvegicus</i>	1	1	1	1	1	1				6
<i>P. longirostris</i>	1	1	1	1	1	1	1	1	1	9
Total	5	6	6	6	6	6	5	5	3	48

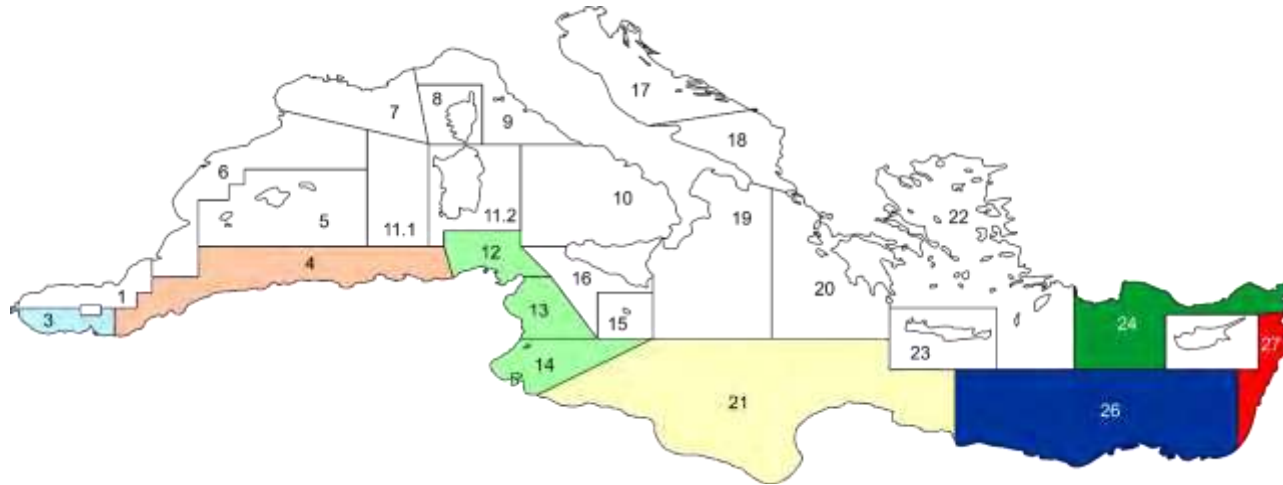


Figure 1 - Pilot sampling sites for non-EU GSAs; one sample of 50 individuals per species per GSA is required

Table 4 - List of the provisional overall sites in the EU GSAs, considering extra samples attributed to each GSA to compensate the eventual failure of sampling in the non-EU countries (light blue, pink, green, orange, blue, red)

Species	GSA1	GSA3	GSA4	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA12	GSA13	GSA14	GSA15	GSA16	GSA17	GSA18	GSA19	GSA20	GSA21	GSA22	GSA23	GSA24	GSA25	GSA26	GSA27	Total	
<i>A. foliacea</i>				2	3		2	3	3	3+1++1				1+1	3+1		2	3+1	2		2+1	1		1+1			38	
<i>A. antennatus</i>	1+1			2+1	3	2	1	3	2	3+1				1+1	2+1		2	3+1	2		1+1	1			1+1			38
<i>M. merluccius</i>	1+1			2+1	3	1	1	2	2	3+1				1+1	2+1	2	2	2+1	1		2+1	1			1+1+1			38
<i>M. barbatus</i>	1+1			1+1	2	2	1	2	2	2+1				1+1	2+1	3	3	2+1	1		2+1	1			1+1+1			38
<i>N. norvegicus</i>	1+1			1+1	3	1	1	2	3	4+1				1+1	3+1	2	2	2+1	2		3	1						38
<i>P. longirostris</i>	1+1			1+1	2	2	1	2	2	2+1				1+1	3+1	2	2	2+1	2		2+1	1			1+1+1			38
Total	10	0	0	14	16	8	7	14	14	24	0	0	0	12	21	9	13	20	10	0	17	6	0	13	0	0	228	

Sampling in GSA1

Table 5 - List of the provisional overall sites for the genetic studies in GSA1: Number of samples per species and geographic subarea

Species	GSA1	
	N° of samples	subarea
<i>A. foliacea</i>		
<i>A. antennatus</i>	2	1a+1b
<i>M. merluccius</i>	2	1a+1b
<i>M. barbatus</i>	2	1a+1b
<i>N. norvegicus</i>	2	1a+1b
<i>P. longirostris</i>	2	1a+1b
total	10	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!



Figure 2 - Pilot sampling sites: GSA1 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA5

Table 6 - List of the provisional overall sites for the genetic studies in GSA5: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA5	
	N° samples	subarea
<i>A. foliacea</i>	2	5a+5 (b)c
<i>A. antennatus</i>	3	5a+5 b +5c
<i>M. merluccius</i>	3	5a+5b+5c
<i>M. barbatus</i>	2	5a+5 (b)c
<i>N. norvegicus</i>	2	5a+5 (b)c
<i>P. longirostris</i>	2	5a+5b
total		14

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

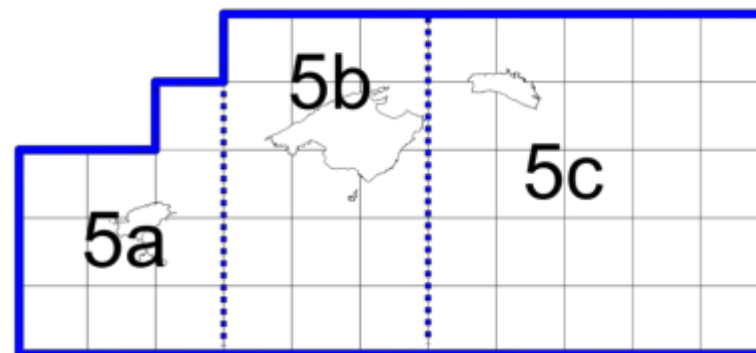


Figure 3 - Pilot sampling sites: GSA5 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA6

Table 7 - List of the provisional overall sites for the genetic studies in GSA6: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA6	
	N° samples	subarea
<i>A. foliacea</i>	3	6a+6b+6c
<i>A. antennatus</i>	3	6a+6b+6c
<i>M. merluccius</i>	3	6a+6b+6c
<i>M. barbatus</i>	2	6a+6(b)c
<i>N. norvegicus</i>	3	6a+6b+6c
<i>P. longirostris</i>	2	6a+6(b)c
total		16

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

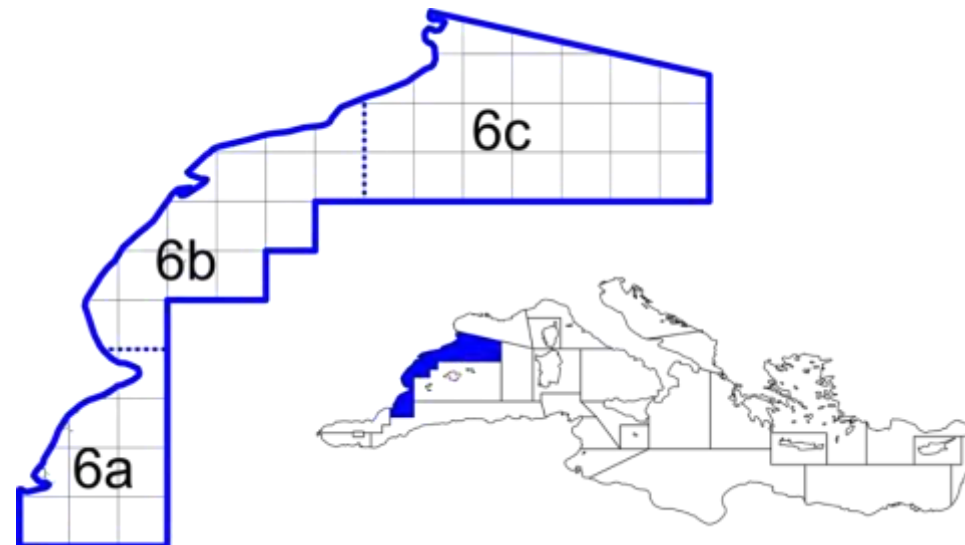


Figure 4 - Pilot sampling sites: GSA6 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA7

Table 8 - List of the provisional overall sites for the genetic studies in GSA7: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA7	
	N° samples	subarea
<i>A. foliacea</i>	0	
<i>A. antennatus</i>	2	7a+7b
<i>M. merluccius</i>	1	7a(b)
<i>M. barbatus</i>	2	7a+7b
<i>N. norvegicus</i>	1	7a(b)
<i>P. longirostris</i>	2	7a+7b
total	8	

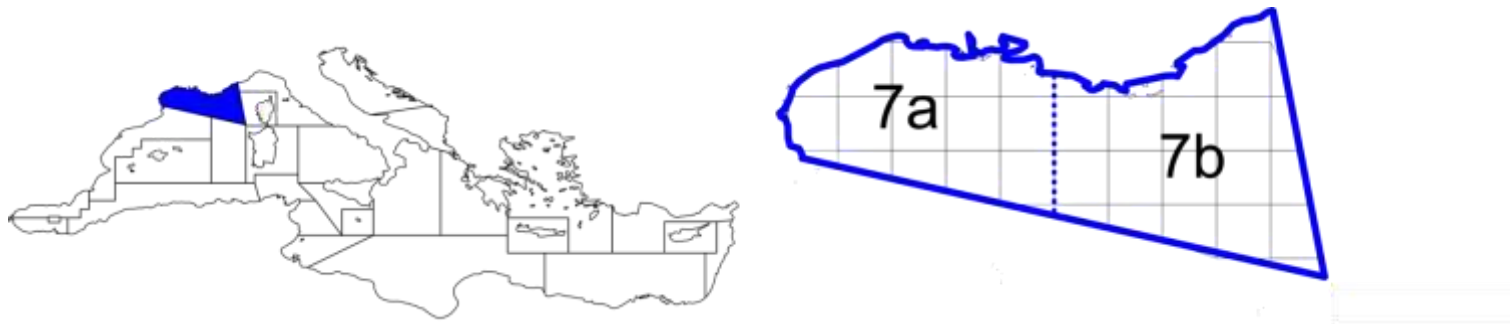


Figure 5 - Pilot sampling sites: GSA7 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA8

Table 9 - List of the provisional overall sites for the genetic studies in GSA8: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA8	
	N° samples	subarea
<i>A. foliacea</i>	2	8a+8b
<i>A. antennatus</i>	1	8a(b)
<i>M. merluccius</i>	1	8a(b)
<i>M. barbatus</i>	1	8a(b)
<i>N. norvegicus</i>	1	8a(b)
<i>P. longirostris</i>	1	8a(b)
total	7	

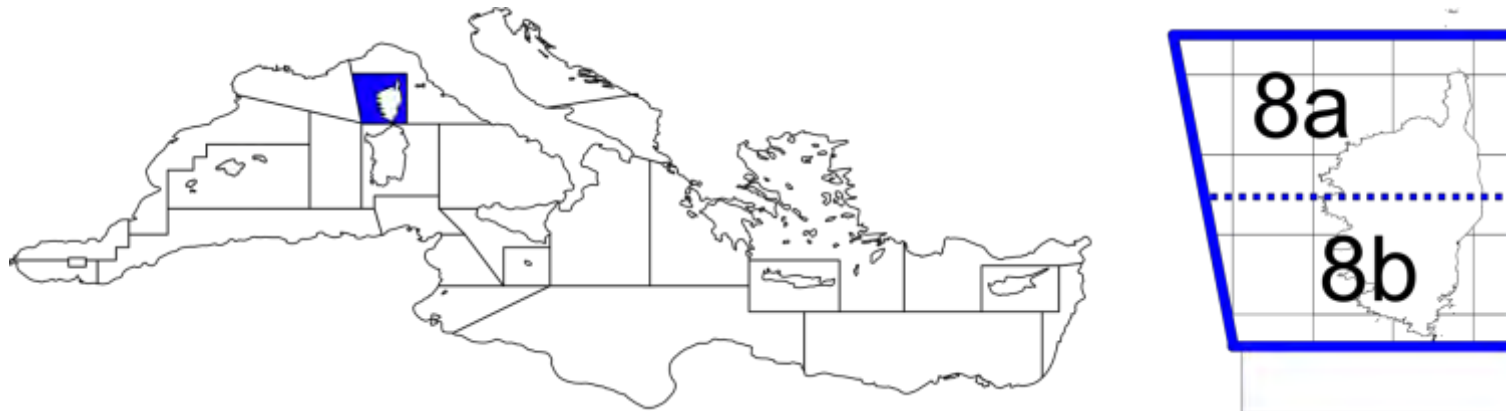


Figure 6 - Pilot sampling sites: GSA8 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA9

Table 10 - List of the provisional overall sites for the genetic studies in GSA9: Number of samples per species and geographic subarea . Between parentheses are indicated alternative subareas.

Species	GSA9	
	N° samples	subarea
<i>A. foliacea</i>	3	9a+ 9b -9c
<i>A. antennatus</i>	3	9a +9b+9c
<i>M. merluccius</i>	2	9a+9(b)c
<i>M. barbatus</i>	2	9a +9(b)c
<i>N. norvegicus</i>	2	9a (b)+9c
<i>P. longirostris</i>	2	9a +9(b)c
total	14	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

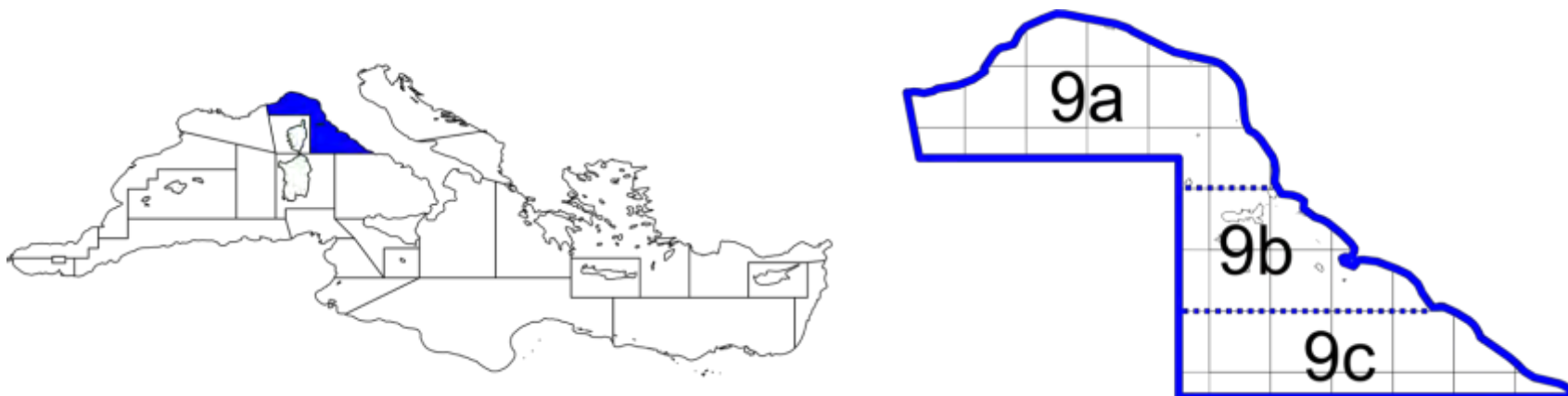


Figure 7 - Pilot sampling sites: GSA9 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA10

Table 11 - List of the provisional overall sites for the genetic studies in GSA10: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA10	
	N° samples	subarea
<i>A. foliacea</i>	3	10a+10b+10c
<i>A. antennatus</i>	2	10a+10(b)c
<i>M. merluccius</i>	2	10a+10(b)c
<i>M. barbatus</i>	2	10a+10(b)c
<i>N. norvegicus</i>	3	10a+10b+10c
<i>P. longirostris</i>	2	10a(b)+10c
total	14	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

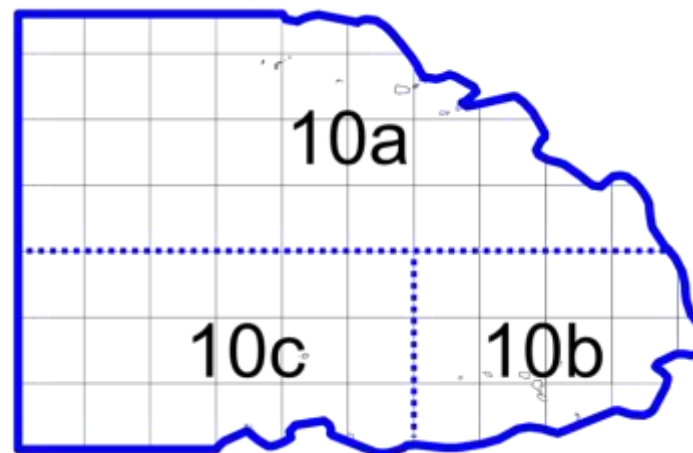


Figure 8 - Pilot sampling sites: GSA10 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA11

Table 12 - List of the provisional overall sites for the genetic studies in GSA11: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA11	
	N° samples	subarea
<i>A. foliacea</i>	5	11a+ 11b+11c+11d+11e
<i>A. antennatus</i>	4	11a(b)+11c+11d+11e
<i>M. merluccius</i>	4	11a(b)+11c+11d+11e
<i>M. barbatus</i>	3	11a(b)+11c+11d(e)
<i>N. norvegicus</i>	5	11a+11b+11c+11d+11e
<i>P. longirostris</i>	3	11a(b)+11c+11d(e)
total	24	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

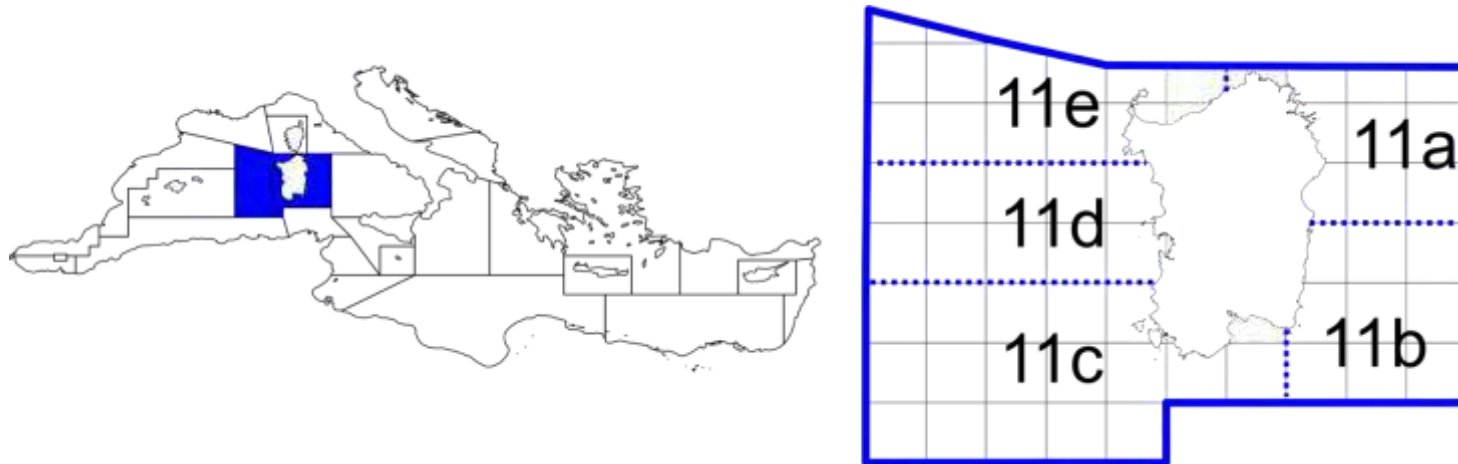


Figure 9 - Pilot sampling sites: GSA11 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA15

Table 13 - List of the provisional overall sites for the genetic studies in GSA15 Number of samples per species and geographic subarea

Species	GSA15	
	N° samples	subarea
<i>A. foliacea</i>	2	15a+15b
<i>A. antennatus</i>	2	15a+15b
<i>M. merluccius</i>	2	15a+15b
<i>M. barbatus</i>	2	15a+15b
<i>N. norvegicus</i>	2	15a+15b
<i>P. longirostris</i>	2	15a+15b
total		12

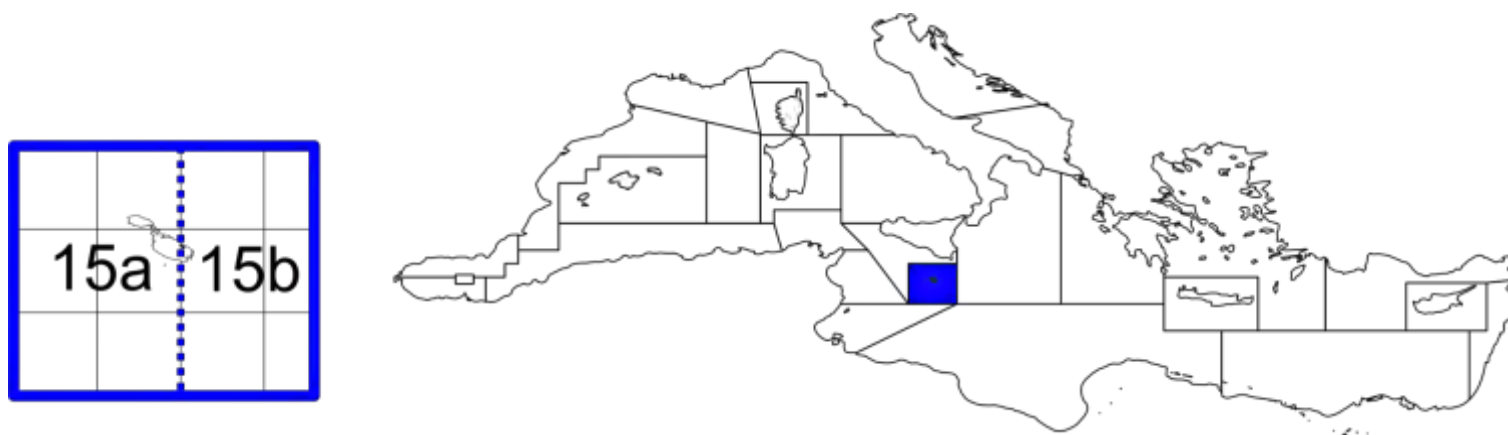


Figure 10 - Pilot sampling sites: GSA15 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA16

Table 14 - List of the provisional overall sites for the genetic studies in GSA16: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA16	
	N° samples	subarea
<i>A. foliacea</i>	4	16a+16b+16c+16d
<i>A. antennatus</i>	3	16a (b) +16c+16d
<i>M. merluccius</i>	3	16a(b)+16c+16d
<i>M. barbatus</i>	3	16a(b)+16c+16d
<i>N. norvegicus</i>	4	16a+16b+16c+16d
<i>P. longirostris</i>	4	16a+16b+16c+16d
total	21	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

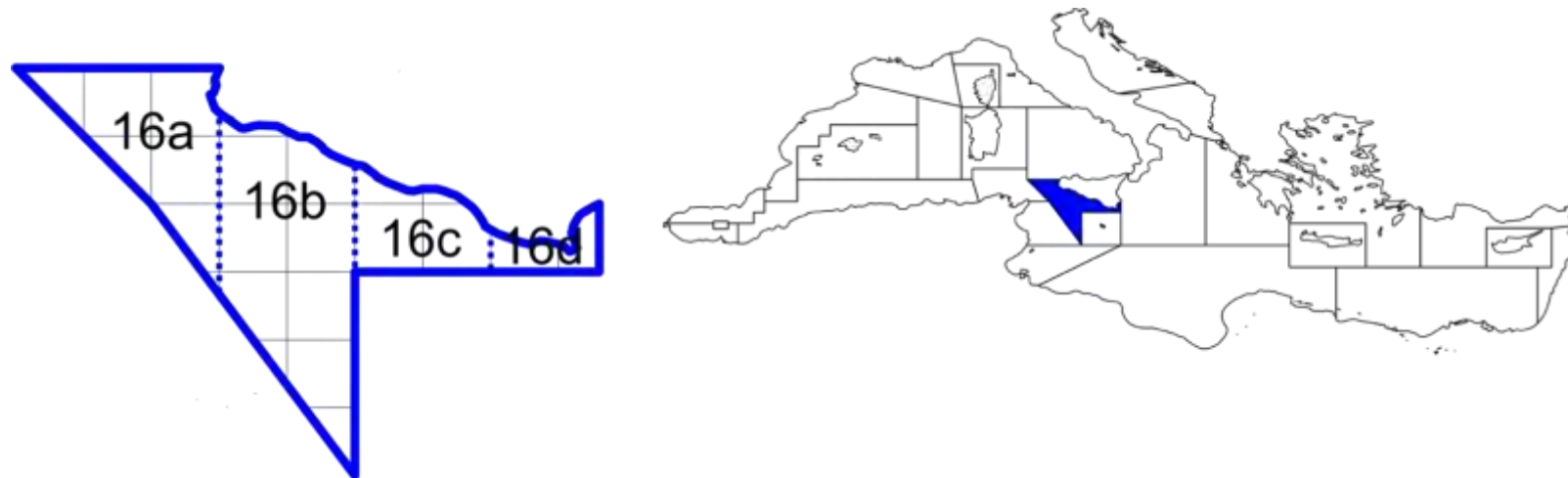


Figure 11 - Pilot sampling sites: GSA16 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA17

Table 15 - List of the provisional overall sites for the genetic studies in GSA17: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA17	
	N° samples	subarea
<i>A. foliacea</i>	0	
<i>A. antennatus</i>	0	
<i>M. merluccius</i>	2	17a(b)+17c
<i>M. barbatus</i> *	3	17a+17b+17c
<i>N. norvegicus</i>	2	17a(b)+17c
<i>P. longirostris</i>	2	17a(b)+17c
total	9	

In **yellow** and bold the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!
 * two samples for the pilot study in GSA17 (17a and 17b)

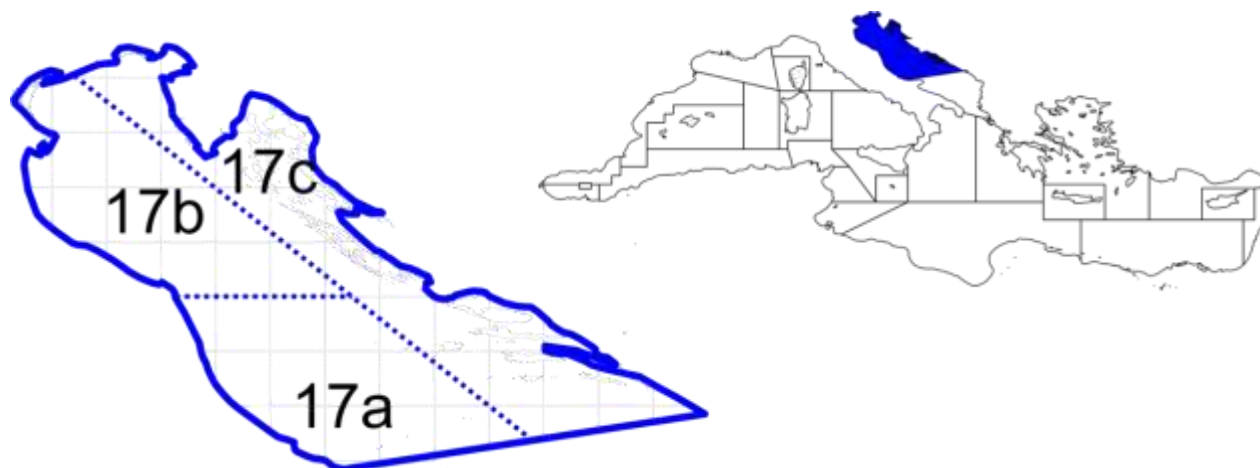


Figure 12 - Pilot sampling sites: GSA17 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA18

Table 16 - List of the provisional overall sites for the genetic studies in GSA18: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA18	
	N° samples	subarea
<i>A. foliacea</i>	2	18a(b)+18c
<i>A. antennatus</i>	2	18a(b)+18c
<i>M. merluccius</i>	2	18a(b)+18c
<i>M. barbatus</i> *	3	18a+18b+18c
<i>N. norvegicus</i>	2	18a(b)+18c
<i>P. longirostris</i>	2	18a(b)+18c
total		13

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!
 * two samples for the pilot study in GSA18 (18a or 18b, and 18c or 18d)



Figure 13 - Pilot sampling sites: GSA18 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA19

Table 17 - List of the provisional overall sites for the genetic studies in GSA19: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA19	
	N° samples	subarea
<i>A. foliacea</i>	4	19a+19b+19c+19d
<i>A. antennatus</i>	4	19a+19b+19c+19d
<i>M. merluccius</i>	3	19a+19b+19c(d)
<i>M. barbatus</i>	3	19a+19b+19c(d)
<i>N. norvegicus</i>	3	19a+19b+19c(d)
<i>P. longirostris</i>	3	19a+19b+19c(d)
total	20	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

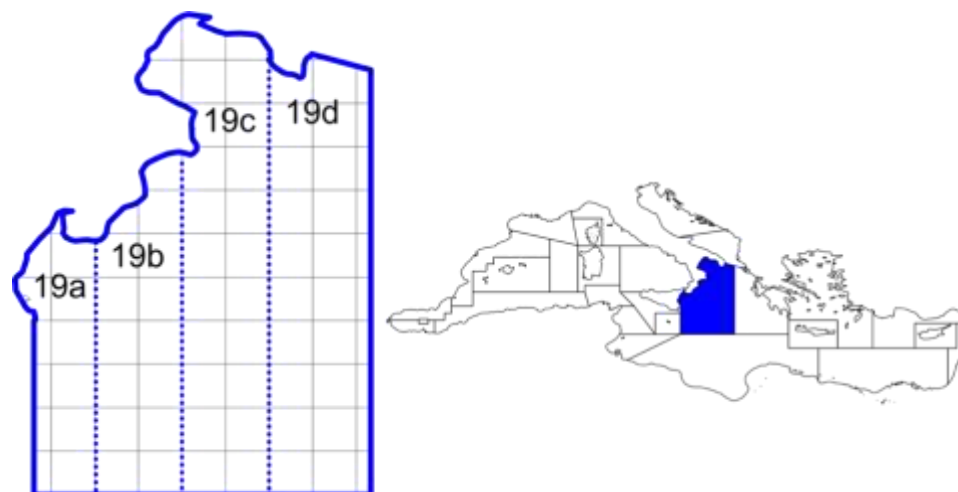


Figure 14 - Pilot sampling sites: GSA19 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA20

Table 18 - List of the provisional overall sites (pilot + additional sites) for the genetic studies in GSA20: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA20	
	N° samples	subarea
<i>A. foliacea</i>	2	20a+20b
<i>A. antennatus</i>	2	20a+20b
<i>M. merluccius</i>	1	20a(b)
<i>M. barbatus</i>	1	20a(b)
<i>N. norvegicus</i>	2	20a+20b
<i>P. longirostris</i>	2	20a+20b
total	10	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!

NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

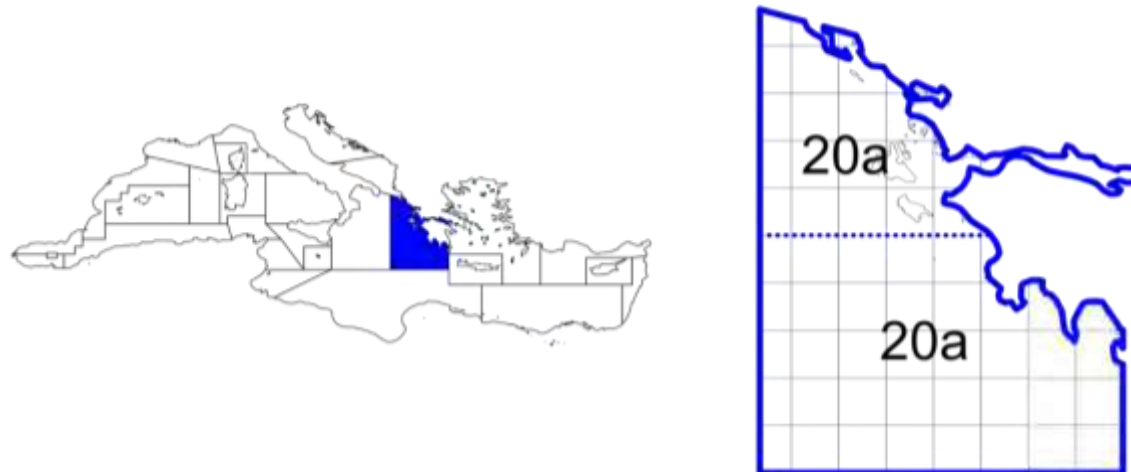


Figure 15 - Pilot sampling sites: GSA20 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA22

Table 19 - List of the provisional overall sites for the genetic studies in GSA22: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA22	
	N° samples	subarea
<i>A. foliacea</i>	3	22a+22b+22c
<i>A. antennatus</i>	2	22a(b)+22c
<i>M. merluccius</i>	3	22a+2b+22c
<i>M. barbatus</i>	3	22a+22b+22c
<i>N. norvegicus</i>	3	22a+22b+22c
<i>P. longirostris</i>	3	22a+22b+22c
total	17	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!

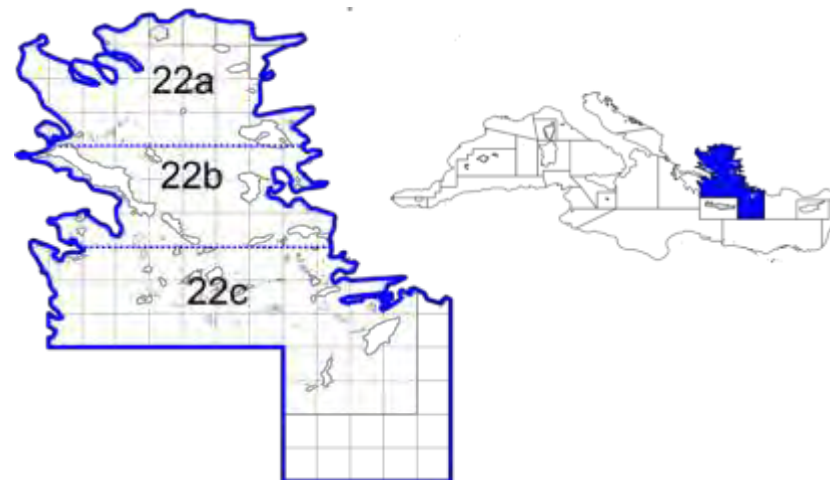


Figure 16 - Pilot sampling sites: GSA22 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA23

Table 20 - List of the provisional overall sites for the genetic studies in GSA23: Number of samples per species and geographic subarea

Species	GSA23	
	N° samples	subarea
<i>A. foliacea</i>	1	23a
<i>A. antennatus</i>	1	23a
<i>M. merluccius</i>	1	23a
<i>M. barbatus</i>	1	23a
<i>N. norvegicus</i>	1	23a
<i>P. longirostris</i>	1	23a
total	6	

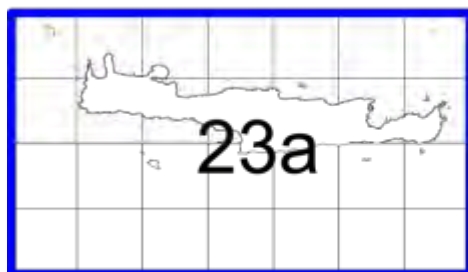


Figure 17 - Pilot sampling sites: GSA23 boundaries and subdivision at the level of the GFCM Statistical Grid

Sampling in GSA25

Table 21 - List of the provisional overall sites for the genetic studies in GSA25: Number of samples per species and geographic subarea. Between parentheses are indicated alternative subareas.

Species	GSA25	
	N° samples	subarea
<i>A. foliacea</i>	2	25a(b)+25c
<i>A. antennatus</i>	2	25a(b)+25c
<i>M. merluccius</i>	3	25a+25b+25c
<i>M. barbatus</i>	3	25a+25b+25c
<i>N. norvegicus</i>	0	
<i>P. longirostris</i>	3	25a+25b+25c
total	13	

In **yellow** and **bold** the subareas where 1 sample should be collected by May for the pilot phase!
 NB: When multiple subareas are **in bold**, it means they are equivalent, please choose just one of these for the pilot sampling and leave the other assigned subareas for the subsequent sampling!



Figure 18 - Pilot sampling sites: GSA25 boundaries and subdivision at the level of the GFCM Statistical Grid

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP1 - Population genetics and phylogeographic studies for identification of biological units of priority species

Deliverable 1.4 – Report with the results of the pilot studies

LEAD: COSTAS TSIGENOPOULOS (HCMR) & LORENZO ZANE (CONISMA)

PARTICIPANTS: LUCA BARGELLONI, RITA CANNAS, ALESSIA CARIANI, LAURA CARUGATI, LEONARDO CONGIU, RACHELE CORTI, ALICE FERRARI, RAFAELLA FRANCH, SILVIA IORI, FRANCESCO MAROSO, ILARIA ANNA MARIA MARINO, RICCARDO MELIS, CHIARA PAPETTI, FEDERICA PIATTONI, ELISABETTA PIAZZA, FAUSTO TINTI (CONISMA)

KATERINA EKONOMAKI, TEREZA MANOUSAKI, IRINI-SLAVKA POLOVINA, VASSO TERZOGLOU, ALEXANDROS TSAKOIANNIS, DIMITRIS TSAPARIS (HCMR)

Date: June 10th, 2020

Contents

Executive Summary	3
Objective	5
Methodology	5
Sampling	5
DNA extraction	8
Library Preparations & Genotyping-By-Sequence (GBS)	11
Bioinformatic Analysis	11
Population genetic analyses	12
Overall Evaluation & Phylogeographic Assessment	15
Bioinformatic analyses results	15
<i>Mullus barbatus</i>	16
<i>Nephrops norvegicus</i>	22
<i>Aristaeomorpha foliacea</i>	28
<i>Merluccius merluccius</i>	34
<i>Parapenaeus longirostris</i>	41
<i>Aristeus antennatus</i>	48
Rarefaction Analysis Results	55
Conclusions and Suggestions	60
References	65

Executive Summary

The Deliverable 1.4 refers to the results of Task 1.4 in which, for the six target species, at least eight locations throughout the Mediterranean Sea were going to constitute the pilot study and their specimens to be genetically analyzed. A number of 50 individuals from each species was collected from the 8 locations and after DNA extraction, the best specimens (in terms of extracted DNA quality) were analyzed (genotyped) through double-digest Random Amplified DNA (ddRAD) sequencing methodology. These pilot locations were identified either based on the results of previous studies, corresponding to genetically distinct populations, or according to availability at the period of summer 2019.

Even though for two fish species the ddRADseq protocol is well-established and tested, we faced major difficulties for the four crustacean species, for which little genomic information exists and no previous attempts have been reported. Additionally, a major impediment was the DNA quality which is a prerequisite for the downstream ddRADseq library preparation and the correct bioinformatic analysis. However, we have produced one ddRADseq library per species, each including some multiplexed specimens from at least 6 sampling locations.

In *M. merluccius*, graphical plots indicated low genetic differentiation between samples with very low but highly significant pairwise F_{st} values. Moreover, an analysis highlighted a strong differentiation of some GSA (namely the Spanish) from the rest of the samples and, to a lesser extent, a further differentiation of samples from the S. Adriatic. This result represents a significant advance in respect to previous published studies, where differences were not detected between Mediterranean samples, unless using outlier loci.

In the red mullet (*Mullus barbatus*), all samples showed a great similarity between them (very low F_{st} values) with the exception of the one coming from the Spanish coast; however, the latter sample performed the worst in the sequencing process and lead to the least number of loci. When current results are compared to those previously reported in the literature, i) we do not encounter the differentiation of the Adriatic Sea from the neighboring regions, and ii) the Spanish sample differentiation (like the Spanish one) has been also advanced but not supported with confidence by our results.

In the deep water rose shrimp (*Parapenaeus longirostris*), pairwise genetic distances (F_{st}) showed a low level of differentiation among samples, with two comparisons involving GSA17ab significant after correction for multiple tests. Similarly, AMOVA pointed out a significant differentiation between specimens from GSA17 from the others, with about 1.6% of the overall genetic variation attributable to this subdivision. Results are concordant with the existence of significant differentiation among Mediterranean samples, already detected with other genetic markers (AFLP and mtDNA).

In the blue and red shrimp (*Aristeus antennatus*), pairwise genetic distances showed no differences among samples, though global F_{st} was small but statistically significant. Results of this preliminary analysis seems to confirm the very low, if any, genetic differentiation among Mediterranean samples detected by allozymes, microsatellites, mtDNA sequences, and AFLPs.

In the Norway lobster (*Nephrops norvegicus*), samples showed a moderate similarity between them with the highest observed between the N. Adriatic sample and all the others. This was also evidenced using other types of analyses and pointed out a significant differentiation between the above sample and all others (~7% differentiation). With practically few recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, current results are the first to show a slight but considerable differentiation of the Adriatic sample.

Finally, for *A. foliacea* all samples showed a great similarity between them with the exception of the two coming from the N. Tyrrhenian coast and South Sardinia; however, the latter two samples performed the best in the sequencing process and lead to the greatest number of loci but additional results for several scenarios of grouping *A. foliacea* samples did not indicate any significant differentiation between groups. Current results refer to samples from the Central and Western Mediterranean where the lack of genetic differentiation among locations has already been reported.

As a summary and based on these provisional results, indications have been given for the future sampling activities in order to have the best quality DNA, stressing in particular the need to collect the tissues immediately after the capture of the animals (within 1-2 hours from the haul), and to strictly avoid the use of any preservative (chemicals) especially in crustaceans. Alternatively, if specimens were frozen and transferred to the laboratories, collectors were advised to dissect and preserve muscle tissues without allowing the samples to thaw.

A rarefaction approach was conducted in the two species for which the ddRAD provided the most reliable results, i.e. *Merluccius merluccius* and *Nephrops norvegicus*. The main aim of this analysis was to investigate if a reduction in the sampling effort (i.e. the number of specimens/sample) might substantially influence the population genetics results. In our case, results showed that the values of heterozygosity and population differentiation statistics (pairwise F_{st}) were very similar when calculated on the full dataset and on reduced datasets obtained for three different scenarios tested. Significance for pairwise F_{st} remained also similar, though it was lost in some comparisons with a very small original F_{st} .

Objective

The main objectives of this deliverable is to report preliminary results coming from a pilot genetic study in which we aimed at i) optimizing laboratory protocols (type of marker to use, sample size, level of sequencing/number of markers); ii) optimizing the final sampling effort by area/species; and iii) analyzing, if possible, the temporal stability, comparing the current results with those of previous genetic studies.

Methodology

A Genotyping-by-Sequencing (GBS) methodology constructing reduced-representation libraries in each species was selected. SNPs markers newly isolated following the double-digest Random Amplified DNA (ddRAD) sequencing are used. Full description of the methodology has been provided in the Inception and 1st Progress Report.

Sampling

A total of 48 sampling sites have been included in the pilot studies.

In Table 1.4.0 and Figure 1.4.0, we compare the sampling sites included in the pilot phase as originally proposed (upper table, and left map) and actually realized (lower table, and right map). In most cases, the foreseen samples for the Western and Central Mediterranean were available, while there were not samples from the Eastern Mediterranean, the Alboran area, the Balearic Islands and the Southern Mediterranean. These areas, of priority importance, will be included anyhow in the full genetic studies.

The details of the samples per species finally included in the pilot studies are described in the following sections (Table 1.4.1).

Table 1.4.0 Sampling sites included in the pilot phase as originally proposed (upper table) and actually realized (lower table). The text and cells in green indicate the sites not foreseen in the proposed list, while the text and cells in red indicate the sites foreseen but not available due to the failure of sampling.

Pilot sampling sites – proposed as in the First Progress Report Annex 6 Table 1																		
Species	GSA1	GSA3	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	total
<i>A. foliacea</i>			1				1	1	1	1			1	1	1			8
<i>A. antennatus</i>	1		1	1			1		1	1			1	1				8
<i>M. merluccius</i>	1	1							1	1		1	1		1		1	8
<i>M. barbatus</i>			1				1				2	2			1		1	8
<i>N. norvegicus</i>	1		1	1			1		1	1	1				1			8
<i>P. longirostris</i>							1	1		1	1	1	1		1	1		8
total	3	1	4	2	0	0	5	2	4	5	4	4	4	2	5	1	2	48

Pilot sampling sites – realized based on the samples available at the end of July 2019																		
Species	GSA1	GSA3	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	total
<i>A. foliacea</i>							1	1		1		1						8
<i>A. antennatus</i>					1						1		1					8
<i>M. merluccius</i>				1			1		1	1			1	1				8
<i>M. barbatus</i>				1			1		1		2	2						8
<i>N. norvegicus</i>				1		1	1		1	1	1	1						8
<i>P. longirostris</i>				1			1	1	1	1	1	1	1					8
total	0	0	0	8	1	1	6	1	12	3	4	7	5	0	0	0	0	48

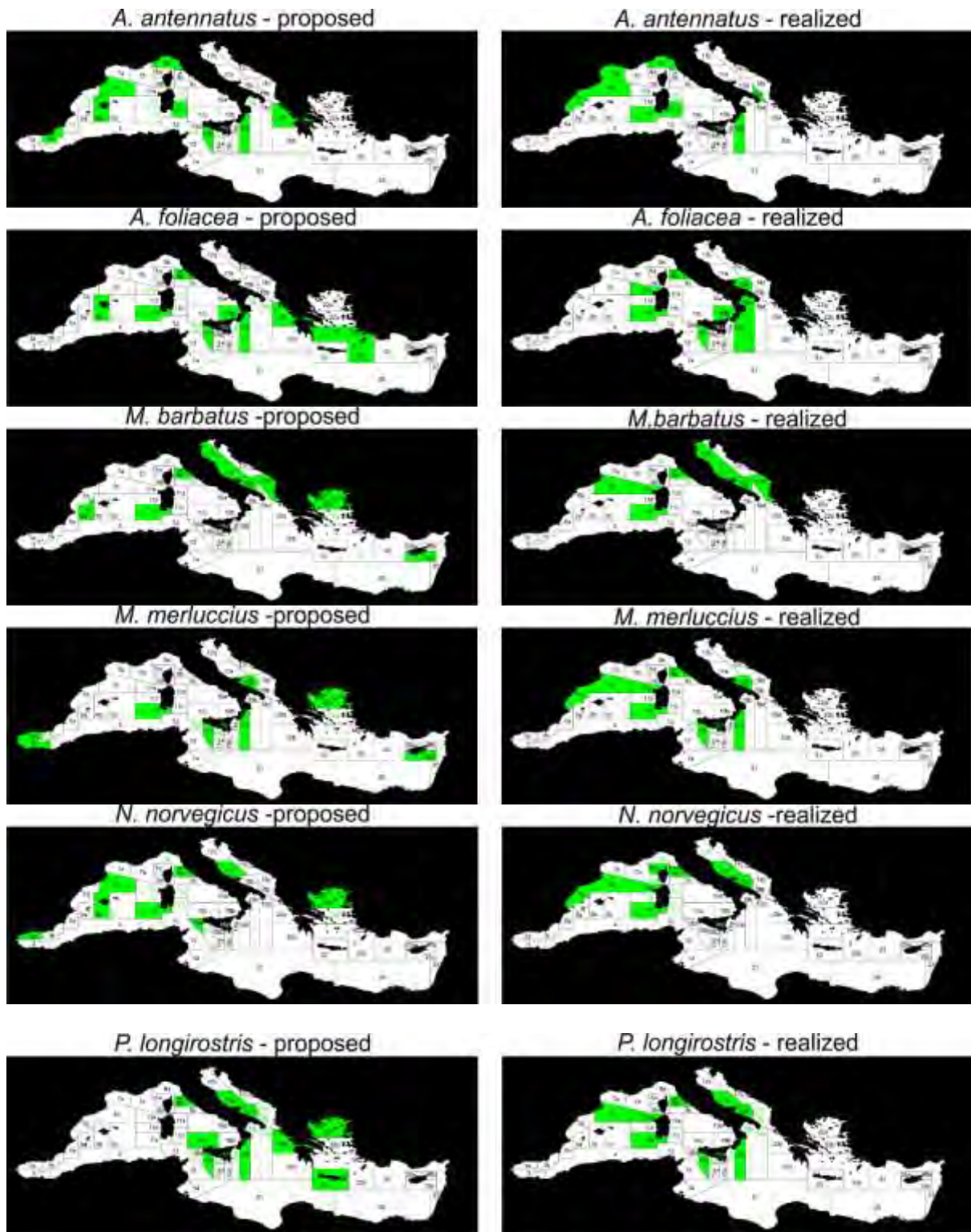


Figure 1.4.0 Maps showing the sites included in the pilot phase as originally proposed (left) and actually realized (right). The text in green indicates alternative subareas for the sampling (see Annex 6 in the First progress report).

DNA extraction

Standard DNA salt extraction protocols by Miller et al., 1988 and Cruz et al. 2017 were used in HCMR and CoNISMa, respectively. Unfortunately, alternative extraction protocols based on commercial kits, successfully used in the preliminary phases, failed to provide high molecular weight (HMW) DNA, probably due to the lower quality of tissues available, the lower efficiency and lower yields of these extraction methods when dealing with problematic samples. The salt extraction protocol is considered a rather cheap approach, but significantly time and labor-consuming, requiring adequate work subdivision among labs and involvement of more time and personnel than previously planned.

In general, after processing more than 400 samples per species, the salting out method provided HMW DNA without RNA contaminants in most cases. DNA extracts have been classified into three (3) quality categories ranging from 1 (high quality) to 3 (not usable for ddRAD). Category 1 (good quality) refers to DNA extracts where only a sharp band of HMW DNA is present (no smear). Category 2 (medium quality) refers to DNA extracts where both a band of HMW DNA and some smear of degraded DNA are present, or to cases that the band of HMW is absent and category 3 where the DNA is degraded or absent. Some samples from specific GSA subareas failed to provide HMW DNA for some species (Table 1.4.1). Moreover, for all the species high intra-area variability was observed, which raised the question if the indications provided for sampling have been respected and/or some other, not identified, factors that could affect DNA extraction results. In particular, the indication to provide tissues from fresh animals was probably not always strictly followed due to logistic problems, when specimens have to be obtained from commercial landings and or from frozen individuals, probably totally unfrozen before processing, despite the opposite clearly required in the protocol. However, in some situations in which the same operator and procedures were applied for the sample processing the results were contrasting, depending on the species.

Table 1.4.1 – List of the DNA extracts for the different species and areas. The cells in green indicate the number of extracts of Category 1- good quality, in light green the extracts of Category 2- medium quality, in orange the extracts of Category 3 the extracts of poor quality, not usable for ddRAD. The last row represents the specimens finally included in the ddRADseq library.

DNA category	Species								
	<i>A. foliacea</i>								
Sub-area GSA	09b(p)	10b	11e	11(p)c	16b	18a	19b	19c	total
1	49	44	50	48	23	10	0	0	224
2	1	2	0	0	27	32	3	47	112
3	0	4	0	2	0	8	47	3	64
ddRAD	50	44	47	50	44	46	0	7	288

A. antennatus									
Sub-area GSA	06b	06c	07a	09p	11ab	11p	18b	19b	total
1	0	1	32	17	30	26	30	38	174
2	15	15	9	20	7	12	13	11	102
3	35	34	9	13	13	12	7	1	124
ddRAD	15	14	41	36	37	38	43	49	273
M. barbatus									
Sub-area GSA	06c	09b	11c	11de	17a	17b	18a	18c	total
1	0	6	0	0	0	1	25	0	32
2	46	44	37	0	50	45	25	0	247
3	4	0	13	50	0	4	0	50	121
ddRAD	46	50	46	0	50	50	46	0	288
M. merluccius									
Sub-area GSA	06b	06c	09bc	11c	11e	16ab	18a	19b	total
1	38	41	47	32	48	45	49	50	350
2	12	6	1	17	1	5	0	0	42
3	0	3	2	1	1	0	1	0	8
ddRAD	33	30	37	29	38	38	41	42	288
N. norvegicus									
Sub-area GSA	06b	06c	08ab	09(a)b	11e	11p(c)	17ab	18a	total
1	10	14	50	50	50	41	47	0	262
2	19	10	0	0	0	3	3	8	43
3	21	26	0	0	0	6	0	42	95
ddRAD	24	22	50	50	50	46	46	0	288
P. longirostris									
Sub-area GSA	06c	09b	11de	11p	16b	17ab	18a	19b	total
1	12	27	43	40	18	36	45	38	259
2	15	13	5	8	17	11	3	11	83
3	23	10	2	2	15	3	2	1	58
ddRAD	27	36	39	38	35	36	39	38	288

Among the three species in charge of HCMR (*M. barbatus*, *A. foliacea* and *N. norvegicus*), *M. barbatus* was the one that performed worse since smears of degraded DNA were present in most samples. As a result, only 6 populations were processed for library construction with some of them included although considered of medium or bad quality. Subareas 11de and 18c failed to provide usable DNA and were excluded. For *N. norvegicus*, the extracted DNA was of high quality in most samples but again in some subareas (GSA6b, 6c, 18a) the quality of DNA was very low or even prohibitive for library construction. Finally, for *A. foliacea* the majority of samples provided DNA of good quality, except for 2 subareas (GSA19b, 19c).

As concerns the three species in charge of CONISMA (*M. merluccius*, *A. antennatus*, and *P. longirostris*), hake (*M. merluccius*) resulted to have high quality DNA with only two extracts of very poor quality. For crustaceans a certain fraction of the samples was not usable for *P. longirostris*, with subareas 06c, 09b and 16b showing about 50% of the specimens with poor quality DNA (cat. 3) and even larger fraction for *A. antennatus*, with subareas 6b and 6c providing DNA classified only in categories 2 and 3 and less than 50% of specimens providing good or medium quality DNA in all subareas, except 19b.

Based on these provisional results, indications have been given for the sampling activities in order to have the best quality DNA, stressing in particular the need to collect the tissues immediately after the capture of the animals (within 1-2 hours from the haul), and to strictly avoid the use of any preservative (chemicals) especially in crustaceans. Alternatively, if specimens were frozen and transferred to the laboratories, collectors were advised to dissect and preserve muscle tissues without allowing the samples to thaw.

The protocols for the DNA extraction, given that the quality of DNA is crucial for ddRAD method, mainly because of the need for accurate quantification and equal representation of all samples in the final library, included the following 6-steps workflow and adopted in this pilot study:

1. DNA extraction;
2. 1st Quality Control through gel electrophoresis images and Nanodrop spectrophotometer measurements;
3. Post-extraction treatment with RNase;
4. 2nd Quality Control: gel image;
5. DNA quantification: Qubit Fluorometer and Qubit dsDNA BR Assay;
6. Final concentration adjustment with dilution and transferring of samples to 96-well plates for library construction.

Library Preparations & Genotyping-By-Sequence (GBS)

For the SNP isolation and identification, some individual DNAs available in all six species have been digested each time by two specific high-fidelity Restriction Enzymes (RE). This is the part of the pilot study which requires the most careful optimization, because the great difference in the genome size of crustaceans compared to fish and the dissimilar GC content (e.g. less frequent cutting sites) among species will result in different performance of different REs. Several Restriction Enzymes (RE) were tested to identify the best combination for the species under investigation and more specifically: *SbfI* (CCTGCA[^]GG), *SphI* (GCATG[^]C), *NlaIII* (CATG[^]) and *PstI* (CTGCA[^]G). A set of adapters was ordered to construct double-digest Random Amplified DNA (ddRAD) libraries and simultaneously genotype 288 samples in a single run.

In May-June, all reagents (enzymes, tags, primers) and materials were ordered and the sequencing center selected and alerted in order to be ready to process the libraries in August- September (according to the original timetable).

The library preparation: the first step in the library design was to evaluate the digestion and to choose the appropriate pair of enzymes for each species. To improve efficiency and reduce costs HCMR also performed the digestion tests on the three species assigned to CONISMA. After the digestion tests, the appropriate pair of enzymes for the crustaceans resulted *SbfI* -*NlaIII* (8-base cutter & 4-base cutter, respectively) and *PstI* - *NlaIII* (6-base cutter & 4-base cutter, respectively), while for the fish species the pair *SbfI*-*SphI* (8-base cutter & 6-base cutter, respectively) i.e. a classical pair of RE used in fish studies in the HCMR lab.

Details on the concentration of enzymes used, amount of DNA, PCR-cycling conditions and gel cut-window chosen are provided in M1.3.

Similarly, the detailed description of the pipeline for the bioinformatic analyses and the first preliminary results for *M. barbatus* and *M. merluccius* are provided in M1.3.

Bioinformatic Analysis

The raw FASTQ files were quality-checked in FastQC 0.11.3. Sequenced reads from each of the species were analyzed separately using STACKS v.2.4 pipeline (Catchen et al. 2013), in order to quality control the reads, identify the genomic loci sequenced, genotype each individual, and conduct basic population genetics analysis. For the quality control, reads were processed with the STACKS component *process_radtags.pl* which removes reads with low quality and reads including Illumina adapters. Then, the surviving high quality forward reads of each sample were used for building *de novo* the loci of that particular sample with the STACKS component *ustacks*. The parameters used were -M 3 (Maximum distance allowed between loci) and -m 3 (Minimum depth of coverage required to create a locus) according to the recommendations from the software manual. Then, a catalogue of loci was

built for each species using the STACKS component *cstacks* with the parameter *-n* equal to 3 (number of mismatches allowed between sample loci when building the catalog). Following the catalogue construction, the loci of each individual were matched to the catalogue through *sstacks*. The next step included the use of *gstacks* to assemble and merge the second read of each pair, call variant sites and identify the genotype of each sample for each catalogue locus. Finally, *populations* was executed by keeping only the loci that were sequenced in >50% of the individuals of each population in at least 4 out of the 6+ populations of each species, with >0.1 minor allele frequency. For each STACKS locus, only one randomly selected SNP was kept.

Through STACKS component populations, we extracted the selected SNPs in vcf, genepop and structure format.

Population genetic analyses

Estimates of genetic diversity within samples in terms of observed (H_o) and expected heterozygosity (H_e) were calculated for each geographical sample with Genetix (Belkir et al 1996-2004).

To estimate the level of differentiation among samples, pairwise and global F_{ST} values were calculated with Arlequin ver. 3.5.1.2. To correct for multiple testing the Benjamini–Yekutieli procedure (Benjamini and Yekutieli 2001), controlling for the false discovery rate (FDR), was used.

The Analysis of Molecular Variance (AMOVA) was performed again with Arlequin, grouping the samples on the basis of *a priori* hierarchical geographical structure. The statistical significance of the resulting values was estimated by comparing the observed distribution with a null distribution generated by 10,000 permutations, in which individuals were redistributed randomly into samples. For each species a different threshold allowed for missing data was set. Specifically, for *M. merluccius* a 0.2 level of allowed missing data, while for *A. antennatus*, *P. longirostris*, *M. barbatus* and *A. foliacea* a value of 0.5 was set, due to the quality of datasets. For *N. norvegicus*, which performed the best quality of dataset the level of allowed missing data was set to 0.02.

The Discriminant Analysis of Principal Components (DAPC) was performed with Adegenet ver. 2.1.1 (R version 3.5.1, R Development Core Team, 2014; <http://www.r-project.org>). The DAPC approach is a multivariate method that does not rely on specific population genetic models. According to this method, genetic data are first transformed using Principal Component Analysis (PCA) into components explaining most of the genetic variation. These components are then used to perform a linear Discriminant Analysis (DA), which provides variables describing genetic groups, minimizing the genetic variance within populations, while maximizing among-population variation.

Prior to running a Discriminant Analysis (DA), an optimum number of principal components was identified using the `optim.a.score ()` function. The DA was then run on the retained principal components using the `dapc ()` function. Finally, after selecting the best number of eigenvalues for the DA analysis, the DAPC results (DAPC scatterplots) were visualized graphically with the `scatter.dapc ()` function.

To determine the number of expected genetic clusters (K) present in the dataset, without any a priori population definition, the `find.clusters()` function included in ADEGENET was initially used to run successive numbers of (K)-means clusters of the individuals, across a range of $K = 1-20$. We identified the best supported number of clusters through comparison of the Bayesian Information Criterion (BIC) for the different values of K.

Finally, the structure files were imported in STRUCTURE v 2.3.4 software (Pritchard et al., 2000; Falush et al 2003, 2007; Hubiz et al., 2009) to evaluate the presence of clusters with the sampled populations. Structure was run with k from 1 to 10 (10 replicates for each k, 100'000 burnin cycles and 1'000'000 iterations). To select the best k parameter representing the number of clusters of the dataset, we used the EVANNO method (Evanno et al. 2005) as implemented in Structure Harvester (Earl & vonHoldt 2012).

A rarefaction analysis was carried out to test the effect of reducing the sample size of the populations on some downstream population genetic analysis. We used *Merluccius merluccius* and *Nephrops norvegicus* as test datasets, being the ones with the most reliable data so far obtained (see results section). Three scenarios were tested for each species: i) all the populations with the sample size of the smallest one (26 for *M. merluccius* and 16 for *N. norvegicus*), ii) all the populations reduced at 75% of their size; iii) all the populations reduced at 50% of their size.

<i>M. merluccius</i>	samples	31	30	33	26	35	34	39	40
	pop	06b	06c	09bc	11c	11e	16ab	18a	19b
SCENARIOS									
smallest		26	26	26	26	26	26	26	26
reduction 50%		16	15	17	13	18	17	20	20
reduction 25%		23	23	25	20	26	26	29	30

<i>N. norvegicus</i>	samples	23	16	50	50	50	44	31
	pop	06b	06c	8ab	09ab	11e	11p	17ab
Scenarios								
smallest		16	16	16	16	16	16	16
reduction 50%		12	8	25	25	25	22	16
reduction 25%		17	12	38	38	38	33	23

For each scenario, 100 replicates were analysed with different subsamples of the full dataset. The statistics calculated were: a) mean observed heterozygosity per population (using R's package *hierfstat*); b) expected heterozygosity per population (using *adegenet*'s `hs()` function); c) percentage of individuals correctly assigned to their origin population by

adegenet's DAPC (using the number of principal components achieving the lowest MSE according to `xval()` function); d) F_{st} matrix with significance calculated over 1000 permutations (using R's *strataG* package and `pairwise.test()` function). For heterozygosities, DAPC assignment and F_{st} values the average, minimum and maximum across 100 replicates were reported. For F_{st} p-values, the percentage of tests providing significant results (with threshold p-value of 0.05) was reported for each pairwise comparison. To allow a proper comparison, also the full dataset was analysed with the same methods and protocol (no intervals available for the "full" scenario as there was only one sampling possibility). The rarefaction analysis was run automatically by using a custom made R script.

Overall Evaluation & Phylogeographic Assessment

Bioinformatic analyses results

For each species, some 600 million sequences were retained after demultiplexing from each run of the Illumina HiSeq400 instrument.

Stacks analyses showed a great variation in the number of sequenced loci among the six species (see Table 1.4.2 for an overview). In particular, for *A. foliacea*, the analysis resulted in a catalogue containing 1,204,691 loci with effective per-sample coverage: mean=22.2x, stdev=21.1x, min=5.2x, max=128.5x. For *N. norvegicus* the final catalogue contained 749,028 loci with effective per-sample coverage: mean=64.3x, stdev=34.8x, min=6.5x, max=170.2x. For *M. barbatus* the final catalogue contained 47,709 loci with effective per-sample coverage: mean=195.6x, stdev=71.1x, min=5.1x, max=373x. For *M. merluccius* the final catalogue contained 101,854 loci with effective per-sample coverage: mean=67.8x, stdev=22.9x, min=5.5x, max=127.6x. For *P. longirostris* the final catalogue contained 970,243 loci with effective per-sample coverage: mean=7.4x, stdev=3.3x, min=3.9x, max=29.3x. For *A. antennatus* the final catalogue contained 1,761,996 loci with effective per-sample coverage: mean=5.6x, stdev=1.0x, min=4.3x, max=12.4x.

Table 1.4.2. Stacks pipeline output summary per species

Species	> # of populations	Total # of loci	> % samples	Final SNPs	N° samples
<i>M. barbatus</i>	4	47,709	50	580	210
<i>N. norvegicus</i>	4	749,028	80	1,393	264
<i>A. foliacea</i>	4	1,204,691	50	3,437	182
<i>M. merluccius</i>	8	101,854	80	734	268
<i>P. longirostris</i>	8	970,243	50	1,045	226
<i>A. antennatus</i>	8	1,761,996	50	1,253	197

Mullus barbatus

Using all individuals, we constructed the "catalogue" which comprised some 48K ddRAD loci. The number of (Ustacks) loci varied significantly among specimens (from 6 to 13,738) and among populations on average (732 in 06c to 2,846 in 17a).

After several steps, described in M1.3, a panel of 580 higher quality SNPs, present in at least 50% of the samples and four populations, was used for all downstream differentiation analysis.

Table 1.4.3. Stacks pipeline output summary for *M. barbatus*

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
06c	369,890.09	666	4,000,818	732.63	6	3,617
09b	1,776,197.14	13,892	34,414,012	1,990.50	63	11,715
11c	3,240,896.46	24,388	40,883,238	2,473.17	102	13,738
17a	2,934,908.36	12,113	15,225,479	2,846.58	50	6,426
17b	2,665,168.22	27,678	17,132,727	2,430.28	128	7,344
18a	2,081,036.61	12,614	20,888,213	1,954.43	60	7,875
ALL	2189713.72	666	40,883,238	2,085.90	6	13,738

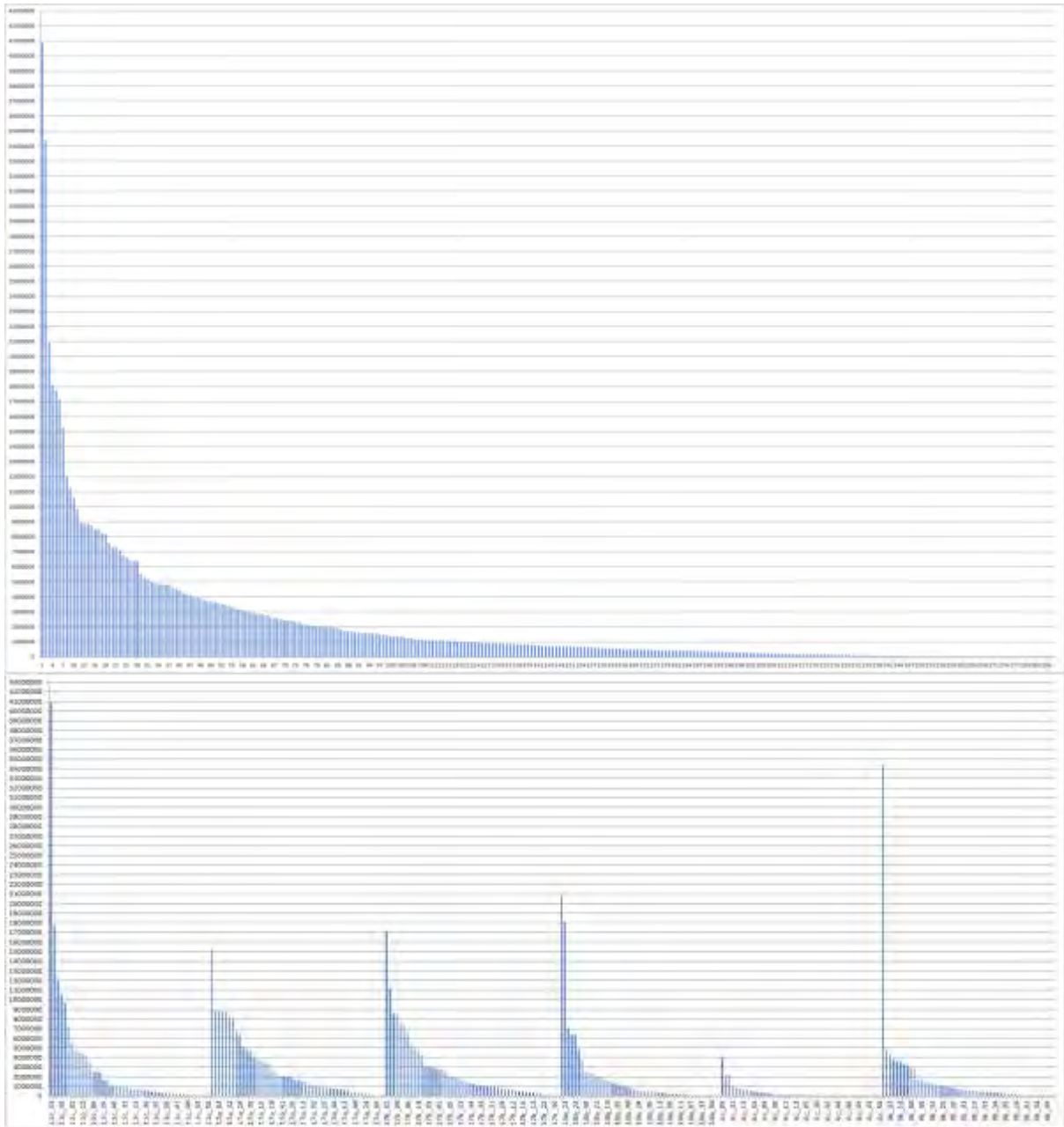
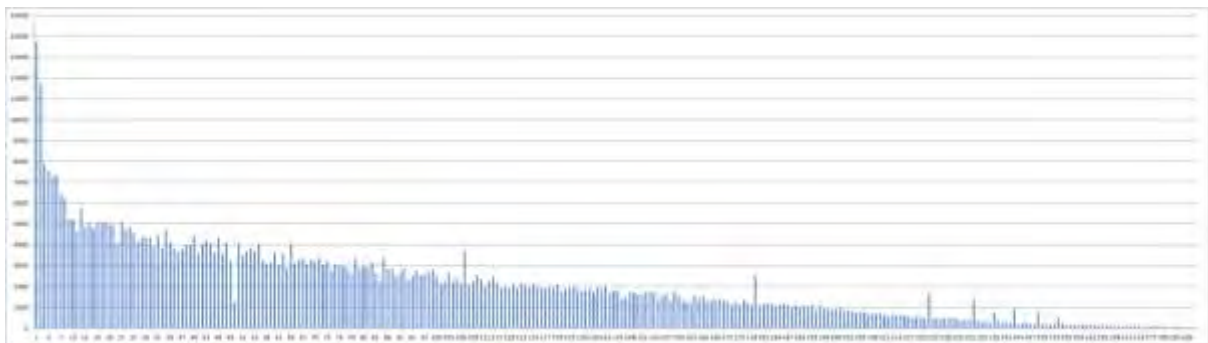


Fig. 1.4.1: Retained reads after demultiplexing for (A) the whole *M. barbatus* dataset, and (B) each population separately



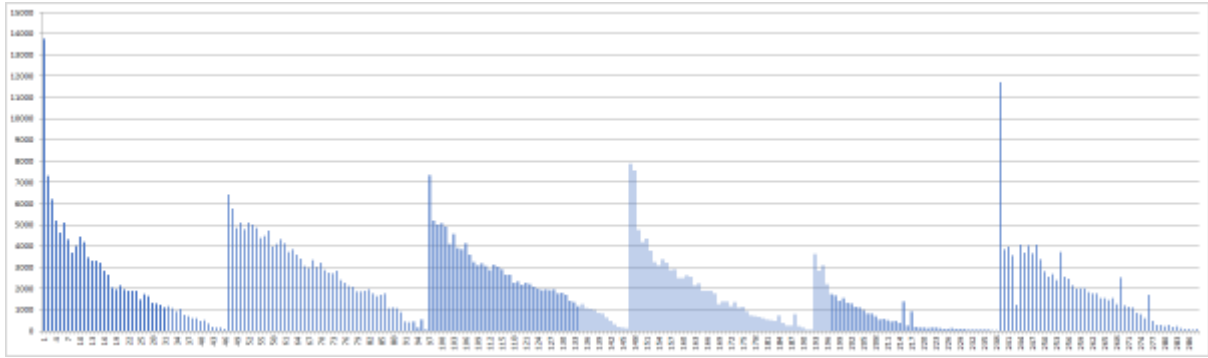


Fig. 1.4.2: ustacks loci for the (A) whole *M. barbatus* dataset, and (B) each population separately

Stacks results:

We removed 46,735 loci that did not pass sample/population constraints from 47,709 loci. We then kept 974 loci, composed of 306,058 sites; 12,725 of those sites were filtered, 580 variant sites remained.

Number of loci with PE contig: 974.00 (100.0%);

Mean length of loci: 304.23bp (stderr 1.19);

Number of loci with SE/PE overlap: 142.00 (14.6%);

Mean length of overlapping loci: 290.18bp (stderr 1.49); mean overlap: 24.80bp (stderr 0.27);

Mean genotyped sites per locus: 305.91bp (stderr 1.16).

Observed heterozygosities were generally low (ranging from 0.1076 to 0.1340) and smaller than the unbiased expected (0.364 on average).

Table 1.4.4: Summary of genetic variability estimates across sampling locations in *M. barbatus*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e).

Samples	H_o	H_e
06c	0.1076	0.3136
09b	0.1106	0.3294
11c	0.1170	0.3270
17a	0.1340	0.3317
17b	0.1274	0.3304
18a	0.1212	0.3344

All samples showed a great similarity between them (F_{st} values from negative to 0.0049, and STRUCTURE plots) with the exception of the one coming from GSA06c (Spanish coast, see the DAPC plot); however, the latter sample performed the worst in the sequencing

process and lead to the least number of loci ie. 732.63 on average when this number for the other samples ranged from 1954.43 to 2846.558 (see Table 1.4.3).

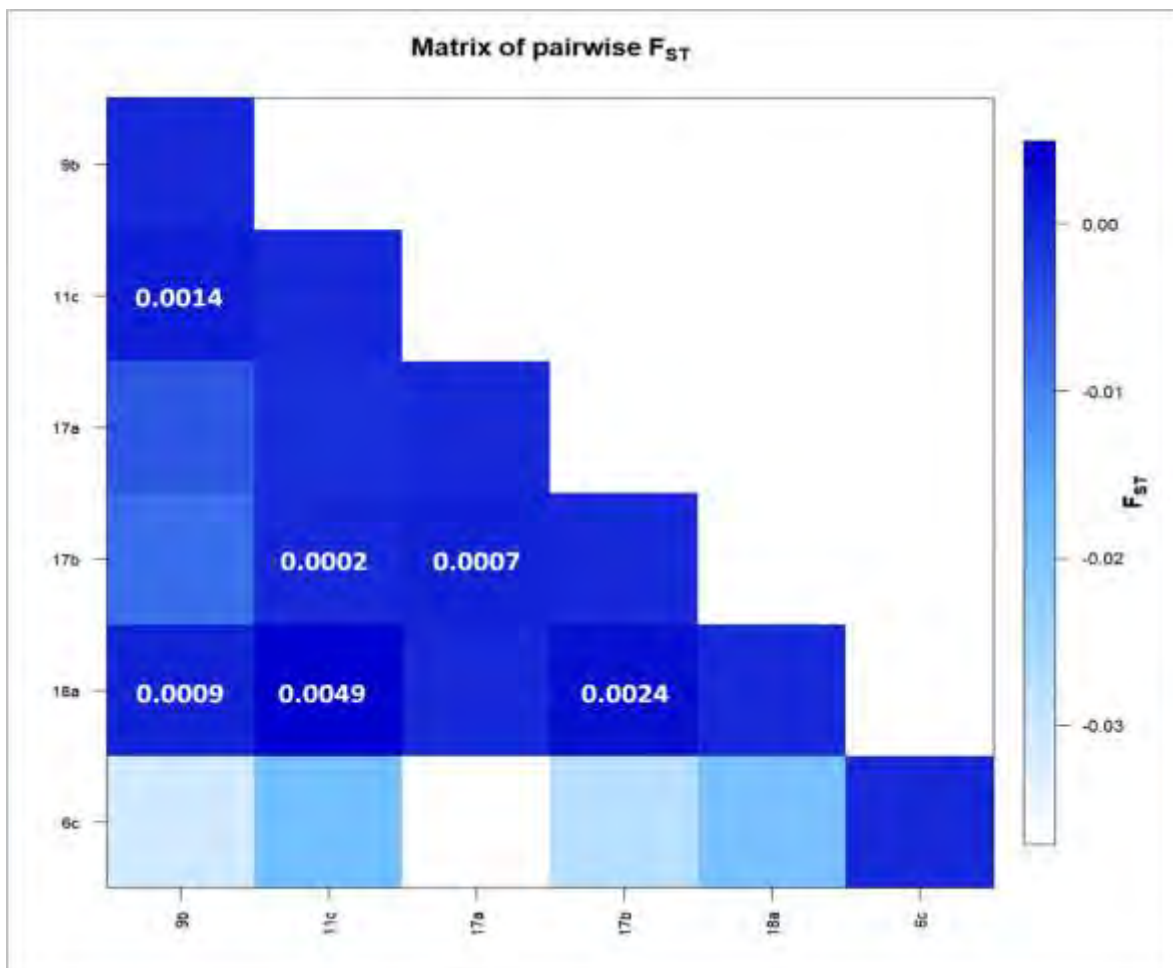


Fig. 1.4.3: Matrix of pairwise F_{st} values for *M. barbatus* based on 580 polymorphic SNPs with less than 50% missing data per locus in the data set. Only positive F_{st} values are reported since negative ones are considered zero. No value was statistically significant (p values >0.05)

Concerning the AMOVA results and various scenarios of groupings that checked (table not provided), the variation among groups was in most cases negative and not significant (p value >0.05) indicating lack of significant differentiation.

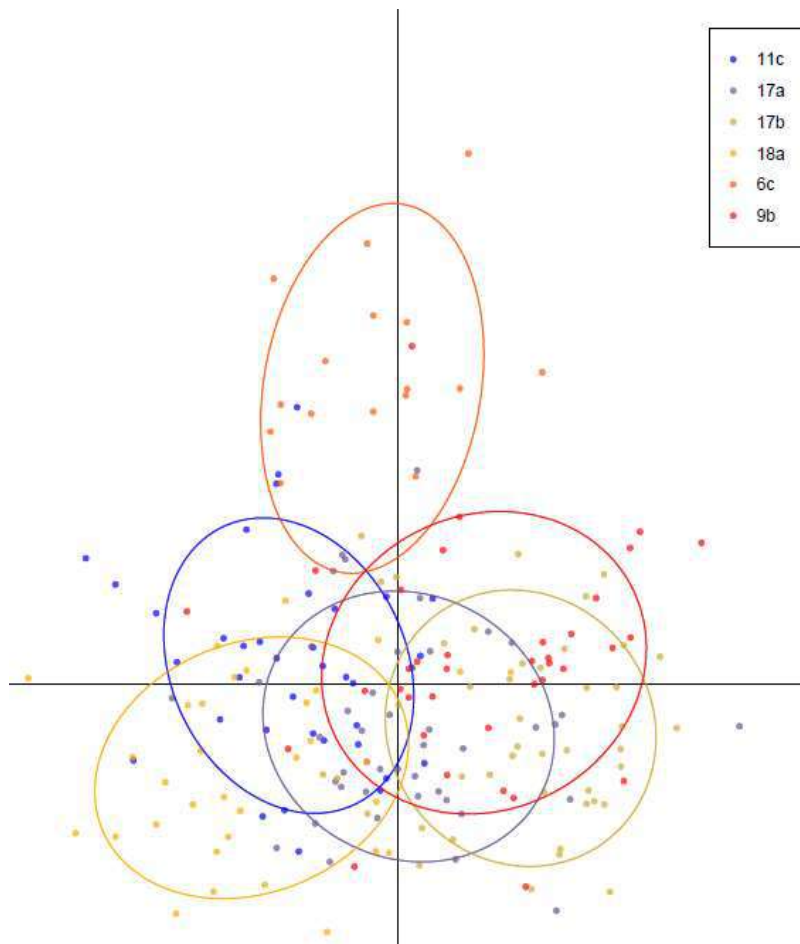


Fig. 1.4.4: DAPC plot for *M. barbatus* based on 580 polymorphic SNP present in >50% in the whole sampling set, and in >4 samples.

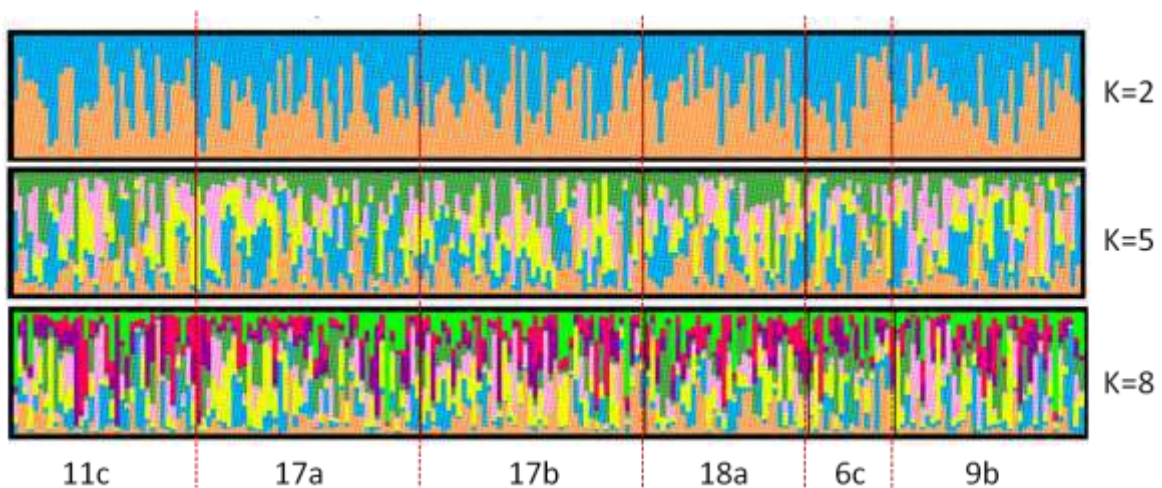


Fig. 1.4.5: Proportions of membership to each of K inferred clusters for *M. barbatus* individuals

When current results are compared to those previously reported in the literature (see D1.2), we do not encounter the differentiation of the Adriatic Sea (present samples 17a and 17b)

from the neighboring regions reported in Maggio et al. (2009) and Matic-Skoko et al. (2018); moreover, the Spanish sample differentiation has been advanced by Galarza et al. (2009) although the sampling effort was limited, with large areas not covered at all.

In conclusion, the *M. barbatus* study requires further analysis to explore the parameter space to extract robust conclusions on the population structure among the studied populations.

Nephrops norvegicus

For the Norway lobster (*Nephrops norvegicus*), the preliminary study was conducted with samples from 8 localities (GSA06b, 06c, 08ab, 09ab, 11c, 11e, 17ab and 18a). The majority of samples provided DNA of high quality in most samples; in only two cases (GSA06b, 06c), approximately half of the specimens were used, and in only one (GSA18a) the DNA quality was very low and this sample was excluded from the library construction.

The number of (Ustacks) loci varied significantly among specimens (from 177 to 41,911) and among populations on average (8,910 in 17ab to 23,603 in 11e).

Using all individuals, we constructed the "catalogue" which comprised some 749K ddRAD loci. After several steps described in M1.3, a panel of 1,393 higher quality SNPs, present in at least 80% of the samples and four populations, was used for all downstream differentiation analysis.

Table 1.4.5. Stacks pipeline output summary for *N. norvegicus*

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
06b	1236835.54	156578	2839694	13,737.37	3922	27,008
06c	835864.91	26453	3289626	9,464.75	745	21,437
08ab	2968118.92	1193171	5065412	20,887.18	14334	30,582
09ab	2356807.16	490916	4933509	18,275.44	8988	30,483
11e	3655603.72	748193	5886414	23,603.94	10821	32,443
11p	2593385.32	745118	7371844	20,077.06	9828	41,911
17ab	782885.06	15025	3260207	8,910.45	177	22,200
ALL	2253100.58	15025	7371844	17321.01	177	41911

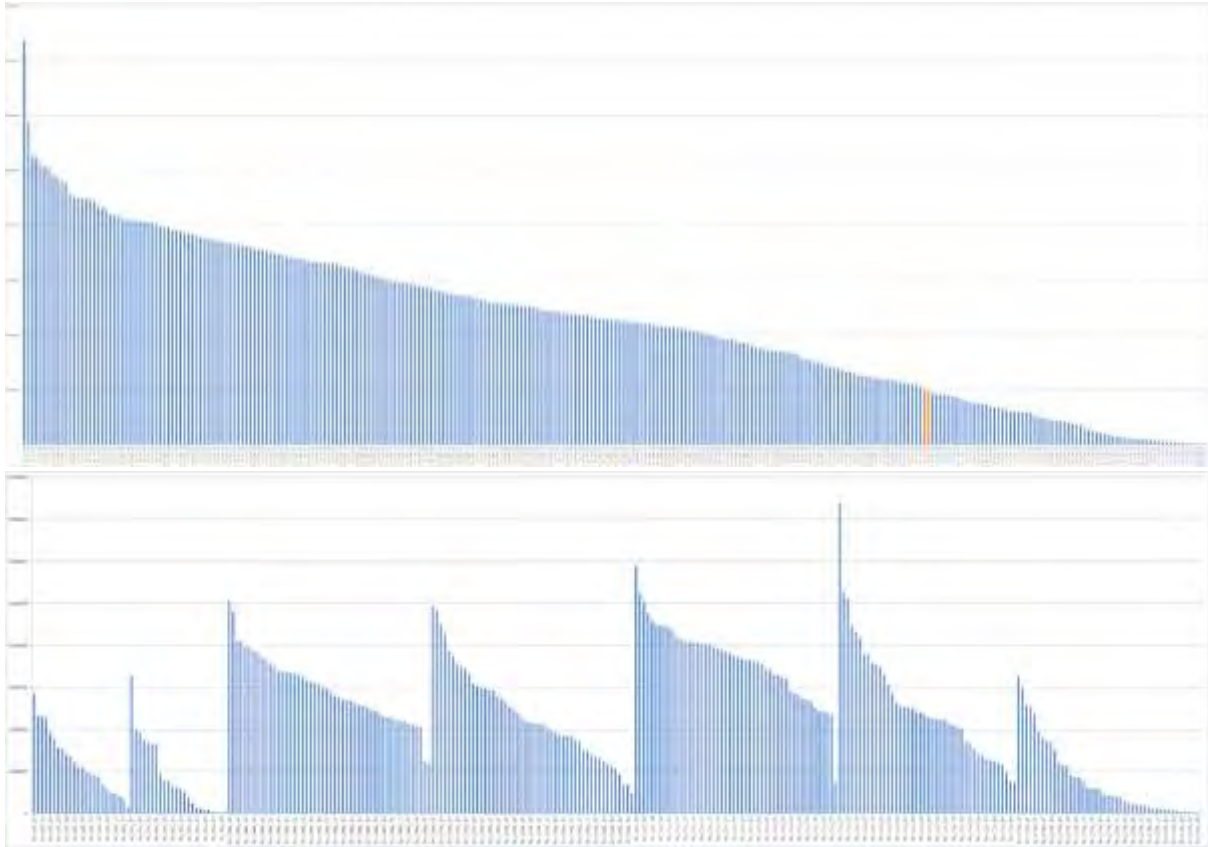
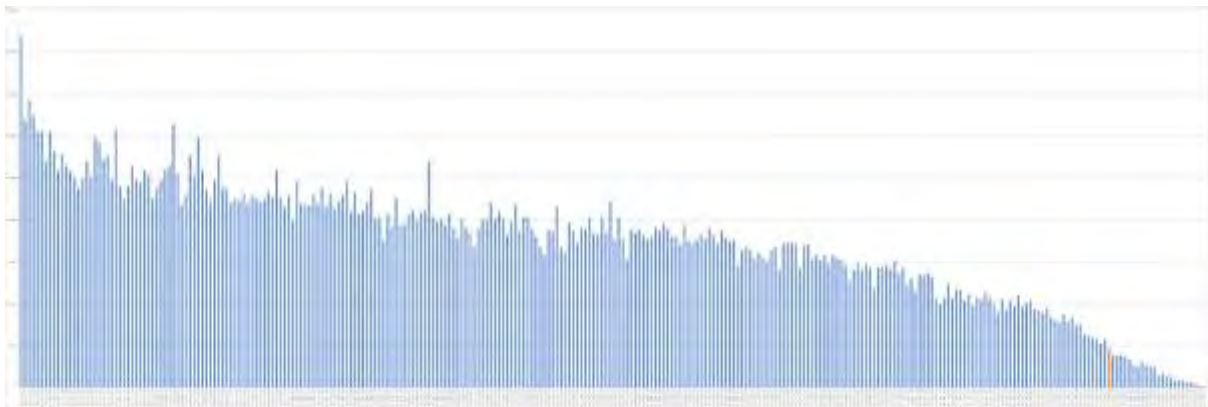


Fig. 1.4.6: Retained reads after demultiplexing for (A) the whole *N. norvegicus* dataset, and (B) each population separately



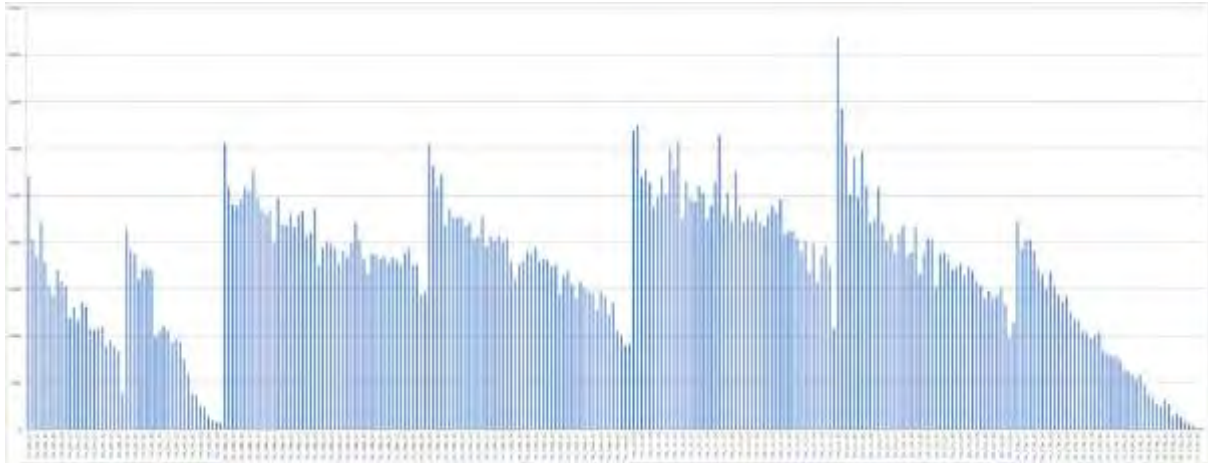


Fig. 1.4.7: ustacks loci for the (A) whole *N. norvegicus* dataset, and (B) each population separately

Stacks results:

We removed 745,500 loci that did not pass sample/population constraints from 749,028 loci; we finally kept 3,528 loci, composed of 1,255,803 sites; 146,424 of those sites were filtered, 1393 variant sites remained.

Number of loci with PE contig: 3528.00 (100.0%);

Mean length of loci: 345.95bp (stderr 0.87);

Number of loci with SE/PE overlap: 177.00 (5.0%);

Mean length of overlapping loci: 372.74bp (stderr 1.66); mean overlap: 20.99bp (stderr 0.24);

Mean genotyped sites per locus: 346.64bp (stderr 0.87).

Observed heterozygosities were among the highest encountered in the whole study (ranging from 0.2474 to 0.2867) but remained smaller than the unbiased expected (0.3381 on average).

Table 1.4.6: Summary of genetic variability estimates across sampling locations in *N. norvegicus*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e).

Samples	H_o	H_e
06b	0.2597	0.3376
06c	0.2503	0.3343
09ab	0.2867	0.3432
11e	0.2736	0.3366
11p	0.2777	0.3400
17ab	0.2474	0.3373

In the Norway lobster, samples showed a moderate similarity between them; the F_{ST} values ranged from 0.006 to 0.0927 (Fig. 1.4.8) with the highest observed between GSA17ab (N. Adriatic) sample and all the others. The latter is more evident in both the DAPC and STRUCTURE plots (see Figs. 1.4.9 and 1.4.10). However, the GSA17ab sample had the least number of loci ie. ~9K on average when this number for the best four samples was above 18K and for the two smaller samples (GSA06b and o6c) ranged from 9.5 to 13.7K.

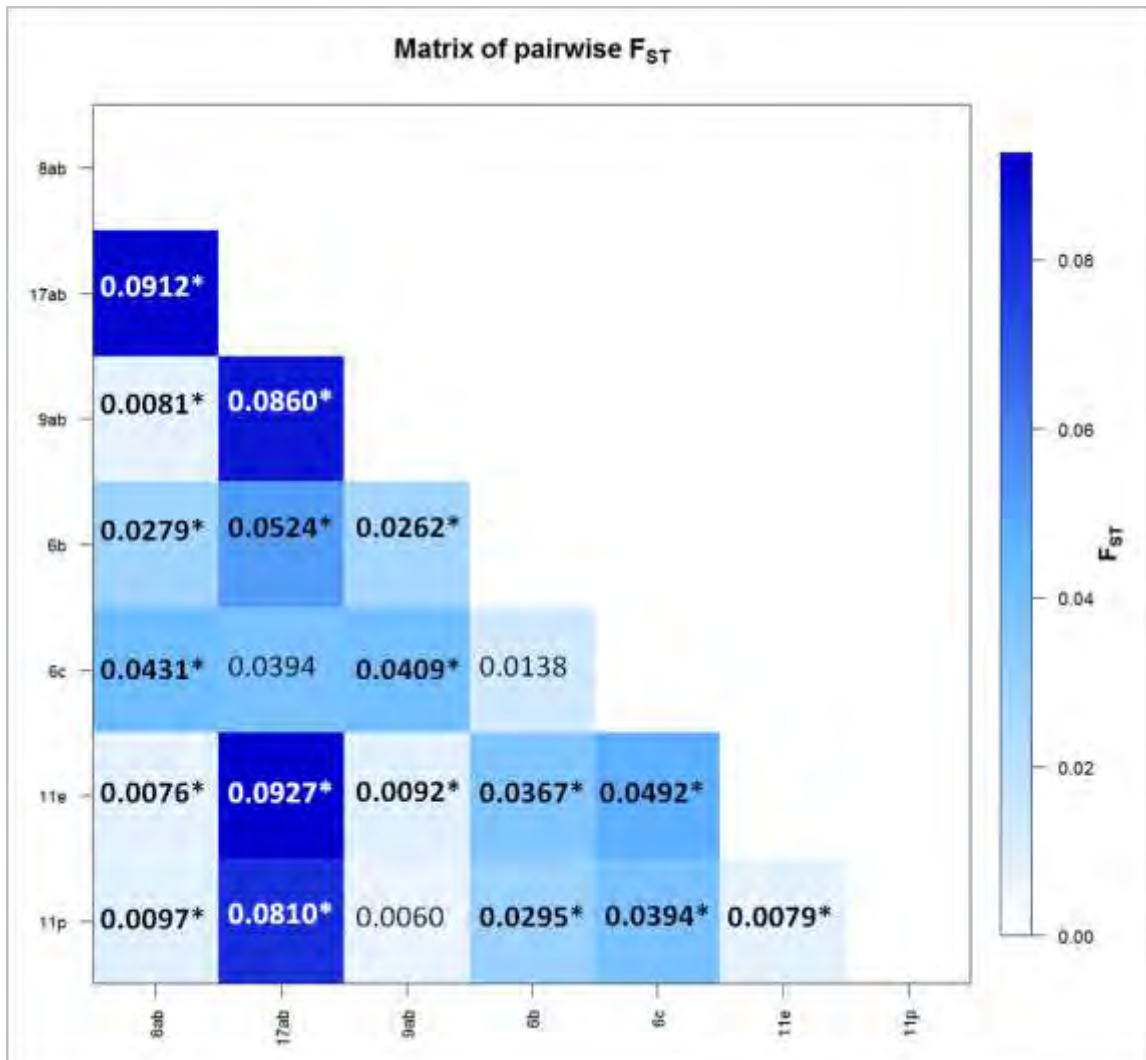


Fig. 1.4.8: Matrix of pairwise F_{ST} values for *N. norvegicus* based on 1,393 polymorphic SNPs with less than 2% missing data per locus in the data set. Values with asterisk are statistically significant after Benjamini–Yekutieli correction.

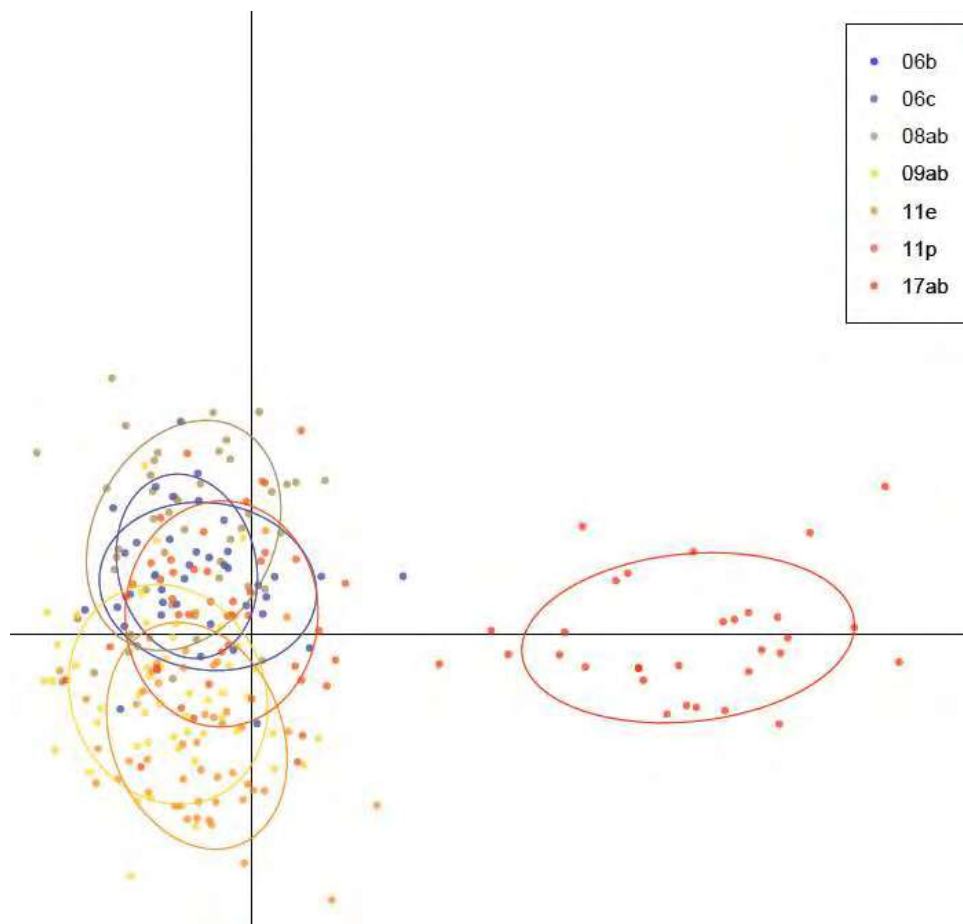


Fig. 1.4.9: DAPC plot for *N. norvegicus* based on 1,393 polymorphic SNP present in >80% of specimens in the whole sampling set, and in >4 samples.

However, AMOVA results for several scenarios of grouping in the Norway lobster samples clearly indicated a moderately significant differentiation between the GSA17 sample and all others (~7% differentiation).

Table 1.4.7: AMOVA results for several scenarios of grouping *N. norvegicus* samples.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Group 1	8ab,9ab,6b,6c,11e,11p	8ab,9ab	8ab,9ab	6b,6c
Group 2	17ab	6b,6c	6b,6c,11e,11p	8ab,9ab,11e,11p
Group 3		11e,11p	17ab	17ab
Group 4		17ab		
Variation %				

Among groups	6.93	3.17	2.23	5.16
Among populations within groups	1.61	0.86	1.91	0.80
Within populations	91.46	95.97	95.86	94.04
Fixation Indices				
FST	0.08540	0.04031	0.04144	0.05964
p-value	0.000	0.000	0.000	0.000
FSC	0.01726	0.00886	0.01954	0.00846
p-value	0.000	0.000	0.000	0.000
FCT	0.06934	0.03174	0.02234	0.05161
p-value	0.000	0.000	0.000	0.000

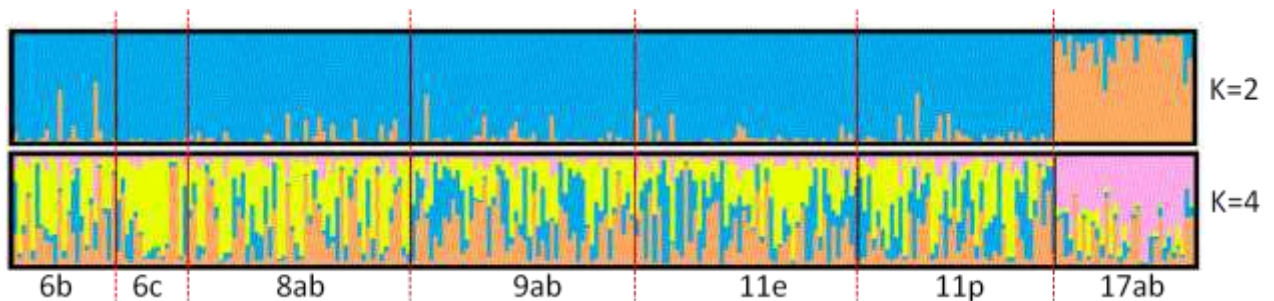


Fig. 1.4.10: Proportions of membership to each of K inferred clusters for *N. norvegicus* individuals

With practically no recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, current results are the first to show a slight but considerable differentiation of the Adriatic sample. In conclusion, the Norway lobster study is the best shown here for the shrimp species and is expected to provide interesting results without any further optimizations.

Aristaeomorpha foliacea

For the giant red shrimp (*Aristaeomorpha foliacea*), the preliminary study was conducted with samples from 8 localities (GSA09b, 10b, 11c, 11e, 16b, 18a, 19b and 19c). The majority of samples provided DNA of good quality, except for two subareas: GSA19b (not finally included in the library preparation) and 19c; therefore, we included samples from the remaining seven locations into library construction, and for 19c only 7 specimens which were considered of quality.

The number of (Ustacks) loci varied significantly among specimens (from 94 to 206,636) and among populations on average (483 in 19c to 52,870 in 09b).

Using all individuals, we constructed the "catalogue" which comprised some 1.2M ddRAD loci, a number which is considered huge. After several steps described in M1.3, and finally excluding the GSA19c sample due to the very low number of reads, a panel of 3,437 higher quality SNPs, present in at least 50% of the samples and four populations, was used for all downstream differentiation analysis; therefore, the estimates below are for only 6 sampling locations.

Table 1.4.8. Stacks pipeline output summary for *A. foliacea*

	Retained Loci			ustacks Loci		
	Average	Min	Max	Average	Min	Max
09b	5770524.10	108002	20949641	52,869.48	2,218	147,959
10b	1792859.73	37317	11808649	22,000.36	209	102,565
11e	432665.38	8088	4103956	8,301.80	72	52,735
11p	5031079.49	32461	23395395	50,659.31	519	206,636
16b	288508.25	25401	1795372	6,342.50	101	32,138
18a	218013.19	8780	1332815	4,678.00	94	24,431
19c	37133.71	14947	86580	483.28	104	1,555
ALL	2251697.72	8088	23395395	23976.42	72	206636

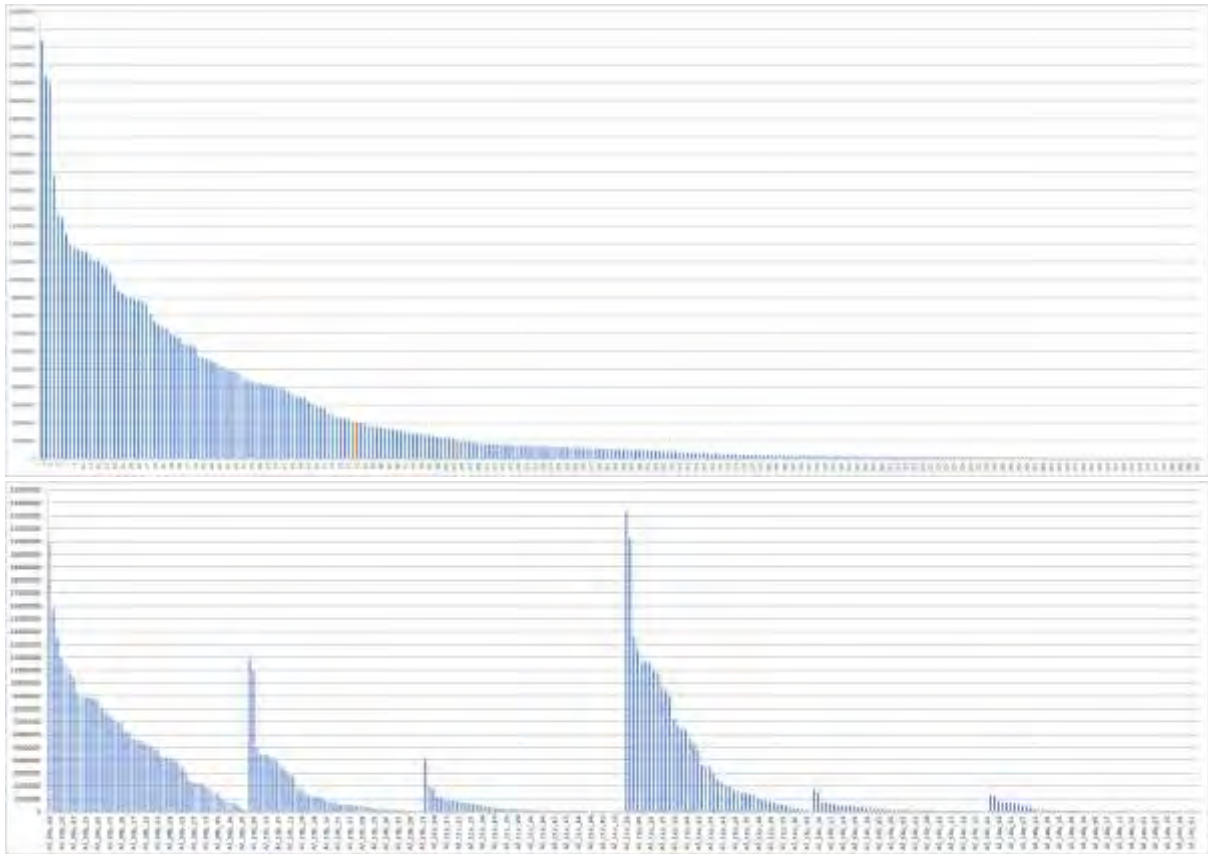


Fig. 1.4.11: Retained reads after demultiplexing for (A) the whole *A. foliaceae* dataset, and (B) each population separately

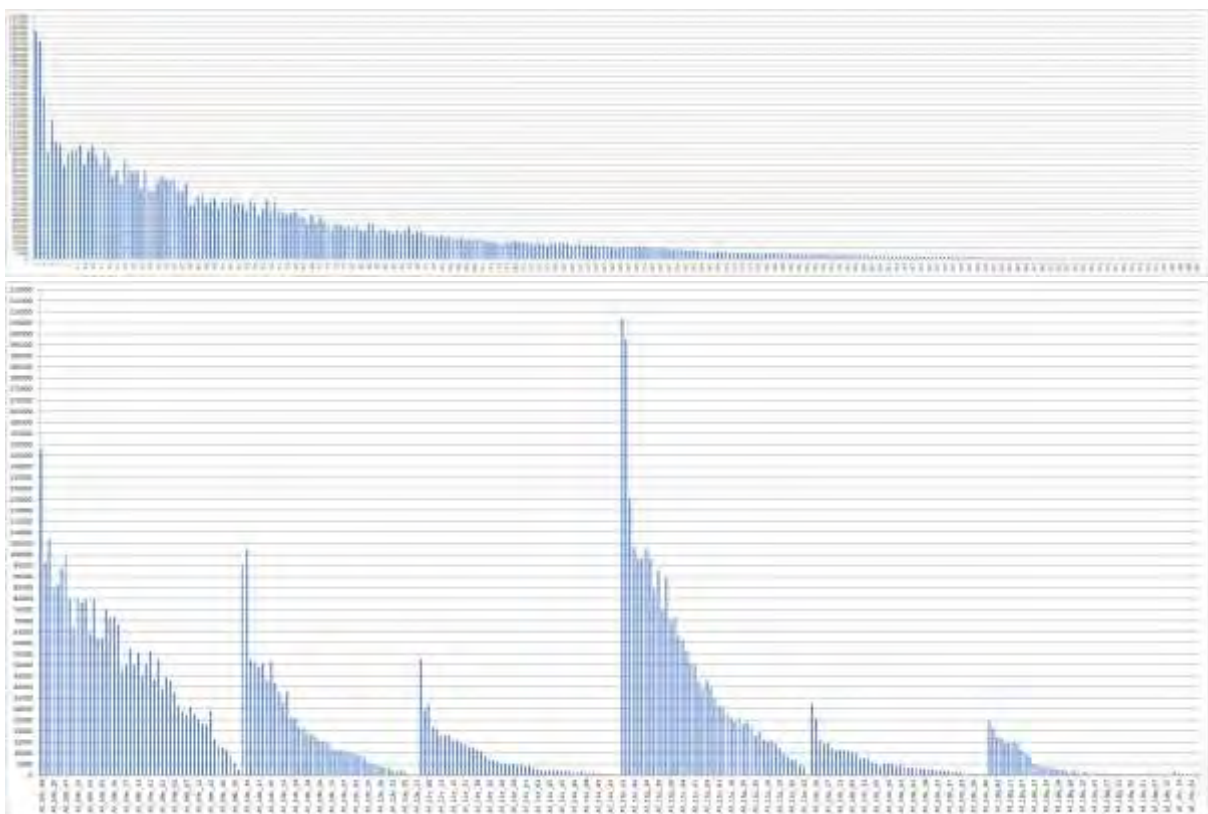


Fig. 1.4.12: ustacks loci for the (A) whole *A. foliacea* dataset, and (B) each population separately

Stacks results:

We removed 1199629 loci that did not pass sample/population constraints from 1,204,691 loci, and we finally kept 5,062 loci, composed of 2,063,113 sites; 216,496 of those sites were filtered, 3,437 variant sites remained.

Number of loci with PE contig: 5,062.00 (100.0%);

Mean length of loci: 397.57bp (stderr 1.30);

Number of loci with SE/PE overlap: 1,792.00 (35.4%);

Mean length of overlapping loci: 374.81bp (stderr 1.39); mean overlap: 30.69bp (stderr 0.36);

Mean genotyped sites per locus: 401.04bp (stderr 1.28).

Observed heterozygosities were generally low (ranging from 0.0968 to 0.2281) and smaller than the unbiased expected (0.3247 on average).

Table 1.4.9: Summary of genetic variability estimates across sampling locations in *A. foliacea*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e).

Samples	H_o	H_e
09b	0.2281	0.3394
10b	0.1802	0.3353
11e	0.1266	0.3233
11p	0.2063	0.3381
16b	0.0968	0.3062
18a	0.1131	0.3059

All samples showed a great similarity between them; this is evident in the F_{st} values which ranged from negative to 0.0069 (Fig. 1.4.13) and the STRUCTURE plots (1.4.15) with the exception of the two coming from GSA09b (N. Tyrrhenian coast) and 11c (south Sardinia) (see also the DAPC plot, Fig. 1.4.14). However, in this case the latter two samples performed the best in the sequencing process and lead to the greatest number of loci ie. >50K on average when this number for the other four samples ranged from 4,678 to 22,000. Moreover, AMOVA results (see Table 1.4.10) for several scenarios of grouping *A. foliacea* samples did not indicate any significant differentiation between groups.

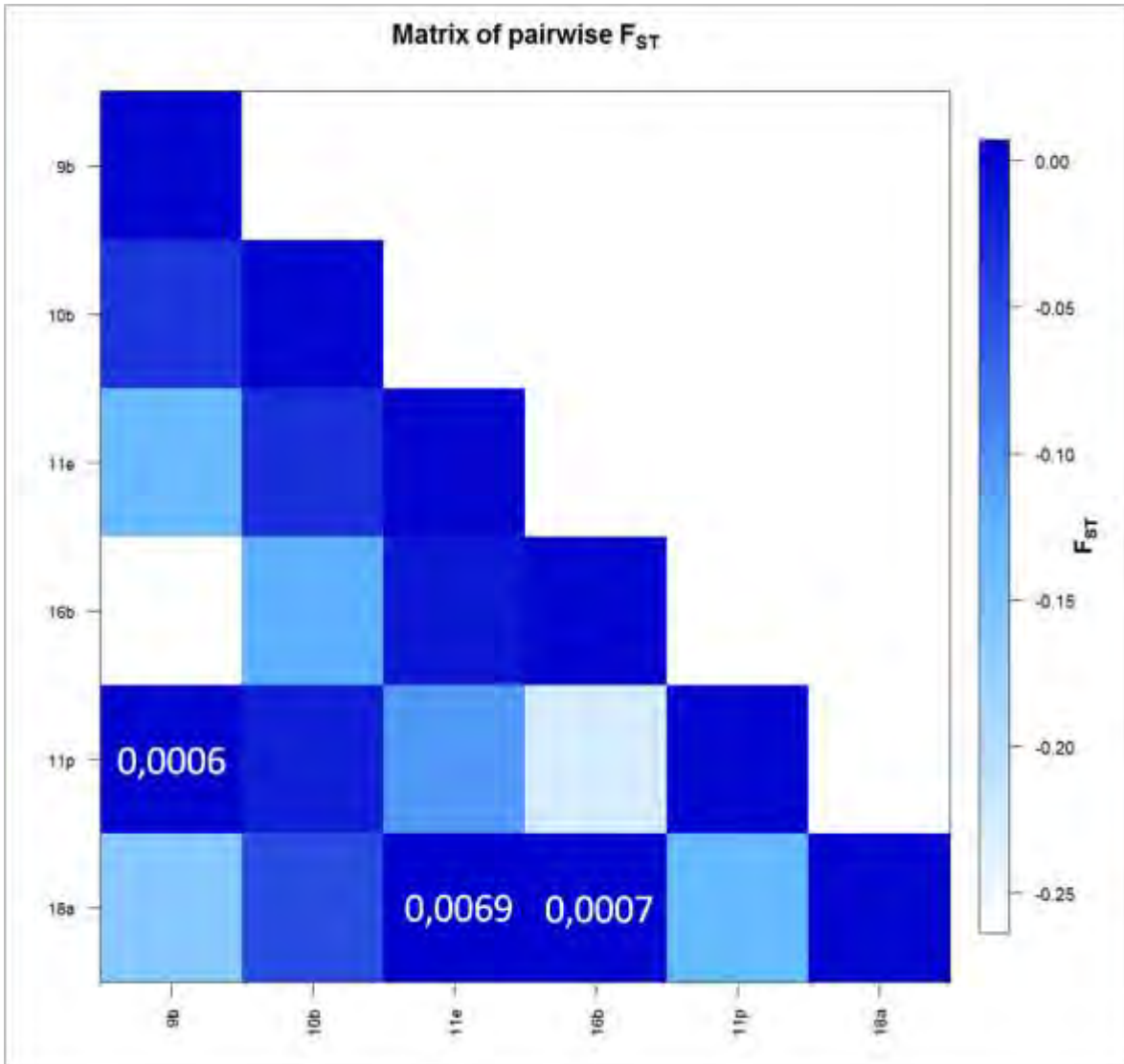


Fig. 1.4.13: Matrix of pairwise F_{ST} values for *A. foliaceae* based on 3,437 polymorphic SNPs with less than 50% missing data per locus in the data set. Only positive F_{ST} values are reported since negative ones are considered practically zero. No value was statistically significant (p values >0.05)

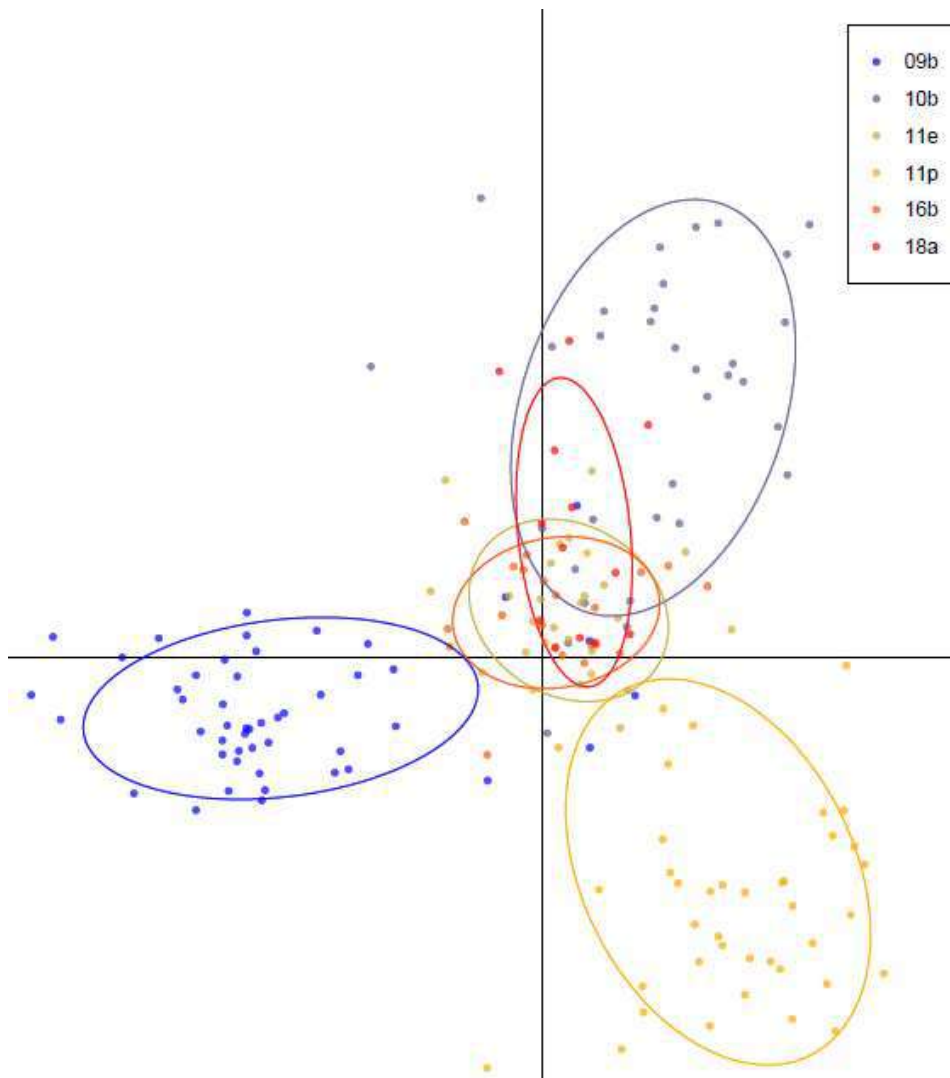


Fig. 1.4.14: DAPC plot for *A. foliacea* based on 3,437 polymorphic SNP present in >50% in the whole sampling set, and in >4 samples.

Table 1.4.10: AMOVA results for several scenarios of grouping *A. foliacea* samples.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Group 1	9b	9b	9b,10b,16b	9b
Group 2	11p	10b	11e	10b
Group 3	10b,11e,16b,18a	11p	11p	11e,11p
Group 4		11e,16b,18a	18a	16b
Group 5				18a
Variation %				
Among groups	-2.19	-3.83	-0.02	-0.57

Among populations within groups	3.89	5.65	1.93	2.46
Within populations	98.30	98.17	98.08	98.11
Fixation Indices				
FST	0.01700	0.01823	0.01913	0.01890
p-value	0.000	0.000	0.000	0.000
FSC	0.03808	0.05446	0.01932	0.02443
p-value	0.000	0.264	0.000	0.000
FCT	-0.02191	-0.03831	-0.00020	-0.00568
p-value	1.000	1.000	0.000	0.000

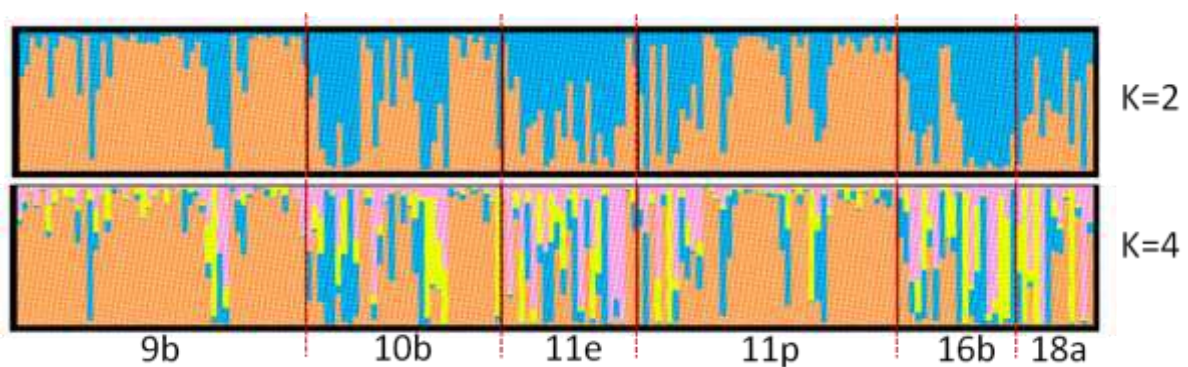


Fig. 1.4.15: Proportions of membership to each of K inferred clusters for *A. foliacea* individuals

Current results refer to samples from the Western and Central Mediterranean where the lack of genetic differentiation among stocks has already been reported (Cannas et al 2012; Marcias et al., 2010; Fernandez et al., 2011a; Fernandez et al., 2013b).

In conclusion, our common belief is that the *A. foliacea* study is expected to provide interesting results when a further optimization of the library preparation will take place in order from one hand to substantially decrease the total number of SNPs recovered and on the other hand increase the number of common SNPs for downstream genetic analyses.

Merluccius merluccius

A total of 288 samples (from eight populations) were genotyped following a ddRAD protocol, with enzymes *SbfI* and *SphI*, selecting fragments from around 200 to around 470 bp. Multiplexed libraries were run on two Illumina lanes, with 150 bp pair-end protocol. R1 from the two lanes and R2 from the two lanes were merged, obtaining a total of 549 millions reads in R1+R2.

509 out of the 549 millions of reads had valid barcodes; only 7.22% of the reads had invalid barcodes. After demultiplexing, on average 1,769 M reads per sample were obtained (range from 6158 to 5,346 M). After eliminating low quality reads and reads without a RAD site at the beginning, on average 98,5% reads were retained (see figure 1.4.16).

Samples were processed with *ustacks* considering a stacks of unique reads only if at least 3 identical reads were found and allowing for up to 3 mismatches to consider two tags as part of the same locus. On average, 10651 tags were found in each sample, with average coverage of 67,8X (table 1.4.11).

Low number of reads seem to affect around 50-20 samples in which a low number of tags has been found (figure 1.4.17). A tag catalog was built with all the samples of the populations. A maximum of 3 mismatches were allowed to merge tags into the same catalog locus. The final number of tags in the catalog is around 100k. The increase in catalog size is linear, therefore it is expected that using more samples will produce a bigger catalog (more tags and therefore more SNPs).

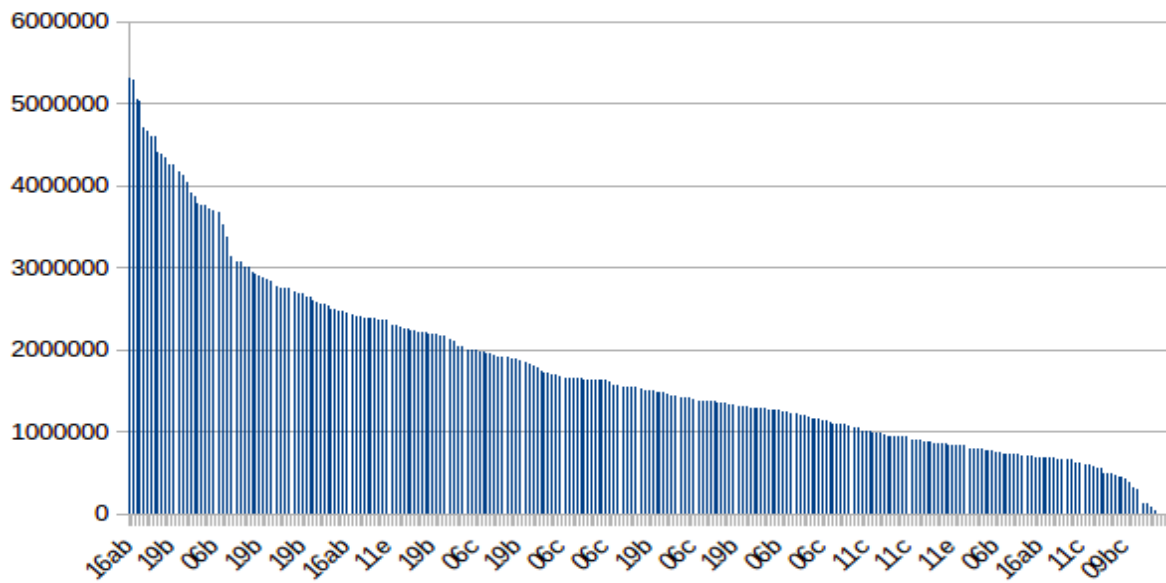
Stacks' module *populations* and *vcftools* were used to filter the initial SNP panel. The 20 "worst" individuals were filtered out to avoid a reduction of final SNP number when filtering for SNPs shared by at least 80% of samples. Only SNPs shared by at least 80% of the individuals and with at least 15X coverage were kept. A total of 734 high quality SNPs were retained (for 268 samples), and used for all downstream differentiation analysis.

Observed heterozygosities were generally low (ranging from 0.2215 to 0.2345) and smaller than the unbiased expected (ranging from 0.2494 to 0.2566; table 1.4.12). Samples showed a significant differentiation among them (F_{st} values from 0.0003 to 0.0114, and AMOVA scenarios) with the population samples from GSA06b, GSA06c and GSA19b showing the highest F_{st} values (figure 1.4.18). The best scenario identified by the AMOVA analysis, based on F_{ct} maximization, is the one in which the total genetic variance is partitioned into three groups, the first including GSA06b, the second GSA06c, and the third one the remaining sites (table 1.4.13). This result represents a significant advance respect to previous published studies, where differences were not detected between Mediterranean samples, unless using outlier loci (Milano et al. 2014).

Nevertheless, such differentiation is low, and not evident considering DAPC (figure 1.4.19) and STRUCTURE analysis (figure 1.4.20). DAPC identified one homogeneous group with the R's *adegenet*'s *find.cluster* function suggesting 1 as the most likely number of clusters in the dataset. STRUCTURE showed a mode at $k = 3$ following the Evanno's method, without any clear geographic localization of the inferred genetic clusters. In conclusion, the *M. merluccius* pilot study provides interesting results and the finding of significant to weak genetic structure within Mediterranean represents a significant advance respect to previous published studies, where differences were not detected at this geographic scale, unless using outlier loci (Milano et al. 2014).

Table 1.4.11: Stacks pipeline output summary for *Merluccius merluccius*

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
06b	1848127,2	47847	4672442	11137,6	2130	22725
06c	1920908,5	689562	4604155	10777,1	6765	19043
09bc	1482734,2	86501	3512812	9939,1	3190	20228
11c	1619669,9	278910	5018105	10100,3	5349	18658
11e	1806589,5	6233	5042656	10997,3	57	22237
16ab	1671673,4	7831	5307458	10272,5	37	23337
18a	1833960,0	5964	4714389	11159,8	238	24076
19b	1812298,8	135891	5284911	10718,2	3869	19705
ALL	1749756,7	5964	5307458	10650,9	37	24076



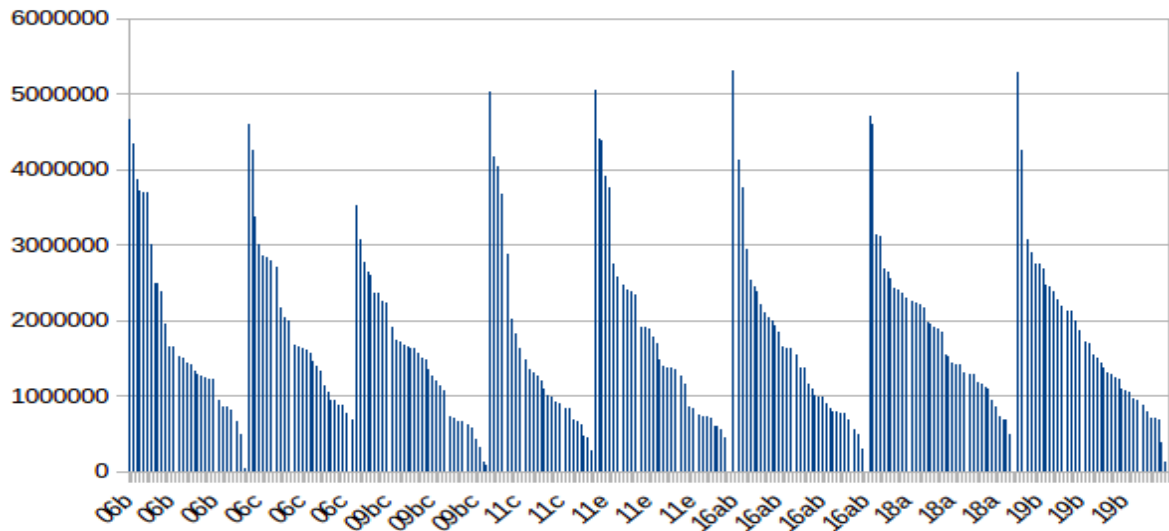
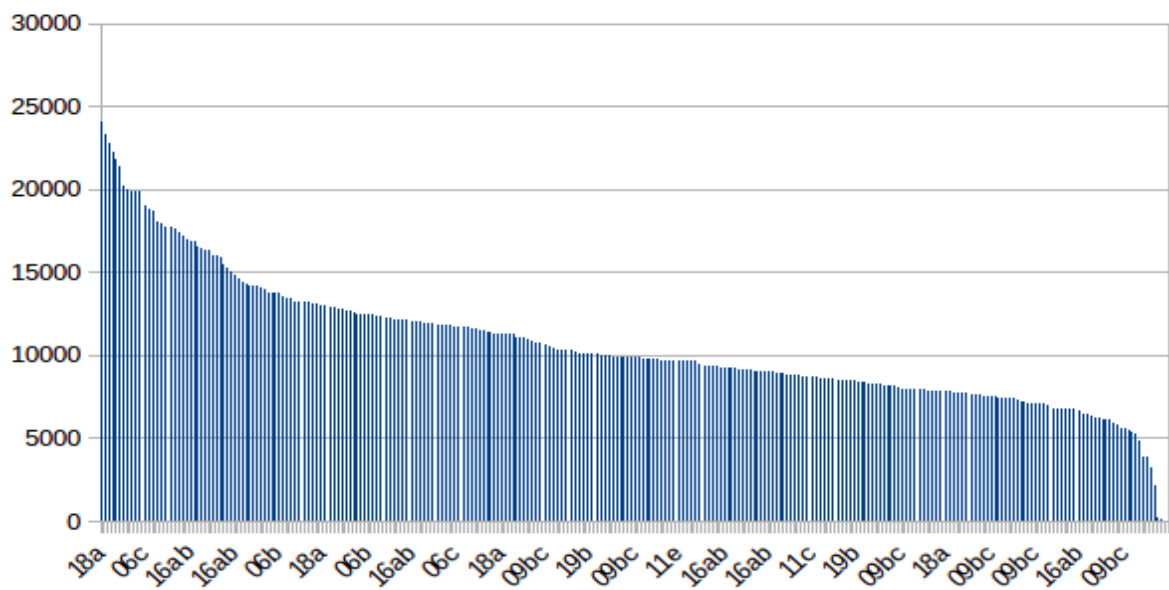


Fig. 1.4.16: Retained reads after demultiplexing for (A) the whole *M. merluccius* dataset, and (B) each population separately.



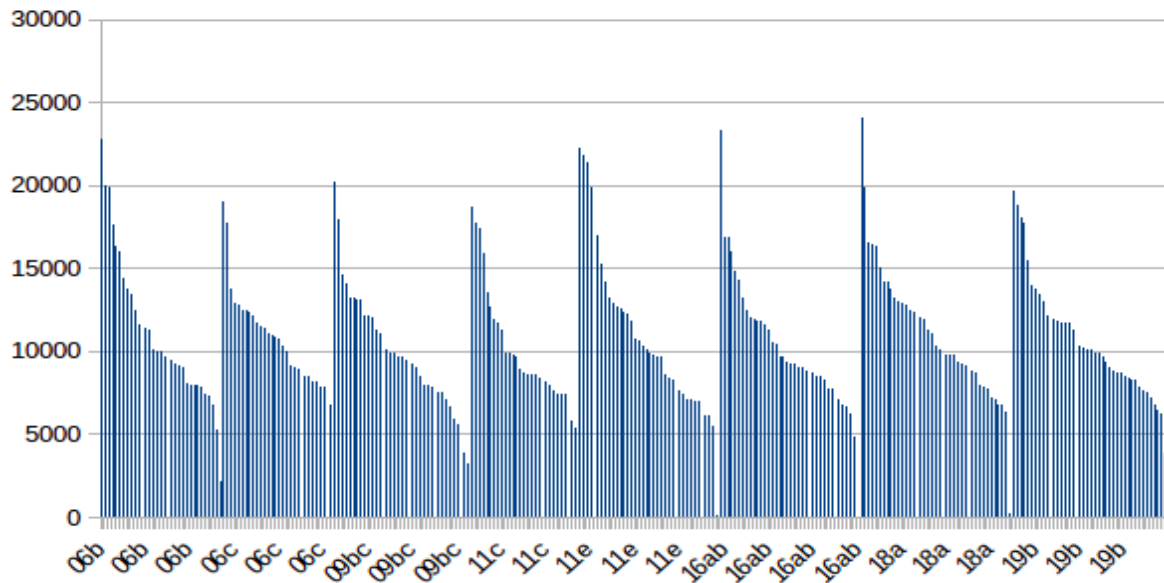


Fig. 1.4.17: ustacks loci for the (A) whole *M. merluccius* dataset, and (B) each population separately.

Table 1.4.12: Summary of genetic variability estimates across sampling locations in *M. merluccius*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e).

Samples	H_o	H_e
06b	0.2215	0.2494
06c	0.2240	0.2523
09bc	0.2267	0.2521
11c	0.2345	0.2566
11e	0.2268	0.2532
16ab	0.2294	0.2562
18a	0.2246	0.2552
19b	0.2242	0.2527

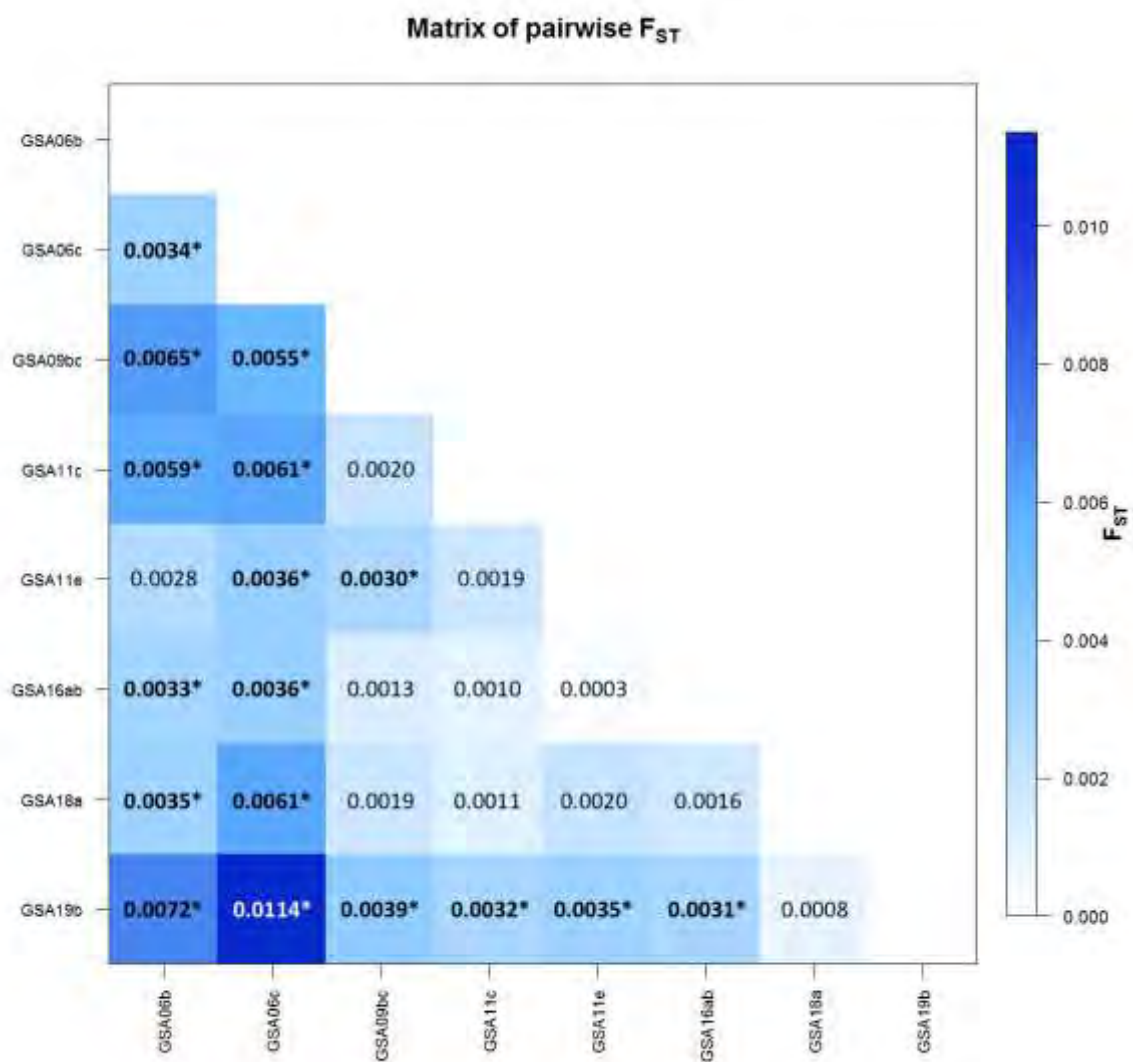


Fig. 1.4.18: Matrix of pairwise F_{ST} for *M. merluccius* based on 734 polymorphic SNPs present in >80% of the whole sampling set. Significant pairwise F_{ST} values, after Benjamini–Yekutieli correction, are reported in bold.

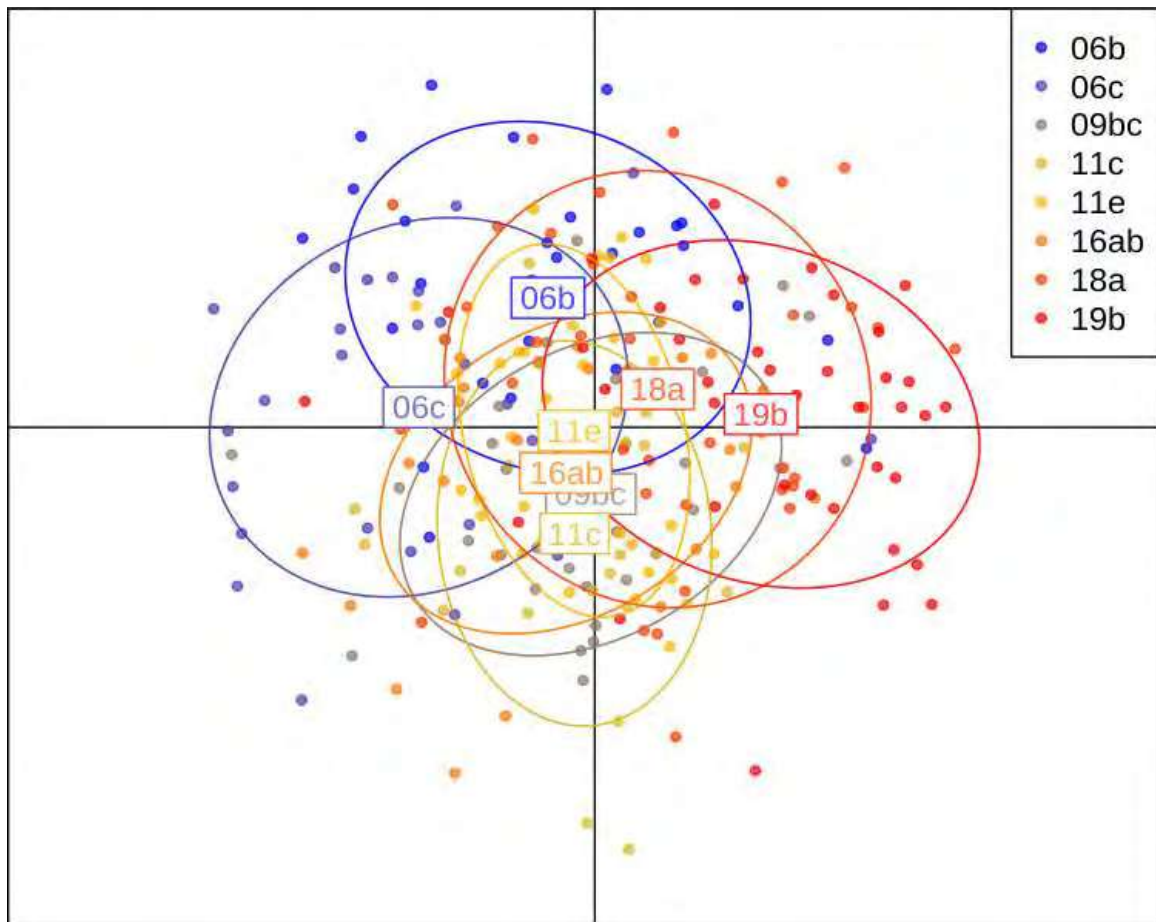


Fig. 1.4.19: DAPC plot for *M. merluccius* based on 20 principal components summarizing the information of 734 polymorphic SNP present in >80% of the whole sampling set.

Table 1.4.13: AMOVA results for several scenarios of grouping *M. merluccius* samples.

	Scenario 1	Scenario 2	Scenario 3
Group 1	06b, 06c	06b	06b
Group 2	09bc, 11c, 11e, 16ab, 18a, 19b	06c	06c
Group 3		09bc, 11c, 11e, 16ab, 18a, 19b	18a, 19b
Group 4			09bc, 11c, 11e, 16ab
Variation %			
Among groups	0.25	0.31	0.26

Among populations within groups	0.46	0.43	0.36
Within populations	99.29	99.26	99.38
Fixation Indices			
FST	0.00713	0.00734	0.00626
P-value	0.000	0.000	0.000
FSC	0.00459	0.00429	0.00364
P-value	0.000	0.000	0.002
FCT	0.00255	0.00306	0.00263
P-value	0.000	0.000	0.000

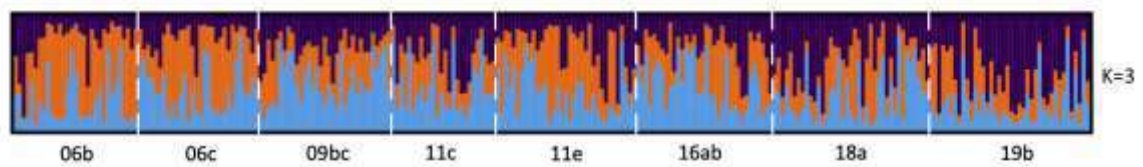


Fig. 1.4.20: Proportions of membership to each K inferred clusters for *M. merluccius* individuals. Best K=3.

Parapenaeus longirostris

A total of 288 samples (from eight populations) were genotyped following a ddRAD protocol, with enzymes *PstI* and *NlaIII*, selecting fragments from around 200 to around 480 bp. Multiplexed libraries were run on two Illumina lanes, with 150 bp pair-end protocol. R1 from the two lanes and R2 from the two lanes were merged. A total of 687.7 millions reads in R1+R2 was obtained.

The ddRADseqTool (Mora-Marqu ez et al. 2017) was used to simulate a ddRAD library using the same enzyme pair and the same size selection used with the target species, to estimate the number of retrieved tags. For the analysis, a fish genome was used, whose length was 831 Mbp. A total of 230'000 tags of size between 200 and 480 bp were estimated. Hypothesizing a 3-4 times bigger size for the unknown *P. longirostris* genome, a total number of fragments around 1 million is expected.

After demultiplexing the 687.7 million reads, on average 2.388 M reads per sample were obtained. After eliminating low quality reads and reads without the RAD sites at the beginning, on average 96,5% reads were retained (range from 18k to 26M, Table 1.4.14). Several samples were characterized by a low number of retained reads (figure 1.4.21) with about 46 specimens producing less than 500,000 reads.

Samples were processed with ustacks considering a stacks of unique reads only if at least 3 identical reads were found and allowing for up to 3 mismatches to consider two tags as part of the same locus. On average 77,846 tags were found in each sample, with average coverage of 7.4X (Table 1.1.14), and, as expected by the uneven distribution of retained reads, 45 individuals providing less than 20,000 loci.

A tag catalog was built with all the samples of the populations. A maximum of 3 mismatches were allowed to merge tags into the same catalog locus. The total number of tags in the catalog is around 970k.

Stacks' module populations and vcftools were used to filter the initial SNP panel. Only the individuals genotyped with more than 760k filtered reads were retained. Only SNPs shared by at least 50% of the individuals and with at least 8X coverage were kept. A total of 1,045 SNPs were retained for 226 samples. Coverage was low.

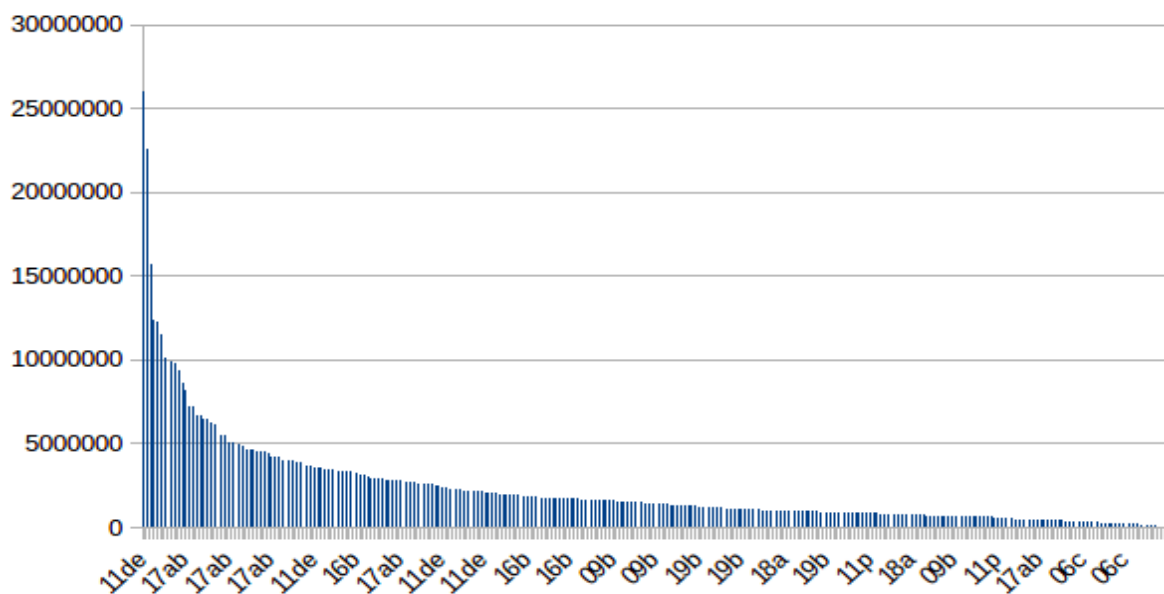
Observed heterozygosities were similar among localities, ranging from 0.2153 to 0.2376, and slightly smaller than expected heterozygosities, which varied between 0.2299 and 0.248 (table 1.4.15). Pairwise F_{st} showed a low level of differentiation among samples, with only two comparisons involving GSA17ab significant after correction for multiple tests (GSA17ab vs GSA06b and GSA17ab vs GSA11de; figure 1.4.23). However, global F_{st} values were small but statistically significant, and AMOVA pointed out a significant differentiation between specimens from GSA17 from the others, with about 1.6% of the overall genetic variation attributable to this subdivision (table 1.4.16). This result was also evident from DAPC plots (figure 1.4.24), which clearly separated GSA17ab from the other samples, and from STRUCTURE analysis (figure 1.4.25), which indicated K=2 as the best grouping and a different genetic constitution of Adriatic samples.

Results of the pilot study confirms the existence of significant differentiation among Mediterranean samples, which was previously detected with AFLP and mtDNA (Lo Brutto et al. 2013). It is also particularly promising, when considering that Lo Brutto and colleagues (2013) detected the deepest differentiation when comparing Aegean samples, still not included in our analysis, with other Mediterranean sites. However, a further optimization of the library preparation is needed to provide the highest resolution and the most robust

results, given that the present dataset is characterized by a low coverage and high level of missing data. In particular, modification of the size selection step to reduce the total number of loci is currently undergoing.

Table 1.4.14: Stacks pipeline output summary for *Parapenaeus longirostris*

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
06c	605468,7	18659	2108874	24480,3	267	90493
09b	3112906,1	278595	22517535	98814,9	8759	293973
11de	3618679,6	89746	26008838	111038,8	2622	334309
11p	2104796,3	99718	8596632	75724,4	2916	213975
16b	1974846,9	188389	15660432	65633,1	5159	281313
17ab	3375919,1	39381	12349915	116172,8	244	237545
18a	985479,2	127019	3344188	43087,9	3066	137850
19b	2079568,8	610144	12250402	74943,1	23539	259846
ALL	2283448,2	18659	26008838	77895,8	244	334309



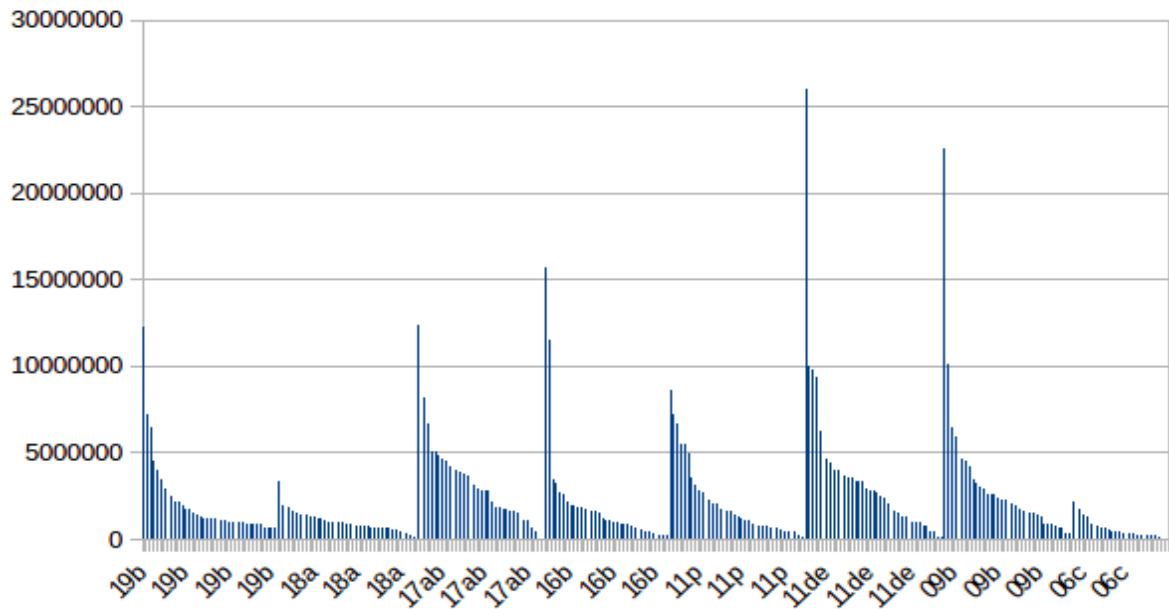
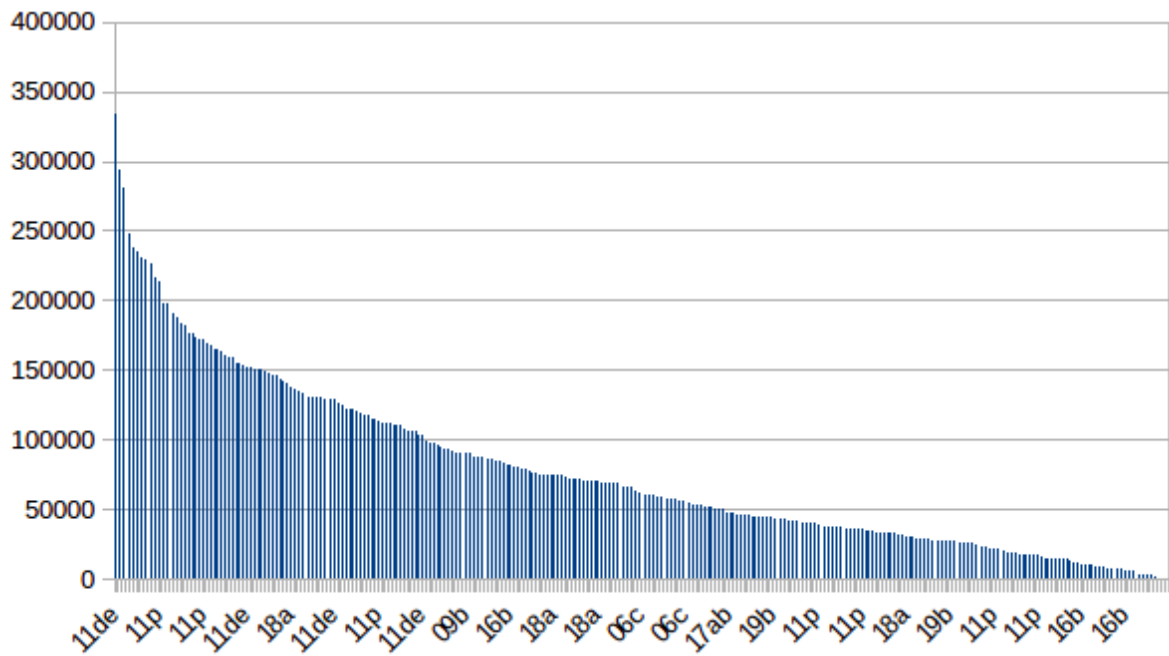


Fig. 1.4.21: Retained reads after demultiplexing for (A) the whole *P. longirostris* dataset, and (B) each population separately



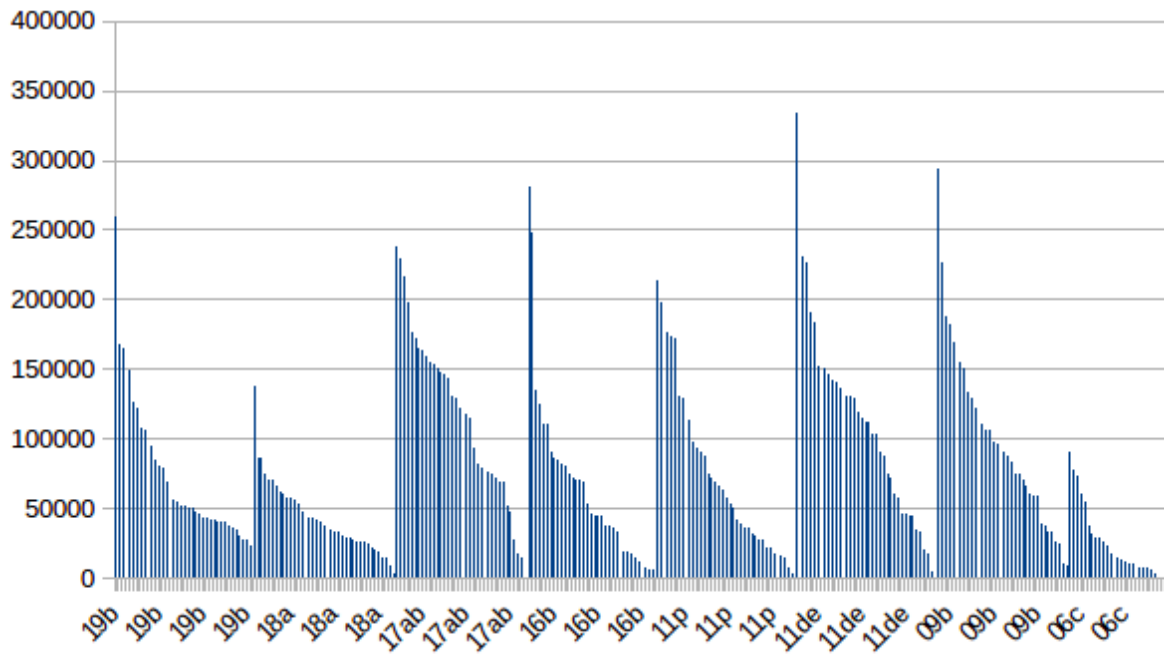


Fig. 1.4.22: ustacks loci for the (A) whole *P. longirostris* dataset, and (B) each population separately

Table 1.4.15: Summary of genetic variability estimates across sampling locations in *P. longirostris*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e).

Sample	H_o	H_e
06c	0.2244	0.2353
09bc	0.2243	0.2326
11de	0.2208	0.2299
11p	0.2153	0.2311
16b	0.2204	0.2307
17ab	0.2376	0.248
18a	0.2223	0.2451
19b	0.2246	0.2401

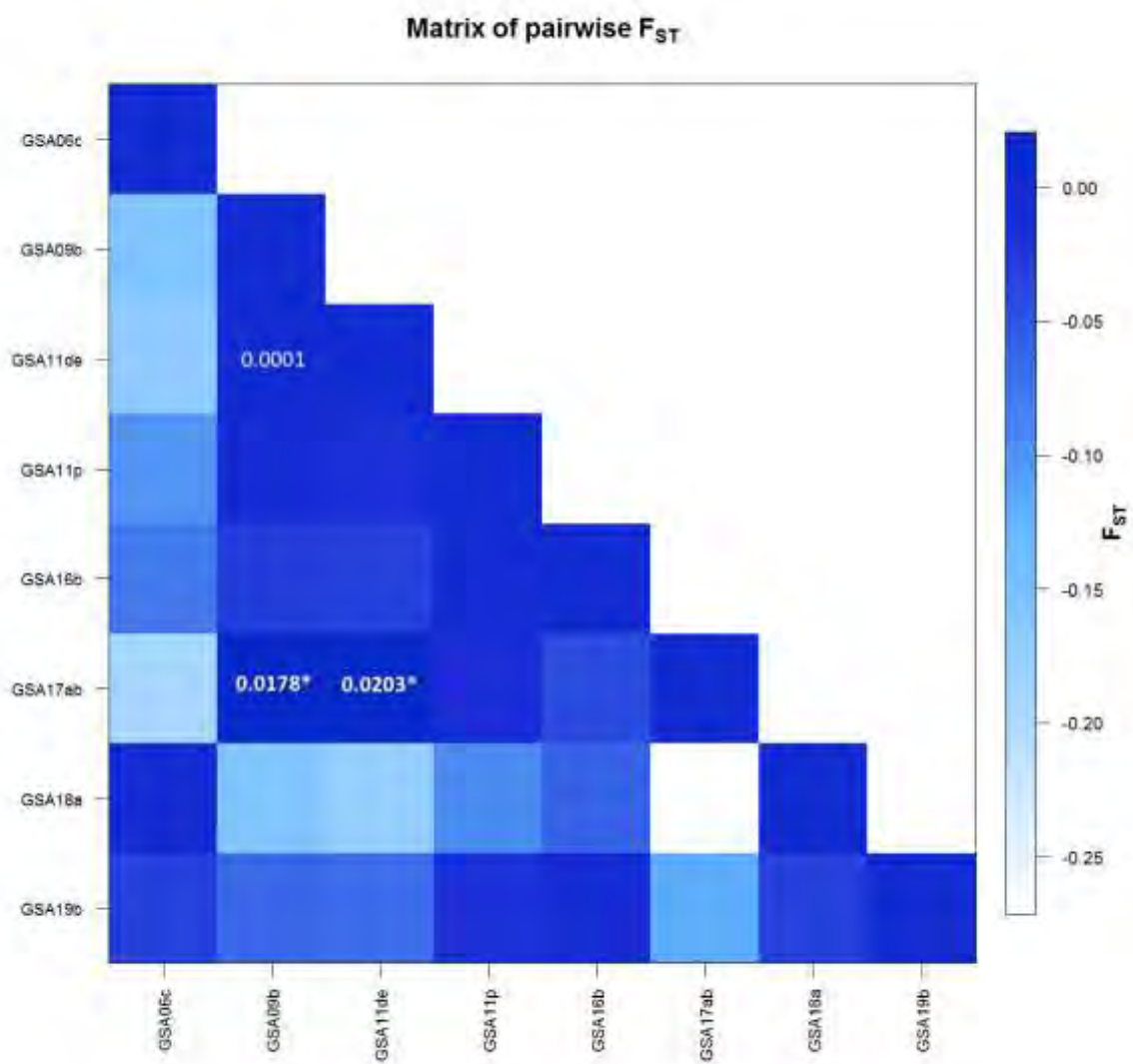


Fig. 1.4.23: Matrix of pairwise F_{ST} for *P. longirostris* based on 1045 polymorphic SNPs present in >50% of the whole sampling set. Significant pairwise F_{ST} values, after Benjamini–Yekutieli correction, are reported in bold.

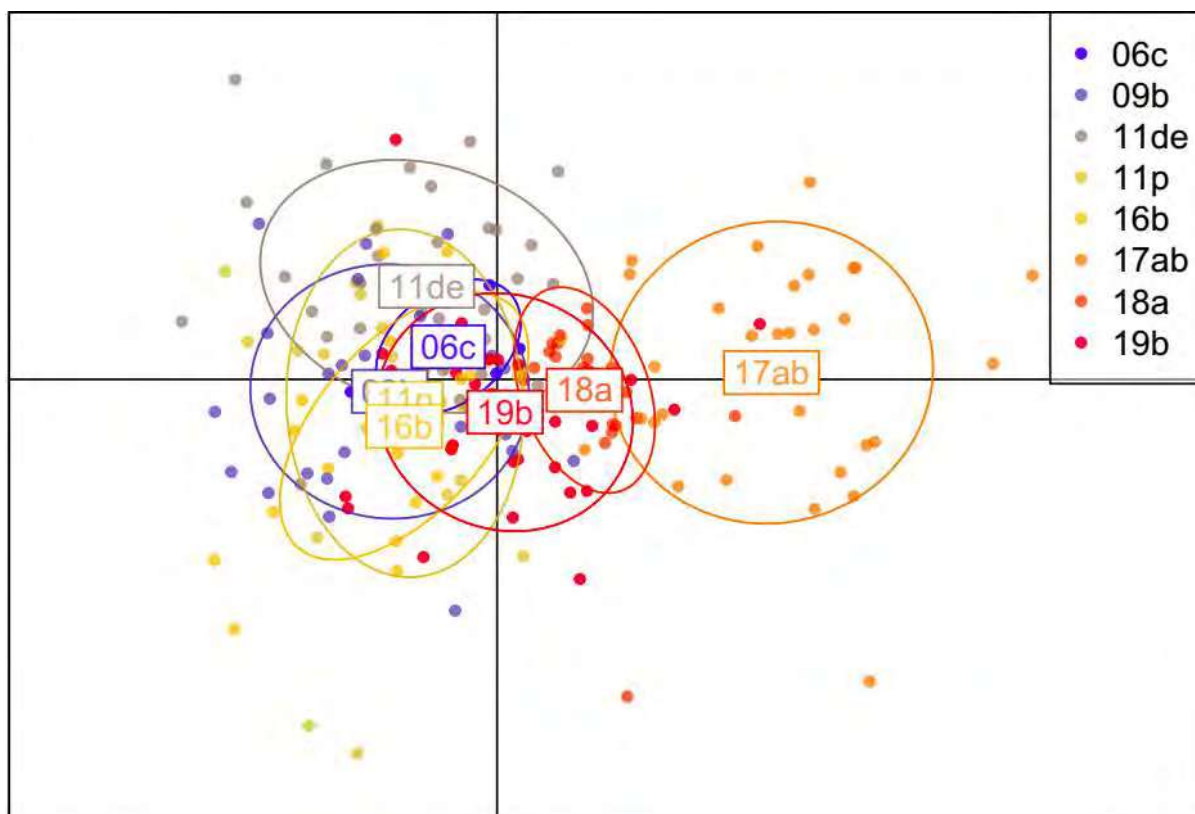


Fig. 1.4.24: DAPC plot for *P. longirostris* based on 1,045 polymorphic SNP present in >50% of the whole sampling set.

Table 1.4.16: AMOVA results for several scenarios of grouping *P. longirostris* samples.

	Scenario 1	Scenario 2	Scenario 3
Group 1	17ab	17ab, 18a, 19b	17ab, 18a, 19b
Group 2	06c, 09b, 11de, 11p, 16b, 18a, 19b	06c, 09b, 11de, 11p, 16b	11de, 11p
Group 3			06c, 09b, 16b
Variation %			
Among groups	1.59	1.59	1.11
Among populations within groups	0.78	0.47	0.50
Within populations	97.63	97.93	98.40

Fixation Indices			
FST	0.02370	0.02065	0.00460
P-value	0.000	0.000	0.000
FSC	0.00797	0.00478	0.00352
P-value	0.000	0.000	0.000
FCT	0.01586	0.01594	0.00108
P-value	0.000	0.000	0.000

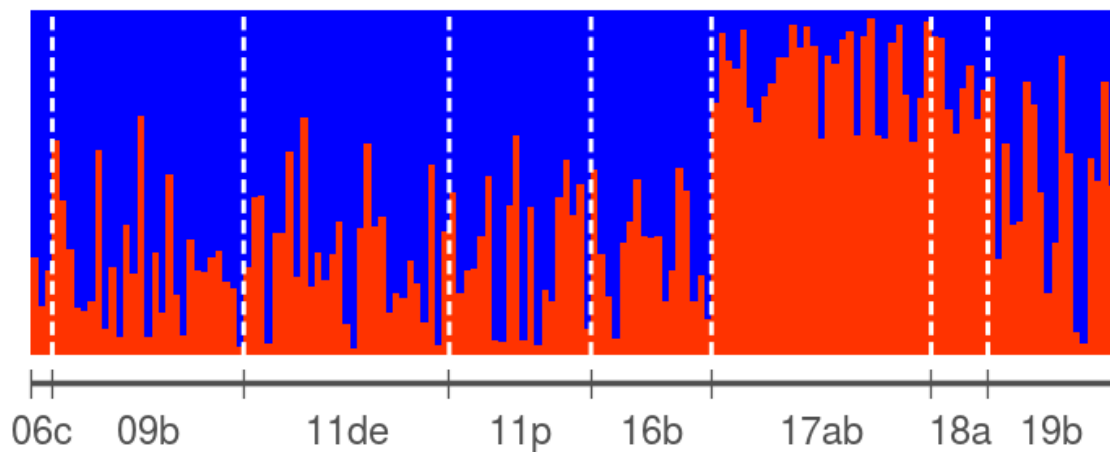


Fig. 1.4.25: Proportions of membership to each K inferred clusters for *P. longirostris*. Best K=2, consensus plot over 10 replicates. For this analysis, a further reduced dataset (153 individuals genotyped at 978 loci) was used to eliminate the bias due to missing data

Aristeus antennatus

A total of 273 samples (from eight populations) were genotyped following a ddRAD protocol, with enzymes *PstI* and *NlaIII*, selected fragments from around 200 to around 480 bp. Multiplexed libraries were run on two Illumina lanes, with 150 bp pair-end protocol. R1 from the two lanes and R2 from the two lanes were merged, obtaining a total of 693.4 millions reads in R1+R2.

ddRADseqTool (Mora-Marqu ez et al. 2017) was used to simulate a ddRAD library using the same enzyme pair and the same size selection used with the target species, to estimate the number of retrieved tags.

For the analysis, a fish genome was used, whose length was 831 Mbp. A total of 230,000 tags of size between 200 and 480 bp were estimated. Considering a genome size up to 5 times bigger for *A. antennatus*, a total number of fragments around 1.5-2 million is expected. After demultiplexing, on average 2.421 M reads per sample were obtained (range from 9K to 13M). After eliminating low quality reads and reads without a RAD site at the beginning, on average 95,3% reads were retained (figure 1.4.26).

Samples were processed with *ustacks* considering a stacks of unique reads only if at least 3 identical reads were found and allowing for up to 3 mismatches to consider two tags as part of the same locus.

On average 98'600 tags were found in each sample, with average coverage of 5.6X (table 1.4.17).

A tag catalog was built with 4 samples per population (32 samples in totals) to reduce analysis time (figure 1.4.27).

A maximum of 3 mismatches were allowed to merge tags into the same catalog locus.

The final number of tags in the catalog is around 1,761,996.

Stacks' module *populations* and *vcftools* were used to filter the initial SNP panel. Only the individuals genotyped with more than 1M filtered reads were retained. Only SNPs shared by at least 50% of the individuals and with at least 6X coverage were kept. A total of 1253 SNPs were retained (for 197 samples). Coverage was low. All the loci retained were used for the downstream differentiation analysis, unless otherwise noted.

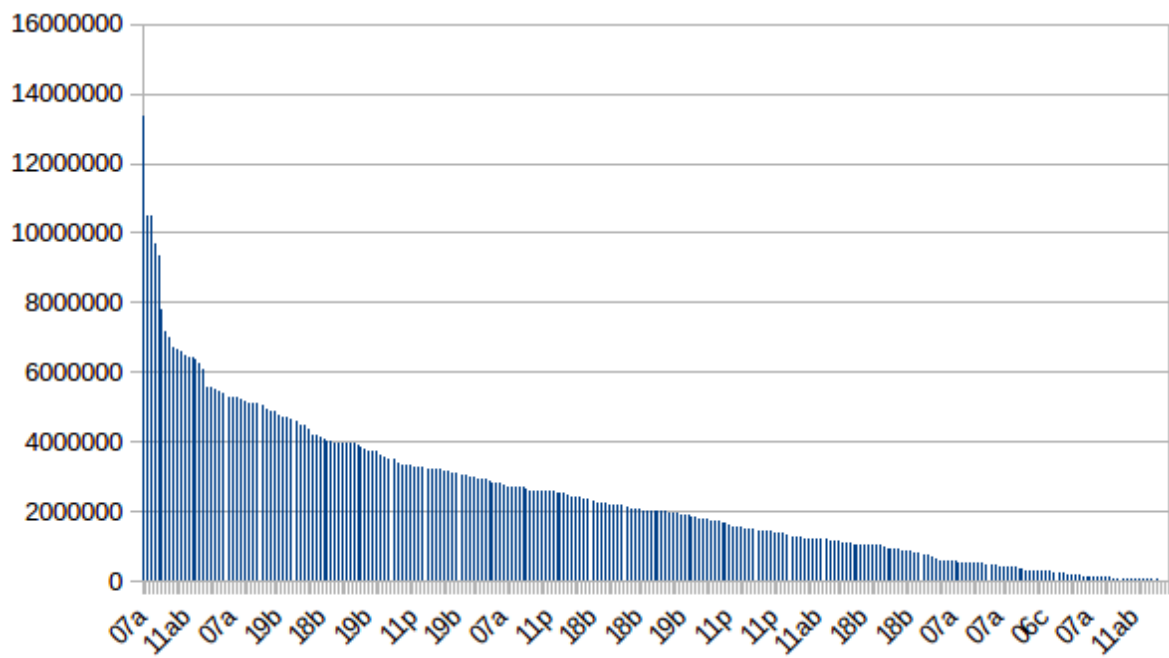
Observed heterozygosities were generally low (ranging from 0.2533 to 0.2612) and slightly higher than the unbiased expected (ranging from 0.2432 to 0.2566; table 1.4.18), with an excess of heterozygotes. All samples showed a great similarity between them (F_{st} values all negative, figure 1.4.28), with a global $F_{st} = 0.0005$, though significant (table 1.4.19).

Such general homogeneity is supported both by DAPC (figure 1.4.29) and STRUCTURE analysis (figure 1.4.30). DAPC identified one homogeneous group with the R's *adegenet's* *find.cluster* function suggesting 1 as the most likely number of clusters in the dataset, and STRUCTURE showed the best likelihood values for K=1.

Table 1.4.17: *Stacks* pipeline output summary for *Aristeus antennatus*

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
06b	889134,1	34425	2590423	27537,6	30	95535
06c	1592801,0	32005	5068477	61706,1	185	231474

07a	3037859,0	31840	13348899	125737,9	34	469802
09p	2627212,8	87300	5436824	110044,7	1216	250072
11ab	2259718,0	9246	9316863	89492,1	22	395190
11p	2971772,7	40481	9689917	124601,4	286	388279
18b	1835493,7	95195	4636558	70907,3	1088	209876
19b	2654667,3	21730	10504251	110099,1	71	414355
ALL	2421295,3	9246	13348899	98600,4	22	469802



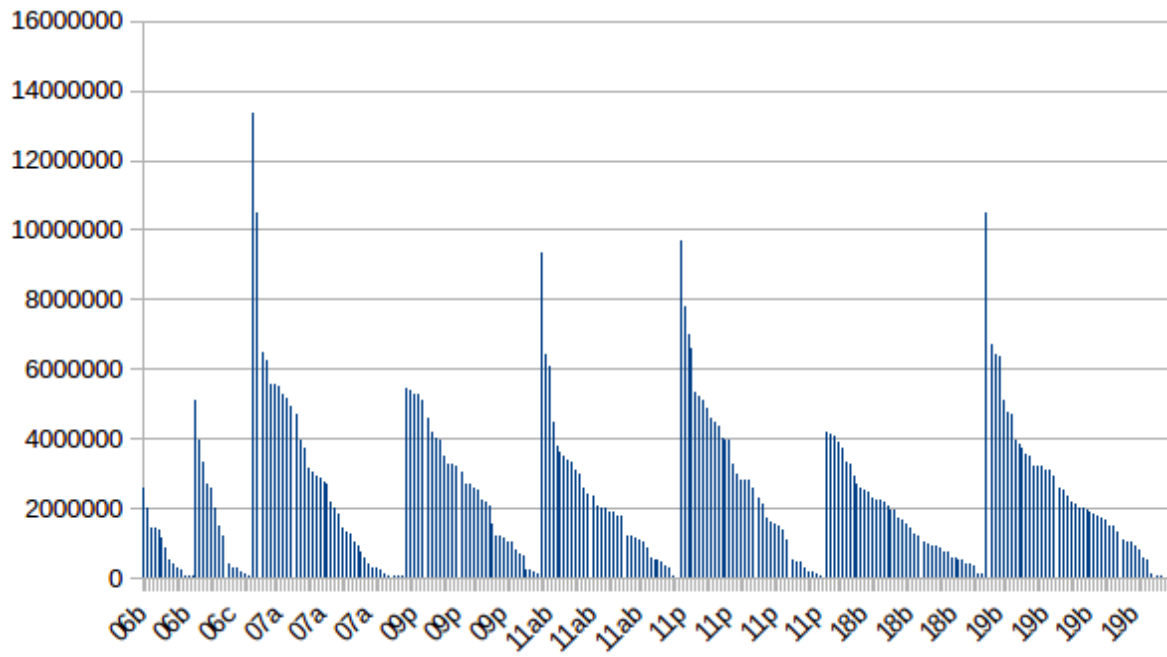
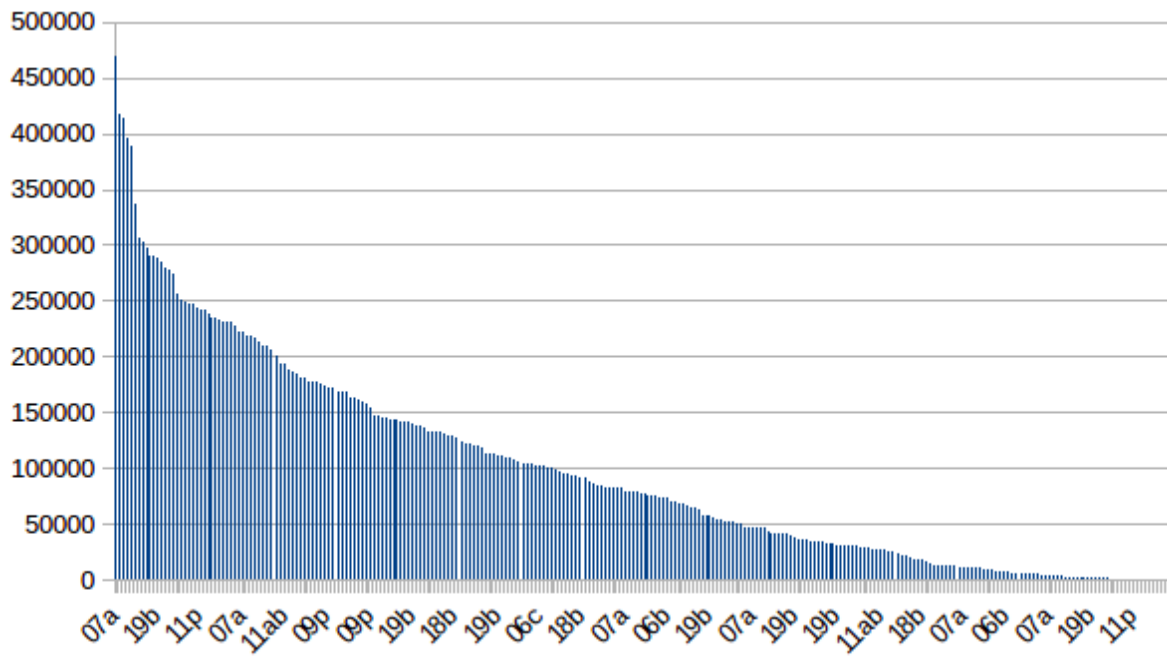


Fig. 1.4.26: Retained reads after demultiplexing for (A) the whole *A. antennatus* dataset, and (B) each population separately



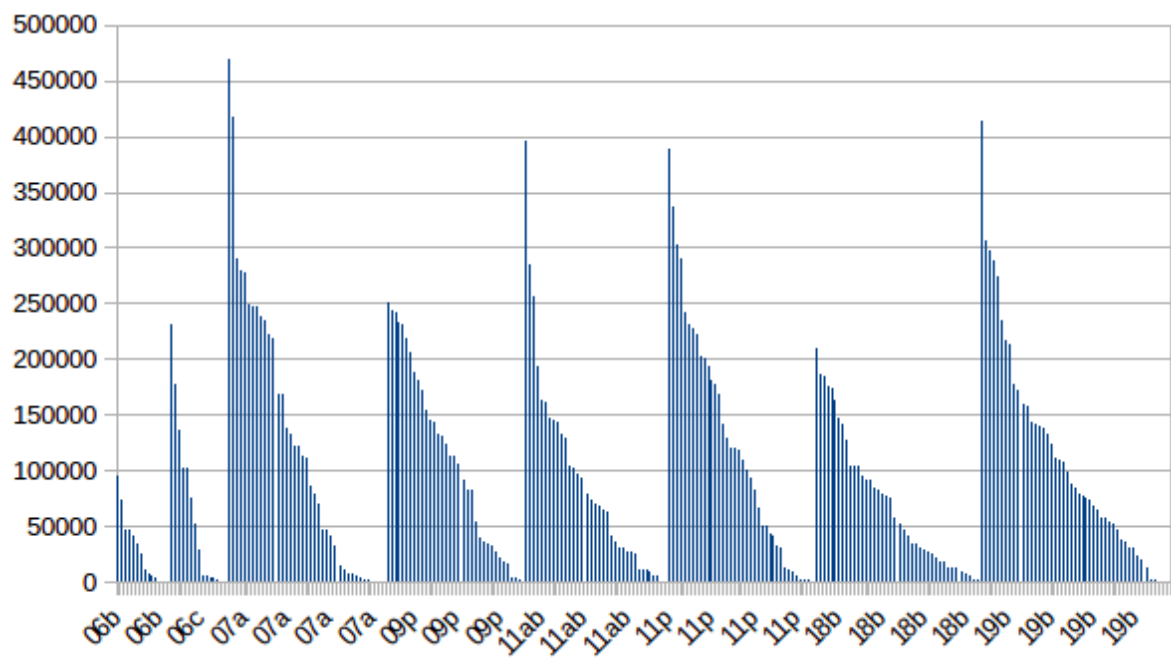


Fig. 1.4.27: ustacks loci for the (A) whole *A. antennatus* dataset, and (B) each population separately.

Table 1.4.18: Summary of genetic variability estimates across sampling locations in *A. antennatus*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e).

Sample	H_o	H_e
06b	0.2583	0.2432
06c	0.2533	0.2468
07a	0.2586	0.2562
09p	0.2582	0.2535
11ab	0.2609	0.2557
11p	0.2610	0.2562
18b	0.2602	0.2509
19b	0.2612	0.2566

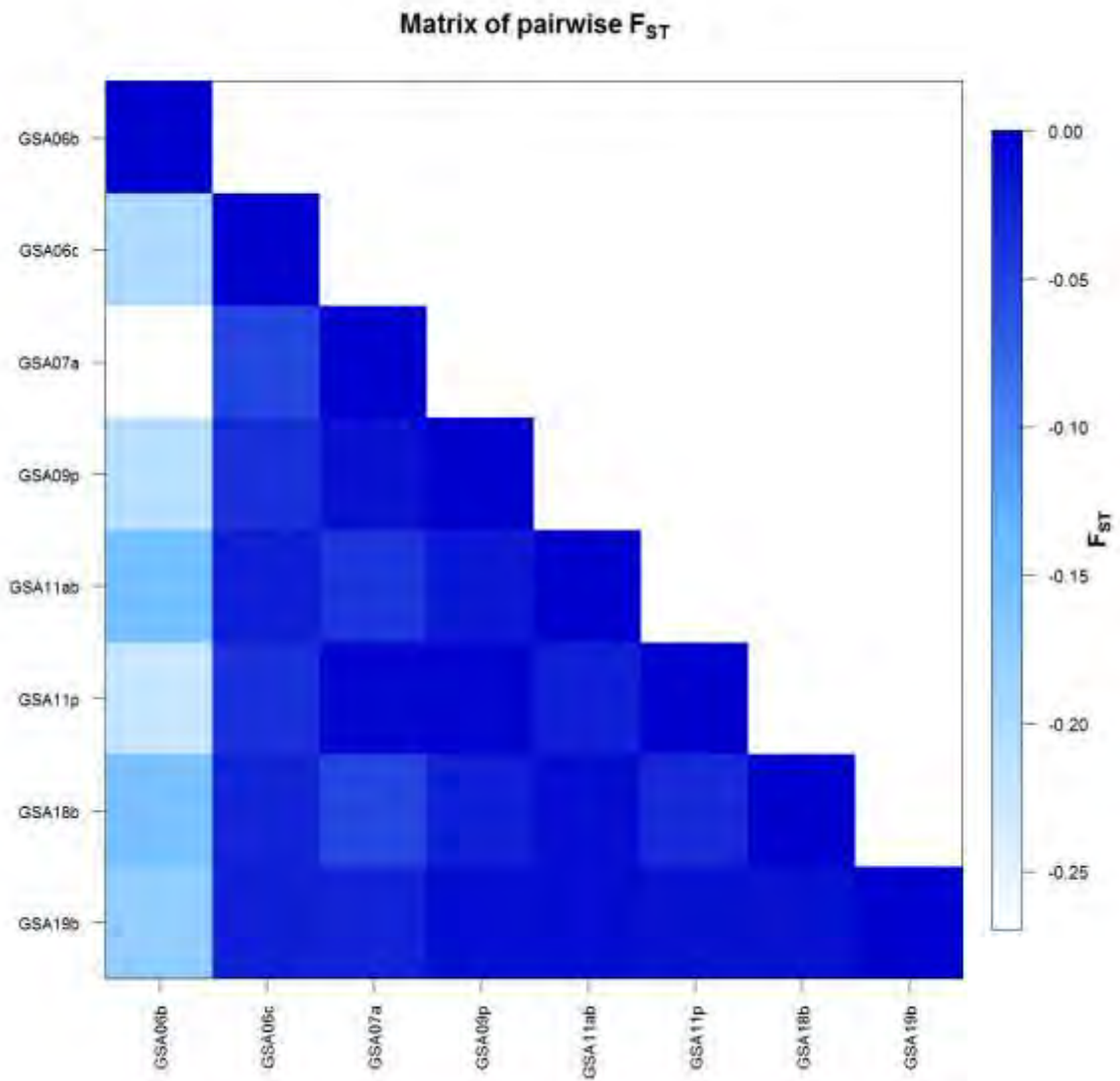


Fig. 1.4.28: Matrix of pairwise F_{ST} for *A. antennatus* based on 1253 polymorphic SNPs present in >50% of the whole sampling set. No value was statistically significant (p values >0.05)

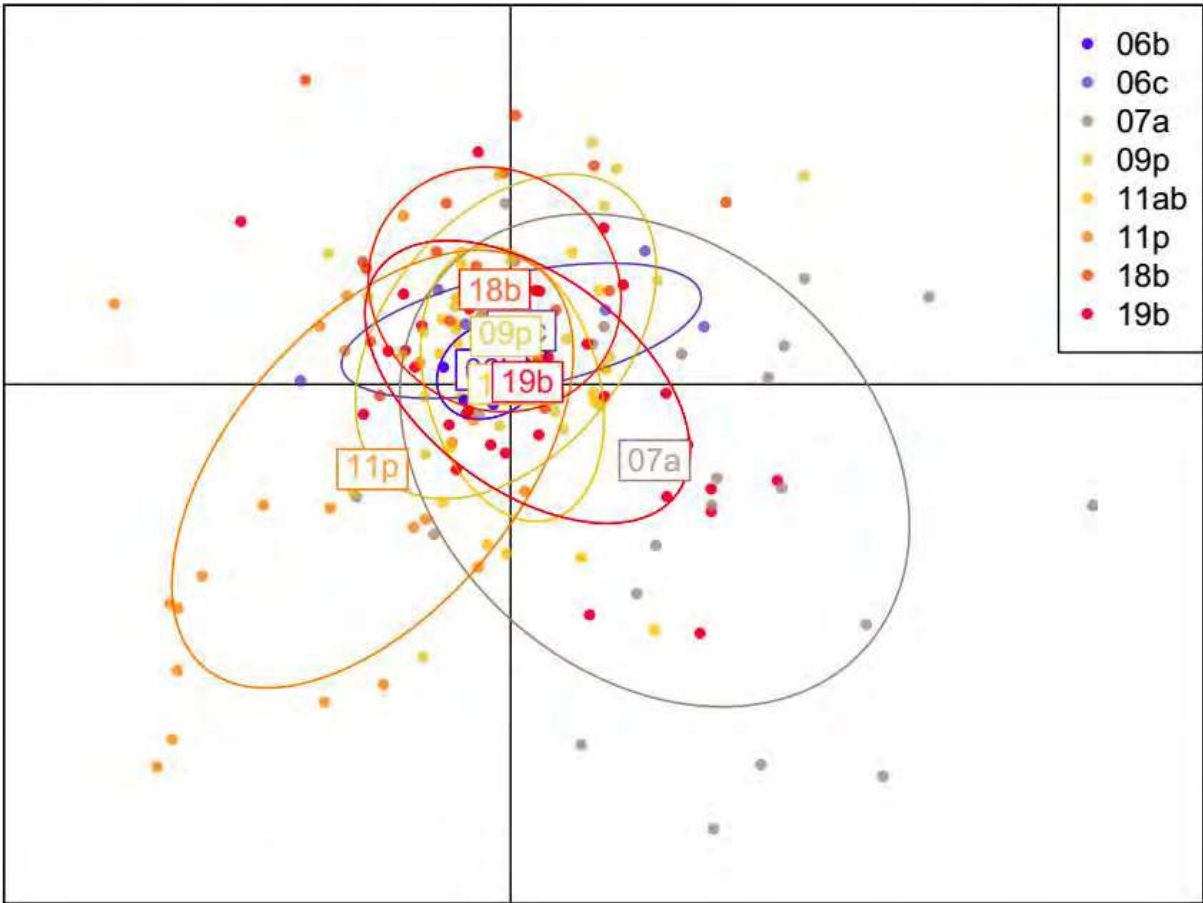


Fig. 1.4.29: DAPC plot for *A. antennatus* based on 1253 polymorphic SNP present in >50% of the whole sampling set.

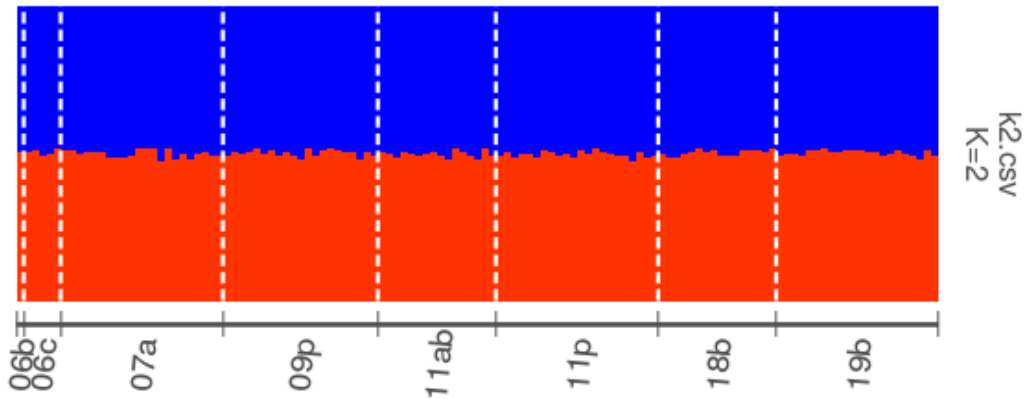


Fig. 1.4.30: Proportions of membership to each K inferred clusters for *A. antennatus*. Best K=2.

Table 1.4.19: AMOVA results for several scenarios of grouping *A. antennatus* samples.

	Scenario 1	Scenario 2
Group 1	06b, 06c, 07a, 09p, 11ab, 11p, 18b, 19b	06b, 06c
Group 2		07a, 09p, 11ab, 11p, 18b, 19b
Group 3		
Variation %		
Among groups		0.10
Among populations within groups	0.05	0.04
Within populations	99.95	99.86
Fixation Indices		
FST	0.0005	0.00141
P-value	0.000	0.058
FSC		0.00044
P-value		0.029
FCT		0.00097
P-value		0.202

Rarefaction Analysis Results

As notified in the previous section, the rarefaction analysis was conducted in only two of the six species studied here, i.e. *Merluccius merluccius* and *Nephrops norvegicus*, because the ddRAD approach provided the most reliable results. In the Table 1.4.20, we show that results highlight that the values of heterozygosity and population differentiation statistics (pairwise F_{st}) are very similar when calculated on the full dataset and on reduced datasets obtained for three different scenarios tested: i) rarefaction of all the samples to the smallest of the original sample sizes, ii) reduction of each sample to 50% of its original sample size, and iii) reduction to 75% of its original sample size.

Table 1.4.20: Rarefaction results for three different scenarios in a) *M. merluccius*, and b) *N. norvegicus*. “Ho” and “He” are the observed and expected heterozygosity, respectively. “DAPC” is the percentage of individuals correctly assigned to their population by DAPC. In the F_{st} matrix, above diagonal, for the full dataset we show if the comparison is significant (*) or not (n.s.), while for the rarefaction scenarios we show the proportion of tests (over 100) that gave significant results. In all cases, the significance threshold was 0.05. For all rarefaction scenarios between brackets is the minimum and maximum for the statistics calculated (again across 100 replicates). For the other scenarios, above diagonal instead of the p-value we show the percentage of replicates that gave p-values below 0.05 (so in practice the higher the number, the higher the significance).

a) *M. merluccius*

Full	06b	06c	09bc	11c	11e	16ab	18a	19b
n° ind	31	30	33	26	35	34	39	40
DAPC	0.354	0.533	0.484	0.461	0.257	0.029	0.102	0.650
Ho	0.221	0.223	0.226	0.234	0.226	0.229	0.224	0.224
He	0.244	0.247	0.247	0.250	0.248	0.251	0.251	0.249
pop	06b	06c	09bc	11c	11e	16ab	18a	19b
06b		n.s.	*	n.s.	n.s.	n.s.	n.s.	*
06c	0.002		*	*	n.s.	n.s.	*	*
09bc	0.005	0.005		n.s.	n.s.	n.s.	*	*
11c	0.002	0.005	0.001		n.s.	n.s.	*	*
11e	0.000	0.002	0.003	0.001		n.s.	n.s.	*
16ab	0.001	0.001	0.001	0.001	-0.00		n.s.	n.s.
18a	0.002	0.003	0.002	0.003	0.000	-0.00		n.s.
19b	0.004	0.005	0.004	0.004	0.003	0.001	0.000	

smallest	06b	06c	09bc	11c	11e	16ab	18a	19b
----------	-----	-----	------	-----	-----	------	-----	-----

n° ind	26	26	26	26	26	26	26	26
DAPC	0.484 (0.115-0.769)	0.582 (0.346-0.846)	0.492 (0.038-0.808)	0.403 (0.115-0.654)	0.337 (0-0.654)	0.129 (0-0.385)	0.211 (0.038-0.538)	0.586 (0.308-0.769)
Ho	0.221 (0.217-0.226)	0.224 (0.221-0.228)	0.227 (0.223-0.232)	0.235 (0.235-0.235)	0.227 (0.221-0.232)	0.23 (0.225-0.234)	0.225 (0.22-0.229)	0.224 (0.219-0.228)
He	0.244 (0.24-0.247)	0.247 (0.243-0.25)	0.246 (0.243-0.25)	0.251 (0.251-0.251)	0.248 (0.244-0.251)	0.25 (0.247-0.254)	0.25 (0.246-0.252)	0.247 (0.244-0.25)
pop	06b	06c	09bc	11c	11e	16ab	18a	19b
06b		0.18	0.86	0.01	0	0.02	0.09	0.39
06c	0.003 (0.001-0.007)		1	1	0.08	0.03	0.58	0.91
09bc	0.005 (0.003-0.011)	0.006 (0.003-0.008)		0	0.17	0.05	0.27	0.72
11c	0.003 (0.001-0.005)	0.005 (0.004-0.007)	0.001 (0-0.003)		0.02	0.01	0.31	0.31
11e	0 (-0.002-0.004)	0.002 (0-0.006)	0.003 (0-0.007)	0.002 (-0.001-0.005)		0	0.02	0.3
16ab	0.002 (-0.001-0.007)	0.001 (0-0.004)	0.002 (-0.001-0.005)	0.002 (0-0.004)	-0.001 (-0.003-0.003)		0	0.02
18a	0.002 (0-0.005)	0.003 (0.001-0.006)	0.002 (-0.001-0.007)	0.004 (0.002-0.005)	0.001 (-0.002-0.006)	-0.001 (-0.003-0.002)		0.02
19b	0.004 (0.001-0.008)	0.005 (0.003-0.008)	0.004 (0.001-0.008)	0.004 (0.001-0.006)	0.003 (0-0.008)	0.001 (-0.002-0.003)	0 (-0.002-0.005)	

50%	06b	06c	09bc	11c	11e	16ab	18a	19b
n° ind	16	15	17	13	18	17	20	20
DAPC	0.491 (0-0.938)	0.608 (0.067-1)	0.505 (0-1)	0.335 (0-0.769)	0.389 (0-0.944)	0.301 (0-0.882)	0.374 (0-0.85)	0.62 (0.1-0.95)
Ho	0.222 (0.211-0.232)	0.224 (0.215-0.234)	0.227 (0.216-0.238)	0.235 (0.226-0.24)	0.227 (0.217-0.237)	0.229 (0.221-0.238)	0.225 (0.217-0.233)	0.225 (0.219-0.231)
He	0.24 (0.233-0.247)	0.242 (0.236-0.25)	0.243 (0.234-0.25)	0.245 (0.238-0.251)	0.244 (0.238-0.251)	0.247 (0.241-0.254)	0.247 (0.243-0.251)	0.245 (0.241-0.249)
pop	06b	06c	09bc	11c	11e	16ab	18a	19b
06b		0.09	0.32	0.08	0.03	0.04	0.13	0.36
06c	0.003 (-0.002-0.011)		0.49	0.29	0.07	0.06	0.26	0.6
09bc	0.005 (0-0.016)	0.006 (0.001-0.014)		0.06	0.12	0.04	0.17	0.53
11c	0.003 (-0.004-0.016)	0.006 (0-0.019)	0.002 (-0.005-0.012)		0.07	0.08	0.15	0.3
11e	0 (-0.005-0.009)	0.002 (-0.003-0.01)	0.003 (-0.003-0.013)	0.001 (-0.007-0.018)		0	0.03	0.34

16ab	0.001 (-0.005-0.012)	0.001 (-0.003-0.008)	0.001 (-0.003-0.01)	0.002 (-0.005-0.02)	-0.001 (-0.005-0.006)		0	0.09
18a	0.002 (-0.002-0.011)	0.003 (-0.001-0.008)	0.003 (-0.001-0.009)	0.004 (-0.002-0.016)	0.001 (-0.004-0.011)	-0.001 (-0.005-0.004)		0.09
19b	0.005 (-0.001-0.016)	0.005 (0-0.013)	0.005 (-0.002-0.015)	0.004 (-0.003-0.015)	0.004 (-0.003-0.014)	0.001 (-0.004-0.01)	0.001 (-0.004-0.007)	

75%	06b	06c	09bc	11c	11e	16ab	18a	19b
n° ind	23	23	25	20	26	26	29	30
DAPC	0.416 (0.043-0.739)	0.54 (0.045-0.909)	0.537 (0.04-0.84)	0.273 (0-0.7)	0.357 (0.038-0.808)	0.168 (0-0.577)	0.25 (0.034-0.69)	0.65 (0.433-0.867)
Ho	0.221 (0.215-0.227)	0.224 (0.22-0.228)	0.227 (0.219-0.232)	0.235 (0.23-0.238)	0.227 (0.222-0.232)	0.229 (0.224-0.233)	0.224 (0.219-0.23)	0.224 (0.221-0.228)
He	0.243 (0.239-0.246)	0.246 (0.241-0.249)	0.246 (0.242-0.249)	0.249 (0.246-0.252)	0.247 (0.245-0.251)	0.25 (0.247-0.254)	0.25 (0.246-0.253)	0.248 (0.245-0.25)
pop	06b	06c	09bc	11c	11e	16ab	18a	19b
06b		0.1	0.7	0.08	0	0.08	0.13	0.55
06c	0.003 (-0.001-0.006)		0.89	0.53	0.03	0.01	0.48	0.9
09bc	0.005 (0.002-0.014)	0.005 (0.003-0.011)		0.02	0.19	0.05	0.29	0.87
11c	0.003 (-0.001-0.012)	0.005 (0.001-0.009)	0.001 (-0.001-0.005)		0.03	0.02	0.29	0.44
11e	0 (-0.002-0.004)	0.002 (-0.001-0.005)	0.003 (0.001-0.007)	0.002 (-0.002-0.007)		0	0.02	0.48
16ab	0.001 (-0.002-0.006)	0.001 (-0.001-0.003)	0.001 (-0.001-0.005)	0.002 (-0.002-0.006)	-0.001 (-0.003-0.003)		0	0
18a	0.002 (0-0.006)	0.003 (0.001-0.006)	0.002 (0-0.005)	0.004 (0.001-0.008)	0.001 (-0.002-0.005)	-0.001 (-0.004-0.002)		0
19b	0.005 (0.002-0.01)	0.005 (0.003-0.008)	0.005 (0.002-0.008)	0.004 (0.001-0.008)	0.004 (0.001-0.007)	0.001 (-0.002-0.003)	0 (-0.002-0.002)	

b) *N. norvegicus*

Full	06b	06c	08ab	09ab	11e	11p	17ab
n° ind	23	16	50	50	50	44	31
DAPC	0	0.438	0.58	0.64	0.3	0.273	0.935
Ho	0.26	0.25	0.274	0.287	0.274	0.278	0.247
He	0.329	0.321	0.333	0.339	0.333	0.336	0.33
pop	06b	06c	08ab	09ab	11e	11p	17ab
06b		n.s.	*	*	*	*	*

06c	-0.00		*	*	*	*	*
08ab	0.02	0.034		*	*	*	*
09ab	0.019	0.032	0.004		*	*	*
11e	0.029	0.04	0.003	0.005		*	*
11p	0.022	0.03	0.005	0.002	0.003		*
17ab	0.041	0.025	0.085	0.08	0.086	0.074	

Smallest	06b	06c	08ab	09ab	11e	11p	17ab
n° ind	16	16	16	16	16	16	16
DAPC	0.308 (0-0.812)	0.565 (0.125-0.875)	0.388 (0.062-0.812)	0.556 (0-0.938)	0.395 (0-0.875)	0.415 (0-0.938)	0.906 (0.688-1)
Ho	0.26 (0.249-0.269)	0.25 (0.25-0.25)	0.275 (0.263-0.284)	0.287 (0.279-0.296)	0.274 (0.259-0.289)	0.278 (0.269-0.291)	0.248 (0.213-0.277)
He	0.324 (0.32-0.328)	0.321 (0.321-0.321)	0.323 (0.317-0.329)	0.33 (0.325-0.335)	0.324 (0.316-0.333)	0.327 (0.321-0.333)	0.32 (0.31-0.327)
pop	06b	06c	08ab	09ab	11e	11p	17ab
06b		0.01	1	0.91	1	1	1
06c	-0.002 (-0.006-0.01)		1	1	1	1	0.71
08ab	0.019 (0.007-0.032)	0.029 (0.021-0.036)		0.57	0.58	0.73	1
09ab	0.017 (0.001-0.032)	0.027 (0.011-0.041)	0.004 (-0.001-0.011)		0.77	0.2	1
11e	0.028 (0.015-0.049)	0.036 (0.024-0.044)	0.003 (-0.001-0.009)	0.005 (-0.001-0.01)		0.51	1
11p	0.02 (0.008-0.037)	0.025 (0.015-0.038)	0.005 (-0.002-0.013)	0.002 (-0.003-0.006)	0.003 (-0.003-0.013)		1
17ab	0.041 (0.016-0.091)	0.026 (0.01-0.061)	0.08 (0.038-0.141)	0.075 (0.027-0.143)	0.083 (0.037-0.151)	0.071 (0.034-0.138)	

50%	06b	06c	08ab	09ab	11e	11p	17ab
n° ind	12	8	25	25	25	22	16
DAPC	0.122 (0-0.833)	0.239 (0-0.75)	0.577 (0.28-0.92)	0.602 (0.12-0.88)	0.508 (0.12-0.84)	0.399 (0-0.955)	0.908 (0.688-1)
Ho	0.259 (0.245-0.277)	0.251 (0.221-0.281)	0.274 (0.268-0.281)	0.287 (0.281-0.293)	0.273 (0.264-0.283)	0.279 (0.27-0.286)	0.246 (0.223-0.274)
He	0.319 (0.313-0.325)	0.304 (0.292-0.311)	0.328 (0.325-0.332)	0.335 (0.332-0.339)	0.328 (0.324-0.332)	0.331 (0.326-0.336)	0.32 (0.312-0.326)
pop	06b	06c	08ab	09ab	11e	11p	17ab
06b		0.02	1	0.98	1	1	0.9

06c	-0.002 (-0.014-0.029)		1	0.99	1	1	0.23
08ab	0.023 (0.006-0.041)	0.038 (0.01-0.078)		0.94	0.82	0.97	1
09ab	0.021 (0.004-0.048)	0.037 (0.009-0.076)	0.004 (0.001-0.007)		0.97	0.24	1
11e	0.032 (0.011-0.052)	0.044 (0.013-0.08)	0.003 (0.001-0.007)	0.005 (0.002-0.009)		0.73	1
11p	0.024 (0.011-0.047)	0.034 (0.011-0.069)	0.005 (0-0.011)	0.002 (-0.001-0.005)	0.004 (-0.001-0.009)		1
17ab	0.041 (0.013-0.084)	0.022 (0-0.075)	0.088 (0.037-0.138)	0.083 (0.03-0.141)	0.09 (0.034-0.149)	0.077 (0.03-0.126)	

75%	06b	06c	08ab	09ab	11e	11p	17ab
n° ind	17	12	38	38	38	33	23
DAPC	0.026 (0-0.353)	0.224 (0-0.667)	0.579 (0.289-0.868)	0.634 (0.395-0.895)	0.406 (0.105-0.895)	0.365 (0-0.909)	0.927 (0.826-1)
Ho	0.259 (0.248-0.269)	0.25 (0.238-0.267)	0.274 (0.269-0.278)	0.287 (0.282-0.291)	0.274 (0.27-0.279)	0.278 (0.274-0.282)	0.248 (0.237-0.264)
He	0.325 (0.322-0.328)	0.315 (0.312-0.319)	0.331 (0.329-0.334)	0.338 (0.336-0.34)	0.331 (0.329-0.334)	0.334 (0.331-0.337)	0.326 (0.322-0.331)
pop	06b	06c	08ab	09ab	11e	11p	17ab
06b		0.02	1	1	1	1	1
06c	-0.002 (-0.008-0.015)		1	1	1	1	0.54
08ab	0.021 (0.011-0.037)	0.035 (0.014-0.054)		1	1	1	1
09ab	0.02 (0.008-0.036)	0.034 (0.014-0.053)	0.004 (0.002-0.006)		1	0.46	1
11e	0.03 (0.018-0.05)	0.042 (0.018-0.059)	0.003 (0.001-0.005)	0.005 (0.003-0.007)		0.96	1
11p	0.023 (0.014-0.036)	0.032 (0.012-0.051)	0.005 (0.002-0.007)	0.002 (0-0.003)	0.003 (0.001-0.007)		1
17ab	0.039 (0.015-0.075)	0.024 (0.01-0.055)	0.084 (0.053-0.108)	0.08 (0.047-0.11)	0.086 (0.053-0.11)	0.074 (0.044-0.098)	

Furthermore, significance for pairwise F_{st} remained practically similar, since we obtained non significant values with all the rarefacted datasets in population comparisons for which the full dataset indicated genetic homogeneity. Similarly, the majority of the rarefacted dataset provided significant results in comparisons where the full dataset indicated genetic differentiation, though, in some cases, significance was lost for comparisons with a very small original F_{st} value.

Based on the above results and our initial sampling plan, i.e. to obtain at least 50 specimens/locality, we are confident that even with the reduction in the samples i) included in the library preparations due to low DNA quality, and ii) those not finally performing well

after the demultiplexing and bioinformatic analyses, we shall end up with a sufficient number of specimens to use into our full biogeographic study.

Conclusions and Suggestions

The pilot study allowed to obtain ddRAD sequencing data for all the six species considered, and it provided important lessons about the use of this methodology. The ddRAD approach produced promising results, identifying more differences than those previously reported among Mediterranean populations of the European hake, providing the first evidence for differentiation of the Adriatic sample in Norway lobsters and, in general, producing results in line with the expectations from literature, when available. The approach, however, was critically dependent on the DNA quality and can require specific optimization when applied to new species.

For the European hake (*Merluccius merluccius*), preliminary study was conducted with samples from 8 localities (GSA06b, 06c, 09bc, 11c, 11e, 16ab, 18a and 19b). DNA extraction results showed that the majority of samples provided DNA of high quality, with the few poor quality specimens randomly scattered among localities. ddRAD was then performed on 288 individuals from all the 8 localities, using a pair of restriction enzymes (RE) composed by an 8-base cutter and a 6-base cutter, which is typical in fish studies. Sequencing at the commercial sequencing provider produced the expected number of reads (549 millions) with on average 1.90 million reads/specimen. Using all the individuals, we constructed the "catalogue" which comprised about 102K ddRAD loci. After filtering and extensive data checking, a panel of 734 high quality SNPs, present in at least 80% of the samples and in all the 8 population samples, was retained and used for subsequent genetic analyses; filtering required eliminating several bad quality specimens leading to the inclusion of 268 individuals in the final analysis. Observed heterozygosities were similar among samples and varied between 0.2215 to 0.2345, whereas expected heterozygosities were slightly higher ranging from 0.2494 to 0.2566. Both DAPC and STRUCTURE plots indicated low differentiation between samples. Pairwise F_{st} s were small (ranging from 0.0003 to a maximum value of 0.0114), but highly significant in most of the comparisons (17 out of 28). AMOVA analysis highlighted the strongest differentiation of GSA06b and 06c from the rest of the samples and, to a lesser extent, a further differentiation of samples from GSA18a and 19b. This result represents a significant advance respect to previous published studies, where differences were not detected between Mediterranean samples, unless using outlier loci (Milano et al. 2014).

For the red mullet (*Mullus barbatus*), preliminary study was conducted with samples from 8 localities (GSA06c, 09b, 11c, 11de, 17a, 17b, 18a and 18c). DNA extraction results have unfortunately showed that the majority of these 400 specimens yielded low quality material to work with and therefore the choice of the samples to include in our library preparations was difficult but straightforward. Due to low quality of the extracted DNA, we had to exclude samples from GSA11de and 18c, and worked on samples from the remaining six locations

which were processed for library construction, although some of them were considered of medium or bad quality. As in the case of *M. merluccius* above, the appropriate pair of enzymes chosen was an 8-base cutter & 6-base cutter which is a classical pair of RE used in fish studies. Sequencing at the commercial sequencing provider produced the expected number of reads (~660 million) with on average 2.18 million reads/fish. Using all individuals, we constructed the "catalogue" which comprised some 48K ddRAD loci. After several steps, a panel of 580 higher quality SNPs, present in at least 50% of the samples and four populations, was used for all downstream differentiation analysis. Observed heterozygosities were generally low (ranging from 0.1076 to 0.1340) and smaller than the unbiased expected (0.364 on average). All samples showed a great similarity between them (F_{st} values from negative -and practically zero- to 0.0049, and STRUCTURE plots) with the exception of the one coming from GSA06c (Spanish coast, see the DAPC plot); however, the latter sample performed the worst in the sequencing process and lead to the least number of loci, i.e. 732.63 on average when this number for the other samples ranged from 1954.43 to 2846.558. When current results are compared to those previously reported in the literature (see D1.2), we do not encounter the differentiation of the Adriatic Sea (present samples 17a and 17b) from the neighboring regions reported in Maggio et al. (2009) and Matic-Skoko et al. (2018); moreover, the Spanish sample differentiation has been advanced by Galarza et al. (2009) although the sampling effort was limited, with large areas not covered at all. In conclusion, the *M. barbatus* study requires further analysis to explore the parameter space to extract robust conclusions on the population structure among the studied populations.

For the deep water rose shrimp (*Parapenaeus longirostris*), the preliminary study was conducted with samples from 8 localities (GSA06c, 09b, 11de, 11p, 16b, 17ab, 18a and 19b). Most of the samples provided high or medium quality DNA extracts, except for locality 06c where about half of the specimens (23/50) were of poor quality. ddRAD was performed on 288 individuals from all the 8 localities, with a RE combination of a 6-base cutter and a 4-base cutter. Sequencing produced about 687 million reads with approximately 2.38 million reads for each specimen. Using all the individuals, we constructed the "catalogue" which comprised about 970K ddRAD loci; this number was very high, due the previously unknown big genome size of the species, resulting in low coverage, and requiring discarding several loci and individuals. In fact, after filtering, a panel of 1,045 higher quality SNPs, present in at least 50% of the samples and eight populations, was used for all the following population genetics analysis and 226 individuals were included in the final analysis. Observed heterozygosities were similar among localities, ranging from 0.2153 to 0.2376, and slightly smaller than expected heterozygosities, which varied between 0.2299 and 0.248. Pairwise F_{st} showed a low level of differentiation among samples, with two comparisons involving GSA17ab significant after correction for multiple tests. Similarly, global F_{st} were small but statistically significant, and AMOVA pointed out a significant differentiation between specimens from GSA17 from the others, with about 1.6% of the overall genetic variation attributable to this subdivision. This result was also evident from DAPC plots. Results of this preliminary analysis seems to confirm the existence of significant differentiation among Mediterranean samples, detected with AFLP and mtDNA (Lo Brutto et al. 2013). It is also particularly promising, when considering that Lo Brutto and colleagues (2013) detected the deepest differentiation when comparing Aegean samples, still not included in our analysis, with other Mediterranean sites. However, a further optimization of the library preparation is needed to provide the highest resolution and the most robust results, given that the present dataset is characterized by a low coverage and high level of missing data. In particular,

modification of the size selection step to reduce the total number of loci is currently undergoing.

For the blue and red shrimp (*Aristeus antennatus*) the preliminary study was conducted with samples from 8 localities (GSA06b, 06c, 07a, 09p, 11ab, 11p, 18b and 19b). Many of the samples (124 out of 400 tested) provided poor quality DNA extracts, with the majority of specimens from localities 06b and 06c not suitable for ddRAD analysis. ddRAD was performed on 273 individuals from all the 8 localities, but with sample size limited to 15 and 14 specimens (for GSA06b and 06c), using a RE combination of a 6-base cutter and a 4-base cutter. Sequencing produced about 693 million reads with approximately 2.4 million reads for each specimen. Using all the individuals, we constructed the "catalogue" which comprised about 1,761K ddRAD loci. Similarly to *P. longirostris*, this number was extremely high, due to the previously unknown big genome size of the species, resulting in very low coverage, and requiring discarding several loci and individuals. In fact, after filtering, a panel of 1,253 higher quality SNPs, present in at least 50% of the samples and eight populations, was used for all the following population genetics analysis, but only 197 individuals were included in the final analysis. Observed heterozygosities were similar among localities, ranging from 0.2533 to 0.2612, and were slightly higher than the expected heterozygosity values ranging between 0.2432 and 0.2566. Pairwise F_{st} showed no differences among samples, though global F_{st} were small but statistically significant. The small differentiation was also evident from DAPC plots. Results of this preliminary analysis seems to confirm the very low, if any, genetic differentiation among Mediterranean samples detected by allozymes (Sarda et al. 1998), microsatellites (Cannas et al. 2012), mtDNA sequences (Fernández et al. 2013b, Maggio et al. 2009, Marra et al. 2015; Roldán et al. 2009; Sardà et al. 2010) and AFLP (Lo Brutto et al. 2012). However, a much higher resolution and more robust results can be obtained by further optimization of the ddRAD protocol, which might reveal previously hidden subtle differences. In fact, even more than for *P. longirostris*, the present dataset is characterized by a low coverage and a high level of missing data, and it is currently being improved by introducing a modification of the size selection step to reduce the total number of loci of the catalogue, increase the average coverage and consequently the number of retained SNPs.

For the Norway lobster (*Nephrops norvegicus*), the preliminary study was conducted with samples from 8 localities (GSA06b, 06c, 08ab, 09ab, 11c, 11e, 17ab and 18a). The majority of samples provided DNA of high quality in most samples; in only two cases (GSA06b, 06c), approximately half of the specimens were used, and in only one (GSA18a) the DNA quality was very low and this sample was excluded from the library construction. The appropriate pair of enzymes chosen was a 8-base cutter & 4-base cutter. Sequencing at the commercial sequencing provider produced the expected number of reads (~648 million) with on average 2.25 million reads/specimen. Using all individuals, we constructed the "catalogue" which comprised some 749K ddRAD loci. After several steps, a panel of 1,393 higher quality SNPs, present in at least 80% of the samples and four populations, was used for all downstream differentiation analysis. Observed heterozygosities were among the highest encountered in the whole study (ranging from 0.2474 to 0.2867) and smaller than the unbiased expected (0.3381 on average). In the Norway lobster, samples showed a moderate similarity between them (F_{st} values were not negative and ranged from 0.006 to 0.0927 with the highest observed between GSA17ab (N. Adriatic) sample and all the others (see the DAPC and STRUCTURE plots); however, the GSA17ab sample had the least number of

loci, i.e. ~9K on average, when this number for the best four samples was above 18K and for the two smaller samples (GSA06b and o6c) ranged from 9.5 to 13.7K. However, AMOVA results for several scenarios of grouping in the Norway lobster samples clearly indicated a moderately significant differentiation between the GSA17 sample and all others (~7% differentiation). With practically few recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, current results are the first to show a slight but considerable differentiation of the Adriatic sample. In conclusion, the Norway lobster study is the best shown here for the crustacean species and is expected to provide interesting results without any further optimization.

For the giant red shrimp (*Aristaeomorpha foliacea*), the preliminary study was conducted with samples from 8 localities (GSA09b, 10b, 11c, 11e, 16b, 18a, 19b and 19c). The majority of samples provided DNA of good quality, except for two subareas: GSA19b (finally not included in the library preparation) and 19c; therefore, we included samples from the remaining seven locations into library construction, and for 19c only 7 specimens which were considered of quality. The appropriate pair of enzymes chosen was a 8-base cutter & 4-base cutter. Sequencing at the commercial sequencing provider produced the expected number of reads (~648 million) with on average 2.25 million reads/specimen. Using all individuals, we constructed the "catalogue" which comprised some 1.2M ddRAD loci, a number which is considered huge. After several steps, and finally excluding the GSA19c sample due to the very low number of reads, a panel of 3,437 higher quality SNPs, present in at least 50% of the samples and four populations, was used for all downstream differentiation analysis; therefore, the estimates below are for only 6 sampling locations. Observed heterozygosities were generally low (ranging from 0.0968 to 0.2281) and smaller than the unbiased expected (0.3247 on average). All samples showed a great similarity between them (F_{st} values from negative to 0.0069, and STRUCTURE plots) with the exception of the two coming from GSA09b (N. Tyrrhenian coast) and 11c (South Sardinia) (see the DAPC plot); however, in this case the latter two samples performed the best in the sequencing process and lead to the greatest number of loci i.e. >50K on average when this number for the other four samples ranged from 4,678 to 22,000. Moreover, AMOVA results for several scenarios of grouping *A. foliacea* samples did not indicate any significant differentiation between groups. Current results refer to samples from the Central and Western Mediterranean where the lack of genetic differentiation among locations has already been reported (Cannas et al 2012; Marcias et al., 2010; Fernandez et al., 2011a; Fernandez et al., 2013b). In conclusion, our common belief is that the *A. foliacea* study is expected to provide interesting results when a further optimization of the library preparation will take place in order from one hand to substantially decrease the total number of SNPs recovered and on the other hand increase the number of common SNPs for downstream genetic analyses.

The rarefaction approach was conducted in the two species for which the ddRAD provided the most reliable results, i.e. *Merluccius merluccius* and *Nephrops norvegicus*. The main aim of this analysis was to investigate if a reduction in the sampling effort (i.e. the number of specimens/sample) might substantially influence the population genetics results. Results showed that the values of heterozygosity and population differentiation statistics (pairwise F_{st}) were very similar when calculated on the full dataset and on reduced datasets obtained for three different scenarios tested: i) rarefaction of all the samples to the smallest of the original sample sizes, ii) reduction of each sample to 50% of its original sample size, and iii)

reduction to 75% of its original sample size. Significance for pairwise F_{st} remained also similar, since we obtained non significant values with all the rarefacted datasets in population comparisons for which the full dataset indicated genetic homogeneity. Similarly, the majority of the rarefacted dataset provided significant results in comparisons where the full dataset indicated genetic differentiation, though, in some cases, significance was lost for comparisons with a very small original F_{st} value.

Moreover, the following information has been gained when working on the above six species in the last 12 months:

1. There seems to be no clear association of the total number of tags and the number of polymorphic SNP identified with the classification of a specimen as of high or medium DNA quality in the library preparation. We suspect that the difference might most probably attributed to the quantification method, and more attention should be also paid there.

2. Size-window for ddRAD libraries needs to be adjusted for each species. More specifically, we suggest for:

- The *M. barbatus* study has to be further optimized from 300-600bp instead of 400-700bp.
- For *P. longirostris*, a further optimization of the library preparation is needed to provide the highest resolution and the most robust results, given that the present dataset is characterized by a low coverage and high level of missing data. In particular, modification of the size selection step to reduce the total number of loci is currently undergoing
- In *A. antennatus*, a much higher resolution and more robust results can be obtained by further optimization of the ddRAD protocol, which might reveal previously hidden subtle differences. In fact, even more than for *P. longirostris*, the present dataset is characterized by a low coverage and a high level of missing data, and it is currently being improved by introducing a modification of the size selection step to reduce the total number of loci of the catalogue, increase the average coverage and consequently the number of retained SNPs.
- For *N. nephrops*, the current study is the best shown here for the crustacean species and is expected to provide interesting results without any further optimization.
- In *A. foliacea*, interesting results are expected when a further optimization of the library preparation will take place in order from one hand to substantially decrease the total number of SNPs recovered and on the other hand increase the number of common SNPs for downstream genetic analyses

3. Finally, for future studies and in other species, we suggest a faster and straightforward optimization approach focused in a couple of populations, with more enzyme combinations, and less coverage to have the first insight of how a new species is behaving.

References

- Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Ann. Stat.*, 29, 1165–1188.
- Cannas, R., S. Marcias, F. Sacco, A. Cau & A. M. Deiana, 2012. First isolation and characterization of genomic SSR markers for the giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827). *Genet Mol Res* 11(3):2745-8 doi:10.4238/2012.August.17.1.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores & W. A. Cresko, 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol* 22(11):3124-3140 doi:10.1111/mec.12354.
- Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E. and Orlando, L. (2017), Experimental conditions improving in-solution target enrichment for ancient DNA . *Mol Ecol Resour*, 17: 508-522. doi:[10.1111/1755-0998.12595](https://doi.org/10.1111/1755-0998.12595)
- Excoffier, L. & Lischer, H. E. L. Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10, 564–567 (2010).
- Falush, D., M. Stephens & J. K. Pritchard, 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567-1587.
- Falush, D., M. Stephens & J. K. Pritchard, 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7(4):574-578 doi:10.1111/j.1471-8286.2007.01758.x.
- Fernández, M. V., F. Maltagliati, F. G. Pannacciulli & M. I. Roldán, 2011. Analysis of genetic variability in *Aristaeomorpha foliacea* (Crustacea, Aristeidae) using DNA-ISSR (Inter Simple Sequence Repeats) markers. *Comptes Rendus Biologies* 334(10):705-712 doi:<https://doi.org/10.1016/j.crv.2011.07.005>.
- Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán, 2013b. Multilocus Comparative Phylogeography of Two Aristeid Shrimps of High Commercial Interest (*Aristeus antennatus* and *Aristaeomorpha foliacea*) Reveals Different Responses to Past Environmental Changes. *PLOS ONE* 8(3):e59033 doi:10.1371/journal.pone.0059033
- Galarza, J. A., G. F. Turner, E. Macpherson & C. Rico, 2009. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). *Canadian Journal of Fisheries and Aquatic Sciences* 66(9):1478-1490 doi:10.1139/f09-098.
- Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi: 10.1093/bioinformatics/btn129.
- Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr52.
- Lischer HEL and Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28: 298-299.
- Lo Brutto, S., T. Maggio & M. Arculeo, 2013. Isolation by distance (IBD) signals in the deep-water rose shrimp *Parapenaeus longirostris* (Lucas, 1846) (Decapoda, Panaeidae) in the Mediterranean sea. *Marine Environmental Research* 90:1-8 doi:10.1016/j.marenvres.2013.05.006.
- Maggio, T., S. Lo Brutto, F. Garoia, F. Tinti & M. Arculeo, 2009. Microsatellite analysis of red mullet *Mullus barbatus* (Perciformes, Mullidae) reveals the isolation of the Adriatic Basin in the Mediterranean Sea. *ICES Journal of Marine Science* 66(9):1883-1891 doi:10.1093/icesjms/fsp160
- Mastretta-Yanes, A., N. Arrigo, N. Alvarez, T. H. Jorgensen, D. Piñero & B. C. Emerson, 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* 15(1):28-41 doi:10.1111/1755-0998.12291.
- Marcias, S., F. Sacco, A. Cau & R. Cannas, 2010. Microsatellite markers for population genetic studies of the giant red shrimp *Aristaeomorpha foliacea* (Crustacea, Decapoda). *Rapp Comm int Mer Médit* 39:386.

- Matic-Skoko, S., T. Segvic-Bubic, I. Mandic, D. Izquierdo-Gomez, E. Arneri, P. Carbonara, F. Grati, Z. Ikica, J. Kolutari, N. Milone, P. Sartor, G. Scarcella, A. Tokac & E. Tzanatos, 2018. Evidence of subtle genetic structure in the sympatric species *Mullus barbatus* and *Mullus surmuletus* (Linnaeus, 1758) in the Mediterranean Sea. *Sci Rep* 8:14 doi:10.1038/s41598-017-18503-7
- Meirmans, P. G. & P. H. Van Tienderen, 2004. genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* 4(4):792-794 doi:10.1111/j.1471-8286.2004.00770.x.
- Milano, I., M. Babbucci, F. Panitz, R. Ogden, R. O. Nielsen, M. I. Taylor, S. J. Helyar, G. R. Carvalho, M. Espiñeira, M. Atanassova, F. Tinti, G. E. Maes, T. Patarnello & L. Bargelloni, 2011. Novel tools for conservation genomics: Comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. *PLoS ONE* 6(11) doi:10.1371/journal.pone.0028008
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16(3):1215.
- Miller, J. M., R. M. Malenfant, P. David, C. S. Davis, J. Poissant, J. T. Hogg, M. Festa-Bianchet, and D. W. Coltman. 2013. Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* 112:240.
- Mora-Márquez F, García-Olivares V, Emerson BC, López de Heredia U. ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. *Mol Ecol Resour.* 2017;17(2):230-246. doi:10.1111/1755-0998.12550
- Paris, J. R., J. R. Stevens & J. M. Catchen, 2017. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution* 8(10):1360-1373 doi:10.1111/2041-210X.12775.
- Peakall, R. & P. E. Smouse, 2012. GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537-2539 doi:10.1093/bioinformatics/bts460.
- Pritchard, J. K., M. Stephens & P. Donnelly, 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155(2):945.
- Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8(1):103-6 doi:10.1111/j.1471-8286.2007.01931.x.
- Zheng, X. W., D. Levine, J. Shen, S. M. Gogarten, C. Laurie & B. S. Weir, 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24):3326-3328 doi:10.1093/bioinformatics/bts606.

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP1 - Population genetics and phylogeographic studies for identification of biological units of priority species

Deliverable 1.5.1 – Data acquired from comprehensive genetic studies for *M. merluccius*, *A. antennatus*, and *A. foliacea*

LEAD: COSTAS TSIGENOPOULOS (HCMR) & LORENZO ZANE (CONISMA)

PARTICIPANTS: LUCA BARGELLONI, RITA CANNAS, ALESSIA CARIANI, LAURA CARUGATI, LEONARDO CONGIU, RACHELE CORTI, ALICE FERRARI, RAFAELLA FRANCH, SILVIA IORI, FRANCESCO MAROSO, ILARIA ANNA MARIA MARINO, RICCARDO MELIS, CHIARA PAPETTI, FEDERICA PIATTONI, ELISABETTA PIAZZA, FAUSTO TINTI (CONISMA)

DIMITRIS TSAPARIS, VASSO TERZOGLU, TEREZA MANOUSAKI, KATERINA EKONOMAKI, PETROULA BOTSIDOU, XENIA SARROPOULOU (HCMR)

Date: March 15th, 2021

Contents

Executive Summary	3
Objective	5
Methodology	5
Sampling	5
DNA extraction	5
Library Preparations & Genotyping-By-Sequence (GBS)	5
Bioinformatic Analysis	6
Population genetic analyses	6
Overall Evaluation & Phylogeographic Assessment	7
<i>Aristaeomorpha foliacea</i>	7
<i>Aristeus antennatus</i>	19
<i>Merluccius merluccius</i>	30
Conclusions	41
References	43

Executive Summary

The Deliverable 1.5.1 refers to the results of Task 1.4 in which, for three out of the six target species, the totality of the locations sampled throughout the Mediterranean Sea was genetically analyzed. The total number of collected specimens from each species was DNA extracted, and after taken into account their DNA quality, the best specimens were included in double-digest Random Amplified DNA (ddRAD) libraries construction and analyzed (genotyped) on the HiSeq4000 Illumina sequencing platform.

In general, the ddRADseq protocol performed better in the fish species (*M. merluccius*) while we faced some difficulties for the two crustacean species, for which little genomic information exists and no previous attempts have been reported.

For the giant red shrimp (*Aristaeomorpha foliacea*), the complete study was conducted with specimens from 32 localities (GSAs and sub-GSAs). In total, 1,692 specimens were sampled from the Balearic Islands to the Cypriot and Egyptian waters. Unfortunately, DNA quality was poor for some samples which finally led us to exclude 21% of them from library preparations and downstream analyses. Using all individuals genotyped, the species "catalogue" comprised 2,393,590 ddRAD loci, and filtering out samples with low number of stack loci (<4,000) and for SNPs present in at least 70% of the samples, we ran all phylogeographic analyses with a dataset composed of 771 samples (for 30 localities) and 443 higher quality SNPs. Our results point out an evident lack of genetic differentiation and are generally in agreement with previous studies conducted at smaller geographic scales and less extended sampling points in the Mediterranean Sea. Therefore, using the ddRADseq approach which provided lots of hundreds of polymorphic markers from an extensive sampling plan from the West (Balearic islands) to the East of the Mediterranean Sea (Cyprus and Egypt), we measured very low pairwise population differentiation metrics (F_{st} values), the absence of clear population structure in the Mediterranean, and the greatest part of the identified genetic variation attributed to differences among individuals in the populations, and much less among groups; in particular, the highest values encountered are found significant for the three major groups of the Western, the Central and the Eastern Mediterranean.

For the blue and red shrimp (*Aristeus antennatus*), a total of 1,471 specimens were sampled from 31 sampling locations. Quality of DNA extracts was poor and a total of 1043 individuals were used for ddRAD. After filtering, a total of 1253 SNPs were retained for 886 samples. All samples showed a great similarity between them with very small F_{st} values, though the global F_{st} was statistically significant. Unsupervised genetic clustering, showed slight differentiation, especially when comparing samples from Western and Eastern Mediterranean. As for *Aristaeomorpha foliacea*, much of the genetic variation was distributed among individuals in the populations, with a slight support for three groups corresponding to Western Mediterranean Sea, Central Mediterranean Sea, and Eastern Mediterranean Sea. Grouping samples based on Mediterranean basins, and measuring the genetic differentiation between groups of population samples using F_{ct} metrics, homogeneity was found among samples from Balearic, Tyrrhenian and Ionian Sea, while a more pronounced structure was detected among samples from Adriatic, Aegean, Levantine seas. In brief, results confirm a very low genetic differentiation among Mediterranean samples detected in published studies, but highlight an underlying significant structure due to the higher power of the markers used. The existence of significant genetic differentiation between Western (GSA01 to GSA12), Central (GSA13 to GSA20) and Eastern (GSA22 to GSA26) Mediterranean Sea, should be taken into account in the future and additional investigations should aim to further elucidate differences at the small scale.

In *Merluccius merluccius*, a total of 1,728 samples (from 41 sampling locations) were genotyped. A total of 665 high quality SNPs was retained (for 1,667 samples) and used for all downstream differentiation analyses. DNA quality was good for most samples and finally about 90% of the samples were used in library preparations. Comparisons between samples showed a significant genetic differentiation. The strongest differentiation was between Atlantic and Mediterranean samples, but also within the Mediterranean populations were structured following a West to East pattern. Analysis identified four groups corresponding to main subdivisions, the first including the sample from the Atlantic Ocean, the second including samples from the Western Mediterranean Sea, the third including samples from the Central Mediterranean Sea, and the fourth one including samples from the Eastern Mediterranean Sea. In addition, arranging population samples in 8 groups according to their geographic origin from West-to-East, provided highly significant evidence of geographic subdivision. Therefore, *M. merluccius* full study provides interesting results and the finding of significant genetic structure within Mediterranean represents a significant advance respect to previous published studies, where such differences were detected only using outlier genetic loci. The differentiation between Western (GSA01 to GSA12), Central (GSA13 to GSA20) and Eastern (GSA22 to GSA27) Mediterranean samples is now fully supported, and additional differences should be taken into account and monitored, particularly in the Central and Eastern Part of the basin.

The application of ddRADseq to this comprehensive genetic study of two aristeid decapod crustacean species (*Aristaeomorpha foliacea* and *Aristeus antennatus*) and one fish (*Merluccius merluccius*) evidenced a much lower level of differentiation of the two crustaceans, that is in line with previous studies. This imply that to obtain solid support for these crustaceans a very high number of loci and a big sample is needed. Our study evidenced two critical steps that can increase the efficiency of the ddRADseq for crustaceans, the first related to DNA quality and the second to the optimization of the protocol for library preparation and sequencing. For DNA quality, a particular care should be taken during sample collection and preservation, to reduce the number of DNA extracts not suitable to ddRAD sequencing. For the optimization of the protocol for library preparation and sequencing we found that particular care should be devoted to DNA extraction itself, selection of restriction enzymes, size-selection of fragments for high throughput sequencing, level of multiplexing, and/or filtering for bioinformatic analysis. We advise, for future applications of ddRADseq in new species, particularly crustaceans, to perform a small scale optimization study to address these issues in advance.

Objective

The main objectives of this deliverable is to report final results for three out of the six species included in the study: for the European hake (*Merluccius merluccius*), the blue and red shrimp (*Aristeus antennatus*) and the giant red shrimp (*Aristaeomorpha foliacea*). Based on the knowledge acquired from the pilot genetic study (see D1.4) in which we optimized the laboratory protocols and bioinformatic pipelines, we processed the totality of the specimens sampled for each species across the Mediterranean Sea, performed an unprecedented biogeographic analysis, always comparing the results for each species to those of previous genetic studies.

Methodology

A Genotyping-by-Sequencing (GBS) methodology constructing reduced-representation libraries in each species was selected. SNPs markers newly isolated following the double-digest Random Amplified DNA (ddRAD) sequencing are used for the first time in these six species and are one of the rare example of application in Mediterranean marine species (Maroso et al., 2018) . Full description of the methodology has been provided in the Inception, the 1st Progress Report, and deliverable 1.4.

Sampling

The number and location of sampling sites differed per species analysed and are detailed in the following sections and in the Tables 1.5.1.1, 1.5.1.6 and 1.5.1.12.

DNA extraction

Standard DNA salt extraction protocols were followed and further details on the methodology can be found in the Deliverable 1.4, included in the 2nd Progress Report (Annex 4). In brief, even though alternative extraction protocols based on commercial kits have also been employed, finally the salt extraction protocol (rather cheap but significantly time and labor-consuming approach) was selected to get high molecular weight (HMW) DNA. DNA extracts have been classified into three (3) quality categories ranging from 1 (high quality to be used in priority for ddRAD) to 3 (very poor quality and not usable for ddRAD). Since the pilot study, for sampled sub-GSAs which were previously identified of poor quality samples, we succeeded to collect new specimens that were included in this analysis.

Library Preparations & Genotyping-By-Sequence (GBS)

Having evaluated and optimized the digestion with the appropriate pair of enzymes for each species, for the crustaceans we chose the enzymes *SbfI* -*NlaIII* (8-base cutter & 4-base cutter, respectively) and *PstI* – *NlaIII* (6-base cutter & 4-base cutter, respectively), while for the fish species the pair *SbfI*-*SphI* (8-base cutter & 6-base cutter, respectively).

Details on the concentration of enzymes used, amount of DNA, PCR-cycling conditions and gel cut-window chosen are provided in Milestone 1.3 available within the Interim Report (Annex 5).. However, based on preliminary results obtained during the pilot study, the size

selection procedure has been modified for *A. antennatus* to reduce the number of loci. In this case, rather than classical gel-based size selection, a Blue Pippin Automated Size Selection machine has been used, fixing average size of recovered fragments to 600 bp (range 543-657 bp). In addition, for this species only, multiplexing was reduced to 144 inds per library, instead of the original 288 multiplexing used in the pilot study, to increase read coverage.

Bioinformatic Analysis

Details on the methodology followed are provided in the deliverable D1.4. Briefly, the raw FASTQ files were quality-checked in FastQC 0.11.3 and sequenced reads from each of the species were analyzed separately using STACKS v.2.4 pipeline (Catchen et al. 2013), with the surviving high quality forward reads of each sample were used for building *de novo* the loci of that particular sample with the STACKS component *ustacks*. Then:

- a catalogue of loci was built for each species using the STACKS component *cstacks*,
- the loci of each individual were matched to the catalogue through *sstacks*,
- we used *gstacks* to assemble and merge the second read of each pair, call variant sites and identify the genotype of each sample for each catalogue locus,
- finally, *populations* was executed to produce the final input files for population genetic analysis.

For each STACKS locus, only one randomly selected SNP was kept.

Population genetic analyses

Details on the methodology followed are again provided in D1.4. Briefly, estimates of genetic diversity within samples in terms of observed (H_o) and expected heterozygosity (H_e) were calculated for each geographical sample with Genetix (Belkir et al., 1996-2004). The level of differentiation among samples, pairwise and global F_{ST} values were calculated with Arlequin ver. 3.5.1.2 and corrections for multiple testings were performed with the Benjamini–Yekutieli procedure (Benjamini and Yekutieli 2001). With the same software, we performed the Analysis of Molecular Variance (AMOVA), grouping the samples on the basis of *a priori* hierarchical geographical structure. The statistical significance of the resulting values was estimated by comparing the observed distribution with a null distribution generated by 10,000 permutations, in which individuals were redistributed randomly into samples. For each species a different threshold allowed for missing data was set. Specifically, for *M. merluccius* a 0.2 level of allowed missing data, while for *A. antennatus*, and *A. foliacea* a value of 0.3 was set, due to the quality of datasets.

The Discriminant Analysis of Principal Components (DAPC) was performed with Adegenet ver. 2.1.1 (R version 3.5.1, R Development Core Team, 2014; <http://www.r-project.org>), the DAPC scatterplots results were visualized graphically and the best supported number of clusters through comparison of the Bayesian Information Criterion (BIC) for the different values of K was identified.

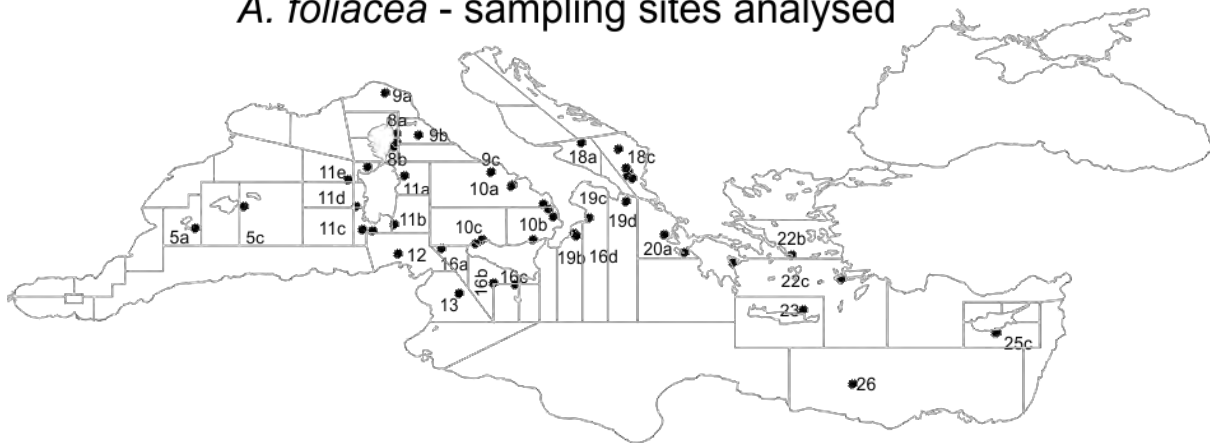
Finally, with the STRUCTURE v 2.3.4 software (Pritchard et al., 2000; Falush et al 2003, 2007; Hubiz et al., 2009) we evaluated the presence of clusters within the sampled populations, running the software with k from 1 to 10. When the most likely k was >1 based on likelihood evaluation, the best k value was selected using the EVANNO method (Evanno et al. 2005) as implemented in Structure Harvester (Earl & vonHoldt 2012).

Overall Evaluation & Phylogeographic Assessment

Aristaeomorpha foliacea

For the giant red shrimp (*Aristaeomorpha foliacea*), the complete study was conducted with specimens from 32 localities (GSAs and sub-GSAs). In total, 1,692 specimens were sampled from the Balearic Islands (5a to 5c) in the W. Mediterranean to the Cypriot (25c) and Egyptian (26a) waters in the E. Mediterranean.

A. foliacea - sampling sites analysed



DNA extraction was done as previously described and a little more than one third of the samples provided DNA of good quality (37.8%) whereas 41.4% of the samples were of medium quality and 20.9% were not used since the DNA quality was low (see Table 1.5.1.1).

Table 1.5.1.1. DNA quality across samples

GSA	Total number of samples	DNA quality			Number of samples in Libraries
		Good	Medium	Low	
5a	50	31	18	1	41
5bc	21	19	2	0	21
8a	35	30	5	0	35
8b	43	38	4	1	41
9a	50	30	20	0	50
9b	50	41	9	0	50
9c	50	33	12	5	45
10a	50	41	9	0	41
10b	50	38	7	5	44
10c	34	17	17	0	34
11a	50	44	6	0	50
11b	50	45	4	1	49
11c	100	58	18	24	74
11d	50	47	3	0	50
11e	50	50	0	0	50
12a	50	15	29	6	44

13a	50	2	30	18	31
16a	46	1	32	13	33
16b	50	0	38	12	38
16c	50	17	33	0	44
18a	50	2	45	3	46
18c	50	5	15	30	20
19b	100	4	31	65	27
19c	50	0	39	11	39
19d	50	10	15	25	25
20a	86	0	78	8	42
20b	50	0	0	50	0
22b	77	15	20	42	35
22c	50	0	43	7	43
23a	50	5	44	1	49
25c	50	1	43	6	44
26a	50	0	31	19	31
Sum	1692	639	700	353	1266

Specimens from all sampled GSAs were included in the ddRAD library preparation (Table 1.5.1.2) at various numbers, with the exception of GSA20b.

Table 1.5.1.2. Number of specimens (in total) in each library

1st library(pilot)	2nd library	3rd library	4th library	5th library
288 specimens	252 specimens	252 specimens	254 specimens	220 specimens
medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium quality DNA

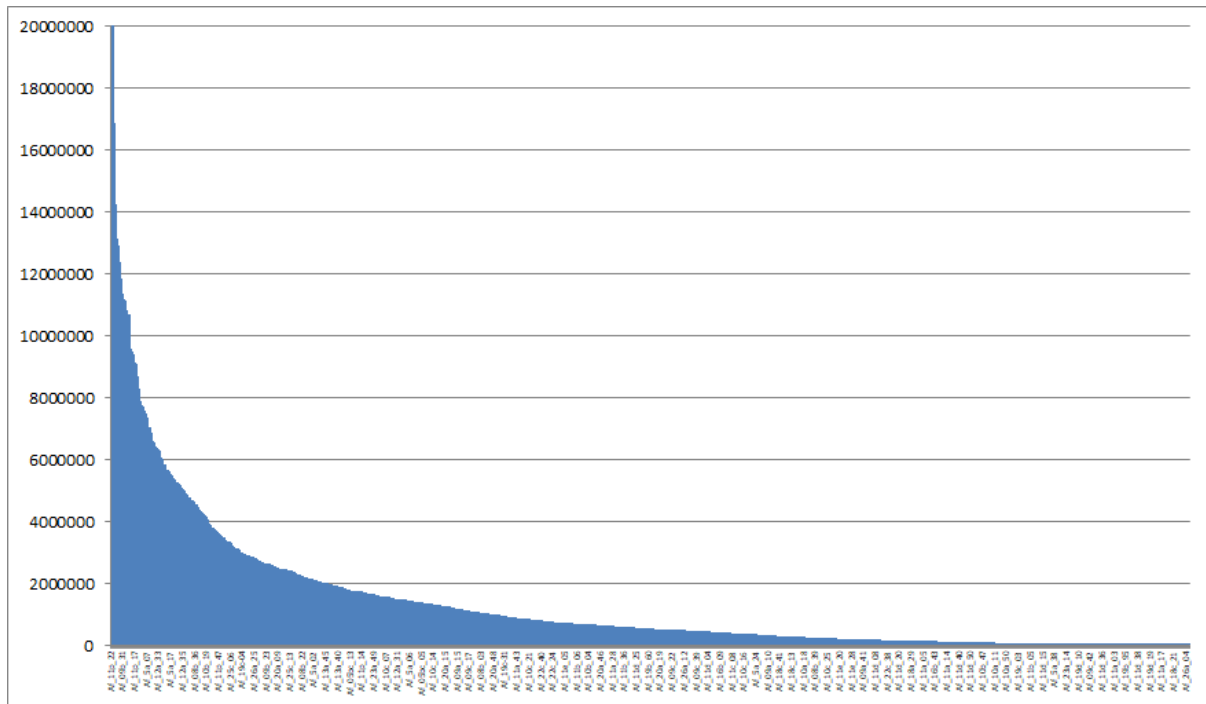
All samples were analyzed following a ddRAD protocol, with enzymes *SbfI* (CCTGCA[^]GG), and *NlaIII* (CATG[^]), selected fragments from around 320 to around 580 base pairs. Multiplexed libraries were run on six Illumina lanes (the first one twice), with 150 bp paired-end protocol. When R1 & R2 reads were merged, we obtained a total of 1.8 billion reads.

After demultiplexing, on average 1.3 million reads per sample were obtained (range from 3,534 to 23M). The number of Ustacks loci (considering a stack of unique reads) had an average of 32,334 but varied significantly among specimens (from 17 to 244,066) and among populations on average (1,217 in 19d to 69,625 in 09b) (see Figures 1.5.1.1, 1.5.1.2 and Table 1.5.1.3).

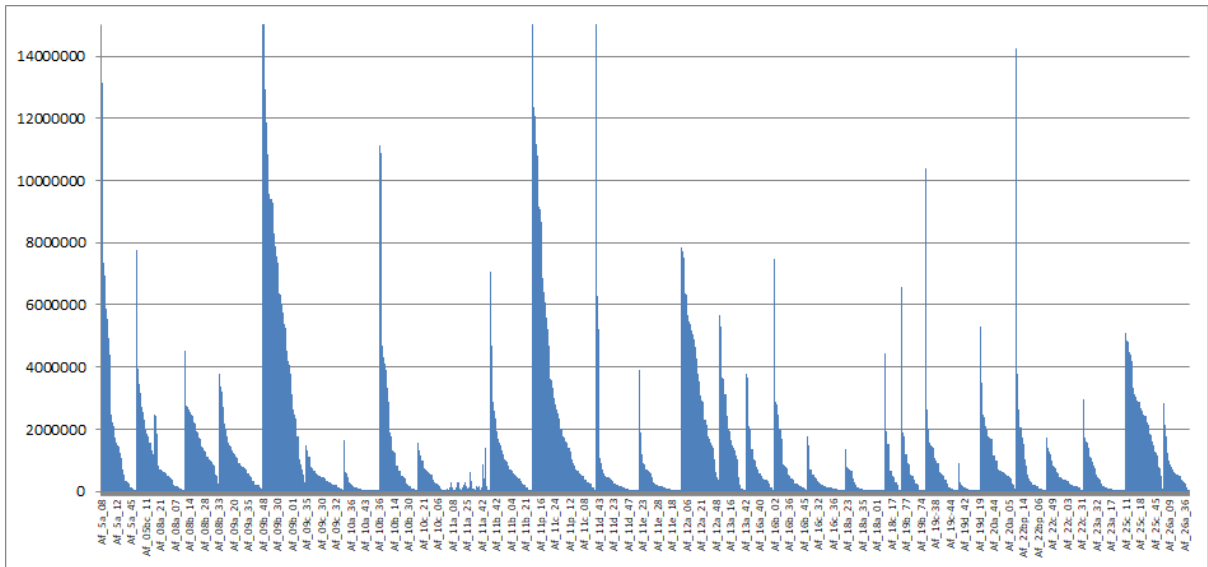
Table 1.5.1.3. Stacks pipeline output summary for *A. foliacea*

GSA	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
5a	2460520	8381	13142758	28844	233	142978
5bc	2591519	152412	7765332	31109	3465	64876
8a	594208	26102	2450446	10420	888	32904
8b	1632859	73735	4540023	10068	628	22501
9a	1084658	29741	3768066	16086	757	38955
9b	5998558	118275	21479914	69625	4387	189912
9c	468567	29845	1492995	3468	231	9349
10a	159245	4827	1655406	2346	105	16779
10b	1798248	15517	11144355	26697	355	122815
10c	521505	3534	1556675	9519	17	21729
11a	166997	4162	1385946	3540	103	20533
11b	1160431	17040	7054575	11864	483	44039
11c	3620539	22626	22716959	41279	619	244066
11d	1136426	11147	18509866	9320	154	105872

11e	434053	7875	3889785	9774	95	58774
12	3506866	265847	7857872	19058	3086	40654
13	1829004	7365	5666024	9049	88	23901
16a	934134	10675	3776392	4733	81	15440
16b	981017	25377	7485678	6411	226	35201
16c	288415	5997	1772331	7644	129	36807
18a	220441	8170	1363474	5661	125	29043
18c	839557	4273	4449794	5612	116	18751
19b	788396	7829	6581946	5573	120	23787
19c	1008130	12144	10392969	4850	147	33085
19d	125737	17492	906196	1217	190	5104
20a	1206295	80580	5283353	10759	1521	26750
22b	1137759	10421	14260181	10313	67	116674
22c	495420	19846	1729640	3794	184	11307
23a	487319	6680	2967236	4965	31	26183
25c	2433762	8123	5095978	14052	51	27605
26	723930	5462	2821372	4138	38	12911

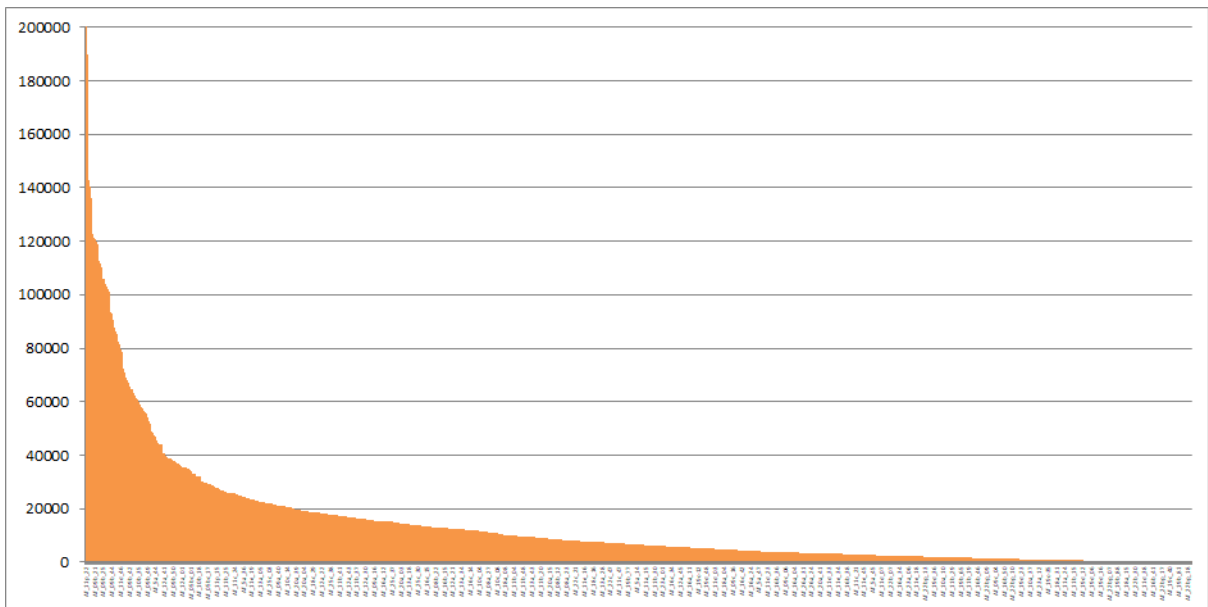


A)

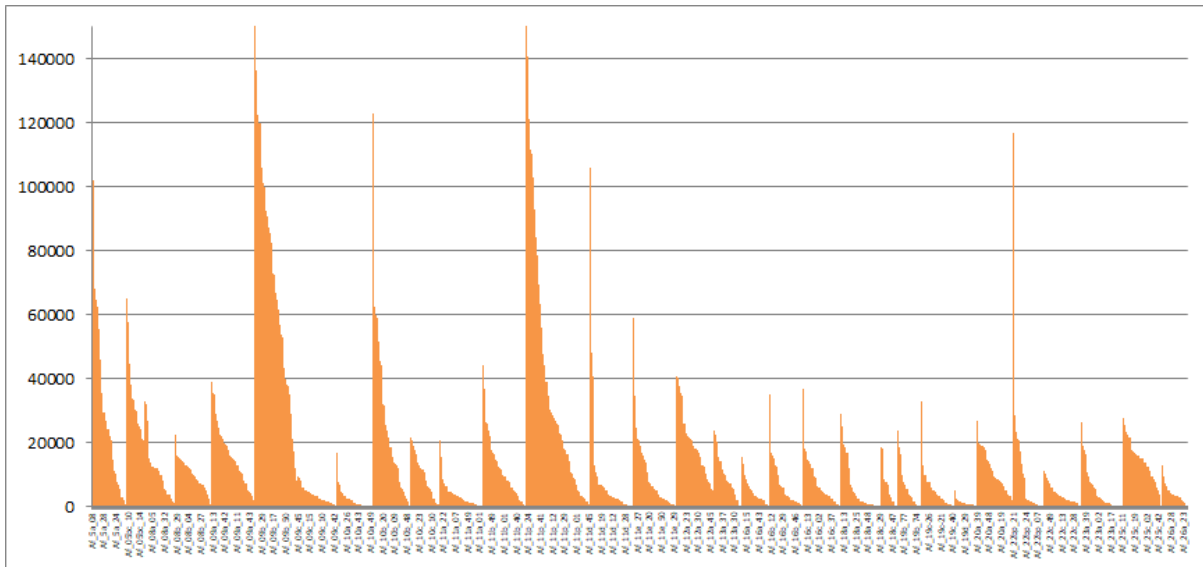


B)

Fig. 1.5.1.1: Retained reads after demultiplexing for (A) the whole *A. foliaceae* dataset, and (B) each population separately.



A)



B)

Fig. 1.5.1.2: ustacks loci for the (A) whole *A. foliaceae* dataset, and (B) each population separately.

Stacks results:

Using all individuals, we constructed the «catalogue» which comprised 2,393,590 ddRAD loci, a number which is considered huge, with the following characteristics:

- Number of loci with PE contig: 691.00 (100.0%);
- Mean length of loci: 365.94bp (stderr 1.30);
- Number of loci with SE/PE overlap: 343.00 (49.6%);
- Mean length of overlapping loci: 366.93bp (stderr 3.01); mean overlap: 25.66bp (stderr 0.32);
- Mean genotyped sites per locus: 371.56bp (stderr 2.93).

We removed 2,392,899 loci that did not pass sample/population constraints from the first number above (2,393,590), and we finally kept 691 loci, composed of 259,777 sites; 45,632 of those sites were filtered, 443 variant sites remained. Stacks' module *populations* and *vcftools* were used to filter the initial SNP panel. Only SNPs shared by at least 70% of the individuals and minor allele frequency larger than 0.1.

After several steps described in milestone M1.3, we proceeded forming two datasets: the first by excluding all samples with less than 4,000 stack loci ("relaxed" dataset with 772 samples), and a second excluding all samples with less than 6,000 loci ("stringent" dataset with 662 samples). We then run the analyses to filter out for SNPs present in at least 70% of the samples to finally have a panel of 734 and 443 higher quality SNPs, respectively, which were used for all downstream differentiation analysis. Below, we show only the results of the more "relaxed" analysis for the sake of simplicity.

Observed heterozygosities were generally low (ranging from 6.9 to 18.6%) and smaller than the expected (23.2 to 38.1%) (Table 1.5.1.4). This phenomenon is also present in *A. Antennatus* (see below). Low H_o values are found, not exclusively, in samples with low number of specimens which seem to have also a smaller number of polymorphic loci from the dataset of 443 SNPs. In the pilot study, observed heterozygosities were also low (but ranging from 9.7 to 22.8%); Interestingly, the expected heterozygosities were higher than those estimated in the final analysis (32.5 versus 28.1% on average).

Table 1.5.1.4: Summary of genetic variability estimates across sampling locations in *A. foliacea*. Reported are the average observed (Ho) and average expected heterozygosity (He).

Sub-GSAs	Samples in the Analysis	Ho	He
5a	33	0.1585	0.2407
5c	20	0.1858	0.2505
8a	25	0.1276	0.2516
8b	36	0.0821	0.2568
9a	46	0.1467	0.2409
9b	50	0.1724	0.2325
9c	15	0.0716	0.3301
10a	7	0.1616	0.3807
10b	37	0.1493	0.2369
10c	25	0.1267	0.2634
11a	15	0.1112	0.3093
11b	39	0.1154	0.2503
11c	62	0.1628	0.2414
11d	21	0.1209	0.2774
11e	28	0.1319	0.2529
12	43	0.1228	0.2442
13	23	0.0813	0.2694
16a	14	0.0728	0.3354
16b	17	0.0903	0.2880
16c	25	0.1243	0.2700
18a	16	0.1456	0.2856
18c	10	0.1186	0.3463
19b	12	0.1260	0.3150
19c	17	0.0692	0.3161
20a	35	0.1032	0.2558
22b	13	0.1730	0.2915
22c	15	0.0744	0.3217
23a	18	0.1215	0.2834
25c	41	0.0977	0.2485
26a	12	0.0706	0.3566

All populations showed a great similarity between them. This is evident in the F_{st} values which ranged from negative to 0.0475 (Fig. 1.5.1.3), using an uncorrected probability threshold of 0.05, and a Benjamini-Hochberg correction. Even when we tried to group subGSAs with a small number of specimens (Fig. 1.5.1.4) in order to form populations with larger numbers of individuals, the F_{st} values remained low and not significant.

The comparison of pairwise F_{st} results of final analysis (based on 443 polymorphic SNPs with less than 70% missing data per locus in the data set) to those from the pilot study (based on 3,437 polymorphic SNPs with less than 50% missing data per locus in the data set) indicate similar patterns of low differentiation. The six populations analysed in the pilot study (GSA09b, 10b, 11c, 11e, 16b, and 18a) had shown a great similarity between them (F_{ST} values from negative to 0.0069 between 11e and 18a). In the final set the F_{st} distances between the above-mentioned populations ranged from negative to 0.0115 (11e-18a).

Matrix of pairwise F_{ST}

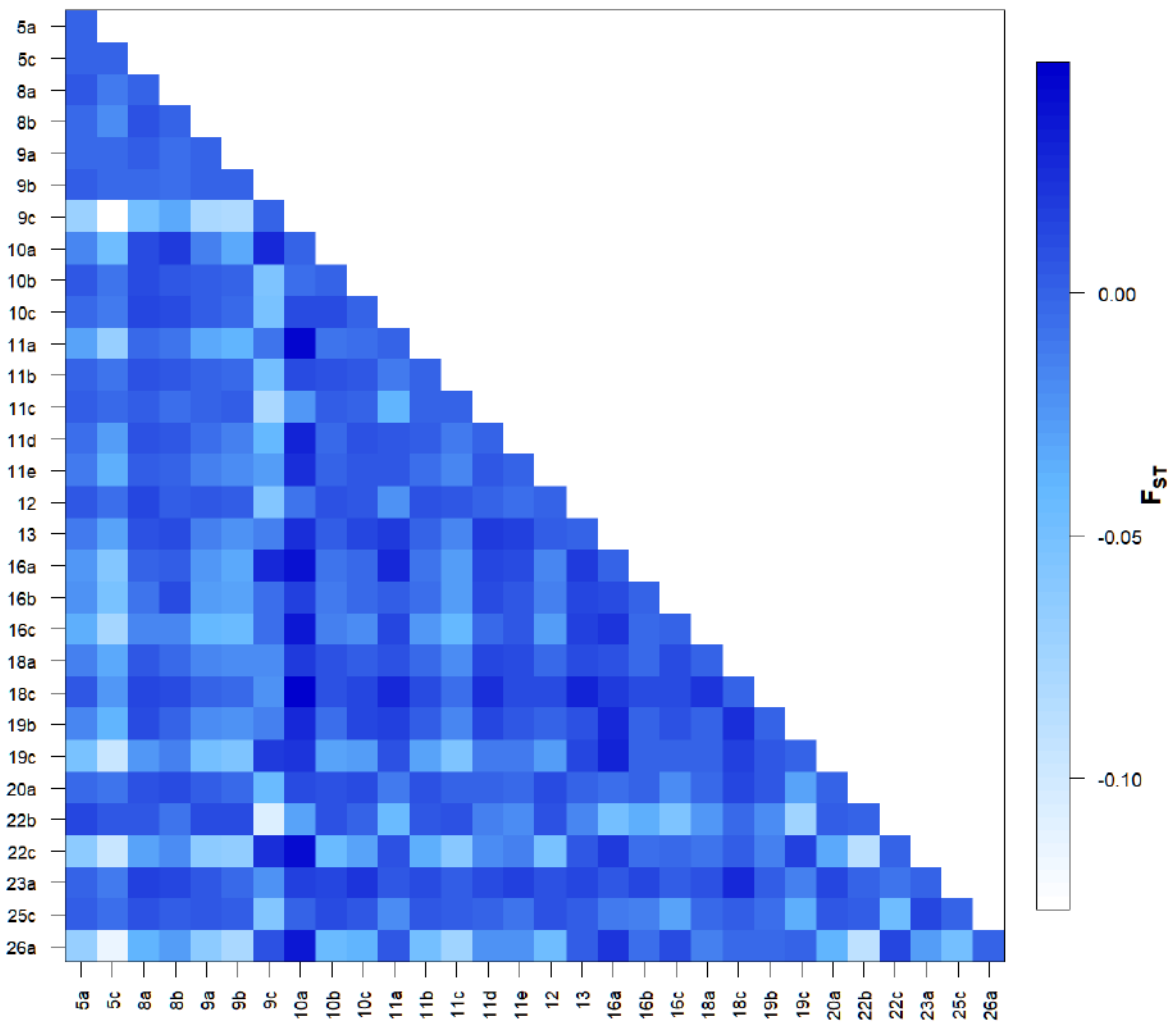


Fig. 1.5.1.3: Matrix of pairwise F_{ST} for *A. foliacea* based on 443 polymorphic SNPs present in >70% of the whole sampling set. No value was statistically significant after Benjamini-Hochberg correction.

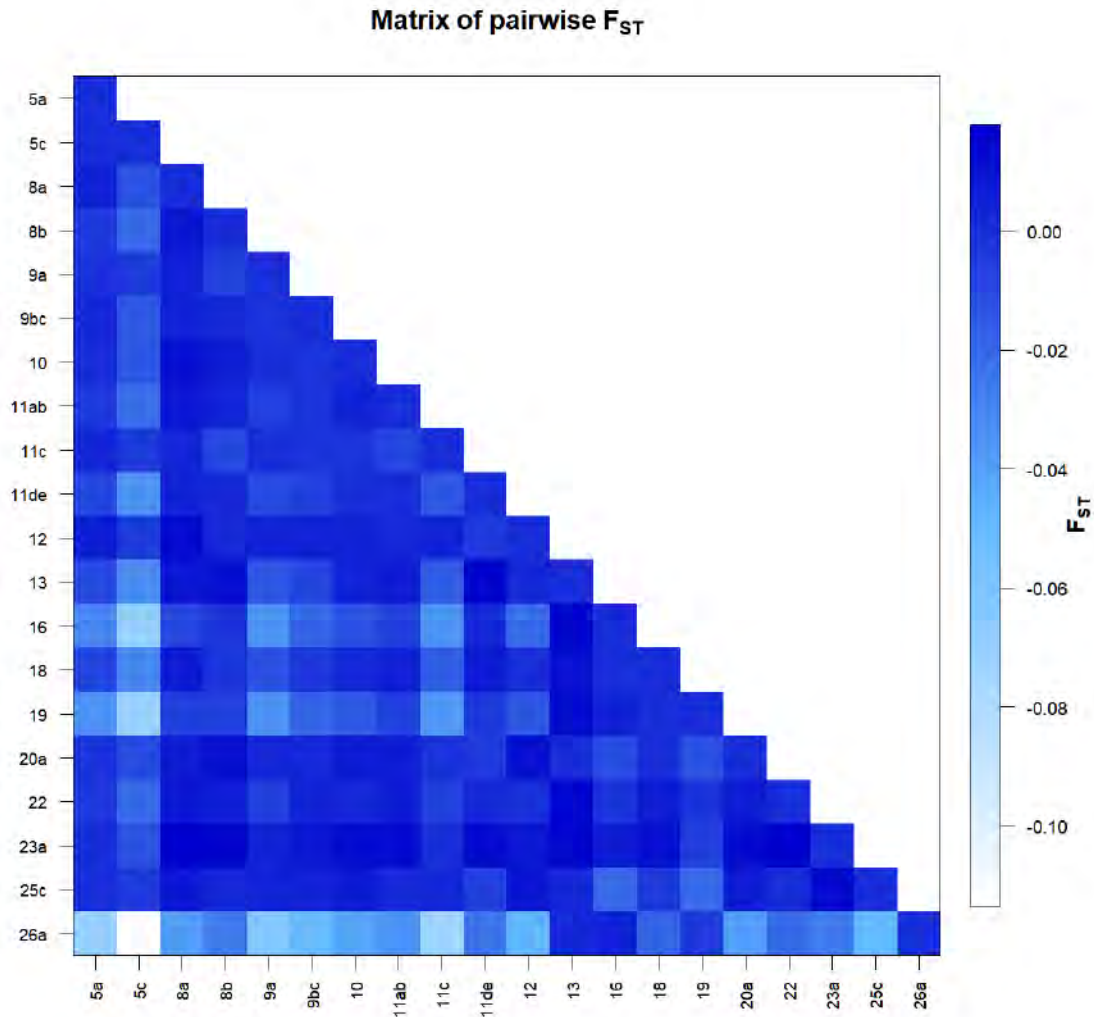


Fig. 1.5.1.4: Matrix of pairwise F_{ST} for *A. foliacea* based on 443 polymorphic SNPs present in >70% of the whole sampling set and after merging neighboring populations of small size. No value was statistically significant after Benjamini– Hochberg correction.

Moreover, AMOVA results (see Table 1.5.1.5) for several scenarios of grouping *A. foliacea* populations did not indicate any significant genetic structure but rather a panmictic situation in the Mediterranean Sea. AMOVA showed a small global $F_{st} = 0.01723$, that was statistically significant ($P < 0.0001$). Slightly lower global $F_{st} = 0.011$ was calculated after merging neighboring populations of low sample size.

In hierarchical AMOVA, three alternative scenarios of grouping populations based on Mediterranean basins, were tested: a) the first proposing two groups and a split between Western (up to the GSA20) and Eastern Mediterranean, b) the second proposing three groups including samples from the W. Mediterranean (up to the Siculo-Tunisian Strait, GSA12), the Central Mediterranean Sea (up to GSA20b and the Ionian Sea) and the Levantine Sea, and c) the third one proposing five groups including the Balearic (till GSA11e), the Tyrrhenian Sea (till GSA12), the Adriatic, the Ionian and the Aegean-Levantine samples from the Eastern Mediterranean Sea Table 1.5.1.5). All scenarios indicated the existence of very weak differentiation (statistically significant in scenarios 2 & 3) among proposed groups (FCT values 0.00007 to 0.00082).

Table 1.5.1.5: AMOVA results for several scenarios of grouping *A. foliacea* samples. In parentheses the values after merging neighboring populations (see Fig. 1.5.1.4)

	Scenario 1- Two groups	Scenario 2 – Three groups	Scenario 3 – Five groups
Group 1	West-Central 5a,5c,8a,8b,9a, 9b,9c,10a,10b,10c 11a,11b,11c,11d,11e, 12,13,16a,16b,16c, 18a,18c,19b,19c,20a	West 5a,5c,8a,8b, 9a,9b,9c,10a, 10b,10c,11a, 11b,11c,11d, 11e,12	Balearic 5a,5c,8a,8b, 9a,11c,11d, 11e
Group 2	East 22b,22c,23a,25c,26a	Central 13,16a,16b,16c, 18a,18c,19b,19c,20a	Tyrrhenian and Tunisia 9b,9c,10a,10b, 10c,11a,11b,12
Group 3		East 22b,22c,23a,25c,26a	Ionian 13,16a,16b,16c, 19b,19c,20a
Group 4			Adriatic 18a,18c
Group 5			Aegean-Levantine 22b,22c,23a,25c,26a
Variation %			
Among groups	0.006 (0.082)	0.082 (0.124)	0.007 (0.052)
Among populations within groups	1.722 (1.085)	1.680 (1.040)	1.718 (1.065)
Within populations	98.272 (98.833)	98.238 (98.837)	98.275 (98.883)
Fixation Indices			
FST	0.01728 (0.01167)	0.01762 (0.01163)	0.01725 (0.01116)
p-value	<0.001 (<0.001)	<0.001 (<0.001)	<0.001 (<0.001)
FSC	0.01722 (0.01086)	0.01682 (0.01041)	0.01718 (0.01065)
p-value	<0.001 (<0.001)	<0.001 (0.021)	<0.001 (0.012)
FCT	0.00006 (0.00082)	0.00082 (0.00124)	0.00007 (0.00052)
p-value	0.05941 (0.003)	<0.001(<0.001)	<0.001 (0.001)

Such weak differentiation is evident both by STRUCTURE (Figure 1.5.1.5) and DAPC analysis (Figure 1.5.1.6). In STRUCTURE software, the best *k* value that was selected using the EVANNO method was 2 but with most individuals assigned to both clusters (admixed). In fact, STRUCTURE's results at higher values of *k* support the absence of genetic differentiation since the same pattern of assignment of each individual in more than one clusters is observed (Figure 1.5.1.5).

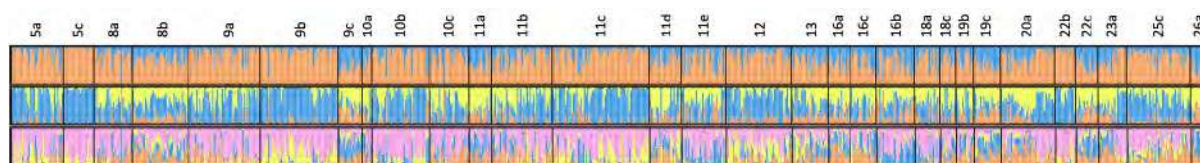


Fig. 1.5.1.5: Proportions of membership to each of K (2, 3 and 4) inferred clusters for *A. foliacea* individuals

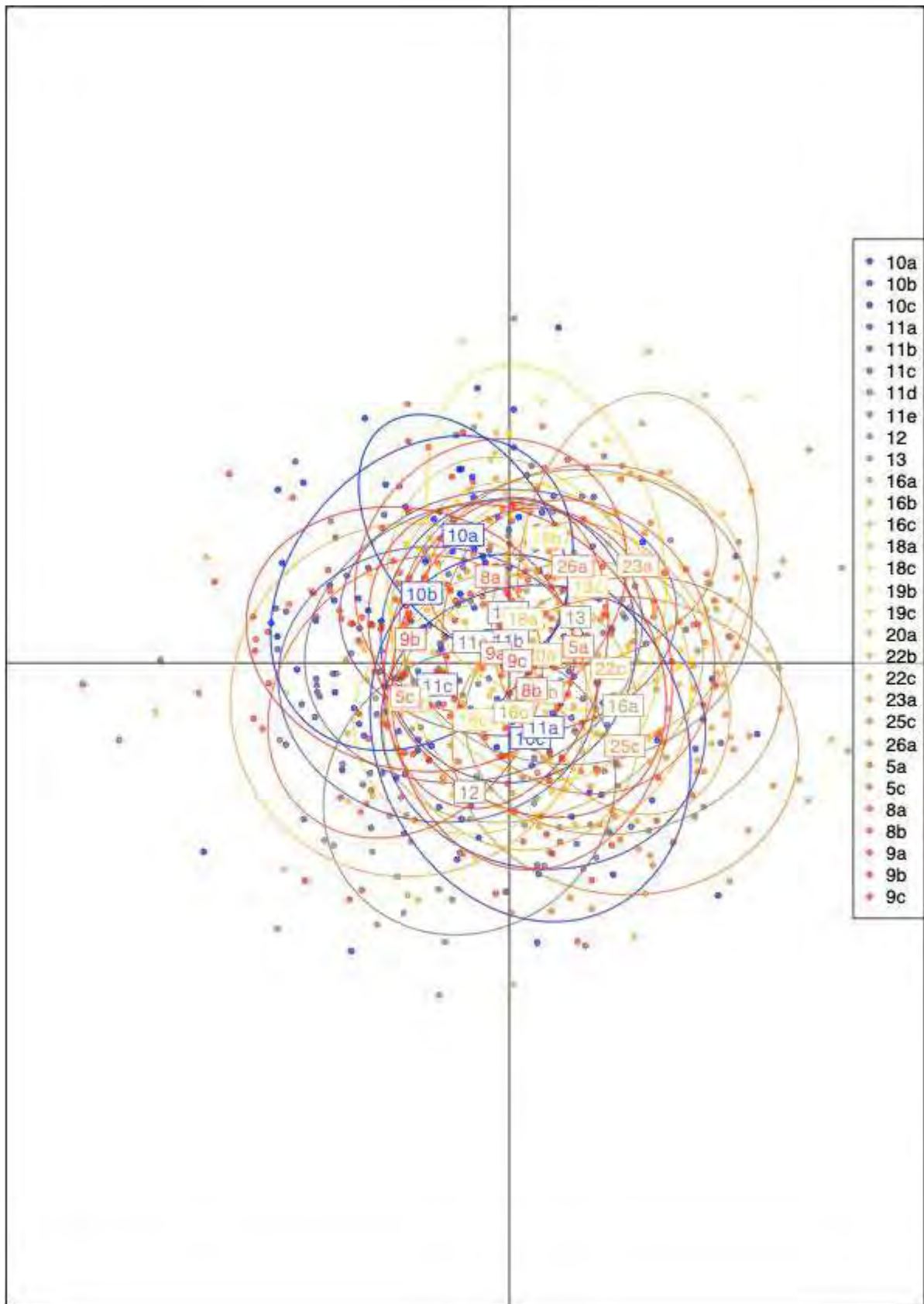


Fig. 1.5.1.6: DAPC plot for *A. foliacea* based on 443 polymorphic SNP genotyped in >70% of the filtered sampling set.

Current results refer to samples from different areas throughout the Mediterranean Sea where the species occurs and point out an evident lack of genetic differentiation; This has already been reported at smaller geographic scales and less extended sampling points (Cannas et al 2012; Marcias et al., 2010; Fernandez et al., 2011a; Fernandez et al., 2013b).

More particularly, Marcias et al. (2010) had already analysed 6 populations with 2 samples from W. Sardinia, 2 from the E. Sardinia (GSA11), and 2 samples off Sicily (Messina, GSA10 and Sicily Channel, GSA16) using 6 microsatellite loci. They reported a substantial genetic homogeneity and no signs of recent bottlenecks, suggesting the existence of a high gene flow connecting all populations with pairwise F_{st} values being low and not significant. Furthermore, AMOVA clearly showed that genetic variability was largely due to differences among individuals (99.31%) rather than to differences among populations (0.69%) as well as Structure analysis further confirmed this absence of population structure ($K=1$). The PCA performed with ADEGENET showed a substantial genetic homogeneity among populations separated by hundreds of kilometres suggesting that western Mediterranean populations could represent a unique panmictic stock. The results from this study are in full agreement with the results from the current ddRADseq approach, ie. Low pairwise F_{st} values, genetic variability largely due to differences among individuals in the populations rather than to differences among populations, and absence of population structure using the Structure analysis.

Fernandez et al (2011), sampled and analysed six locations across the Mediterranean Sea (and one location in Mozambique Channel in the Western Indian Ocean) by means of Inter Simple Sequence Repeats (ISSR) markers; One at the Balears (GSA5), east of Corsica (GSA8), two south of Sicily (GSA16), one in Ionian (GSA20), and one in the Aegean (GSA22). Average estimates of genetic diversity did not significantly differ among sampled localities, and heterozygosity values ($H = 0.105 \pm 0.015$) were low. Similarly, the AMOVA allocated > 98% of genetic variability to the within- sample component; However, the F-statistics value associated with the remaining part of variance was significant. Furthermore, when localities were grouped according to geographical regions (East versus West Mediterranean), no genetic variation was assigned to the among-region level, even when in the comparison we include the Mozambique channel. Last, the STRUCTURE cluster analyses did not detect geographically or genetically distinct groups. Replicate runs yielded consistent results and the uppermost hierarchical structure present in the entire dataset was detected for $K = 2$ by the K statistics, and the distribution of the two clusters across individuals did not show any geographical meaning.

Furthermore, in Fernandez et al (2013a) the same six samples as above were analysed with the addition of an extra Australian one using mitochondrial DNA (COI gene) sequences. Pairwise F_{ST} comparisons within the Mediterranean indicated no genetic differentiation between most locations ($F_{ST}=0$ to 0.004) with the exception of high values detected for the comparisons of Ionian-Aegean against the Balears-Tyrrhenian-Sicily ($F_{ST}=0.206$ to 0.382). This time significantly high molecular variance among all localities (85.0% of variance among samples, $\Phi_{ST}=0.850$, $P<0.005$) resulting from genetic differences among the Mediterranean Sea, Mozambique Channel and North Western Australia (92.3% of variance among regions, $\Phi_{CT}=0.923$, $P=0.036$). Within the Mediterranean, although at a lesser extent, molecular variance was also significant (16.9% of variance among samples, $\Phi_{ST}=0.169$, $P<0.005$), suggesting that a degree of genetic differentiation was present among the Mediterranean local samples. However, when Mediterranean samples were grouped in western and eastern basins, the variation among groups was not statistically significant (23.3%, $\Phi_{CT}=0.233$, $P=0.061$). Interestingly, from the four haplogroups (HG1-HG4) detected, two were restricted to the Mediterranean localities and with minor presence in the Mozambique Channel (HG1, HG2). Within the Mediterranean Sea, one of the haplogroups (HG1) was mostly presented in the most western samples (>75%), whilst the other haplogroup (HG 2) was mostly present in the East Mediterranean samples (>65% in the Ionian and the Aegean Seas).

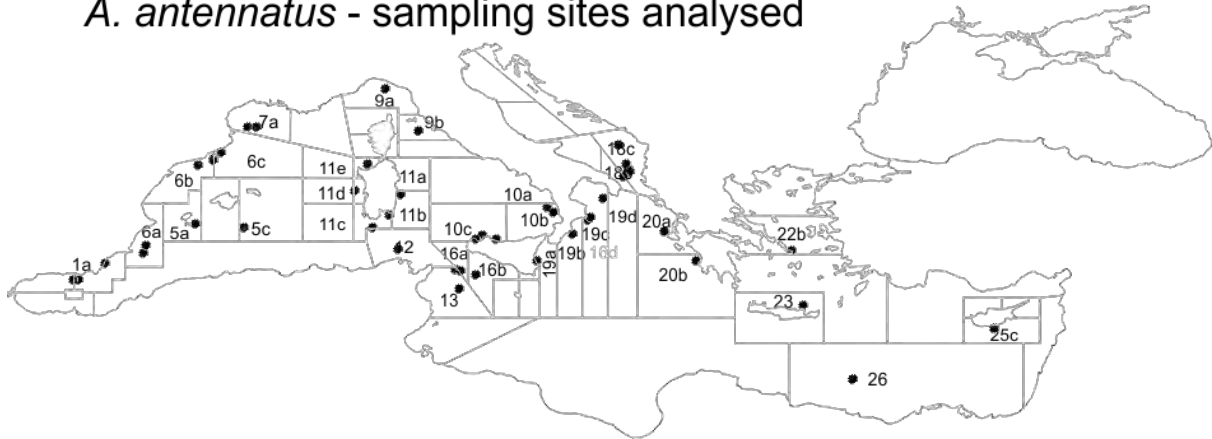
Finally, Fernandez et al (2013b) kept on analyzing some of the above samples using this time two nuclear genes (PEPCK and NaK) and one mitochondrial (COI) gene sequences to argue that two major phylogroups were detected. One group corresponds to Australia and the second includes the Mediterranean and the Indian Ocean sample, where MED appears monophyletic.

Therefore, our study is in accordance with results from all the above genetic studies aiming to reveal the biogeographic structure in *A. foliacea* in the Mediterranean Sea. In brief, using the ddRADseq approach we estimate very low pairwise F_{st} values in samples coming from the West (Balearic Islands) to the East of the Mediterranean Sea (Cyprus and Egypt). The greatest part of the genetic variability identified is attributed to differences among individuals in the populations (>98%), i.e. each sample or population is composed of specimens which are already quite differentiated between them and not more than to those specimens coming from other populations. Much less genetic variability (approx. 1.7%) is found among the 31 populations, and last when a number of groups is defined *a posteriori* some small but statistically significant variation appears (0.006 to 0.082) for some of them. The latter highest value for among groups variation (0.082) is found significant for the three major groups of the western, the central and the Eastern Mediterranean (see scenario 2 in Table 1.5.1.5). Last, the model-based Bayesian approach using STRUCTURE advocated for the absence of population structure in the Mediterranean.

Aristeus antennatus

For the blue and red shrimp (*Aristeus antennatus*), the complete study was conducted with specimens from 31 localities (GSAs and sub-GSAs). In total, 1471 specimens were sampled from the Alboran Sea (1a) in the W. Mediterranean to the Egyptian (26a) waters in the E. Mediterranean.

A. antennatus - sampling sites analysed



We encountered many problems related to the poor quality of DNA extracts. In particular, the majority of specimens from localities 16ab, 20a, 20b, 22b and 23a were not suitable for ddRAD analysis, thus reducing to few individuals the sample size for these GSA. Samples from GSA06b and 06c provided bad results in the pilot study DNA extractions and were resampled. ddRAD was performed on 1043 individuals including all the samples with good DNA quality and the majority of samples with medium DNA quality, some of which were excluded based on the insufficient concentration obtained from DNA extraction (Table 1.5.1.6).

Table 1.5.1.6: DNA quality across samples.

GSA	Total number of samples	DNA quality			Number of samples in Libraries
		Good	Medium	Bad	
01a	50	43	4	3	47
05a	50	19	25	6	38
05c	50	32	14	4	35
06a	49	48	1	0	47
06b pilot*	(50)	0	15	35	14
06c pilot*	(50)	1	15	34	15
06b final*	50	37	6	7	43
06c final*	50	33	14	3	42
07a	50	32	9	9	41
09p	50	17	20	13	36
09b	25	15	6	4	18
10a	50	41	6	3	45
10bc	37	34	2	1	36
11ab	50	30	7	13	35
11c	50	46	2	2	48
11d	50	45	4	1	48
11e	50	38	5	7	41
11p	50	26	12	12	38
12a	50	14	19	17	24
13a	50	11	25	14	23
16ab	50	2	6	42	8
18b	50	30	13	7	43

18c	35	14	18	3	32
19a	50	19	21	10	35
19b	50	38	11	1	49
19c	50	32	14	4	46
19d	50	33	11	6	43
20a	25	2	15	8	14
20bp	54	3	8	43	5
22b	12	1	7	4	5
23a	33	9	11	13	15
25c	1			1	-
26a	50	21	23	6	34
Sum	1471	766	369	336	1043

*samples from GSA06b and 06c provided bad results in the DNA extractions for the pilot study and were resampled

Specimens from all sampled GSAs were included in the ddRAD library preparation at a various number, with the exception of GSA25c which comprised a single individual (Table 1.5.1.7).

Table 1.5.1.7. Number of specimens (specim.) in each library. Totals include replicated control samples.

1st library (pilot)	1st-bis library	2nd library	3rd library	4th library	5th library	6th library	7th library	8th library
273 specim.	144 specim.	144 specim.	144 specim.	144 specim.	144 specim.	144 specim.	144 specim.	48 specim.
Medium to bad quality DNA	medium to bad quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA

All the samples were genotyped following a ddRAD protocol, with enzymes *PstI* and *NlaIII*. Following the low coverage in the first library used for the pilot study, the size selection window was optimized to reduce the number of loci (library 1-bis). Adapter-ligated fragments from around 543 to around 657 bp (which correspond to genomic fragments from around 420 to around 540 bp) were selected using a Blue Pippin Automated Size Selection machine, respect to the previous gel-based size selection which targeted genomic fragments from around 200 to around 480 bp. In addition, the number of multiplexed individuals per library was reduced to 144. Due to the change of size selection and level of multiplexing, the results obtained from the first library were discarded and the corresponding individuals have been included in libraries 1-bis to 8.

ddRADseqTool (Mora-Marqu ez et al. 2017), a software that allows the generation of in silico double-digested fragments to optimize ddRADseq experiments, was used to simulate a ddRAD library using the same enzyme pair and the same size selection used with the target species, to estimate the number of retrieved tags. For the analysis, a fish genome was used, whose length was 831 Mbp. A total of 45,000 tags of size between 420 and 540 bp were estimated. Considering a genome size up to 5 times bigger for *A. antennatus*, a total number of fragments around 200,000-250,000 is expected.

Multiplexed libraries were run on eight Illumina lanes, with 150 bp paired-end protocol. R1 R2 reads were merged, obtaining a total of 2,090 billion reads in R1+R2.

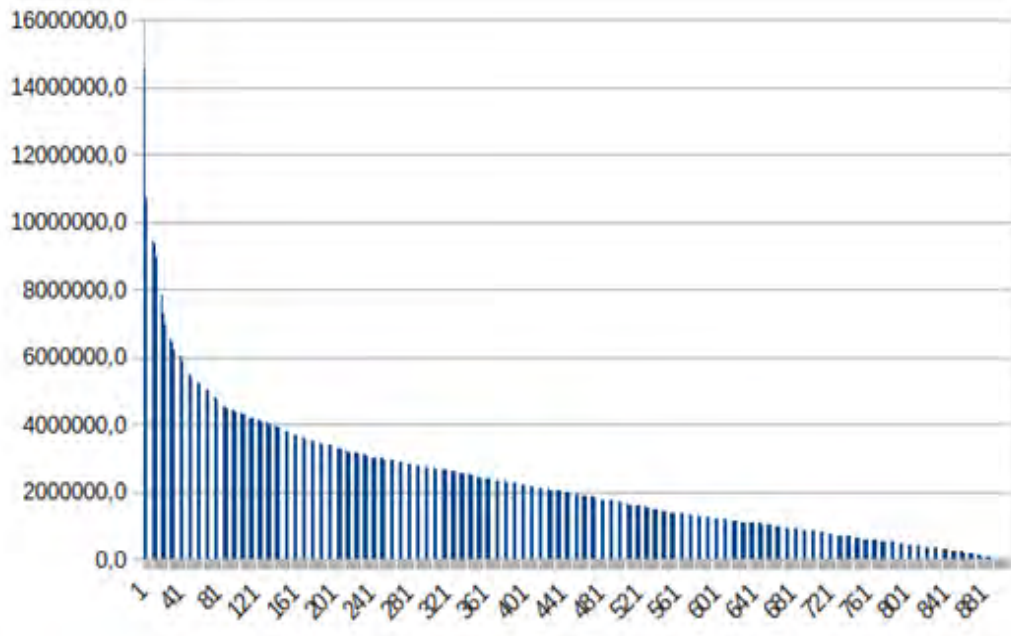
After demultiplexing, on average 2.299 M reads per sample were obtained, with a range from 877 to 15M (Table 1.5.1.8).

After eliminating low quality reads and reads without a RAD site at the beginning, on average 95,3% reads were retained (figure 1.5.1.7).

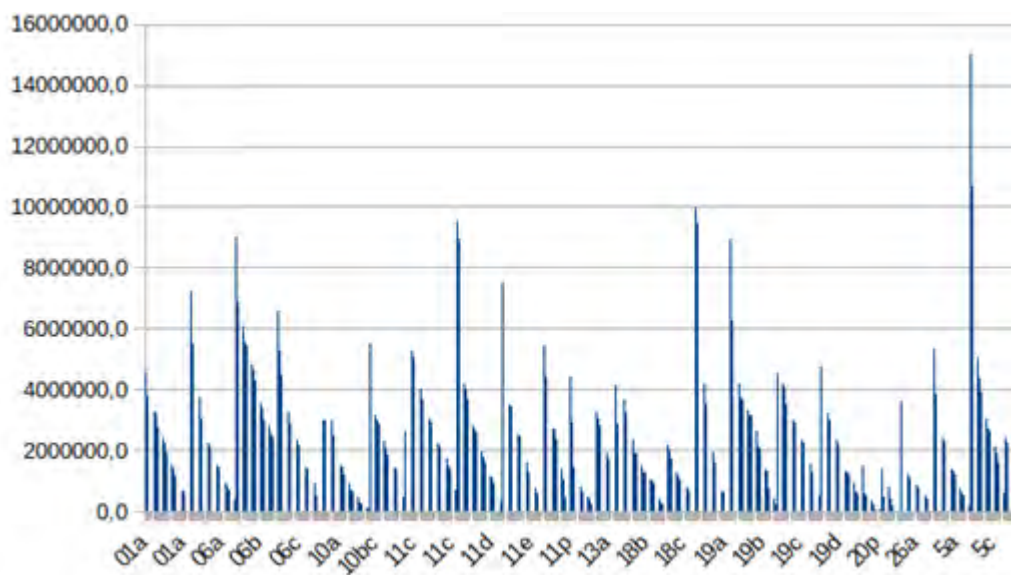
Samples were processed with *ustacks* considering a stack of unique reads only if at least 3 identical reads were found and allowing for up to 3 mismatches to consider two tags as part of the same locus. On average 166'177 tags were found in each sample, with average coverage of 9.8X (Table 1.5.1.8). A tag catalog was built with 60 samples (the two or three with the highest number of reads from each population). A maximum of 3 mismatches were allowed to merge tags into the same catalog locus. The final number of tags in the catalog is around 5.4M.

Table 1.5.1.8: Stacks pipeline output summary for *Aristeus antennatus*

GSA	Sample	Retained reads				Stacks loci			
		Total	Average	MIN	MAX	Total	Average	MIN	MAX
01a	47	99097457	2108456	265158	5189623	7177987	152723	31689	277276
05a	38	65066378	1712273	5249	5339027	4834414	127221	643	284791
05c	35	141503968	4042970	578182	15027996	8706645	248761	62167	665508
06a	47	113430630	2413417	330223	8801426	7604397	161795	37291	571763
06b	14	194823337	4530775	2066419	9590313	12265297	285239	146088	626969
06c	15	115785612	2756800	499721	7837878	7944750	189160	55668	361199
06b	43	194823337	4530775	2066418	9590312	12265297	285239	146088	626969
06c	42	115785612	2756800	499721	7837878	7944750	189160	55668	361199
07a	41	436526	72754	877	251817	55385	9230	16	31933
09p	36	26482226	1471234	258606	2923770	2136649	118702	31422	203850
09b	18	11354266	2838566	2360661	2997871	775678	193919	183091	197530
10a	45	69131597	1536257	5961	13579196	5540166	123114	475	760965
10bc	36	78792020	2188667	376366	5470680	5646206	156839	42790	306688
11ab	35	8665012	1733002	199713	2652170	646222	129244	25329	187606
11c	48	140373689	2924451	709823	6670584	9140800	190433	73383	359798
11d	48	137586758	2866390	110152	9513636	8814789	183641	13304	502559
11e	41	93852440	2289083	46881	7475907	6444588	157185	5130	393617
11p	38	73922407	2549048	38568	6929932	5141530	177294	4964	358567
12a	24	26579616	1107484	20414	4396554	2355994	98166	2395	325167
13a	23	60262557	2620111	291079	10738131	4599886	199995	33159	534109
16ab	8	21733241	2716655	1938628	4133802	1540356	192544	155278	248958
18b	43	67358977	1566487	6679	3925665	5566417	129451	715	256525
18c	32	39030551	1219704	92776	3373825	3334380	104199	11818	227176
19a	35	103261815	2950337	105599	9992151	7208768	205964	11899	544328
19b	49	146098647	2981605	205262	9401064	9777325	199537	25137	533344
19c	46	122223363	2657029	318102	6525343	8293740	180298	37223	340363
19d	43	85904429	1997777	60382	4758045	6189788	143948	7599	285090
20a	14	7051548	503682	63631	1506627	739141	52795	7107	139084
20bp	6	371834	92958	1212	175894	43902	10975	11	21578
22b	5	2874825	574964	1968	1370639	289684	57936	107	120316
23a	15	5501003	366733	1280	2081907	532490	35499	9	162504
25c	-								
26a	34	31293913	920409	51492	3606831	3020797	88847	5830	268985



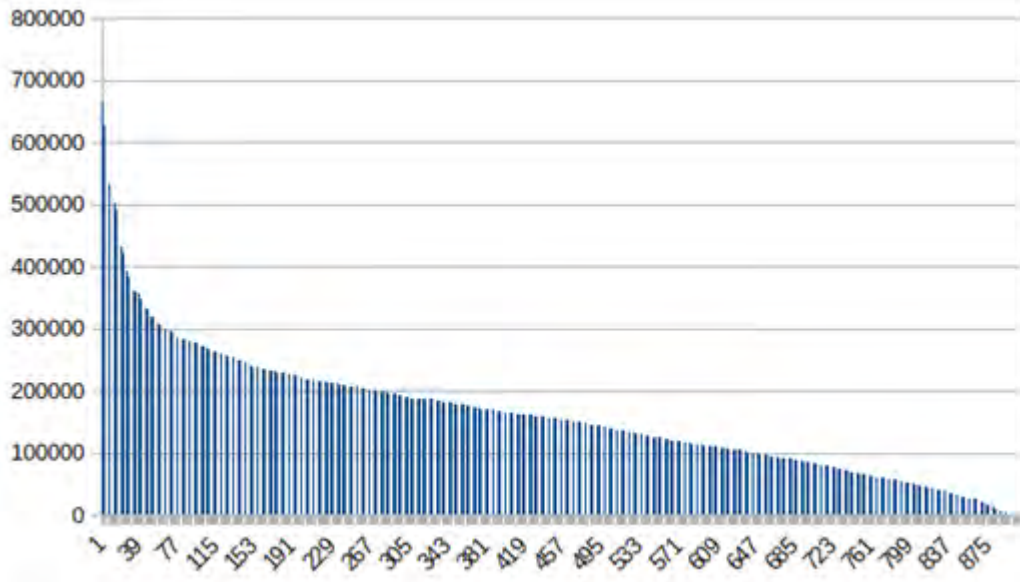
A)



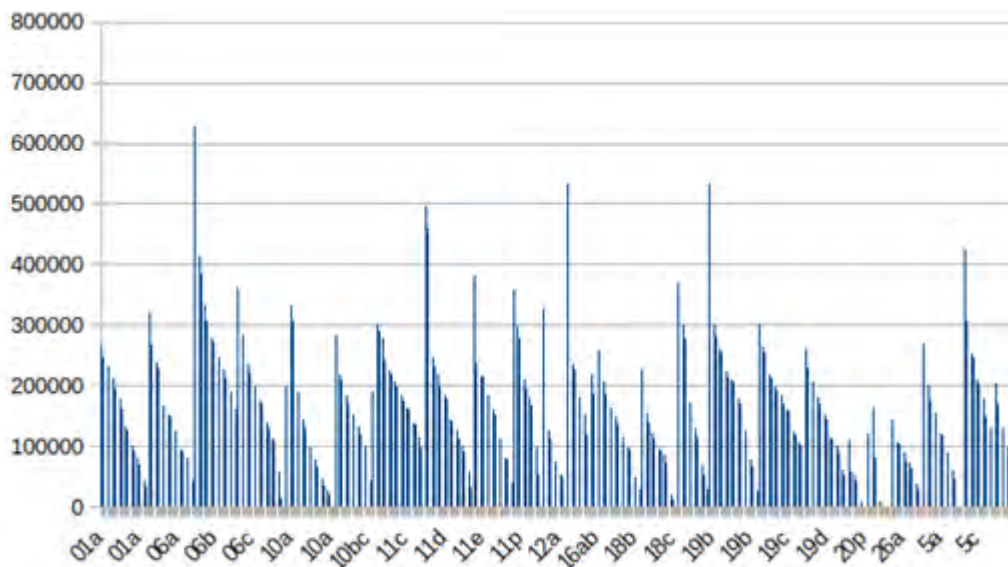
B)

Fig. 1.5.1.7: Retained reads after demultiplexing for (A) the whole *A. antennatus* dataset, and (B) each population separately

A tag catalog was built with 60 samples to reduce analysis time (figure 1.5.1.8). A maximum of 3 mismatches were allowed to merge tags into the same catalog locus. The final number of tags in the catalog is 5'321'403.



A)



B)

Fig. 1.5.1.8: ustacks loci for the (A) whole *A. antennatus* dataset, and (B) each population separately.

Stacks' module *populations* and *vcftools* were used to filter the initial SNP panel. Only SNPs shared by at least 70% of the individuals and with at least 10X coverage were kept. A total of 1253 SNPs were retained for 1043 samples. At this stage a sequence coverage problem became evident for about 140 individuals from the first library sequenced with the optimized approach (Library 01-bis, containing individuals from GSA 06b, 06c, 07a, 09p, 11ab and 11p), which contained less reads than expected biasing overall results. For this reason, samples from this library were provisionally removed, and remaining individuals from GSA09p and GSA11ab were pooled with individuals from GSA09b and 11p, respectively. In addition, due to the bad quality of DNA resulting in low sample size also GSA20a and 20b were pooled.

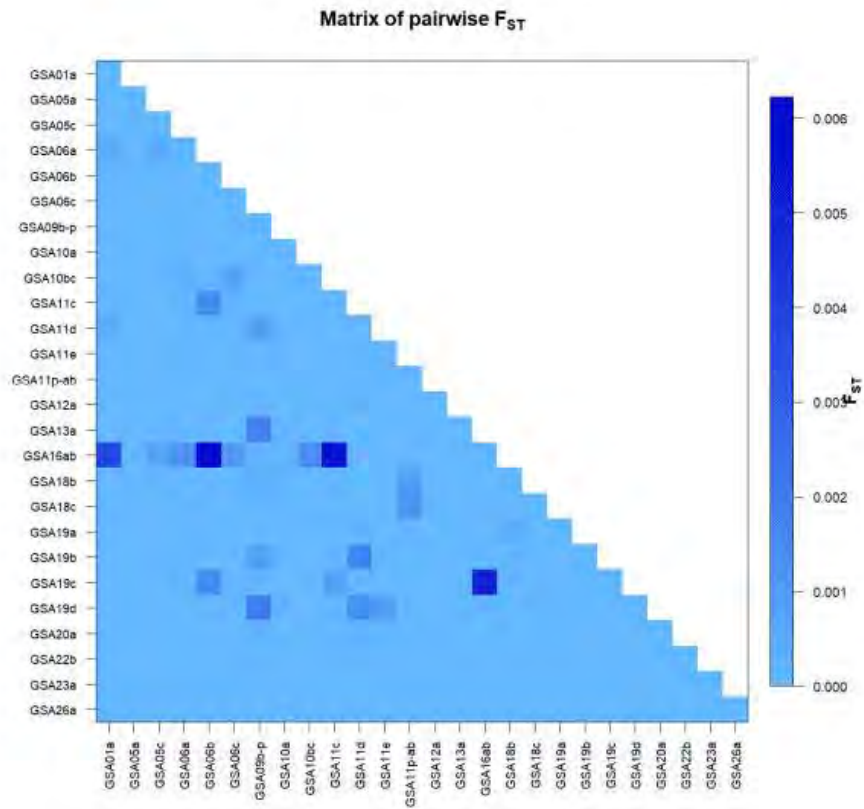
This led to a number of 886 samples included in this report, with ongoing efforts to increase the number of reads through resequencing of library 01-bis.

Observed heterozygosities were generally low (ranging from 0.1191 to 0.1527) and slightly smaller than the unbiased expected ones (ranging from 0.1464 to 0.1865; table 1.5.1.9), with an excess of homozygotes.

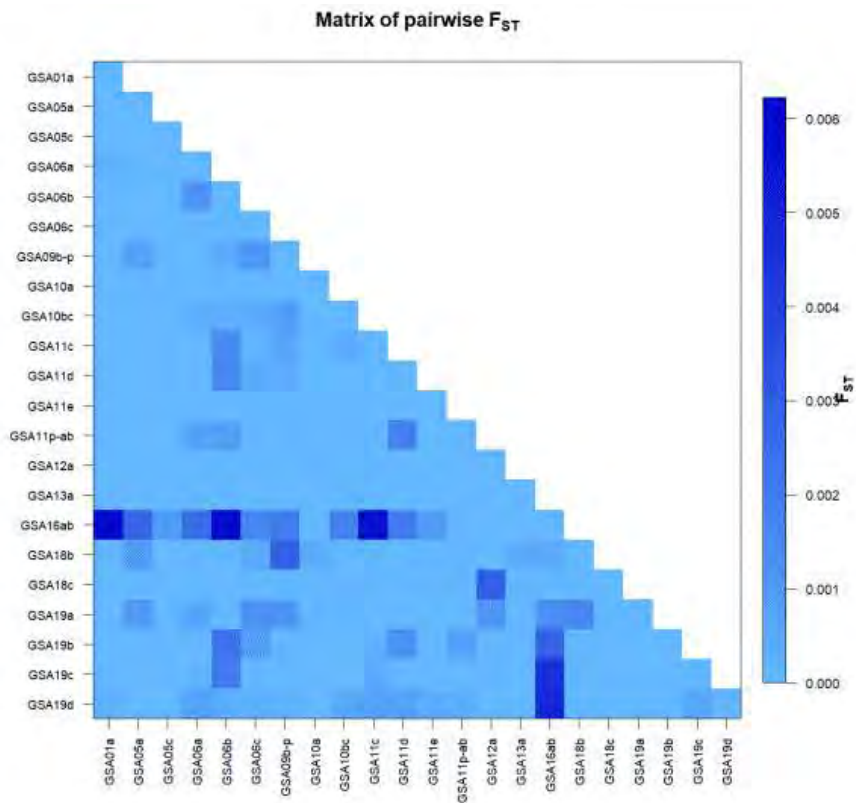
Table 1.5.1.9: Summary of genetic variability estimates across sampling locations in *A. antennatus*. Reported are the average observed (Ho) and average unbiased expected heterozygosity (He).

Samples	Ho	He
01a	0.1444	0.1831
05a	0.1430	0.1824
05c	0.1489	0.1839
06a	0.1408	0.1808
06b	0.1457	0.1847
06c	0.1446	0.1830
09b-p	0.1458	0.1865
10a	0.1367	0.1798
10bc	0.1426	0.1807
11c	0.1392	0.1805
11d	0.1467	0.1794
11e	0.1449	0.1783
11p-ab	0.1465	0.1721
12a	0.1455	0.1793
13a	0.1475	0.1761
16ab	0.1443	0.1837
18b	0.1368	0.1832
18c	0.1401	0.1865
19a	0.1483	0.1725
19b	0.1420	0.1808
19c	0.1412	0.1809
19d	0.1416	0.1778
20a-b-p	0.1527	0.1805
22b	0.1191	0.1726
23a	0.1264	0.1464
26a	0.1437	0.1821

All samples showed a great similarity between them. F_{st} values ranged from negative values to 0.00622 (figure 1.5.1.9). Six out of 325 comparisons were significant using an uncorrected probability threshold of 0.05, but none of them remained significant after Benjamini-Hochberg correction. Correspondingly, AMOVA showed a small global $F_{st} = 0.00415$, that was nonetheless statistically significant ($P < 0.00001$). In other words, populations are very similar but a significant differentiation occurs on a global level among them.



A)



B)

Fig. 1.5.1.9. (A) Matrix of pairwise F_{ST} for *A. antennatus* based on 1,253 polymorphic SNPs present in >70% of the whole sampling set, and (B) after removing the populations with the highest levels of missing data (20a-b-p, 22b, 23b and 26b). No value was statistically significant after Benjamini-Hochberg correction.

Hierarchical AMOVA, indeed, confirmed the existence of significant though weak differentiation, as indicated the existence of a significant percentage of genetic variation (measured by the differentiation metrics F_{ct}) explained by differentiation among groups. The best scenario explaining population structure, maximizing F_{ct} value, corresponded to three groups, the first including samples from the Western Mediterranean Sea, the second including samples from the Central Mediterranean Sea, and the third one including samples from the Eastern Mediterranean Sea ($F_{ct}=0.00145$, $P<0.001$; table 1.5.1.10). We also performed a pairwise F_{ct} analysis among macroareas, grouping samples based on Mediterranean basins. Results showed homogeneity among samples from Balearic, Sardinia and Ionian Sea, while a more pronounced differentiation of samples from Adriatic, Aegean, Levant seas was evident (table 1.5.1.11).

Table 1.5.1.10: AMOVA results for several scenarios of grouping *A. antennatus* samples.

	Scenario 1 – 3 groups	Scenario 2 – 6 groups
Grouping	<p>Group1: West Med (GSA01a, GSA05a,GSA05c,GSA06a,GSA06b,GSA06c,GSA09b -p,GSA10a,GSA10bc, GSA11c,GSA11d,GSA11e,GSA11p-ab,GSA12a)</p> <p>Group2: Central Med (GSA13a,GSA16ab,GSA18b,GSA18c,GSA19a,GSA1 9b,GSA19c,GSA19d,GSA20a-b-p)</p> <p>Group3: East Med (GSA22b, GSA23a,GSA26a)</p>	<p>Group1: Balearic (GSA01a, GSA05a,GSA05c,GSA06a,GSA 06b,GSA06c, GSA11c,GSA11d,GSA11e)</p> <p>Group2: Tyrrhenian and Tunisia (GSA09b- p,GSA10a,GSA10bc,GSA11p- ab,GSA12a)</p> <p>Group3: Ionian and Sicily Strait (GSA13a,GSA16ab, GSA19a,GSA19b,GSA19c,GSA 19d,GSA20a-b-p)</p> <p>Group4: Adriatic (GSA18b,GSA18c)</p> <p>Group5: Aegean (GSA22b,GSA23a)</p> <p>Group6: Levant (GSA26a)</p>
Percentage of Variation		
Among groups	0.15	0.10
Among populations within groups	0.34	0.34
Within populations	99.51	99.55
Fixation indices		
FST	0.00487	0.00446
P-value	<0.001	<0.001
FSC	0.00342	0.00342
P-value	<0.001	<0.001
FCT	0.00145	0.00104
P-value	<0.001	<0.001

Table 1.5.1.11: Pairwise F_{ct} results for *A. antennatus* samples grouped based on Mediterranean basins. Below the diagonal: F_{ct} values; above the diagonal: P values. In bold, F_{st} significant after Benjamini-Hochberg correction.

	Balearic	Tyr-Tun	Ionian-SS	Adriatic	Aegean	Levant
Balearic		0.382	0.018	0.004	0.000	0.000
Tyr-Tun	-0.0000		0.397	0.225	0.000	0.000
Ionian-SS	0.0003	-0.0001		0.832	0.000	0.102
Adriatic	0.0090	0.0003	-0.0008		0.000	0.000
Aegean	0.0178	0.0154	0.0121	0.0127		0.771
Levant	0.0134	0.0121	0.0072	0.0083	-0.0130	

Such weak differentiation is evident both by STRUCTURE (figure 1.5.1.10) and DAPC analysis (figure 1.5.1.11). DAPC identified one homogeneous group with the R's *adegenet*'s *find.cluster* function suggesting 1 as the most likely number of clusters in the dataset, and STRUCTURE showed the best likelihood values for $K=1$, though STRUCTURE's results at higher values of k suggest the presence of genetic differentiation between samples from West, Central and East Mediterranean (figure 1.5.1.11)

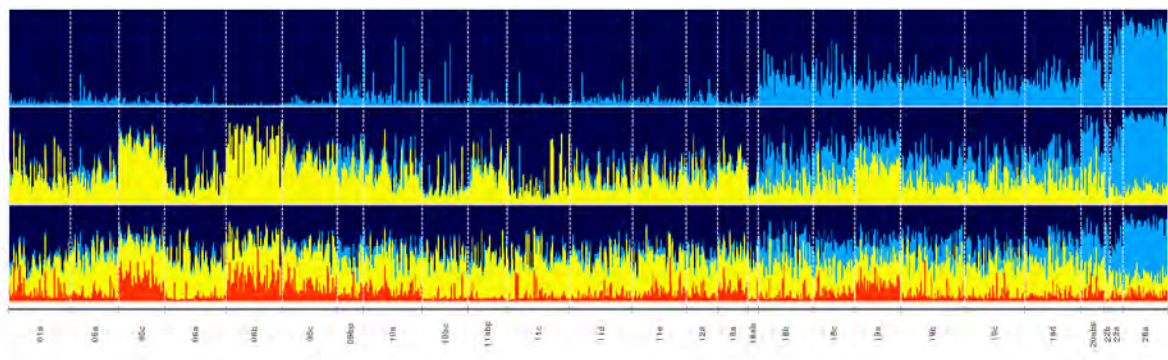
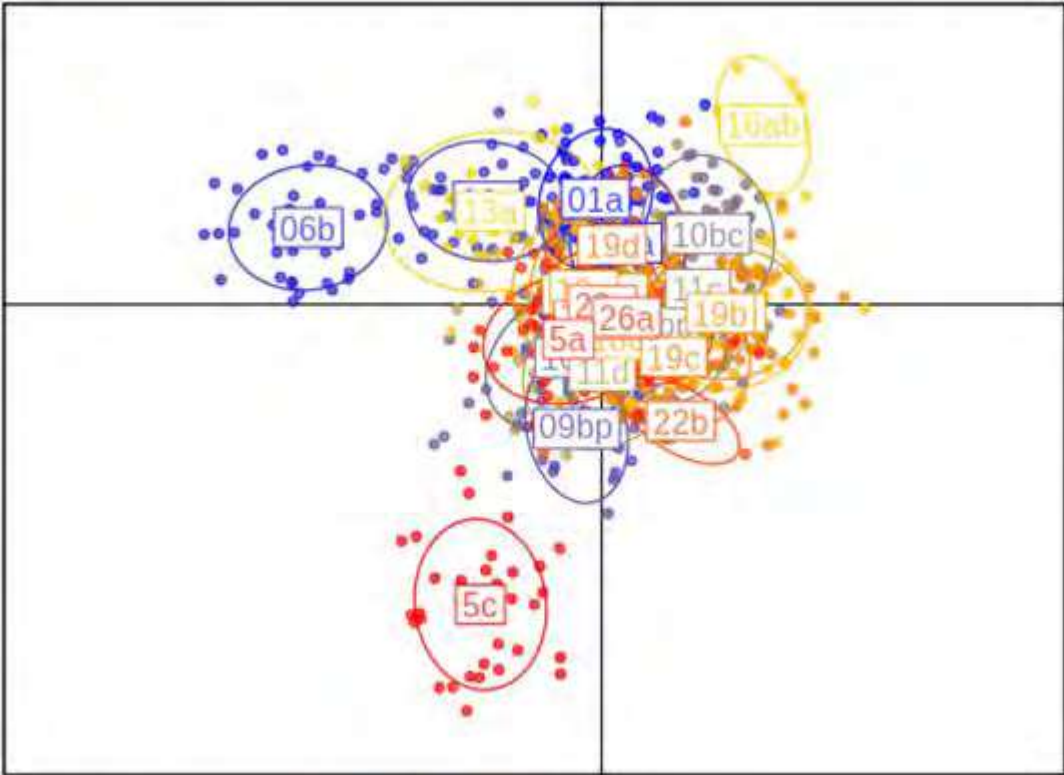
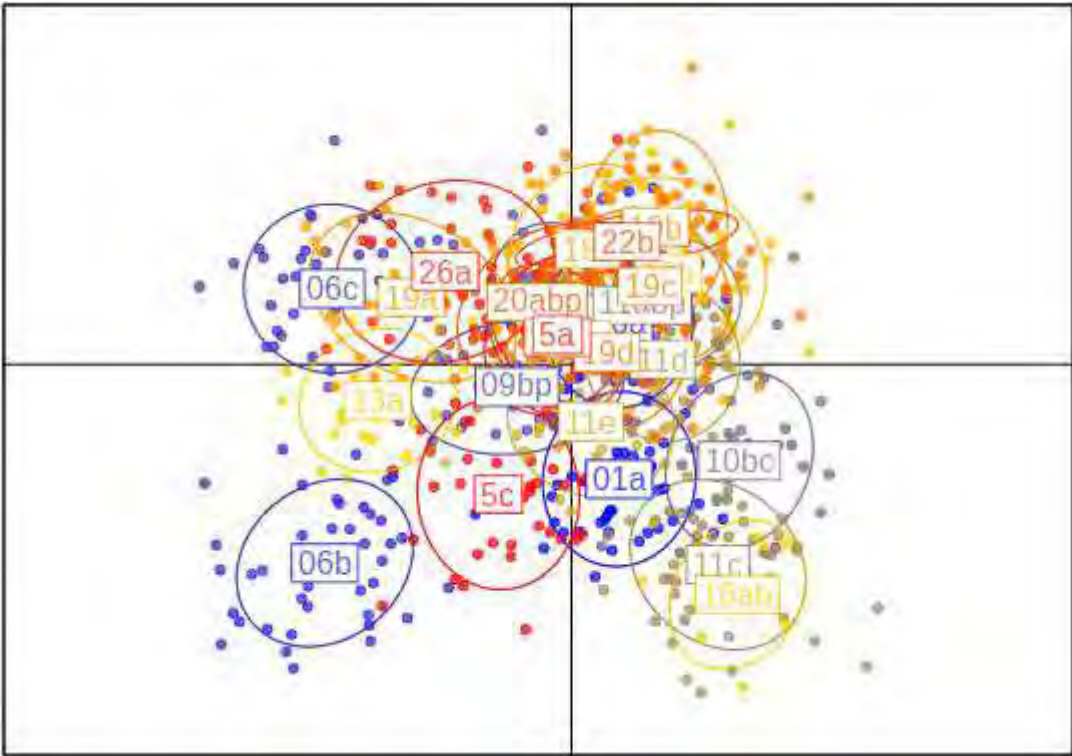


Fig. 1.5.1.10: Proportions of membership to each K inferred clusters for *A. antennatus*.



A)



B)

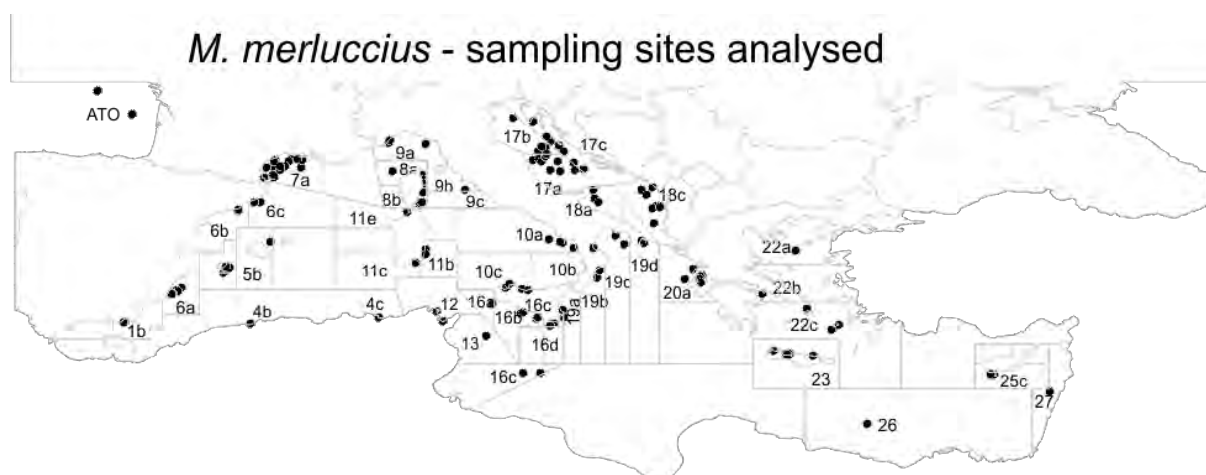
Fig. 1.5.1.11: DAPC plot for *A. antennatus* based on 1,253 polymorphic SNP (A) and on 869 haplotypes (B) present in >70% of the whole sampling set.

Results of this preliminary analysis seems to confirm the very low genetic differentiation among Mediterranean samples detected by allozymes (Sarda et al. 1998), microsatellites (Cannas et al. 2012), mtDNA sequences (Fernández et al. 2011b, 2013b, Maggio et al. 2009, Marra et al. 2015; Roldán et al. 2009; Sardà et al. 2010) and AFLP (Lo Brutto et al. 2012). However, our results reveal previously hidden subtle differences. Indeed, two genetic clusters were identified in two of the reported studies, though results were not confirmed and are in contrast with other papers analysing other sites in the same region with different markers, and reporting a substantial genetic homogeneity within the Mediterranean Sea. Roldan et al. 2009, using mtDNA on three population samples from Palamòs, Genova and Palermo, detected two clusters in the West Mediterranean pointing out a possible distinction of the Genova samples. Fernandez et al. 2011b, using mtDNA and analyzing 506 individuals from 11 population samples including Mediterranean, Atlantic and Indian Ocean detected significant differentiation with a possible distinction between Western and Eastern Mediterranean. These results are also in line with the most recent papers on genetic differentiation of *A. antennatus*. Heras et al (2019), using 12 microsatellites observed little differentiation within Mediterranean, though two differentiated stocks were found, corresponding to western Mediterranean and eastern Mediterranean, with no further differences within western Mediterranean. This result was partially confirmed by Agullo et al. (2020) that using the same set of microsatellites, detected a high level of geographical connectivity among groups in the samples from western Mediterranean collected in two consecutive years (2016 and 2017), though a significant differentiation was observed in 2017 ($F_{st} = 0.0025$, $p < 0.05$). Finally, in a recent report Catanese et al. (2020), developed and *ad hoc* filtering procedure to select, among 115071 SNPs identified by a genotyping by sequencing approach, a panel of 232 loci that detected differentiation among four populations samples, including Atlantic and three Balearic samples; unsurprisingly this panel, as well as a further subselection of 80 SNPs, identified differences among all the four populations samples tested, including the three from western Mediterranean, though his potential to detect differentiation in general remains untested.

Therefore, our results are in accordance with the low level of differentiation detected in published studies of *A. antennatus*, but clearly point out significant differentiation at least between different Mediterranean areas. In particular, though we did not find strong support for differences in the Western Mediterranean, we obtained a clear indication for structuring between West, Central and East Mediterranean, which provided the best scenario of differentiation based on hierarchical AMOVA. In particular, we detected significant pairwise F_{ct} when comparing groups of samples from the eastern and western part of this sea. From this perspective, the ddRADseq approach was successful in identifying significant differences, that can be further evidenced by using outlier analysis and selecting the most discriminating SNPs.

Merluccius merluccius

For the European hake (*Merluccius merluccius*), the complete study was conducted with specimens from 42 localities (GSAs and sub-GSAs).



In total, 1,728 specimens were sampled from the Alboran Sea (1a and 1b) in the W. Mediterranean to the Cypriot (25c) and Egyptian-Gaza (26a and 27b) waters in the E. Mediterranean (Table 1.5.1.12).

Table 1.5.1.12: DNA quality across samples.

GSA	sample number	DNA quality			Number of samples in Libraries
		Good	Medium	Bad	
01b	35	35	0	0	35
04a	32	2	26	4	28
04c	49	31	18	0	49
05b	50	48	2	0	48
06a	24	24	0	0	24
06b	50	38	12	0	55
06c	50	41	6	3	47
07ab	37	37	0	0	37
08a	41	22	13	6	33
08ab	13	12	0	1	12
09a	50	12	31	7	43
09bc	50	47	1	2	48
10a	49	43	2	4	45
10bc	50	49	1	0	32
11ab	50	23	27	0	50
11c	50	32	17	1	48
11e	50	48	1	1	54
12a	50	43	5	2	48
12m	57	56	1	0	56
13a	34	34	0	0	34
16ab	50	45	5	0	54
16c	50	44	5	1	49
16d	22	16	4	2	18
17ab	50	48	0	2	48
17b	50	20	28	2	46
17c	50	49	0	1	45
18a	50	49	0	1	48
18c	50	47	3	0	45

19a	50	42	5	3	47
19b	50	50	0	0	55
19cd	50	41	7	2	45
20a	50	41	9	0	45
22b	50	19	27	4	46
22c	40	29	10	1	39
22p	60	60	0	0	60
23p	73	57	14	2	48
25c	51	49	1	1	50
26a	50	31	18	1	39
27b	49	43	2	4	35
ATO	50	48	0	2	42
Sum	1908	1501	301	64	1728

Specimens from all sampled GSAs were included in the ddRAD library preparation (Table 1.5.1.13) at a various number, with the exception of GSA05a, due to its small sample size and GSA11d.

Table 1.5.1.13. Number of specimens (in total) in each library

1st library(pilot)	2nd library	3rd library	4th library	5th library	6th library
288 specimens	288 specimens	288 specimens	288 specimens	288 specimens	288 specimens
medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA

All the samples included in the ddRAD libraries were genotyped following a ddRAD protocol, with enzymes *SbfI* and *SphI*, selecting fragments from around 200 to around 470 bp. Multiplexed libraries were run on six Illumina lanes, with 150 bp paired-end protocol. R1 and R2 reads were merged, obtaining a total of about 3.5 billion reads in R1+R2.

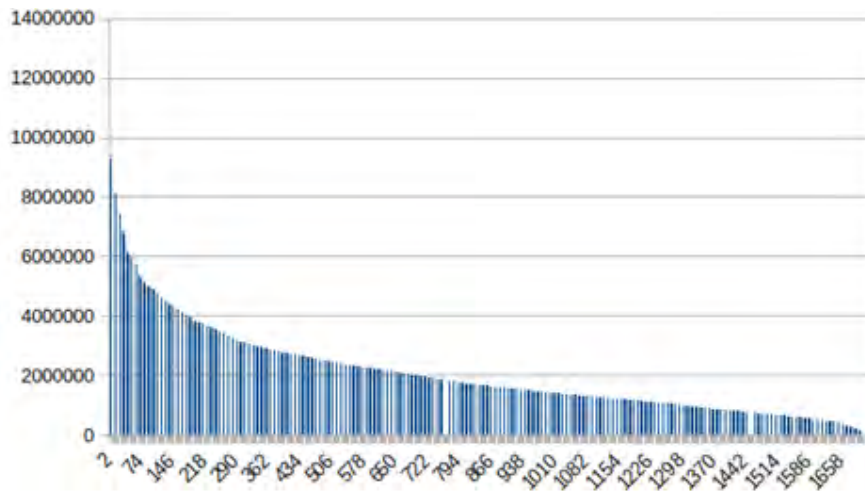
After demultiplexing, on average 2,071 million reads per sample were obtained (range from 12,196,074 to 5,346). After eliminating low quality reads and reads without a RAD site at the beginning, on average 98.75% reads were retained (see figure 1.5.1.12).

Samples were processed with *ustacks* considering a stack of unique reads only if at least 3 identical reads were found and allowing for up to 3 mismatches to consider two tags as part of the same locus. On average, 13,157 tags were found in each sample, with average coverage of 53.3X (table 1.5.1.14).

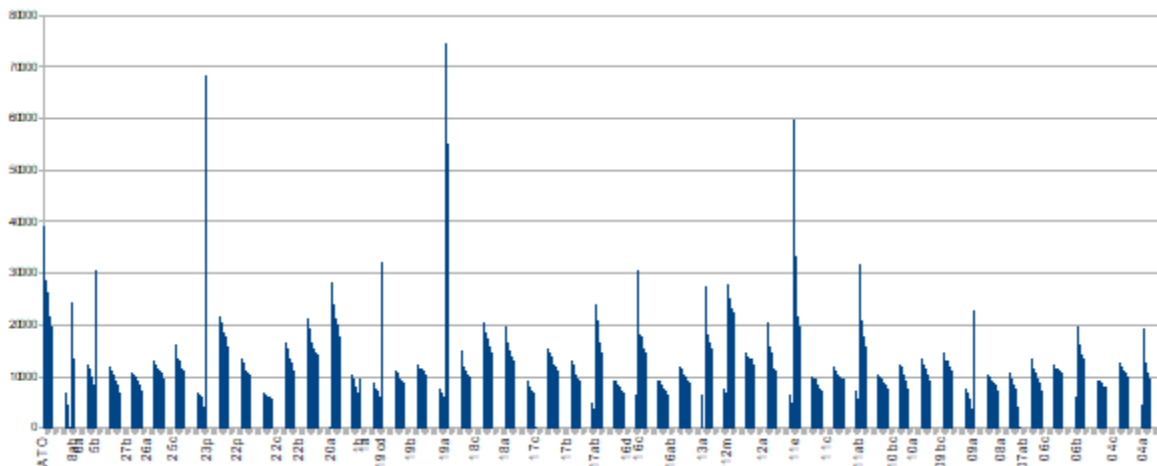
Table 1.5.1.14: Stacks pipeline output summary for *Merluccius merluccius*

GSA	Sample	Retained reads				Stacks loci			
		Total	Average	MIN	MAX	Total	Average	MIN	MAX
01b	31	71171783	2295864,0	5935	5265028	346664	11182,7	91	22437
04a	28	42712627	1525451,0	520201	4319309	271625	9700,9	6385	19301
04c	49	92813533	1894153,7	76090	5084568	502193	10248,8	2898	19629
05b	48	83216432	1733675,7	14752	5205686	644654	13430,3	160	39401
06a	24	71974532	2998938,8	829793	8029479	369347	15389,5	8455	30553
06b	55	93975419	1708644,0	47847	4672442	599612	10902,0	2130	22725
06c	47	91323719	1943057,9	472866	4604155	507626	10800,6	6234	19043
07ab	37	76806646	2075855,3	241900	7729058	564019	15243,8	5250	44178
08a	33	78203214	2369794,4	10000	12103852	391472	11862,8	77	47585
08ab	12	37631349	3135945,8	11002	7887677	178150	14845,8	120	33995
09a	43	59813884	1391020,6	22997	3408372	483288	11239,3	232	23117
09bc	48	78213719	1629452,5	86501	8979028	511681	10660,0	3190	44827
10a	45	94293432	2095409,6	238192	8512911	565326	12562,8	4449	60226
10bc	32	69368478	2167764,9	13863	5726709	389521	12172,5	162	26381
11ab	50	102140544	2042810,9	327647	5639715	608539	12170,8	5831	31643
11c	48	77673318	1618194,1	278910	5018105	481413	10029,4	5349	18658
11e	54	138593422	2566544,9	6233	11491386	729618	13511,4	57	59867
12a	48	107070693	2230639,4	5774	6720606	582958	12145,0	175	27467
12m	56	118634348	2118470,5	501292	10040375	974811	17407,3	7274	64436
13a	34	91175130	2681621,5	555200	7630917	503534	14809,8	6729	40867
16ab	54	104436786	1934014,6	7831	6095937	612201	11337,1	37	39354
16c	49	102856765	2099117,7	7795	7408022	582072	11879,0	108	30795
16d	18	30275950	1681997,2	344732	3382340	173661	9647,8	5555	15634
17ab	48	115396212	2404087,8	586182	5925701	579668	12076,4	5415	26309
17b	46	53653969	1166390,6	128285	3981583	476893	10367,2	3695	27100
17c	45	87434419	1942987,1	541137	7162548	685067	15223,7	6255	34377
18a	48	94328288	1965172,7	5964	4714389	545063	11355,5	238	24076
18c	45	110280190	2450670,9	476421	7791926	781000	17355,6	9533	38173
19a	47	119907115	2551215,2	10392	9559709	1036128	22045,3	147	74749
19b	55	102955926	1871925,9	135891	5852270	595606	10829,2	3869	23576
19cd	45	88351202	1963360,0	11313	6873119	584630	12991,8	156	34846
20a	45	90689740	2015327,6	282666	5554932	716810	15929,1	6389	41088
22b	46	109118491	2372141,1	12117	9045534	797778	17343,0	92	38491
22c	39	68009943	1743844,7	214332	4506977	639897	16407,6	5303	38732

22p	60	107966270	1799437,8	185543	8140832	578748	9645,8	3895	25834
23p	48	146554164	3053211,8	439754	8438228	1175696	24493,7	8093	68215
25c	50	83514774	1670295,5	57127	4975331	500677	10013,5	1553	21965
26a	39	71692694	1838274,2	137954	5199730	444247	11390,9	3802	21982
27b	35	90403929	2582969,4	11226	9289285	409008	11685,9	76	34418
ATO	42	92999350	2214270,2	6162	6055756	596424	14200,6	95	39126



A)



B)

Fig. 1.5.1.12: Retained reads after demultiplexing for (A) the whole *M. merluccius* dataset, and (B) each population separately.

Low number of reads affects around 50 samples in which a low number of tags has been found (figure 1.5.1.13). A tag catalog was built with 200 samples (the five with more reads of each population). A maximum of 3 mismatches were allowed to merge tags into the same catalog locus. The final number of tags in the catalog is around 2.4M.

Stacks' module *populations* and *vcftools* were used to filter the initial SNP panel. Forty one individuals that produced a number of raw reads below 200'000 were filtered out to avoid a reduction of final SNP number when filtering for SNPs shared by at least 80% of samples. Only SNPs shared by at least 70% of the individuals and with at least 10X coverage were kept. After retaining only one random SNP per tag, a total of 665 high quality SNPs were retained (for 1,667 samples) and used for all downstream differentiation analysis; samples from GSA01 (subareas 01a and 01b) were pooled in a single sample leading to a final number of 40 sites. In addition to the SNP dataset, we also generated a haplotype dataset including 670 haplotypes shared by at least 60% of the samples.

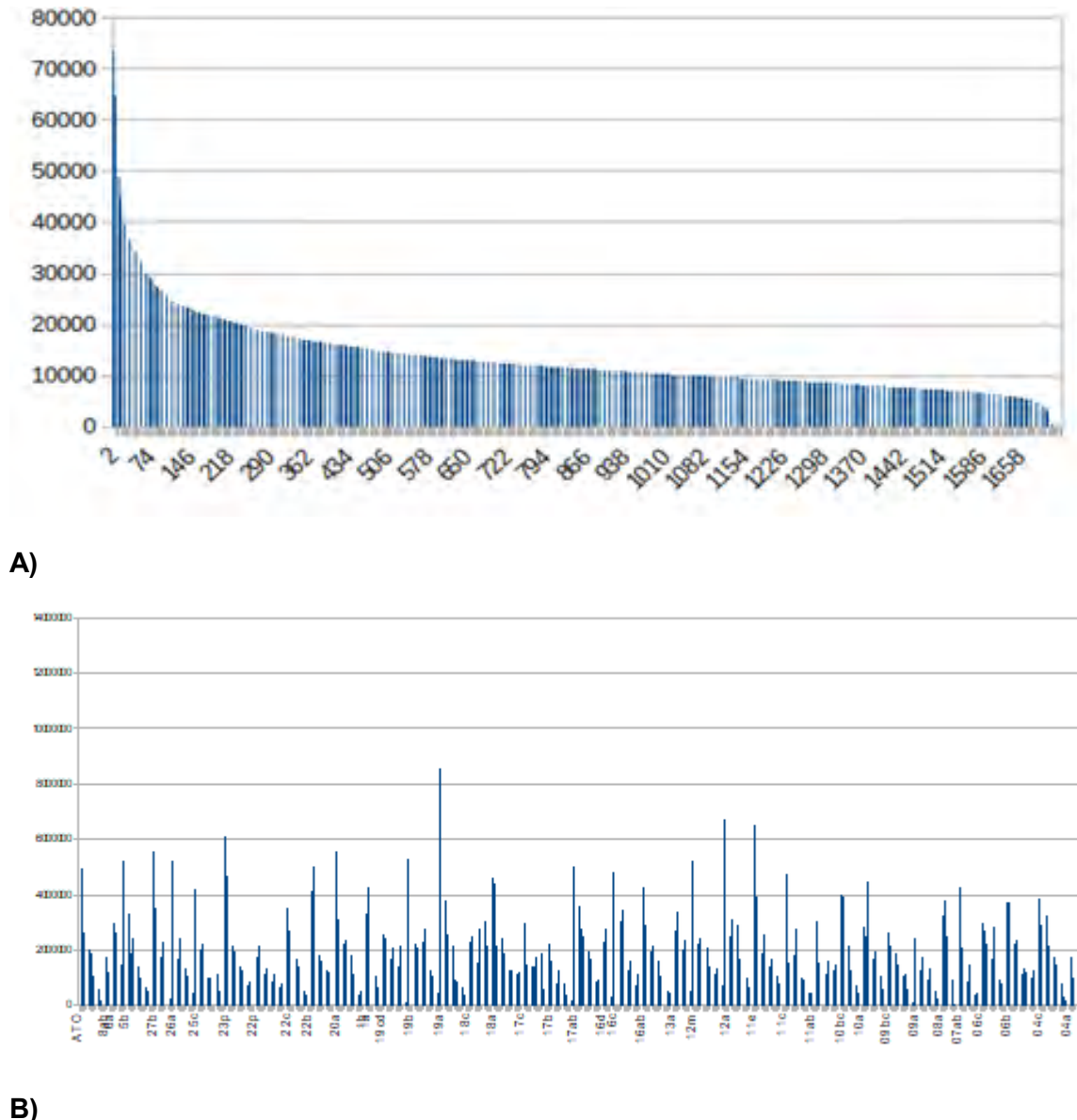


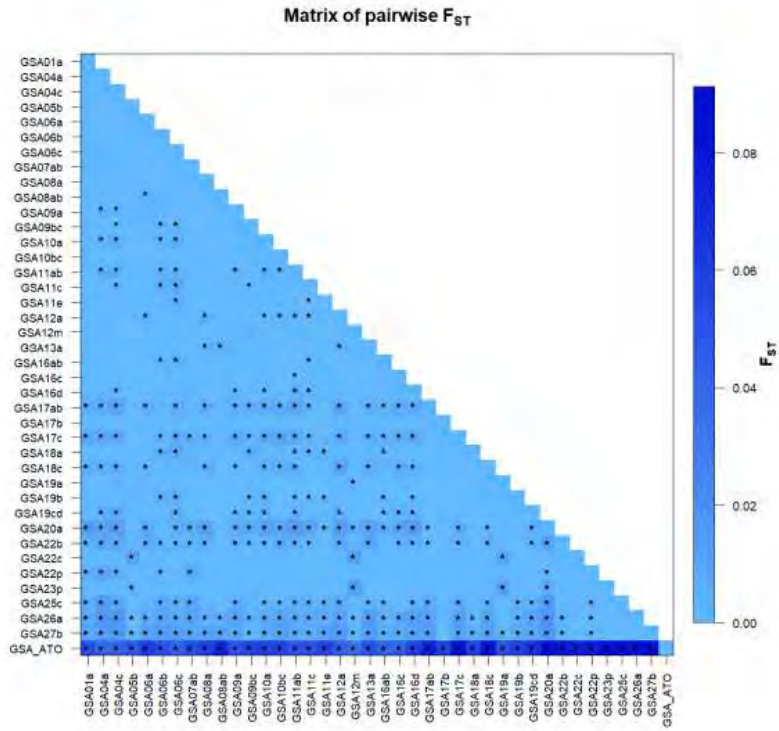
Fig. 1.5.1.13: ustacks loci for the (A) whole *M. merluccius* dataset, and (B) each population separately.

Observed heterozygosities were generally low (ranging from 0.2200 to 0.2705) and smaller than the unbiased expected heterozygosities (ranging from 0.2592 to 0.3020; table 1.5.1.15). Comparisons between samples showed a significant genetic differentiation with F_{st} values ranging from -0.11955 to 0.10207 and 277 out of 780 tests, significant after Benjamini and

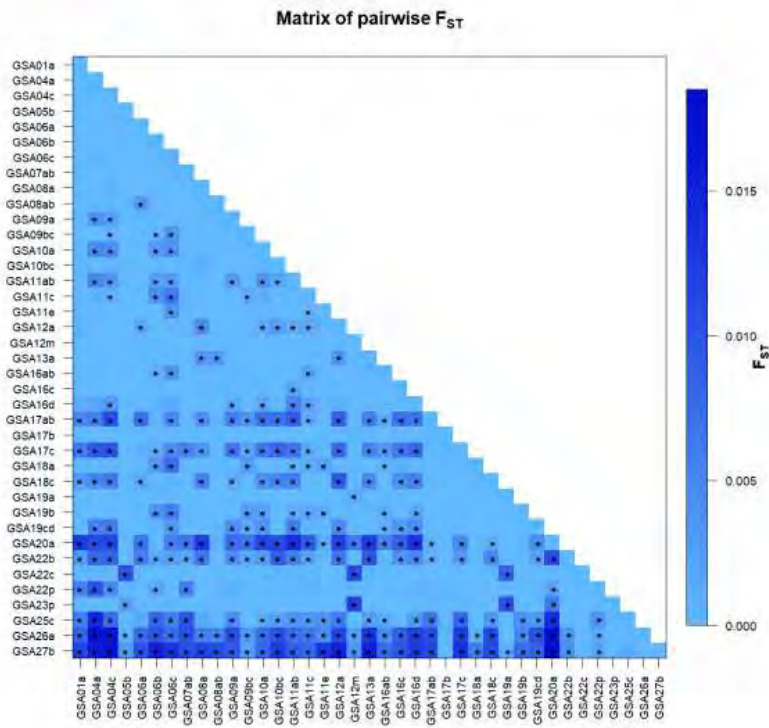
Hochberg (1995) correction for multiple tests. In addition to the Atlantic Ocean sample, included for reference, the population samples from Aegean and Levant seas showed the highest F_{st} values (figure 1.5.1.14). Accordingly, non-hierarchical AMOVA showed a highly significant global F_{st} when including the Atlantic sample ($F_{st} = 0.01358$, $P < 0.00001$) and when considering the Mediterranean samples alone ($F_{st} = 0.01044$, $P < 0.00001$) allowing to confidently reject the null hypothesis of homogeneity in both the cases.

Table 1.5.1.15: Summary of genetic variability estimates across sampling locations in *M. merluccius*. The average observed (H_o) and average unbiased expected heterozygosity (H_e) are reported.

Samples	H_o	H_e
01b	0.2275	0.2679
04a	0.2375	0.2727
04c	0.2404	0.2707
05b	0.2381	0.2722
06a	0.2410	0.2726
06b	0.2341	0.2680
06c	0.2343	0.2692
07ab	0.2457	0.2751
08a	0.2459	0.2791
08ab	0.2705	0.3020
09a	0.2469	0.2703
09bc	0.2371	0.2691
10a	0.2423	0.2698
10bc	0.2372	0.2679
11ab	0.2412	0.2676
11c	0.2376	0.2701
11e	0.2371	0.2681
12a	0.2410	0.2703
12m	0.2429	0.2799
13a	0.2419	0.2741
16ab	0.2396	0.2699
16c	0.2378	0.2680
16d	0.2482	0.2790
17ab	0.2347	0.2634
17b	0.2413	0.2772
17c	0.2349	0.2633
18a	0.2322	0.2673
18c	0.2270	0.2605
19a	0.2411	0.2782
19b	0.2370	0.2721
19cd	0.2399	0.2667
20a	0.2313	0.2632
22a	0.2212	0.2689
22b	0.2319	0.2592
22c	0.2414	0.2736
23a	0.2287	0.2693
25c	0.2200	0.2674
26a	0.2437	0.2688
27b	0.2303	0.2706
ATO	0.2340	0.2643



A)



B)

Fig. 1.5.1.14: (A) Matrix of pairwise F_{ST} for *M. merluccius* based on 665 polymorphic SNPs present in >70% of the whole sampling set, and (B) after removing Atlantic samples. Significant pairwise F_{ST} values, after Benjamini– Hochberg correction (25% FDR), are marked by asterisk.

Hierarchical AMOVA was used to infer the best scenario explaining population structure based on F_{ct} maximization, allowing to identify four groups, the first including the sample from the Atlantic Ocean, the second including samples from the Western Mediterranean Sea, the third including samples from the Central Mediterranean Sea, and the fourth one including samples from the Eastern Mediterranean Sea (table 1.5.1.16). In addition, arranging population samples in 8 groups according to their geographic origin from West-to-East, provided highly significant evidence of geographic subdivision (table 1.5.1.16 and figure 1.5.1.15).

Table 1.5.1.16: AMOVA results for several scenarios of grouping *M. merluccius* samples.

	Scenario 1: 3 groups	Scenario 2: 4 groups	Scenario 3: 8 groups
Grouping	<p>Group1: West Med (GSA01a,GSA04a,GSA04c ,GSA05b,GSA06a,GSA06 b,GSA06c,GSA07ab,GSA0 8a,GSA08ab, GSA09a,GSA09bc,GSA10 a,GSA10bc,GSA11ab,GSA 11c,GSA11e,GSA12a,GSA 12m,GSA_ATO)</p> <p>Group2: Central Med (GSA13a,GSA16ab,GSA1 6c,GSA16d,GSA17ab,GSA 17b,GSA17c,GSA18a,GSA 18c,GSA19a,GSA19b,GSA 19cd,GSA20a)</p> <p>Group3: East Med (GSA22a,GSA22b,GSA22 c,GSA23a,GSA25c,GSA26 a,GSA27b)</p>	<p>Group1: West Med (GSA01a,GSA04a,GSA04c, GSA05b,GSA06a,GSA06b,G SA06c,GSA07ab,GSA08a,G SA08ab,GSA09a,GSA09bc, GSA10a,GSA10bc,GSA11a b,GSA11c,GSA11e,GSA12a ,GSA12m)</p> <p>Group2: Central Med (GSA13a,GSA16ab,GSA16c ,GSA16d,GSA17ab,GSA17b ,GSA17c,GSA18a,GSA18c, GSA19a,GSA19b,GSA19cd, GSA20a)</p> <p>Group3: East Med (GSA22a,GSA22b,GSA22c, GSA23a,GSA25c,GSA26a,G SA27b)</p> <p>Group4: Atlantic (GSA_ATO)</p>	<p>Group1:Balearic (GSA01a,GSA04a,GSA04c, GSA05b, GSA06a,GSA06b,GSA06c, GSA11c,GSA11e)</p> <p>Group2: Tyrrhenian sea and Tunisia (GSA08a,GSA08ab,GSA09a,GSA 09bc,GSA10a,GSA10bc,GSA11a b,GSA12a,GSA12m)</p> <p>Group3: Gulf Lion (GSA07ab)</p> <p>Group4: Ionian and Strait of Sicily (GSA13a,GSA16ab,GSA16c,GSA 16d,GSA19a,GSA19b,GSA19cd, GSA20a)</p> <p>Group5: Adriatic (GSA17ab,GSA17b,GSA17c,GSA 18a,GSA18c)</p> <p>Group6: Aegean (GSA22a,GSA22b,GSA22c,GSA2 3a)</p> <p>Group7: Levant (GSA25c,GSA26a,GSA27b)</p> <p>Group8: Atlantic (GSA_ATO)</p>
Percentage of Variation			
Among groups	0.48	0.91	0.78
Among populations within groups	1.05	0.75	0.69
Within populations	98.47	98.34	98.53
Fixation Indices			
FST	0.01529	0.01661	0.01473
P-value	<0.001	<0.001	<0.001
FSC	0.01051	0.00761	0.00697
P-value	<0.001	0.002	0.002

FCT	0.00483	0.00907	0.00782
P-value	<0.001	<0.001	<0.001

	Balearic	Gulf of Lion	Sardinia	Ionian	Adriatic	Aegean	Levant	Atlantic
Balearic		0.96475	0.11	0.2198	0	0	0	0
Gulf of Lion	-0.00149		0.75317	0.81624	0.01812	0.06663	0	0
Sardinia	0.00025	-0.00082		0.10366	0	0	0	0
Ionian	0.00013	-0.00141	0.00083		0.0001	0	0	0
Adriatic	0.00464	0.00589	0.00527	0.00145		0	0	0
Aegean	0.009	0.0052	0.00731	0.00589	0.00507		0	0
Levant	0.01201	0.00936	0.0097	0.0102	0.01259	0.00378		0
Atlantic	0.06159	0.06136	0.06308	0.06819	0.0872	0.08907	0.09232	

Fig. 1.5.1.15: Matrix of pairwise Fct for *M. merluccius* based on 665 polymorphic SNPs present in >70% of the whole sampling set and considering 8 geographic groups as in table 1.5.13. Below the diagonal: Fct values. Above the diagonal: Fct probabilities. In bold and yellow, Fct significant after Benjamini-Hochberg correction.

The analysis suggests the presence of several levels of genetic structuring in the dataset analyzed. The strongest differentiation is between Atlantic and Mediterranean samples, but also within the Mediterranean populations are structured in a West to East pattern. This is evident from both STRUCTURE (figure 1.5.1.16) and DAPC analysis (figure 1.5.1.17), but with some differences between the results of the two approaches: DAPC identified the Eastern Mediterranean group (GSA from 22 to 26) as the most differentiated one, followed by samples from Adriatic and Central Mediterranean. The remaining populations seem to be structured according to Isolation By Distance. STRUCTURE, on the other side, suggests a stronger differentiation (k=2) when comparing samples from the Western Mediterranean and the western part of Central Mediterranean (GSA from 1 to 16) with those from the East part of Central Mediterranean, the Adriatic sea and the Eastern Mediterranean (GSA from 17 to 26). A second level of differentiation (k=3) involves this second group and separates samples from Central Mediterranean from samples from Eastern Mediterranean.

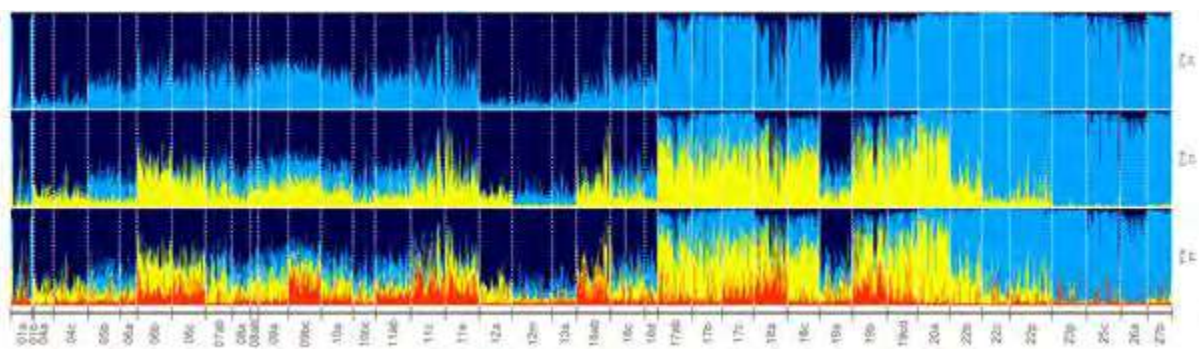


Fig. 1.5.1.16: Proportions of membership to each K inferred clusters for *M. merluccius* individuals (only Mediterranean samples).

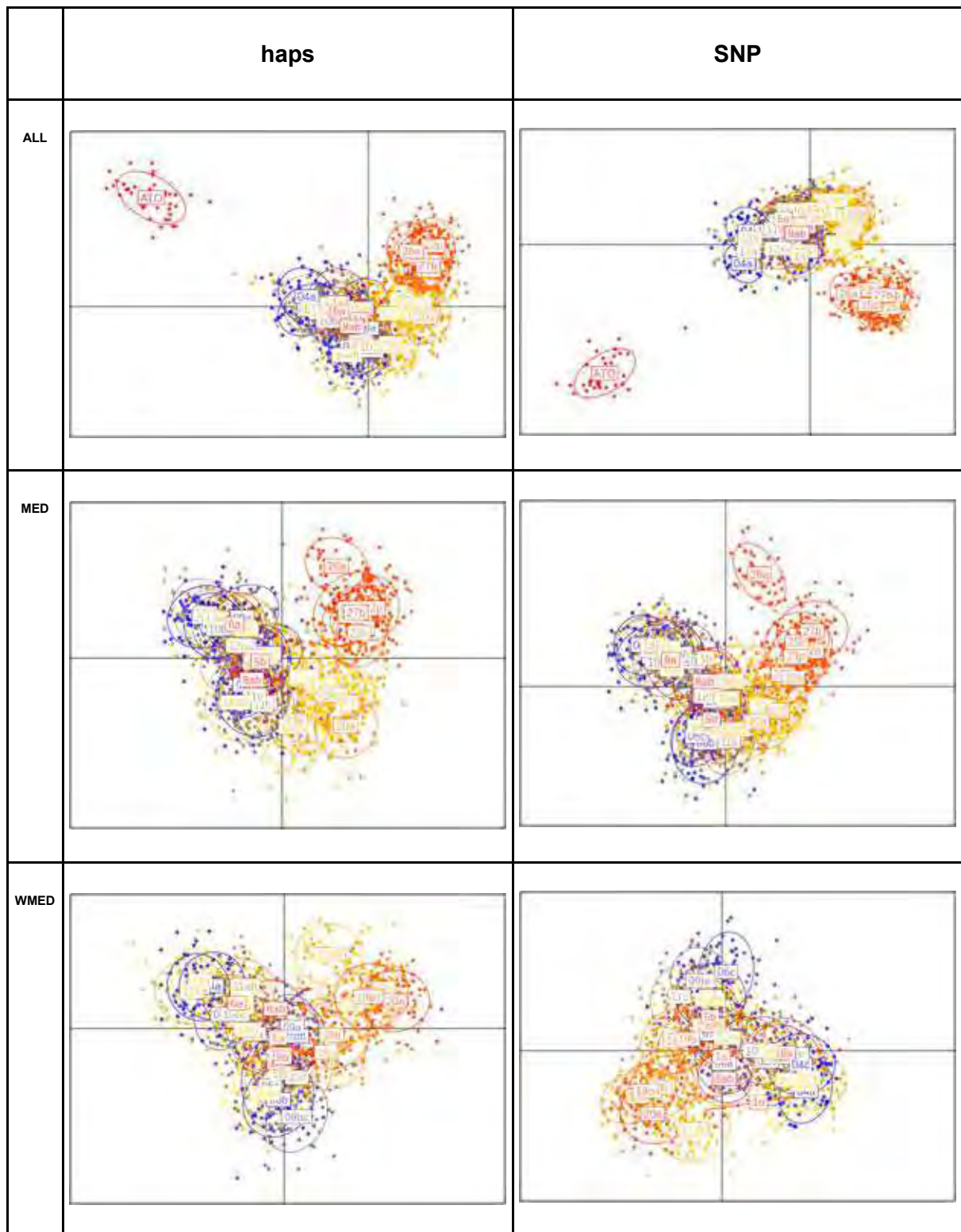


Fig. 1.5.1.17. DAPC plot for *M. merluccius* based on 20 principal components summarizing the information of 665 polymorphic SNP present in >70% of the whole sampling set (right) and 670 haplotypes (left) present in >70% and >60% of the whole sampling set.

Several studies have investigated the European hake population genetics within the Mediterranean Sea using different methods (i.e. allozymes, mtDNA-RFLP, mtDNA sequences, microsatellites and SNPs). In general, many studies identified a substantial homogeneity within the Mediterranean but suggested a strong subdivision of Atlantic and Mediterranean hake stocks (Lo Brutto et al., 1998; Roldan et al., 1999; Lundy et al., 1999; Castillo et al., 2004; Levi et al., 2004; Lo Brutto et al., 2004; Pita et al., 2010; Tanner et al., 2014). On the contrary, other studies (Castillo et al., 2004 using 5 microsatellites; Cimmaruta et al., 2005 using allozymes) described the occurrence of genetic heterogeneity within the Mediterranean Sea. Similarly, Nielsen et al. (2012) using a big number of SNPs with significantly higher resolving power, confirmed the genetic break between Atlantic and Mediterranean populations and Milano et al. (2014), using outliers from the previous study, described a finer-scale significant genetic population structure in the Mediterranean detecting a strong differentiation among Western, Central and Eastern Mediterranean geographical samples.

Therefore, our *M. merluccius* full study provides further support for the differentiation between Atlantic and Mediterranean stocks and further highlights the existence of differences within Mediterranean, mainly arranged in a West to East direction, previously detected only using outlier loci.

Conclusions

The comprehensive genetic study allowed to obtain ddRAD sequencing data for three out of the six species considered. The ddRAD approach identified differences among Mediterranean populations of the European hake, but for the two aristeid decapod crustacean species (*Aristaeomorpha foliacea* and *Aristeus antennatus*) it seems that there is very little differentiation in the Mediterranean basin.

For the giant red shrimp (*Aristaeomorpha foliacea*), the complete study was conducted with specimens from 32 localities and 1,692 specimens sampled from the Balearic Islands to Cyprus and Egypt. Unfortunately, due to the low DNA quality of some specimens we had to proceed with some 80% of the samples, and these individuals were genotyped, the species "catalogue" comprised 2,393,590 ddRAD loci. After implying various quality filters, discarding samples with low number of reads and stacks loci (<4,000), and keeping SNPs present in at least 70% of the samples, we run all phylogeographic analyses with a dataset composed of 771 samples (for 30 localities) and 443 higher quality SNPs. Our results point out an evident lack of genetic differentiation and are generally in agreement with previous studies conducted at smaller geographic scales and less extended sampling points in the Mediterranean Sea. Therefore, using the ddRADseq approach, lots of hundreds of polymorphic markers were found thanks to from an extensive sampling plan from the West (Balearic islands) to the East of the Mediterranean Sea (Cyprus and Egypt). Despite the high remarkable number of marker we estimate very low differentiation between samples, i.e. very low pairwise F_{st} values, The model-based Bayesian approach advocated for the absence of population structure in the Mediterranean Sea. The greatest part of the identified genetic variation (>98%) is attributed to differences among individuals in the sampled locations or populations, and much less among groups. However, the highest value of fixation among the a priori group tested (Table 1.5.1.5, $F_{CT}= 0.00082$, p value <0.001), is found to significantly differentiate the three major groups: the Western group up to the Siculo-Tunisian Strait (GSAs 5 to 12), the central group from the Tunisian coast (Gulf of Hammamet, GSA13) to the Southern Ionian (GSA20) and the Eastern Mediterranean (GSA20 to 26).

For the blue and red shrimp (*Aristeus antennatus*) the preliminary study was conducted with samples from 31 localities using 1253 SNPs. Due to the bad quality of DNA, of the 1471 specimens available 1'043 individuals were suitable for ddRAD analysis and, after filtering a total of 886 samples were retained. The results indicate a great similarity between samples, very small pairwise F_{st} values, though the global F_{st} was statistically significant. Accordingly, unsupervised genetic clustering indicated as most likely the existence of a single genetic group, but three weakly differentiated groups can be identified, corresponding to Western, Central and Eastern Mediterranean Sea (Table 1.5.1.10, $F_{CT}=0.00145$, p-value <0.001). At a smaller scale, significant differentiation of samples from Adriatic, Aegean, Levantine seas was detected (Table 1.5.1.11), where 10 out of 12 pairwise comparisons were significant (F_{CT} range from 0.0083 to 0.0178, p-values after correction for multiple test <0.001). In brief, results confirm a very low genetic differentiation among Mediterranean samples detected in published studies, but highlight an underlying significant, though still weak, structure due to the higher power of the markers used. The existence of significant genetic differentiation between Western (GSA01 to GSA12), Central (GSA13 to GSA20) and Eastern (GSA22 to GSA26) Mediterranean Sea, should be taken into account in the future and additional investigations with high resolution markers should aim to further elucidate differences at the smaller scale such those detected for Adriatic, Aegean, and Levantine seas. With this regard it will be important to increase sample size by careful standardization of collection and preservation of specimens to obtain good quality DNA for ddRAD genetic analysis and to reduce the percentage of discarded specimens.

In *Merluccius merluccius*, a total of 1728 samples were genotyped from 41 sampling locations, and a total of 665 high quality SNPs were retained for 1667 samples in the final dataset.

Comparisons between samples showed a significant genetic differentiation, with the strongest differentiation between Atlantic and Mediterranean samples, but significant structuring within Mediterranean. Mediterranean populations were structured in a West to East pattern. Analysis identified four groups corresponding to main subdivisions, the first including the sample from the Atlantic Ocean, the second including samples from the Western Mediterranean Sea, the third including samples from the Central Mediterranean Sea, and the fourth one including samples from the Eastern Mediterranean Sea (Table 1.5.1.16, $F_{CT}=0.00907$, p-value <0.001). In addition, arranging population samples in 8 groups according to their geographic origin from West-to-East, provided highly significant evidence of geographic subdivision (Fig. 1.5.1.15) with 21 out of 28 significant pairwise comparisons (F_{ct} : range 0.00145 to 0.09232; p-values after correction for multiple tests: range <0.001 to 0.0181) and non-significant values occurring mainly comparing western Mediterranean samples or Western and Central Mediterranean samples. Therefore, *M. merluccius* full study provides interesting results and the finding of significant genetic structure within Mediterranean represents a significant advance respect to previous published studies, where such differences were detected only using outlier genetic loci (i.e. the few loci detecting significantly more differences than what is found using the entire dataset) such in Milano et al. (2010). In particular, the differentiation between Western (GSA01 to GSA12), Central (GSA13 to GSA20) and Eastern (GSA22 to GSA27) Mediterranean samples is now fully supported, and additional differences among groups of GSA or single samples), should be taken into account and monitored, particularly in the Central and Eastern Part of the basin (Fig. 1.1.5.14).

As a concluding remark, the application of ddRADseq to this comprehensive genetic study of two aristeid decapod crustacean species (*Aristaeomorpha foliacea* and *Aristeus antennatus*) and one fish (*Merluccius merluccius*) evidenced a much lower level of differentiation of the two crustaceans, that is in line with previous studies. This imply that to obtain solid support for these crustaceans a very high number of loci and a big sample is needed. Our study evidenced two critical steps in the application of the ddRADseq to crustaceans, the first related to DNA quality and the second to the optimization of the protocol for library preparation and sequencing. For DNA quality, a particular care should be taken during sample collection and preservation, to reduce the number of DNA extract not suitable to ddRAD sequencing. During this study we made a strong effort to standardize protocols for specimen manipulation after collection and for preservation. For crustaceans, we found that the time past from collection to processing was critical and, when we were able to control all the procedure, good quality DNA was obtained when samples were processed fresh, few hours after the collection and placed directly on board in an appropriate volume of ethanol. Yet, the quality of DNA was still variable for specimens collected at different sites and, to a lesser extent, for specimens collected at the same site, implying that further optimization is needed. For the optimization of the protocol for library preparation and sequencing we found that a step of selection of restriction enzyme pairs for the ddRADseq digestion is needed and we indeed ended up with different enzyme pairs for *Aristaeomorpha foliacea* and *Aristeus antennatus*. We also found that, with the available enzyme pairs for ddRADseq and due to the big size of crustaceans genomes, a very high number of loci is obtained, which required further optimization of size-selection of fragments for high throughput sequencing, level of multiplexing, and/or filtering for bioinformatic analysis. Based on these experiences, we believe that future applications of ddRADseq in new species, particularly crustaceans, can be facilitated by performing a preliminary optimization study, on a very small scale, to address these issues in advance.

References

Agullo, M., S. Heras, J. L. Garcia-Marin, M. Vera, L. Planella & M. I. Roldan (2020). Genetic analyses reveal temporal stability and connectivity pattern in blue and red shrimp *Aristeus antennatus* populations. *Scientific Reports* 10(1):10 doi:10.1038/s41598-020-78634-2.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. 1996-2004 GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France)

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics* 29:1165-1188.

Cannas, R., F. Sacco, M. C. Follesa, A. Sabatini, M. Arculeo, S. Lo Brutto, T. Maggio, A. M. Deiana & A. Cau (2012). Genetic variability of the blue and red shrimp *Aristeus antennatus* in the Western Mediterranean Sea inferred by DNA microsatellite loci. *Marine Ecology* 33(3):350-363 doi:10.1111/j.1439-0485.2011.00504.x.

Cannas, R., S. Marcias, F. Sacco, A. Cau & A. M. Deiana (2012). First isolation and characterization of genomic SSR markers for the giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827). *Genetics and Molecular Research* 11(3):2745-8 doi:10.4238/2012.August.17.1.

Castillo, A. G. F., J. L. Martinez & E. Garcia-Vazquez (2004). Fine spatial structure of Atlantic hake (*Merluccius merluccius*) stocks revealed by variation at microsatellite loci. *Marine Biotechnology* 6(4):299-306 doi:10.1007/s10126-004-3027-z.

Catanese, G., Trotta, J. R., Iriondo, M., Grau, A. M., & Estonba, A. (2021). Discovery of SNP markers of red shrimp *Aristeus antennatus* for population structure in Western Mediterranean Sea. *Conservation Genetics Resources*, 13(1), 21-25 doi:10.1007/s12686-020-01178-0.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. and Cresko, W.A. (2013), Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22: 3124-3140. <https://doi.org/10.1111/mec.12354>

Cimmaruta, R., P. Bondanelli & G. Nascetti (2005). Genetic structure and environmental heterogeneity in the European hake (*Merluccius merluccius*). *Molecular Ecology* 14(8):2577-2591 doi:10.1111/j.1365-294X.2005.02595.x.

Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E. and Orlando, L. (2017), Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources* 17:508-522. doi:[10.1111/1755-0998.12595](https://doi.org/10.1111/1755-0998.12595)

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4(2), 359-361.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14(8):2611-2620.

Excoffier, L. & Lischer, H. E. L. (2010). Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Falush, D., M. Stephens & J. K. Pritchard (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567-1587.

- Falush, D., M. Stephens & J. K. Pritchard (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7(4):574-578 doi:10.1111/j.1471-8286.2007.01758.x.
- Fernández, M. V., F. Maltagliati, F. G. Pannacciulli & M. I. Roldán (2011). Analysis of genetic variability in *Aristaeomorpha foliacea* (Crustacea, Aristeidae) using DNA-ISSR (Inter Simple Sequence Repeats) markers. *Comptes Rendus Biologies* 334(10):705-712 doi:https://doi.org/10.1016/j.crv.2011.07.005.
- Fernández, M. V., S. Heras, F. Maltagliati, A. Turco & M. I. Roldán (2011b). Genetic structure in the blue and red shrimp *Aristeus antennatus* and the role played by hydrographical and oceanographical barriers. *Marine Ecology Progress Series* 421:163-171 doi:10.3354/meps08881
- Fernandez, M. V., S. Heras, F. Maltagliati & M. I. Roldan (2013°). Deep genetic divergence in giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827) across a wide distributional range. *Journal of Sea Research* 76:146-153 doi:10.1016/j.seares.2012.08.004.
- Fernández, M. V., S. Heras, J. Viñas, F. Maltagliati & M. I. Roldán (2013b). Multilocus Comparative Phylogeography of Two Aristeid Shrimps of High Commercial Interest (*Aristeus antennatus* and *Aristaeomorpha foliacea*) Reveals Different Responses to Past Environmental Changes. *PLoS ONE* 8(3):e59033 doi:10.1371/journal.pone.0059033
- Galarza, J. A., G. F. Turner, E. Macpherson & C. Rico (2009). Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). *Canadian Journal of Fisheries and Aquatic Sciences* 66(9):1478-1490 doi:10.1139/f09-098.
- Heras, S., L. Planella, J. L. Garcia-Marin, M. Vera & M. I. Roldan (2019). Genetic structure and population connectivity of the blue and red shrimp *Aristeus antennatus*. *Scientific Reports* 9 doi:10.1038/s41598-019-49958-5.
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332.
- Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi: 10.1093/bioinformatics/btn129.
- Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr52.
- Levi, D., B. Patti, P. Rizzo, S. Lo Brutto, N. Parrinello & M. Arculeo (2004). Genetic and morphometric variations of Mediterranean hake, *Merluccius merluccius*, in the Strait of Sicily (central Mediterranean): implications for stock assessment of shared resources. *Italian Journal of Zoology* 71(2):165-170 doi:10.1080/11250000409356568
- Lischer HEL and Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28: 298-299.
- Lo Brutto, S., M. Arculeo, A. Mauro, M. Scalisi, M. Cammarata & N. Parrinello, 1998. Allozymic variation in Mediterranean hake *Merluccius merluccius* (Gadidae). *Italian Journal of Zoology* 65:49-52
- Lo Brutto, S., M. Arculeo & N. Parrinello (2004). Congruence in genetic markers used to describe Mediterranean and Atlantic populations of European hake (*Merluccius merluccius* L. 1758). *Journal of Applied Ichthyology* 20(2):81-86 doi:10.1046/j.1439-0426.2003.00514.x.
- Lo Brutto, S., T. Maggio, A. M. Deiana, R. Cannas & M. Arculeo (2012). Further investigations on populations of the deep-water blue and red shrimp *Aristeus antennatus* (Risso, 1816) (Decapoda, Dendrobranchiata), as inferred from Amplified Fragment Length Polymorphism (AFLP) and mtDNA analyses. *Crustaceana* 85(11):1393-1408 doi:10.1163/15685403-00003131.

- Lo Brutto, S., T. Maggio & M. Arculeo (2013). Isolation by distance (IBD) signals in the deep-water rose shrimp *Parapenaeus longirostris* (Lucas, 1846) (Decapoda, Penaeidae) in the Mediterranean sea. *Marine Environmental Research* 90:1-8 doi:10.1016/j.marenvres.2013.05.006.
- Lundy, C. J., P. Moran, C. Rico, R. S. Milner & G. M. Hewitt (1999). Macrogeographical population differentiation in oceanic environments: a case study of European hake (*Merluccius merluccius*), a commercially important fish. *Molecular Ecology* 8(11):1889-1898 doi:10.1046/j.1365-294x.1999.00789.x.
- Maggio, T., S. Lo Brutto, F. Garoia, F. Tinti & M. Arculeo (2009). Microsatellite analysis of red mullet *Mullus barbatus* (Perciformes, Mullidae) reveals the isolation of the Adriatic Basin in the Mediterranean Sea. *ICES Journal of Marine Science* 66(9):1883-1891 doi:10.1093/icesjms/fsp160
- Marcias, S., F. Sacco, A. Cau & R. Cannas (2010). Microsatellite markers for population genetic studies of the giant red shrimp *Aristaeomorpha foliacea* (Crustacea, Decapoda). *Rapp Comm int Mer Médit* 39:386.
- Marra, A., S. Mona, R. M. Sa, G. D'Onghia & P. Maiorano (2015). Population Genetic History of *Aristeus antennatus* (Crustacea: Decapoda) in the Western and Central Mediterranean Sea. *Plos One* 10(3):16 doi:10.1371/journal.pone.0117272
- Maroso, F., J. E. J. Hillen, B. G. Pardo, K. Gkagkavouzis, I. Coscia, M. Hermida, R. Franch, B. Hellemans, J. Van Houdt, B. Simionati, J. B. Taggart, E. E. Nielsen, G. Maes, S. A. Ciavaglia, L. M. I. Webster, F. A. M. Volckaert, P. Martinez, L. Bargelloni, and R. Ogden (2018). Performance and Precision of Double Digestion RAD (DdRAD) Genotyping in Large Multiplexed Datasets of Marine Fish Species. *Marine Genomics* 39:64–72. doi: <https://doi.org/10.1016/j.margen.2018.02.002>
- Mastretta-Yanes, A., N. Arrigo, N. Alvarez, T. H. Jorgensen, D. Piñero & B. C. Emerson (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* 15(1):28-41 doi:10.1111/1755-0998.12291.
- Matic-Skoko, S., T. Segvic-Bubic, I. Mandic, D. Izquierdo-Gomez, E. Arneri, P. Carbonara, F. Grati, Z. Ikica, J. Kolitari, N. Milone, P. Sartor, G. Scarcella, A. Tokac & E. Tzanatos (2018). Evidence of subtle genetic structure in the sympatric species *Mullus barbatus* and *Mullus surmuletus* (Linnaeus, 1758) in the Mediterranean Sea. *Scientific Reports* 8:14 doi:10.1038/s41598-017-18503-7
- Meirmans, P. G. & P. H. Van Tienderen (2004). genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4(4):792-794 doi:10.1111/j.1471-8286.2004.00770.x.
- Milano, I., M. Babbucci, F. Panitz, R. Ogden, R. O. Nielsen, M. I. Taylor, S. J. Helyar, G. R. Carvalho, M. Espiñeira, M. Atanassova, F. Tinti, G. E. Maes, T. Patarnello & L. Bargelloni (2011). Novel tools for conservation genomics: Comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. *PLoS ONE* 6(11) doi:10.1371/journal.pone.0028008
- Miller SA, Dykes DD, Polesky HF. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16(3):1215.
- Miller, J. M., R. M. Malenfant, P. David, C. S. Davis, J. Poissant, J. T. Hogg, M. Festa-Bianchet, and D. W. Coltman (2013). Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* 112:240.
- Mora-Márquez, F., García-Olivares, V., Emerson, B. C., & López de Heredia, U. (2017). ddradseqtools: a software package for in silico simulation and testing of double-digest RAD seq experiments. *Molecular Ecology Resources*, 17(2), 230-246.

- Nielsen, E., Cariani, A., Aoidh, E. et al. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications* 3, 851 (2012). <https://doi.org/10.1038/ncomms1845>
- Paris, J. R., J. R. Stevens & J. M. Catchen (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution* 8(10):1360-1373 doi:10.1111/2041-210X.12775.
- Peakall, R. & P. E. Smouse (2012). GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537-2539 doi:10.1093/bioinformatics/bts460.
- Pita, A., P. Presa & M. Perez (2010). GENE FLOW, Multilocus assignment and genetic structuring of the European hake (*Merluccius merluccius*). *Thalassas* 26(2):129-133.
- Pritchard, J. K., M. Stephens & P. Donnelly (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155(2):945.
- Roldan, M. I., J. L. Garcia-Marin, F. M. Utter & C. Pla (1998). Population genetic structure of European hake, *Merluccius merluccius*. *Heredity* 81:327-334 doi:10.1038/sj.hdy.6883830
- Roldán, M. I., S. Heras, R. Patellani & F. Maltagliati (2009). Analysis of genetic structure of the red shrimp *Aristeus antennatus* from the Western Mediterranean employing two mitochondrial regions. *Genetica* 136(1):1-4 doi:10.1007/s10709-008-9330-2.
- Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8(1):103-6 doi:10.1111/j.1471-8286.2007.01931.x.
- Sardà, F., C. Bas, M. I. Roldán, C. Pla & J. Leonart (1998). Enzymatic and morphometric analyses in mediterranean populations of the rose shrimp, *Aristeus antennatus* (Risso, 1816). *Journal of Experimental Marine Biology and Ecology* 221(1):131-144 doi:10.1016/S0022-0981(97)00119-6.
- Sardà, F., M. I. Roldán, S. Heras & F. Maltagliati (2010). Influence of the genetic structure of the red and blue shrimp, *Aristeus antennatus* (Risso, 1816), on the sustainability of a deep-sea population along a depth gradient in the Western Mediterranean. *Scientia Marina* 74(3):569-575 doi:10.3989/scimar.2010.74n3569.
- Tanner, S. E., M. Perez, P. Presa, S. R. Thorrold & H. N. Cabral (2014). Integrating microsatellite DNA markers and otolith geochemistry to assess population structure of European hake (*Merluccius merluccius*). *Estuarine Coastal and Shelf Science* 142:68-75 doi:10.1016/j.ecss.2014.03.010
- Zheng, X. W., D. Levine, J. Shen, S. M. Gogarten, C. Laurie & B. S. Weir (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24):3326-3328 doi:10.1093/bioinformatics/bts606.

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP1 - Population genetics and phylogeographic studies for identification of biological units of priority species

Deliverable 1.5.2 – Data acquired from comprehensive genetic studies for M. barbatus, N. norvegicus, and P. longirostris

LEAD: COSTAS TSIGENOPOULOS (HCMR) & LORENZO ZANE (CONISMA)

PARTICIPANTS: LUCA BARGELLONI, RITA CANNAS, ALESSIA CARIANI, LAURA CARUGATI, LEONARDO CONGIU, RACHELE CORTI, ALICE FERRARI, RAFAELLA FRANCH, SILVIA IORI, FRANCESCO MAROSO, ILARIA ANNA MARIA MARINO, RICCARDO MELIS, CHIARA PAPETTI, FEDERICA PIATTONI, ELISABETTA PIAZZA, FAUSTO TINTI (CONISMA)

DIMITRIS TSAPARIS, VASSO TERZOGLU, TEREZA MANOUSAKI, KATERINA EKONOMAKI, PETROULA BOTSIDOU, XENIA SARROPOULOU (HCMR)

Date: May 10th, 2021

CONTENTS

Executive Summary	3
Objective	6
Methodology	6
Sampling	6
DNA extraction	6
Library Preparations & Genotyping-By-Sequence (GBS)	6
Bioinformatic Analysis	7
Population genetic analyses	7
Overall Evaluation & Phylogeographic Assessment	9
<i>Mullus barbatus</i>	9
<i>Nephrops norvegicus</i>	20
<i>Parapenaeus longirostris</i>	29
Conclusions	39
General conclusion	42
References	43

Executive Summary

The Deliverable 1.5.2 comes as a continuation of the results already reported in Deliverable 1.5.1 and finalizes the comprehensive genetic work for the last three out of the six target species. Here, the totality of the locations sampled throughout the Mediterranean Sea was genetically analyzed. The total number of collected specimens from each species was DNA extracted, and after taking into account their DNA quality, the best specimens were included in double-digest Random Amplified DNA (ddRAD) libraries construction and analyzed (genotyped) on the HiSeq4000 Illumina sequencing platform.

Contrary to the deliverable 1.5.1 in which the ddRADseq protocol performed better in the fish species (*M. merluccius*) than in other two crustacean species (the giant red shrimp *Aristaeomorpha foliacea* and the blue and red shrimp *Aristeus antennatus*), in this last piece of work we see that the best results obtained were for the Norway lobster (*Nephrops norvegicus*) than for the other two species, a fish i.e. the red mullet (*Mullus barbatus*) and the deep water rose shrimp (*Parapenaeus longirostris*). For all three species, little genomic information exists and no previous attempts have been reported.

In the red mullet (*Mullus barbatus*), the complete study was conducted with specimens from 38 localities and 2,133 specimens sampled from the W. Mediterranean (GSA1b, N. Alboran Sea) to the South-Eastern Mediterranean Sea (GSA25c, 26 and 27). DNA quality was generally mediocre, with only one third of the samples providing good DNA quality whereas another one third was not used since the DNA quality was bad. The latter led us to finally not include into the final sample-set of 1,373 specimens for the ddRAD library preparation, samples from seven GSAs which had low DNA quality. Using all individuals genotyped, the species "catalogue" comprised 462,836 ddRAD loci, and filtering out samples with low number of stack loci (<3,000) and for SNPs present in at least 70% of the samples, we ran all phylogeographic analyses with a dataset composed of 771 samples (for 30 localities) and 853 high quality SNPs. Our results point out an evident lack of genetic differentiation and are generally in agreement with previous studies conducted at smaller and little more extensive sampling points range in the Mediterranean Sea. Therefore, using the ddRADseq approach which provided lots of hundreds of polymorphic markers from an extensive sampling plan from the West (Balearic Islands) to the East of the Mediterranean Sea (Cyprus and Egypt), we measured very low pairwise F_{st} values, the absence of clear population structure in the Mediterranean, and the greatest part of the identified genetic variation was attributed to differences among individuals in the populations, and much less among groups. In particular, the highest values encountered are found significant for the three major groups of the Western, the Central and the Eastern Mediterranean and less for the West to East Mediterranean differentiation; however, this differentiation is explaining only 1.8 and 1.3%, respectively, of the divergence between groups for the two above scenarios.

In the Norway lobster (*Nephrops norvegicus*), the comprehensive study included 1,537 specimens from 30 localities sampled from the W. Mediterranean (GSA1b, N. Alboran Sea) to the Aegean Sea (GSA22b) in the Eastern Mediterranean Sea. DNA extraction quality was surprisingly high for the shrimp species analyzed in the project (61.1%) and only 18.5% were not used since the DNA quality was bad. Therefore, we finally included 1,152 specimens for

the ddRAD library preparation, and unfortunately samples from two GSAs which had low DNA quality were not represented. The species "catalogue" using all individuals genotyped, comprised 1,682,988 ddRAD loci, and filtering out samples with low number of stack loci (<3,000) and for SNPs present in at least 70% of the samples, we ran all phylogeographic analyses with a dataset composed of 890 samples (for 27 localities) and 730 high quality SNPs. Our results reaffirm the ones we got from the pilot study and showed a significant differentiation of the samples eastern of the Adriatic Sea (GSA17) against the others from the central and Western Mediterranean Sea. Additionally, when testing for alternative scenarios relatively high and significant values were also encountered for the separation of the Adriatic Sea (GSA17 to 19) from the neighboring basins to the west (GSA1 to 11) and the east (GSA22). With practically few recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, current results are the first to show a considerable differentiation of the Adriatic and Eastern Mediterranean samples.

For the deep-water rose shrimp (*Parapenaeus longirostris*), a total of 1,750 specimens were sampled from 35 sampling locations. Quality of DNA extracts was poor and a total of 1,008 individuals from 34 locations were used for ddRAD. After filtering, a total of 1,225 SNPs was retained for 782 individuals. Pairwise F_{st} , a metric of population differentiation among pairs of samples, showed many significant pairwise differentiation, and the global F_{st} value was highly significant. Genetic clustering methods confirmed the existence of differentiation, indicating potential subdivisions of East Mediterranean samples and a slightly different cluster composition for samples West and East of the Strait of Sicily. As for other species analyzed in this project, much of the genetic variation was distributed among individuals in the populations, but a support for three groups was found: a "western-central" one including samples from Western and Central Mediterranean Sea up to the Strait of Sicily, a "central" one including the remaining samples from the Central Mediterranean Sea except the easternmost Ionian sample, and an "eastern" group that included the samples from Eastern Mediterranean Sea and the remaining Ionian. The results of the final study are in line with the results of the only previous study investigating populations genetics at the Mediterranean scale, which detected using mitochondrial DNA a deep differentiation between Aegean samples and Tyrrhenian Adriatic ones. Our results suggest that the major genetic breakpoint is indeed located in the Strait of Sicily and that a possible further subdivision is present at the boundary between Central and Eastern Mediterranean, though not coinciding with the boundary between Ionian and Aegean Sea. The existence of significant genetic differentiation should be taken into account in the future and additional investigations are needed particularly in the Sicily Strait area, where *Parapenaeus longirostris* is an important resource shared by different nations being the main target of trawl fisheries.

The application of ddRADseq to this comprehensive genetic study of the two crustacean species (*Parapenaeus longirostris* and *Nephrops norvegicus*) and one fish (*Mullus barbatus*) evidenced different levels of differentiation for the three species: higher in the Norway lobster which is something novel for the species, medium-high in deep-water rose shrimp, and very low for the red mullet. We remind that in the previous reported genetic study for the two aristeid decapod crustacean species (*Aristaeomorpha foliacea* and *Aristeus antennatus*) and the hake (*Merluccius merluccius*) also evidenced a much lower level of differentiation for the two crustaceans, that is in line with previous studies. This implies that to obtain solid support, especially in species with low genetic diversity and unknown genome size, a very high number of loci and a big sample is needed. Our study, considering all the six species

investigated (the ones included in this Deliverable in the previous D1.5.1), evidenced two critical steps that can increase the efficiency of the ddRAD sequencing: the first related to DNA quality, and the second to the optimization of the protocol for library preparation and sequencing. For DNA quality, a particular care should be taken during sample collection and preservation, to reduce the number of DNA extracts not suitable to ddRAD sequencing. For the optimization of the protocol for library preparation and sequencing we found that particular care should be devoted to DNA extraction itself, selection of restriction enzymes, narrower size-selection of fragments for high throughput sequencing, lower level of multiplexing, and/or stricter filtering for bioinformatic analysis. We advise, for future applications of ddRAD sequencing in new species, particularly crustaceans, to perform a small-scale optimization study to address these issues in advance.

Objective

The main objective of this deliverable is to report final results for last three out of the six species included in the study: for red mullet (*Mullus barbatus*), the Norway lobster (*Nephrops norvegicus*) and the deep water rose shrimp (*Parapenaeus longirostris*). Based on the knowledge acquired from the pilot genetic study (see D1.4) in which we optimized the laboratory protocols and bioinformatic pipelines, we processed the totality of the specimens sampled for each of the three species across the Mediterranean Sea, performed an unprecedented biogeographic analysis, always comparing the results for each species to those of previous genetic studies.

Methodology

A Genotyping-by-Sequencing (GBS) methodology constructing reduced-representation libraries in each species was selected. SNPs markers newly isolated following the double-digest Random Amplified DNA (ddRAD) sequencing are used. Full description of the methodology has been provided in the Inception, the 1st Progress Report, and deliverables 1.4 and 1.5.1.

Sampling

The number and location of sampling sites differed per species analysed and are detailed below in the following sections (Tables 1.5.2.1, 1.5.2.5 and 1.5.1.10).

DNA extraction

Standard DNA salt extraction protocols were followed and further details on the methodology can be found in deliverable 1.4 and 1.5.1. In brief, even though alternative extraction protocols based on commercial kits have also been employed, finally the salt extraction protocol (rather cheap but significantly time and labor-consuming approach) was selected to get high molecular weight (HMW) DNA. DNA extracts have been classified into three (3) quality categories ranging from 1 (high quality to be used in priority for ddRAD) to 3 (very poor quality and not usable for ddRAD). Since the pilot study, for sampled sub-GSAs which were previously identified of poor-quality samples, we succeeded to collect new specimens that were included in this analysis.

Library Preparations & Genotyping-By-Sequence (GBS)

Having evaluated and optimized the digestion with the appropriate pair of enzymes for each species, for the crustaceans we chose the enzymes *SbfI* - *NlaIII* (8-base cutter & 4-base cutter, respectively) in *N. norvegicus* and *PstI* - *NlaIII* (6-base cutter & 4-base cutter, respectively) in *P. longirostris*, while for *M. barbatus* the pair *SbfI*-*SphI* (8-base cutter & 6-base cutter, respectively).

Details on the concentration of enzymes used, amount of DNA, PCR-cycling conditions and gel cut-window chosen are provided in M1.3. Based on preliminary results obtained during the pilot study, the size selection procedure has been modified for *P. longirostris* to reduce

the number of loci. In this case, rather than classical gel-based size selection, a Blue Pippin Automated Size Selection machine has been used, fixing average size of recovered fragments to 600 bp (range 543-657 bp). In addition, for this species only, multiplexing was reduced to 144-159 individuals per library, instead of the original 288 multiplexing used in the pilot study, to increase read coverage.

Bioinformatic Analysis

Details on the methodology followed are provided in D1.4. Briefly, the raw FASTQ files were quality-checked in FastQC 0.11.3 and sequenced reads from each of the species were analyzed separately using STACKS v.2.4 pipeline (Catchen et al. 2013), with the surviving high quality forward reads of each sample were used for building *de novo* the loci of that particular sample with the STACKS component *ustacks*. Then, a catalogue of loci was built for each species using the STACKS component *cstacks*, then the loci of each individual were matched to the catalogue through *sstacks*, then we used *gstacks* to assemble and merge the second read of each pair, call variant sites and identify the genotype of each sample for each catalogue locus, and finally, populations was executed to produce the final input files for population genetic analysis. For each STACKS locus, only one randomly selected SNP was kept.

Population genetic analyses

Details on the methodology followed are provided in Deliverable 1.4. Briefly, estimates of genetic diversity within samples in terms of observed (H_o) and expected heterozygosity (H_e) were calculated for each geographical sample with Genetix (Belkiri et al 1996-2004). The level of differentiation among samples, pairwise and global F_{ST} values were calculated with Arlequin ver. 3.5.1.2 and corrections for multiple testings were performed with the Benjamini–Yekutieli procedure (Benjamini and Yekutieli 2001). With the same software, we performed the Analysis of Molecular Variance (AMOVA), grouping the samples on the basis of *a priori* hierarchical geographical structure (Arlequin suite ver. 3.5, Excoffier & Lischer, 2010). The statistical significance of the resulting values was estimated by comparing the observed distribution with a null distribution generated by 10,000 permutations, in which individuals were redistributed randomly into samples. For each species a different threshold allowed for missing data was set. Specifically, for *P. longirostris* value of 0.5 was set, due to the quality of datasets.

The Discriminant Analysis of Principal Components (DAPC) was performed with Adegenet ver. 2.1.1 (R version 3.5.1, R Development Core Team, 2014; <http://www.r-project.org>), the DAPC scatterplots results were visualized graphically and the best supported number of clusters through comparison of the Bayesian Information Criterion (BIC) for the different values of K was identified.

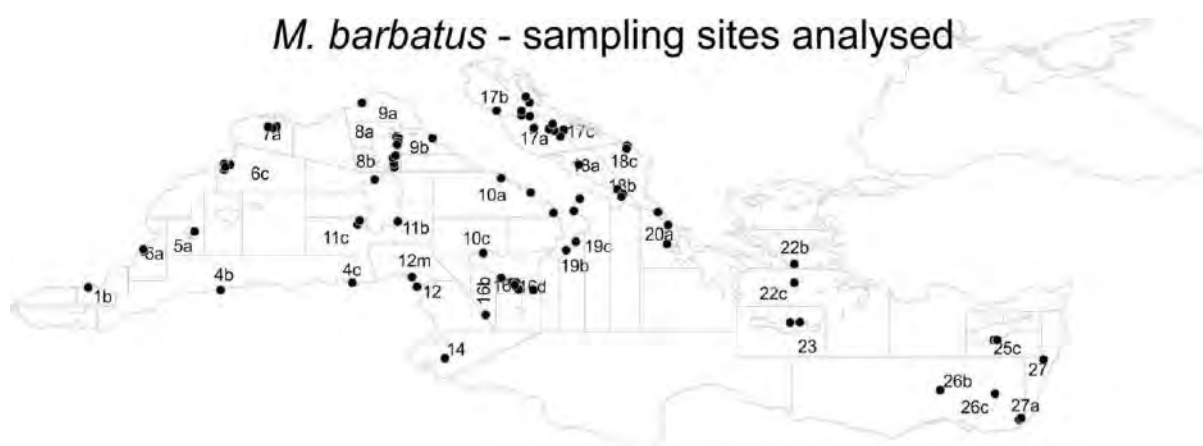
Finally, with the STRUCTURE v 2.3.4 software (Pritchard et al., 2000; Falush et al 2003, 2007; Hubisz et al., 2009) we evaluated the presence of clusters within the sampled populations, running the software with k from 1 to 10. When k was >1 based on likelihood,

the best k value was selected using the EVANNO method (Evanno et al. 2005) as implemented in Structure Harvester (Earl & vonHoldt 2012).

Overall Evaluation & Phylogeographic Assessment

Mullus barbatus

For the red mullet (*M. barbatus*), the complete study was conducted with specimens from 38 localities (GSAs and sub-GSAs). In total, 2,133 specimens were sampled from the N. Alboran Sea (1b) in the W. Mediterranean to the Cypriot (25c) and the South and Eastern Levant Sea (26 and 27) waters in the E. Mediterranean.



DNA extraction was done as previously described and a little more than one third of the samples provided DNA of good quality (35.9%) whereas 32.7% of the samples were of medium quality and 31.4% were not used since the DNA quality was bad (see Table 1.5.2.1).

Table 1.5.2.1. DNA quality across samples in *M. barbatus*

GSA	Total number of samples	DNA quality			Number of samples in Libraries
		Good	Medium	Bad	
01b	48	48	0	0	48
04b	50	0	13	37	13
04c	50	0	4	46	0
05a	50	47	3	0	47
06a	50	48	2	0	48
06c	100	46	50	4	94
07a	51	48	2	1	48
08ab	49	23	26	0	48
09a	50	10	6	34	16
09b	50	6	44	0	50
10a	50	0	33	17	33

10bc	56	0	52	4	52
11b	50	0	43	7	43
11c	50	0	46	4	46
11e	50	0	0	50	0
12a	50	40	6	4	46
12m	74	48	19	7	48
14a	50	10	19	21	29
16b	50	0	0	50	0
16c	50	0	14	36	0
16d	49	0	2	47	0
17a	50	1	49	0	50
17b	50	0	50	0	50
17c	50	44	6	0	50
18a	50	25	25	0	46
18b	50	0	37	13	37
18c	100	47	2	51	47
19b	100	85	13	2	89
19c	50	24	20	6	40
20a	59	9	7	43	16
22b	63	0	4	59	0
22c	50	0	0	50	0
23a	57	0	48	9	47
25c	77	5	35	37	40
26b	50	41	9	0	41
26c	50	40	8	2	40
27a	50	49	0	1	49
27b	50	22	0	28	22
Sum	2,133	766	697	670	1,373

Specimens from all sampled GSAs were included in the ddRAD library preparation (Table 1.5.2.2) at various numbers ranging from 13 (04b) to 94 (06c), with the exception of seven (7) GSAs that had samples which resulted in consistently low DNA quality (04c, 11e, 16b, 16c, 16d, 22b, and 22c).

Table 1.5.2.2. Number of specimens (in total) in each library for *M. barbatus*

1st library(pilot)	2nd library	3rd library	4th library	5th library
288 specimens	288 specimens	288 specimens	288 specimens	221 specimens
medium quality DNA	good quality DNA	medium to good quality DNA	medium to good quality DNA	medium quality DNA

All samples were analyzed following a ddRAD protocol, with enzymes *SbfI* (CCTGCA[^]GG), and *SphI* (GCATG[^]C), selected fragments from around 300 to around 600 base pairs. Multiplexed libraries were run on five Illumina lanes, with 150 bp paired-end protocol. When R1 & R2 reads were merged, we obtained a total of 3 billion reads.

After demultiplexing, on average 2.17 M reads per sample were obtained (range from 1,373 to 59M!). The number of Ustacks loci (considering a stack of unique reads) had an average of 3,770 but varied significantly among specimens (from only 6 to 36,590) and among populations on average (425 in 04b to 7,990 in 18b) (see Figures 1.5.2.1, 1.5.2.2 and Table 1.5.2.3).

Table 1.5.2.3. Stacks pipeline output summary for *M. barbatus*. With blue and red numbers, we show the higher and lower values, respectively.

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
01b	4186249	9379	59166515	5951	156	36590
04b	77023	5094	701457	425	30	2724
05a	1248373	6318	6244600	3162	65	8155
06a	1267548	68943	6211299	4194	990	8351
06c	341054	666	5156052	806	6	6265
07a	2481193	5861	32190811	4819	84	26791
08ab	1317144	23769	12693545	2224	79	9272
09a	1676217	20919	9931048	3938	235	11611
09b	1776197	13892	34414012	2363	64	16636
10a	4217368	307447	33285506	7103	2891	22497
10bc	6131796	74430	51073840	5101	253	17025
11b	1329827	17266	5113239	3988	229	8515
11c	3240896	24388	40883238	3199	103	19221
12a	1453976	15089	8954655	2519	91	6757
12m	1374627	32174	9529845	4121	561	12350
14a	2129052	283549	6734691	5281	2317	9132
17a	2934908	12113	15225479	3660	51	10057
17b	2665168	27678	133258411	3117	131	10714
17c	1620294	78547	9115295	2472	316	7501
18a	2081036	12614	20888213	2451	60	112751
18b	5010729	656194	25582915	7990	3665	17498
18c	954726	3658	5987719	3475	36	8210
19b	1337084	14965	7229957	2865	62	6730
19c	2584116	49268	17104320	4535	482	12333
20a	500445	14899	2893114	2514	189	5992
23a	1899756	30069	4672672	4840	455	7641
25c	2260470	241269	5613376	4828	1840	7175
26b	5014835	41241	23059405	6751	131	13077
26c	773671	88461	2092704	2945	792	4660
27a	1997359	49572	13740988	4704	780	14258
27b	1750238	187716	3627535	3883	1533	5533

ALL	2181721,77	666	133258411	3878,194	6	112751
-----	------------	-----	-----------	----------	---	--------

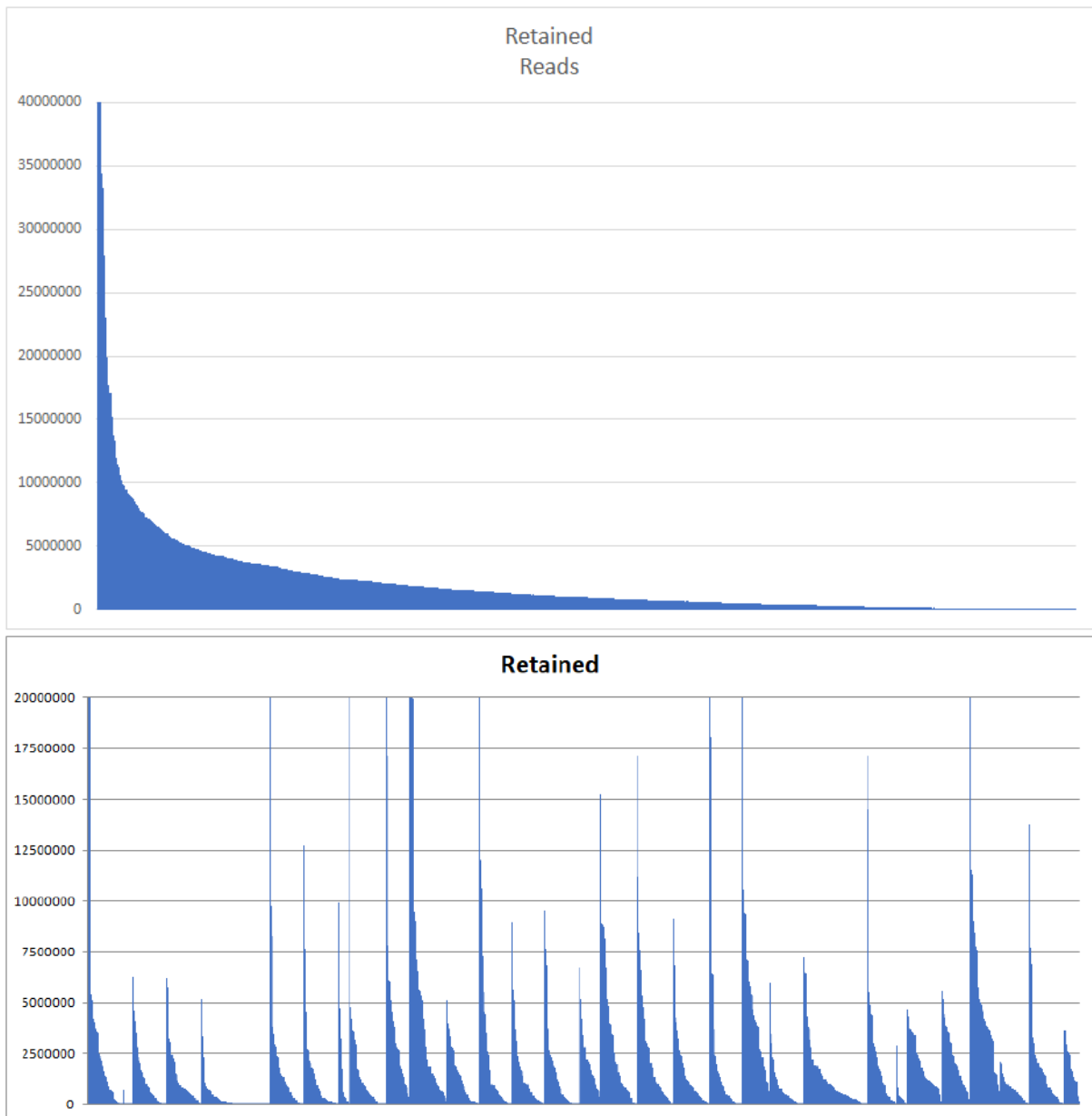


Fig. 1.5.2.1: Retained reads after demultiplexing for (A) the whole *M. barbatus* dataset, and (B) each population separately

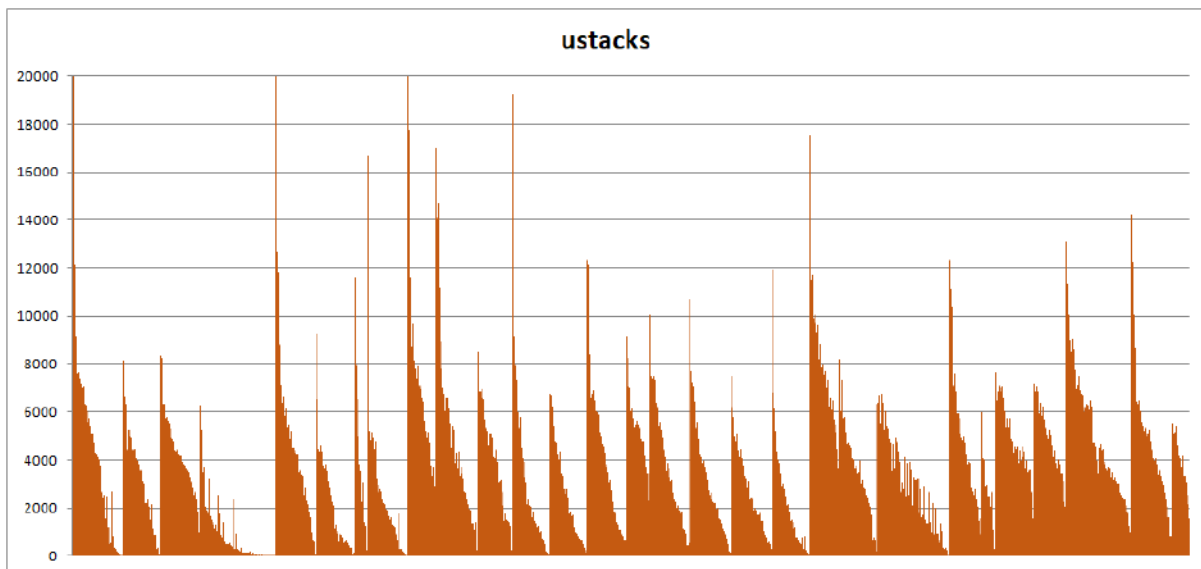
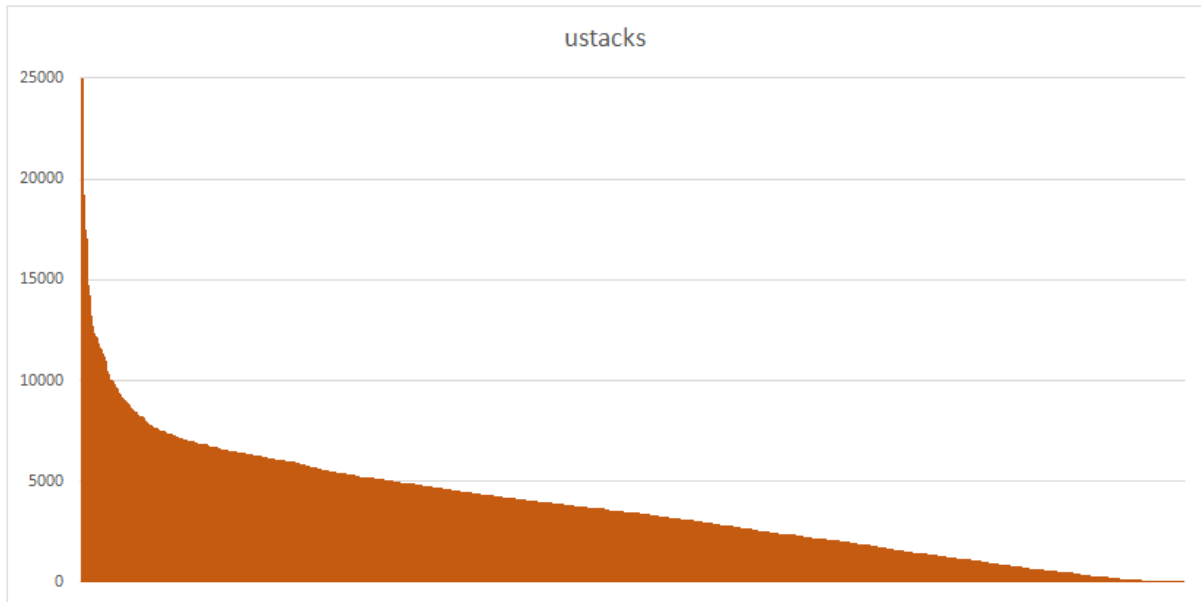


Fig. 1.5.2.2: ustacks loci for the (A) whole *M. barbatus* dataset, and (B) each population separately

Stacks results:

Using all individuals, we constructed the «catalogue» which comprised 462,836 ddRAD loci.

After several steps described in the document Milestone 1.3, we proceeded forming two datasets: the first by excluding all samples with less than 2,000 stack loci (“relaxed” dataset with 953 samples from 31 populations), and a second excluding all samples with less than 3,000 loci (“stringent” dataset with 771 samples from 30 populations). We then run the analyses to filter for SNPs present in at least 70% of the samples to finally have a panel of 615 and 853 higher quality SNPs, respectively, which were used for all downstream differentiation analysis, with the following characteristics:

- Number of loci with PE contig: 1236.00 (100.0%)
- Mean length of loci: 260.53bp (stderr 1.57)
- Number of loci with SE/PE overlap: 548.00 (44.3%)
- Mean length of overlapping loci: 223.85bp (stderr 1.42); mean overlap: 28.74bp (stderr 0.23)
- Mean genotyped sites per locus: 265.33bp (stderr 1.47).

Below, we show only the results of the more “stringent” analysis for the sake of simplicity.

To that purpose too, due to the very low number of specimens in some samples we integrated them into the geographically closest one: this is the case for GSA6c (6 specimens into the 6a), GSA9a (9 specimens into the 9b), GSA18a (12 specimens into the 18b), and GSA20a (4 specimens into the 19c) (see Table 1.5.2.4).

Observed heterozygosities were generally low (ranging from 11.8 to 19.7%, mean 16.8%) and smaller than the expected (24.7 to 31.2%) (Table 1.5.1.4). Low H_o values are found, not exclusively, in samples with low number of specimens which seem to have also a smaller number of polymorphic loci from the dataset of 853 SNPs. In the pilot study, observed heterozygosities were lower (average 12%) but interestingly, the expected heterozygosities were higher than those estimated in the final analysis (32.8 versus 26.6% on average).

Table 1.5.2.4: Summary of genetic variability estimates across sampling locations in *M. barbatus*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e). With red, we show the combined samples from the fusion of two neighboring ones with small sample sizes finally.

sub-GSAs	Samples in the Analysis	H_o	H_e
1b	34	0.1958	0.2576
5a	24	0.1788	0.2589
6ac	45	0.1778	0.2539
7a	34	0.1894	0.2526
8ab	16	0.1344	0.3020
9ab	21	0.1563	0.2768
10a	30	0.1957	0.2558
10bc	36	0.1371	0.2557
11b	29	0.1794	0.2594
11c	17	0.1462	0.3122
12a	16	0.1333	0.3044
12m	29	0.1961	0.2546
14a	27	0.1966	0.2576
17a	21	0.1178	0.2973
17b	28	0.1241	0.2775
17c	19	0.1392	0.2849
18ab	49	0.1897	0.2477

18c	30	0.1820	0.2550
19b	39	0.1469	0.2527
19c20a	30	0.1754	0.2544
23a	44	0.1876	0.2482
25c	37	0.1765	0.2574
26b	40	0.1892	0.2465
26c	20	0.1640	0.2788
27a	38	0.1897	0.2512
27b	18	0.1779	0.2656
Sum	771		

All populations showed a great similarity between them. This is evident in the F_{st} values which ranged from negative to 0.0096 (Fig. 1.5.2.3). All values were statistically not significant, either using an uncorrected probability threshold of 0.05 or a Benjamini-Hochberg correction. Even when we tried to group subGSAs with a small number of specimens (Table 1.5.2.4) in order to form populations with larger numbers of individuals, the F_{st} values remained low and not significant.

The comparison of pairwise F_{st} results of final analysis (based on 853 polymorphic SNPs with less than 30% missing data per locus in the data set) to those from the pilot study (based on 580 polymorphic SNPs but with less than 50% missing data per locus in the data set) indicate similar patterns of low differentiation. The six populations analyzed in the pilot study (GSA06c, 09b, 11c, 17a, 17b, 18a) had shown a great similarity between them (F_{ST} values from negative to 0.0049 between 11c and 18a). In the final set, the F_{st} values between the above-mentioned populations ranged from negative to 0.0115 (11e-18a).

Matrix of pairwise F_{ST}

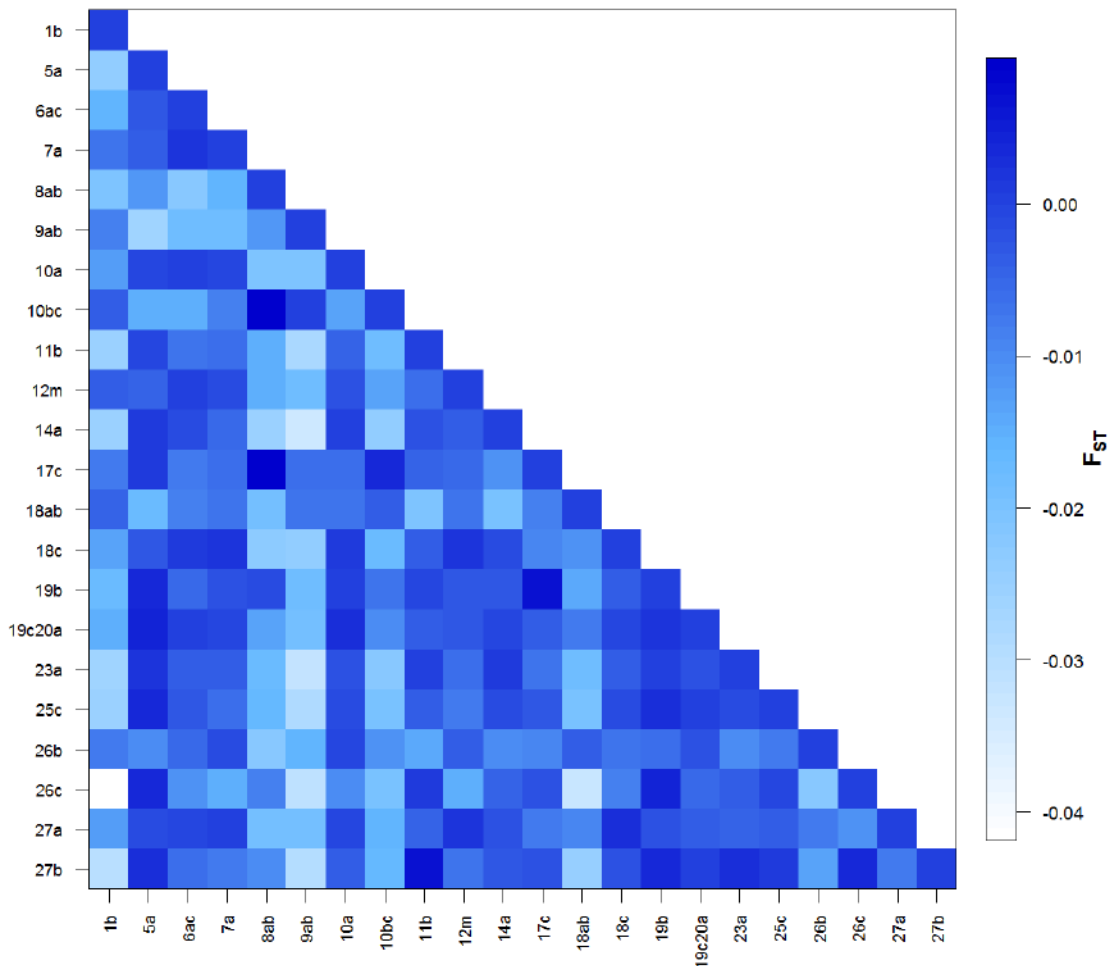


Fig. 1.5.2.3: Matrix of pairwise F_{ST} values for *M. barbatus* based on 853 polymorphic SNPs with less than 30% missing data per locus in the data set (22 populations - 689 specimens). No value was statistically significant after Benjamini– Hochberg correction.

Concerning the AMOVA results and various scenarios of groupings that checked (table not provided), the variation among groups was in most cases negative and not significant (p value > 0.05) indicating lack of significant differentiation.

Table 1.5.2.5: AMOVA results for several scenarios of grouping *M. barbatus* samples

	Scenario 1 - Two groups	Scenario 2 - Three groups	Scenario 3 - Four groups
Group 1	West 1b, 5a, 6ac,7a, 8ab, 9ab,10a, 10bc,11b, 12m,	West 1b, 5a, 6ac,7a, 8ab, 9ab,10a, 10bc,11b, 12m,	West 1b, 5a, 6ac,7a, 8ab, 9ab,10a, 10bc,11b, 12m,
Group 2	East 14a,17c,18ab, 18c,19b, 19c20a,23a,25c, 26b,26c,27a,27b	Adriatic-Ionian 14a,17c,18ab, 18c,19b,19c20a	Ionian 14a,19b,19c20a
Group 3		East 23a,25c, 26b,26c,27a,27b	Adriatic 17c,18ab, 18c
Group 4			East 23a,25c, 26b,26c,27a,27b
Variation %			
Among groups	0.0135	0.0184	-0.0025
Among populations within groups	0.8023	0.7966	0.8113
Within populations	99.1842	99.1850	99.1913
Fixation Indices			
FST	0.00816	0.00815	0.00809
p-value	<0.001	<0.001	<0.001
FSC	0.00802	0.00797	0.00811
p-value	<0.001	<0.001	<0.001
FCT	0.00014	0.00018	-0.00003
p-value	0.051	0.017	ns

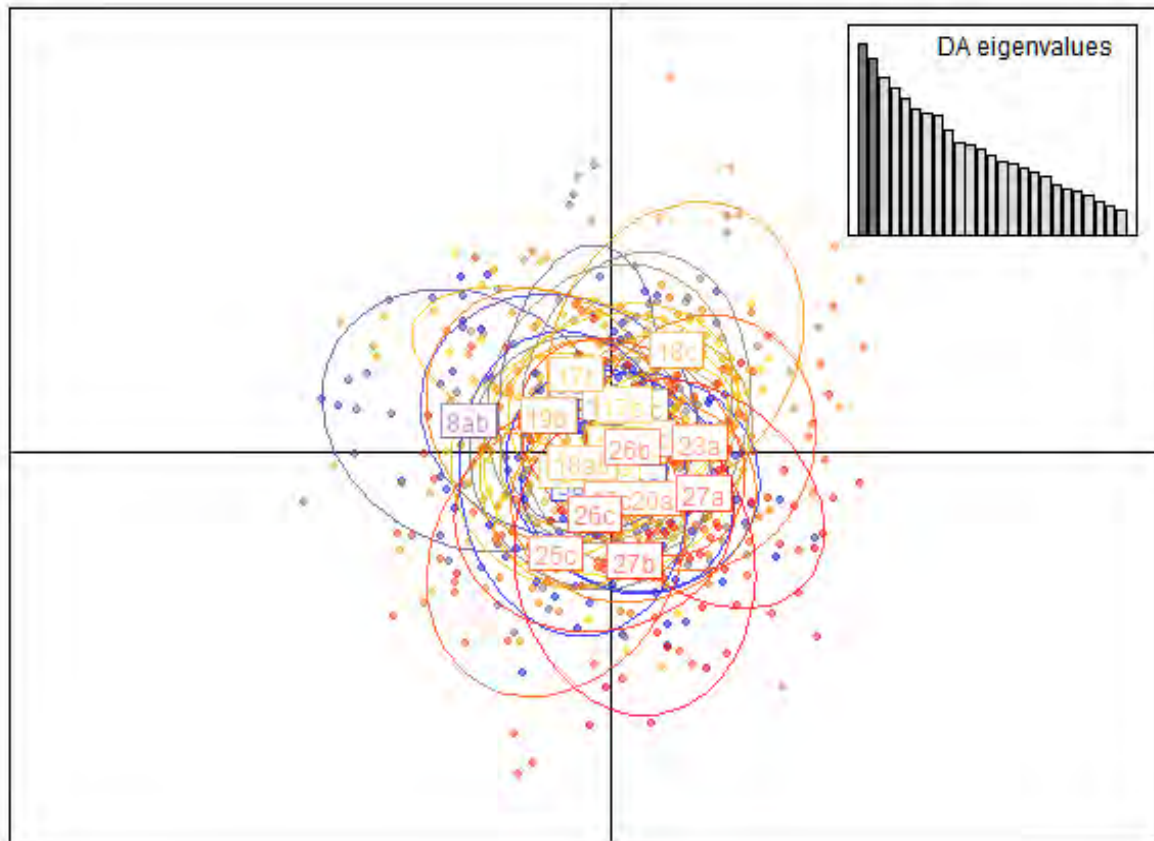


Fig. 1.5.2.4: DAPC plot (retained 100 PCA axes) for *M. barbatus* based on 853 polymorphic SNP present in >70% in the whole data set (771 individuals).

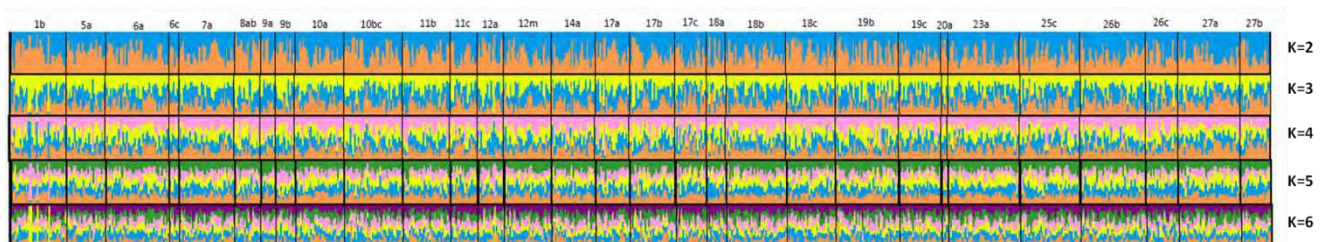


Fig. 1.5.2.5: Proportions of membership to each of K inferred clusters for *M. barbatus* individuals

The red mullet is one of the species processing rich bibliography for phylogeographic studies in the Mediterranean Sea.

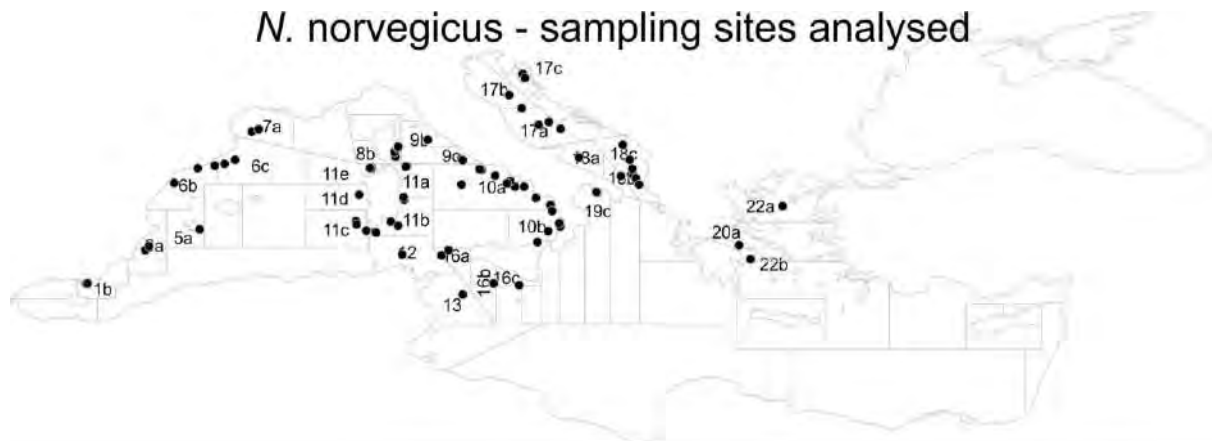
Previous genetic analyses have indicated low genetic diversity between Ionian and Aegean samples (Mamuris et al., 1998a; 1998c; 2001), Spanish samples (Felix-Hackradt et al., 2013) or even from France to the S. Aegean (Arculeo et al., 1999). Subtle but statistically significant genetic differentiation was however reported by Garoia et al (2004) between some Adriatic red mullet populations and by Apostolidis et al. (2009) between some Aegean samples. Galarza et al (2009) in the first performed wide-sampling study argued for a high

genetic similarity between Italian and Greek populations, whereas Levantine ones from Turkey and Catalan and Almerian ones being the most distinct. Interestingly, Maggio et al (2009) report a clear separation of the Adriatic samples from all the other neighboring populations; however, mean F_{st} value was very low ($F_{st} = 0.003$; $p < 0.001$). Last, in the more recent and marker-rich genetic study conducted so far by Matic-Skoko et al. (2018), the differentiation pattern was quite complex as they identified a break in gene flow toward the north and middle Adriatic regions and further to the south, including the Ionian Sea. But no significant pairwise differentiations were observed for populations within the eastern Mediterranean and reduced gene flow was described between the western Mediterranean region (Balearic Sea) and the Adriatic Sea, but not between the western and E. Mediterranean.

When current results are compared to those previously reported in the literature (see above and in the Deliverable 1.2), we do not encounter the differentiation of the Adriatic Sea (present samples 17a and 17b) from the neighboring regions reported in Maggio et al. (2009) and Matic-Skoko et al. (2018).

Nephrops norvegicus

For the Norway lobster (*Nephrops norvegicus*), the complete study was conducted with specimens from 30 localities (GSAs and sub-GSAs). In total, 1,537 specimens were sampled from the N. Alboran Sea (1b) in the W. Mediterranean to the Aegean Sea (22b) waters in the E. Mediterranean.



DNA extraction was conducted as previously described and more than half of the samples provided DNA of good quality (61.1%) whereas 20.4% of the samples were of medium quality and only 18.5% were not used since the DNA quality was bad (see Table 1.5.2.5).

Table 1.5.2.5. DNA quality across samples in *N. norvegicus*

GSA	Total number of samples	DNA quality			Number of samples in Libraries
		Good	Medium	Bad	
01b	50	43	3	4	46
05a	50	25	17	8	42
06a	50	48	2	0	46
06b	100	60	19	21	74
06c	100	62	12	26	68
07a	48	41	7	0	41
08ab	50	50	0	0	50
09b	50	50	0	0	50
09c	50	21	22	7	21
10a	43	42	1	0	42
10b	35	26	5	4	26
11a	50	50	0	0	47
11b	15	15	0	0	15

11c	88	69	7	12	76
11d	50	48	0	2	48
11e	50	50	0	0	50
12	50	0	32	18	32
13	50	0	31	19	31
16a	50	16	17	17	28
16b	50	3	24	23	27
16c	39	17	19	3	34
17ab	50	47	3	0	46
17c	50	30	8	12	38
18a	50	0	8	42	0
18b	60	58	2	0	44
18c	50	22	26	2	43
19c	48	42	5	1	47
20a	48	0	19	29	19
22a	36	5	16	15	21
22b	27	0	8	19	0
Sum	1,537	940	313	284	1,152

Specimens from all sampled GSAs were included in the ddRAD library preparation (Table 1.5.2.6) at various numbers ranging from 15 (11b) to 76 (11c), with the exception of only two (2) GSAs that had samples which resulted in consistently low DNA quality (18a and 22b).

Table 1.5.2.6. Number of specimens (in total) in each library for *N. norvegicus*

1st library(pilot)	2nd library	3rd library	4th library
288 specimens	288 specimens	288 specimens	288 specimens
medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA

All samples were analyzed following a ddRAD protocol, with enzymes *SbfI* (CCTGCA[^]GG), and *NlaIII* (CATG[^]), selected fragments from around 300 to around 700 base pairs. Multiplexed libraries were run on five Illumina lanes, with 150 bp paired-end protocol. When R1 & R2 reads were merged, we obtained a total of 2.5 billion reads.

After demultiplexing, on average 2.17 M reads per sample were obtained (range from 4,584 to 20M). The number of Ustacks loci (considering a stack of unique reads) had an average of 11,281 but varied significantly among specimens (from only 38 to 58,525) and among populations on average (273 in 16b to 23,603 in 11e) (see Figures 1.5.2.6, 1.5.2.7 and Table 1.5.2.7).

Table 1.5.2.7. Stacks pipeline output summary for *N. norvegicus*. With blue and red numbers, we show the higher and lower values, respectively.

	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
01b	1406649	50065	4625279	10726	530	45654
05a	1037544	16385	7185843	4102	143	172306
06a	1540483	61241	5297680	8640	864	22196
06b	3067420	105625	14028976	10944	670	27007
06c	4349283	26453	20609713	12855	744	36853
07a	1244929	20942	6759357	7816	273	25526
08ab	2968118	119371	5065412	20886	14334	30583
09b	2356807	490916	4933509	18275	8989	30481
09c	1061316	24382	3390985	6908	261	17459
10a	2663775	180738	7029895	13710	2292	25153
10b	1880628	192446	5078257	11054	2395	20722
11a	1681646	15307	3808582	10218	142	19136
11b	1776229	5987	11893060	5199	38	30120
11c	2278388	6938	11214867	15456	73	41906
11d	1612204	59727	6949053	8414	640	29924
11e	3655603	748193	5886414	23603	10821	32442
12	1075641	52570	8632424	4194	309	21722
13	1124480	49792	4736122	5100	306	14526
16a	136600	9929	1043429	850	61	4779
16c	340871	5509	2419250	2464	41	13784
16b	33030	4584	212475	273	52	1296
17ab	782885	15025	3260207	8910	177	22198
17c	4166393	115514	1855461	16971	1418	58525
18b	3685164	609976	6362880	17374	5846	24712
18c	2231460	127565	8017457	12128	1336	31208
19c	2858837	22413	16172997	13145	256	58100
20a	885544	49767	4745076	3809	388	14099
22a	1633099	55333	4545719	9835	809	22033
ALL	1911965,21	4584	20609713	10137,82	38	172306

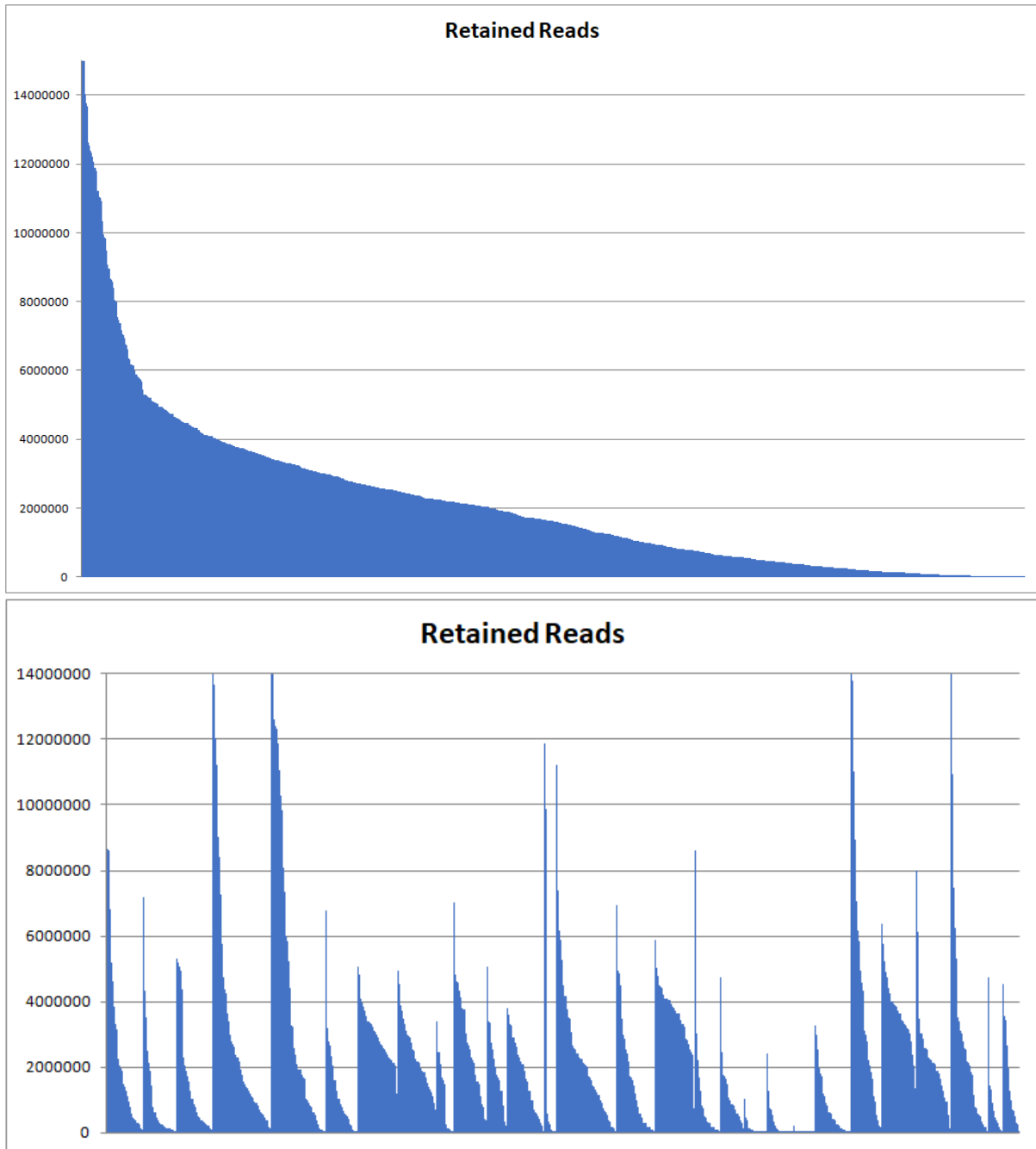


Fig. 1.5.2.6: Retained reads after demultiplexing for (A) the whole *N. norvegicus* dataset, and (B) each population separately

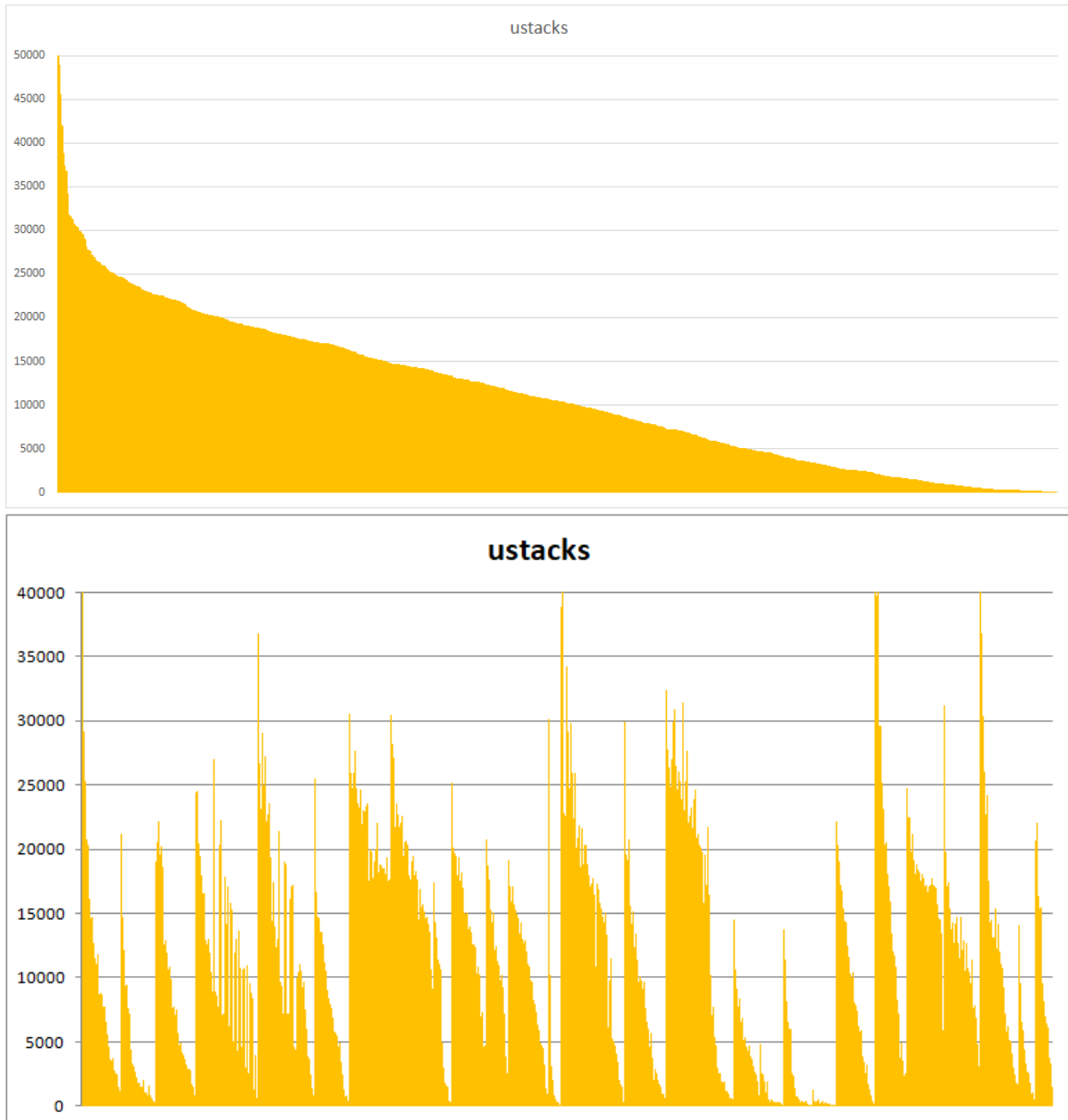


Fig. 1.5.2.7: ustacks loci for the (A) whole *N. norvegicus* dataset, and (B) each population separately

Stacks results:

Using all individuals, we constructed the «catalogue» which comprised 1,682,988 ddRAD loci.

After several steps described in the document Milestone 1.3, we proceeded forming two datasets: the first by excluding all samples with less than 2,000 stack loci (“relaxed” dataset with 948 samples from 27 populations), and a second excluding all samples with less than 3,000 loci (“stringent” dataset with 890 samples from 27 populations). We then run the analyses to filter for SNPs present in at least 70% of the samples to finally have a panel of

586 and 730 higher quality SNPs, respectively, which were used for all downstream differentiation analysis, with the following characteristics:

- Number of loci with PE contig: 1402.00 (100.0%)
- Mean length of loci: 359.67bp (stderr 1.59)
- Number of loci with SE/PE overlap: 58.00 (4.1%)
- Mean length of overlapping loci: 381.53bp (stderr 2.52); mean overlap: 22.86bp (stderr 0.30)
- Mean genotyped sites per locus: 260.60bp (stderr 0.30).

Below, we show only the results of the more “stringent” analysis for the sake of simplicity.

Observed heterozygosities were among the highest encountered in the whole study (ranging from 0.2474 to 0.2867) but remained smaller than the unbiased expected (0.3381 on average).

Table 1.5.2.8: Summary of genetic variability estimates across sampling locations in *N. norvegicus*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e). With red, we show the combined samples from the fusion of two neighboring ones with small sample sizes finally.

sub-GSAs	Samples in the Analysis	H_o	H_e
1b	38	0.1587	0.2675
5a	16	0.1211	0.3262
6a	37	0.1510	0.2612
6b	67	0.1577	0.2547
6c	60	0.1545	0.2540
7a	32	0.1518	0.2654
8ab	50	0.1987	0.2439
9b	50	0.2173	0.2508
9c	12	0.1988	0.3082
10a	41	0.1802	0.2573
10b	23	0.1764	0.2690
11ab	47	0.1603	0.2598
11c	68	0.1843	0.2454
11d	33	0.1572	0.2566
11e	50	0.1963	0.2432
12	11	0.1511	0.3680
13	23	0.0811	0.3391
16ac	10	0.1616	0.3594
17ab	35	0.2109	0.2634
17c	34	0.1507	0.2610
18b	44	0.19913	0.2645

18c	42	0.14888	0.2618
19c	39	0.14927	0.2558
20a	9	0.15661	0.4308
22a	19	0.15378	0.2873
Sum	890		

In the Norway lobster, samples showed a moderate similarity between them; the F_{ST} values ranged from negative to >0.04 (Fig. 1.5.2.8) with the highest observed between the Adriatic and most Eastern ones and the remaining westwards (deeper blue). The latter is more evident in both the DAPC and STRUCTURE plots (see Figs. 1.5.2.9 and 1.5.2.10) with a clear distinction of the populations up to GSA16c and the other eastwards; the inclusion of GSA20a, in the western group in the DAPC might be just an artefact due to the small sample size finally included in the analysis (9 specimens). Moreover, this result holds true irrespectively of the number of populations included in the analyses even when we group samples with a small number of specimens finally included in the analysis.

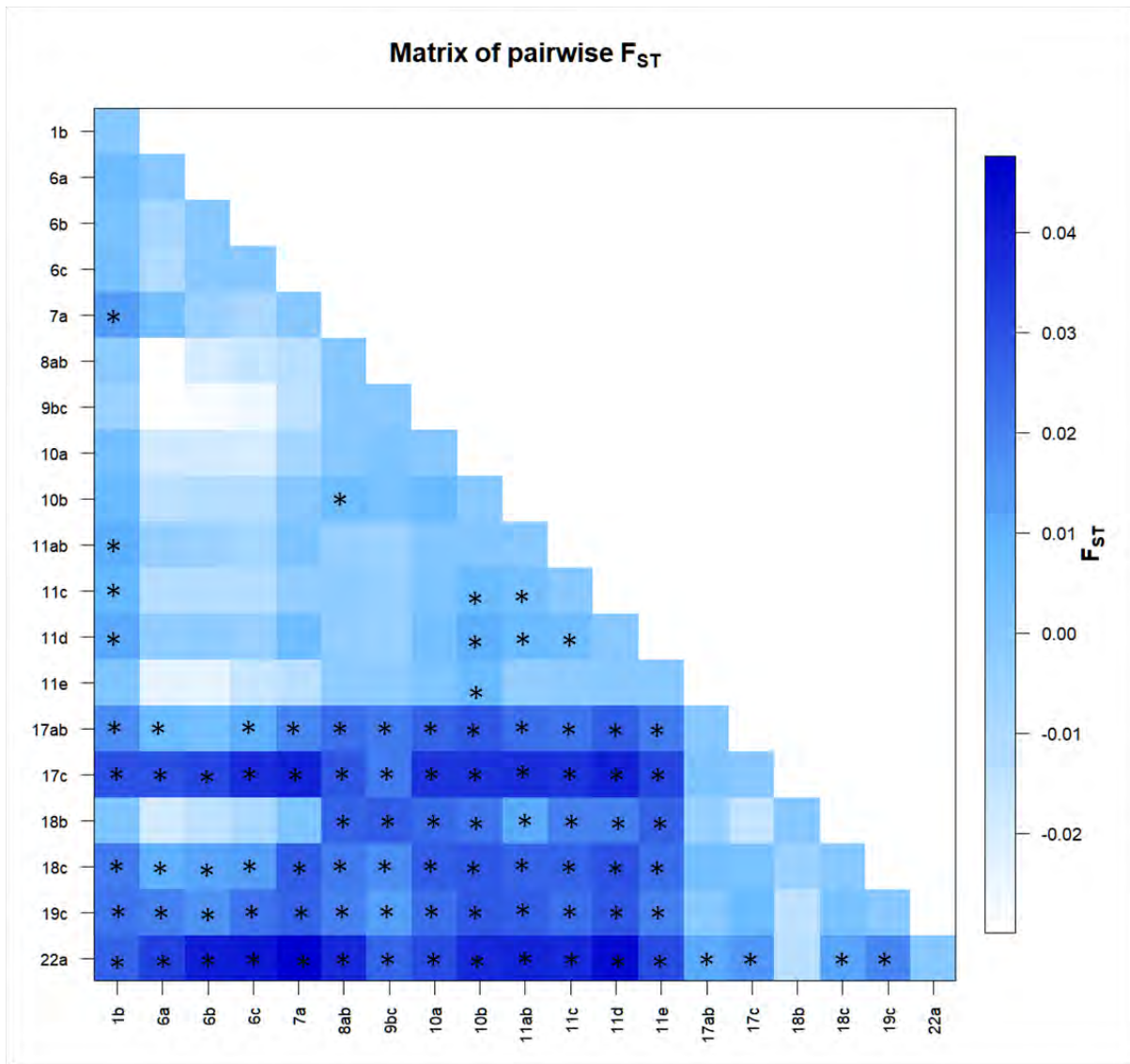


Fig. 1.5.2.8: Matrix of pairwise F_{st} values for *N. norvegicus* based on 730 polymorphic SNPs with less than 30% missing data per locus in the data set (19 populations - 821 specimens). Values with asterisk are statistically significant after Benjamini–Hochberg correction.

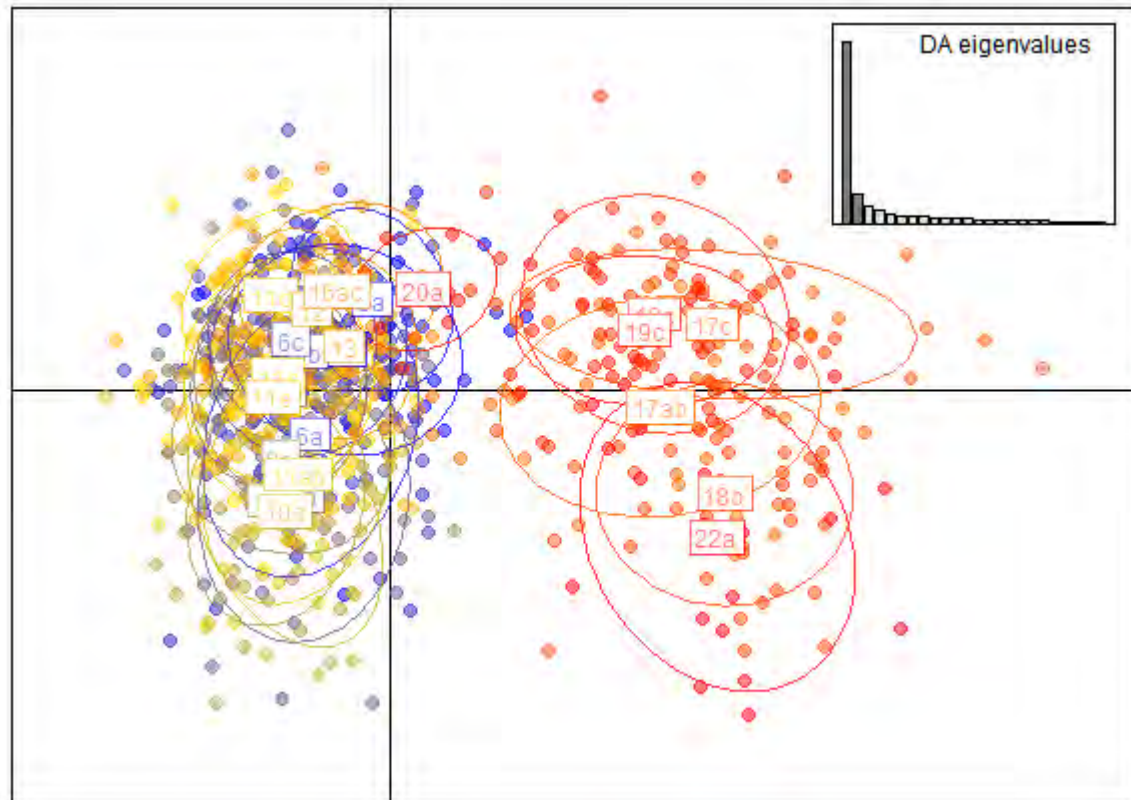


Fig. 1.5.2.9: DAPC plot (retained 100 PCA axes) for *N. norvegicus* based on 730 polymorphic SNP present in >70% of specimens in the whole data set (890 individuals).

Additionally, AMOVA results for several grouping scenarios in the Norway lobster samples clearly indicated a significant differentiation between the West (up to GSA11e) and the Eastern Mediterranean (scenario 1, Table 1.5.2.9). In this scenario testing, Adriatic samples seem to have a key-position showing their distinctiveness and show differentiation when they are grouped together (17ab to 19c, scenario 4) or just up to 18c (scenario 2).

Table 1.5.2.9: AMOVA results for several scenarios of grouping *N. norvegicus* samples.

	Scenario 1 - Two groups	Scenario 2 - Three groups	Scenario 3 - Five groups	Scenario 4 - Three groups
Group 1	West 1b,6a,6b,6c,7a, 8ab, 9bc,10a, 10b,11ab, 11c,11d,11e	West 1b,6a,6b, 6c,7a, 8ab,9bc,10a, 10b,11ab, 11c,11d,11e	West 1b,6a,6b, 6c	West 1b,6a,6b, 6c,7a, 8ab, 9bc,10a, 10b,11ab, 11c,11d,11e

Group 2	East 17ab,17c,18b, 18c,19c,22a	Adriatic 17ab,17c,18b, 18c	Balearic 7a,8ab, 11c,11d,11e	Adriatic 17ab,17c,18b, 18c,19c
Group 3		East 19c,22a	Italian 9bc,10a, 10b,11ab,	Aegean 22a
Group 4			Adriatic 17ab,17c,18b, 18c	
Group 5			East 19c,22a	
Variation %				
Among groups	2.3479	2.2182	1.3577	2.4993
Among populations within groups	0.9124	0.9079	0.8016	0.8231
Within populations	96.7396	96.8739	97.8406	96.6776
Fixation Indices				
FST	0.03260	0.03126	0.02159	0.03322
p-value	<0.001	<0.001	<0.001	<0.001
FSC	0.00934	0.00929	0.00813	0.00844
p-value	<0.001	<0.001	<0.001	<0.001
FCT	0.02348	0.02218	0.01358	0.02499
p-value	<0.001	<0.001	<0.001	<0.001

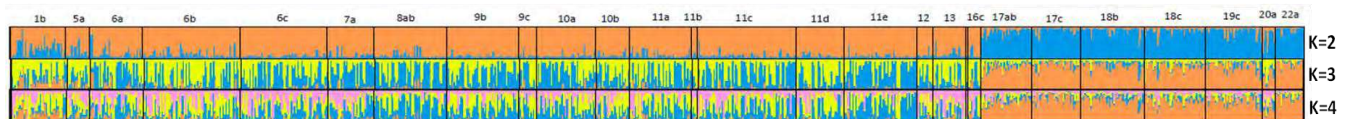
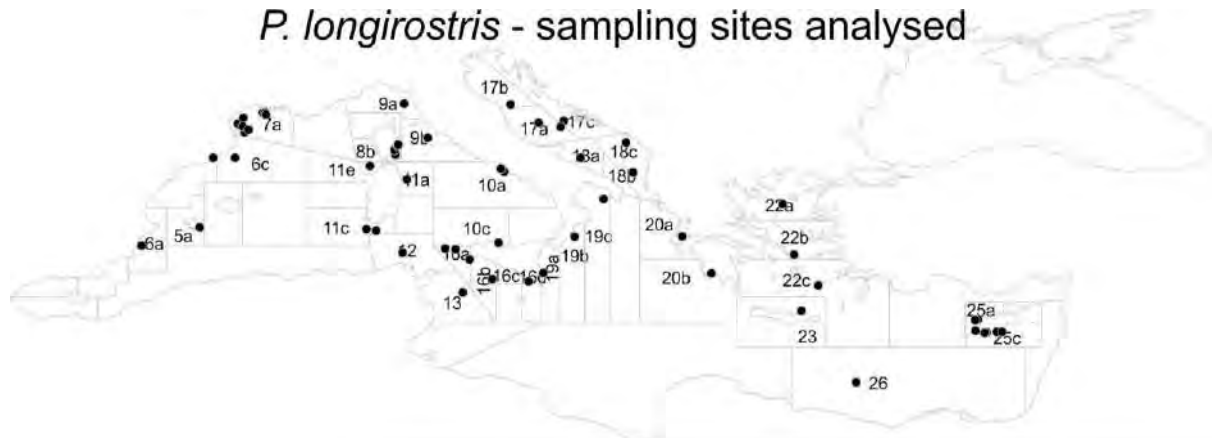


Fig. 1.5.2.10: Proportions of membership to each of K inferred clusters for *N. norvegicus* individuals

With practically no recent studies till now investigating the Norway lobster population genetics within the Mediterranean Sea, current results are the first to show a considerable differentiation of the Adriatic and the E. Mediterranean samples. Passamonti et al (1997) reported a very low genetic differentiation identified based on allozymic data among samples collected from the English Channel in the Atlantic to the Adriatic and the Aegean. However, these results were contradicted by Stamatis et al (2004 and 2006) who identified at least three different genetic pools from the North to the Aegean Sea using either mtDNA RFLPs (2004) and allozymes (2006). Similarly, Maltagliati et al (1998), identified moderate genetic distances and argued for some eight separate stocks from samples they analyzed from Portugal to the Aegean Sea whereas recently Gallagher et al (2019) revealed population structure between a single “eastern” Mediterranean (Ancona) sample and each of the eight other samples from the North Atlantic using mtDNA D-loop sequences.

Parapeneus longirostris

For *Parapeneus longirostris*, the complete study was conducted with specimens from 35 localities (GSAs and sub-GSAs). In total, 1750 specimens were sampled from the Balearic Sea (6a) in the W. Mediterranean to the Egyptian (26) waters in the E. Mediterranean.



We encountered serious problems related to the poor quality of DNA extracts. At several sites, the majority of specimens were not suitable for ddRAD analysis (see for instance GSA16a, 16c, 16d, 23a, 25c), thus strongly reducing the sample size. ddRAD was performed on 1008 individuals including all the samples with good DNA quality (772) and most of the samples with medium DNA quality, some of which were excluded based on the insufficient concentration obtained from DNA extraction (Table 1.5.2.10).

Table 1.5.2.10: DNA quality across samples in *Parapeneus longirostris*

GSA	Total number of samples	DNA quality			Number of samples in Libraries
		Good	Medium	Bad	
05a	50	37	8	5	40
06a	49	44	3	2	47
06c	100	60	15	25	72
07a	50	30	11	9	34
08ab	50	40	10	0	42
09a	50	15	18	17	33
09b	50	27	13	10	35
10ab	50	27	12	11	31
10c	47	9	5	33	14
11a	50	37	9	4	43
11c	50	30	10	10	37
11e	50	43	5	2	39
11p	50	40	8	2	33
12a	50	14	10	26	13
13a	50	12	24	14	17
16a	50	3	7	40	10
16b	50	18	17	15	35
16c	50	4	10	36	11

16d	50	3	7	40	7
17ab	50	36	11	3	36
17c	50	5	31	14	24
18a	50	45	3	2	39
18c	50	19	25	6	36
19a	50	3	22	25	20
19b	50	38	11	1	38
19d	50	18	22	10	22
20a	50	29	13	8	42
20b	50	8	24	18	27
22a	50	21	5	24	26
22b	50	13	29	8	35
22c	50	11	14	25	22
23a	40	2	13	25	9
25a	14	0	5	9	0
25c	50	11	8	31	19
26c	50	20	16	14	20
Sum	1,750	772	454	524	1,008

Specimens from all sampled GSAs were included in the ddRAD library preparation at a various number, with the exception of GSA25c which comprised a single individual (Table 1.5.2.11).

Table 1.5.2.11: Number of specimens (specim.) in each library. Totals include replicated control samples.

1st library (pilot)	1st-bis library	2nd library	3rd library	4th library	5th library	6th library	7th library
288 specim.	144 specim.	144 specim.	144 specim.	159 specim.	159 specim.	159 specim.	159 specim.
Medium to bad quality DNA	medium to bad quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA	medium to good quality DNA

A total of 1008 samples (from 34 populations) were genotyped following a ddRAD protocol, with enzymes *PstI* and *NlaIII*. As for *Aristeus antennatus*, following the low coverage in the first library used for the pilot study, the size selection window was optimized to reduce the number of loci (library 1-bis). Adapter-ligated fragments from around 543 to around 657 bp

(which correspond to genomic fragments from around 420 to around 540 bp) were selected using a Blue Pippin Automated Size Selection machine, respect to the previous gel-based size selection which targeted genomic fragments from around 200 to around 480 bp. In addition, the number of multiplexed individuals per library was reduced to 144. Due to the change of size selection and level of multiplexing, the results obtained from the first library were discarded and the corresponding individuals have been included in libraries 1-bis to 7. The ddRADseqTool (Mora-Marqu ez et al. 2017), a software that allows the generation of *in silico* double-digested fragments to optimize ddRADseq experiments, was used to simulate a ddRAD library using the same enzyme pair and the same size selection used with the target species, to estimate the number of retrieved tags. For the analysis, a fish genome was used, whose length was 831 Mbp. A total of 43'200 tags of size between 420 and 540 bp were estimated. Hypothesizing a 3-4 times bigger size for the unknown *P. longirostris* genome, a total number of fragments around 120'000-150'000 is expected.

Multiplexed libraries were run on seven Illumina lanes, with 150 bp paired-end protocol. R1 R2 reads were merged, obtaining a total of 3.898 billion reads in R1+R2.

After demultiplexing, on average 1.911 M reads per sample were obtained. After eliminating low quality reads and reads without the RAD sites at the beginning, on average 96,5% reads were retained (range from 443 to 81M, Table 1.5.2.12). Several samples were characterized by a low number of retained reads (Figure 1.5.2.11) with about 380 specimens producing less than 500,000 reads.

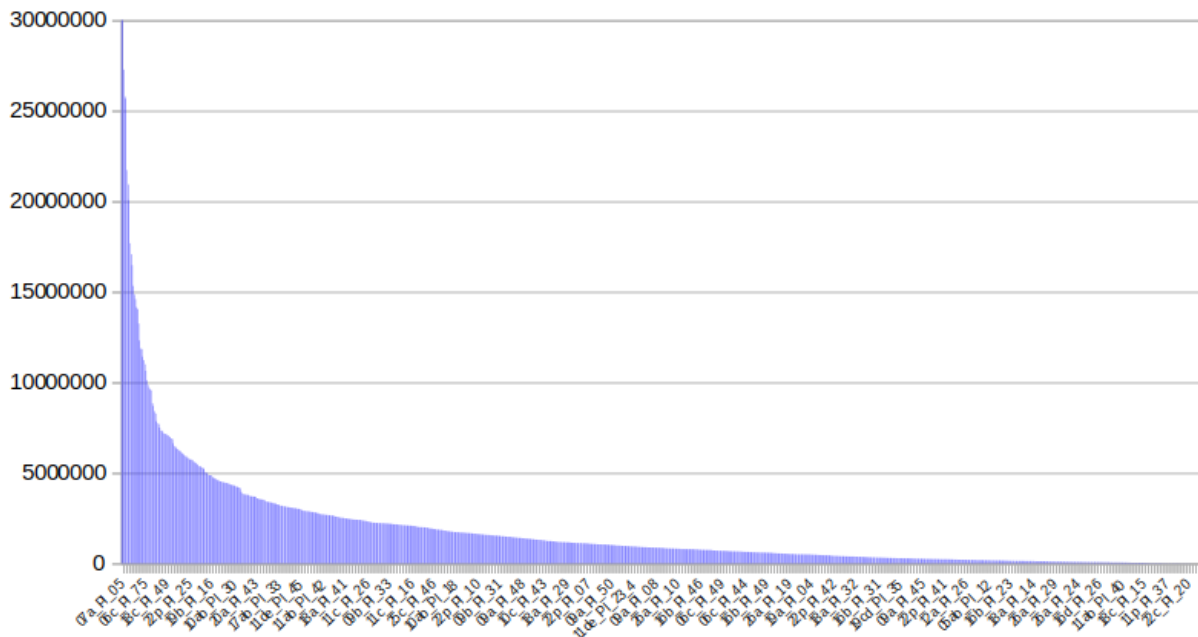
Samples were processed with ustacks considering a stack of unique reads only if at least 3 identical reads were found and allowing for up to 3 mismatches to consider two tags as part of the same locus. On average 32,637 tags were found in each sample, with average coverage of 44.7X (Table 1.5.2.12), and, as expected by the uneven distribution of retained reads, 235 individuals providing less than 5,000 loci.

A tag catalog was built with all the samples of the populations (Figure 1.5.2.12). A maximum of 3 mismatches were allowed to merge tags into the same catalog locus. The total number of tags in the catalog is around 2M.

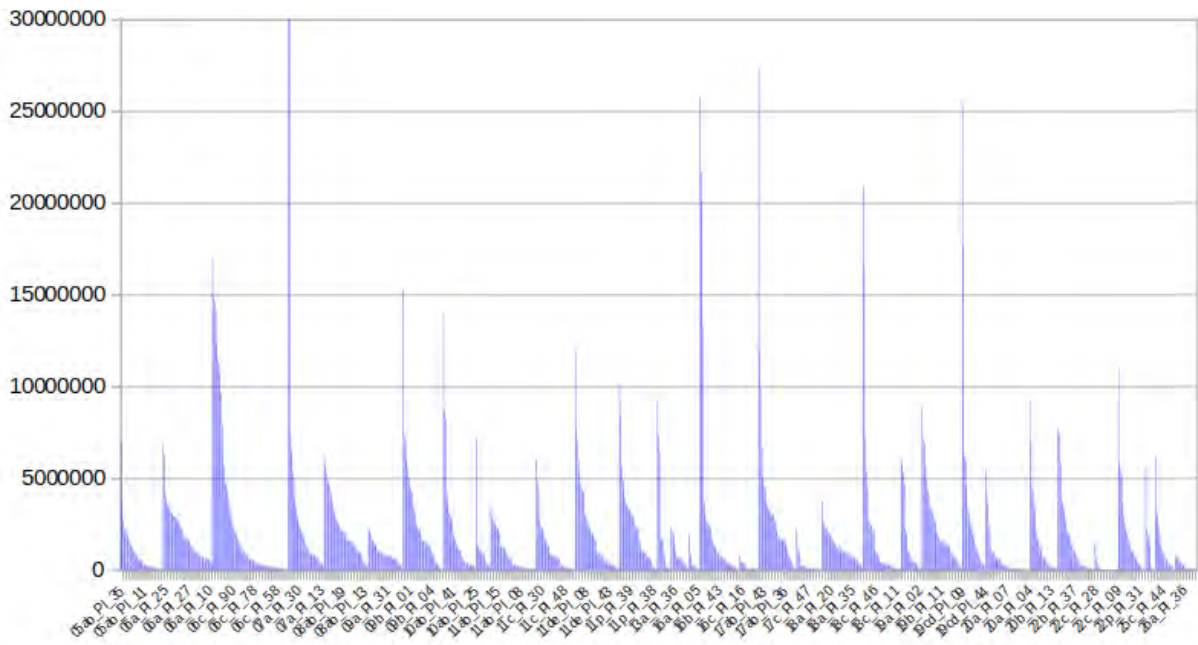
Table 1.5.2.12: Stacks pipeline output summary for *Parapenaeus longirostris*

GSA	Retained Reads			ustacks Loci		
	Average	Min	Max	Average	Min	Max
05ab	970052,8	14648	6968505	23304,5	601	100673
06a	1993054,9	290537	6855709	70374,5	21076	155960
06c	2693616,3	2554	17068875	26244,0	12	125536
07a	5796064,6	176698	81072916	62590,8	6668	456140
08ab	2236826,0	115611	6310205	41417,3	3839	86979
09a	933500,3	87367	2203119	53514,5	9279	94576
09b	2612971,5	10763	15321152	25924,1	214	127233
10ab	2178139,8	68501	14051001	36591,5	2271	140075
10c	1227800,2	150074	7154687	56268,7	12824	195655
11ab	867215,8	443	3550156	13275,4	9	46683
11c	1178278,8	6813	6040404	16405,1	137	62989

11de	2173074,1	6649	11860842	23827,1	87	114336
11p	2387027,9	12551	10107680	27846,6	137	91274
12a	2345157,5	49774	9682809	27659,7	954	90304
13a	835395,1	39509	2316339	20259,9	1653	56239
16a	429548,5	37090	1989142	20080,8	11	91466
16b	3074583,1	15973	25769360	38773,1	306	190818
16c	266632,0	25292	785699	5263,1	561	14307
16d	51633,3	3385	137445	1712,1	28	4843
17ab	3389481,9	54461	27263123	82926,3	3500	348452
17c	298611,6	7488	2214510	8238,0	273	49911
18a	1258852,6	127054	3780680	49197,1	7739	101513
18c	2152171,8	12782	20926997	20055,7	397	118658
19a	1626478,8	8324	6096150	18081,3	138	72122
19b	2331952,7	93782	8838883	64719,9	5618	135453
19cd	3830307,1	228925	25654350	39910,8	8336	143072
20a	679338,8	911	5485767	9924,9	11	56408
20b	1708887,1	42106	9554104	17892,7	662	77507
22b	1595949,8	15254	7693932	16614,1	257	63432
22c	135105,5	763	1551083	1720,6	9	18023
22p	1997358,9	4161	10990284	42926,8	193	145408
23a	1468405,0	4172	5705001	16216,4	78	55610
25c	1273181,1	5367	6178306	25789,3	107	75019
26a	264264,8	35729	858603	9816,3	1727	28301
ALL	1910527,3	443	81072916	32637,0	9	456140

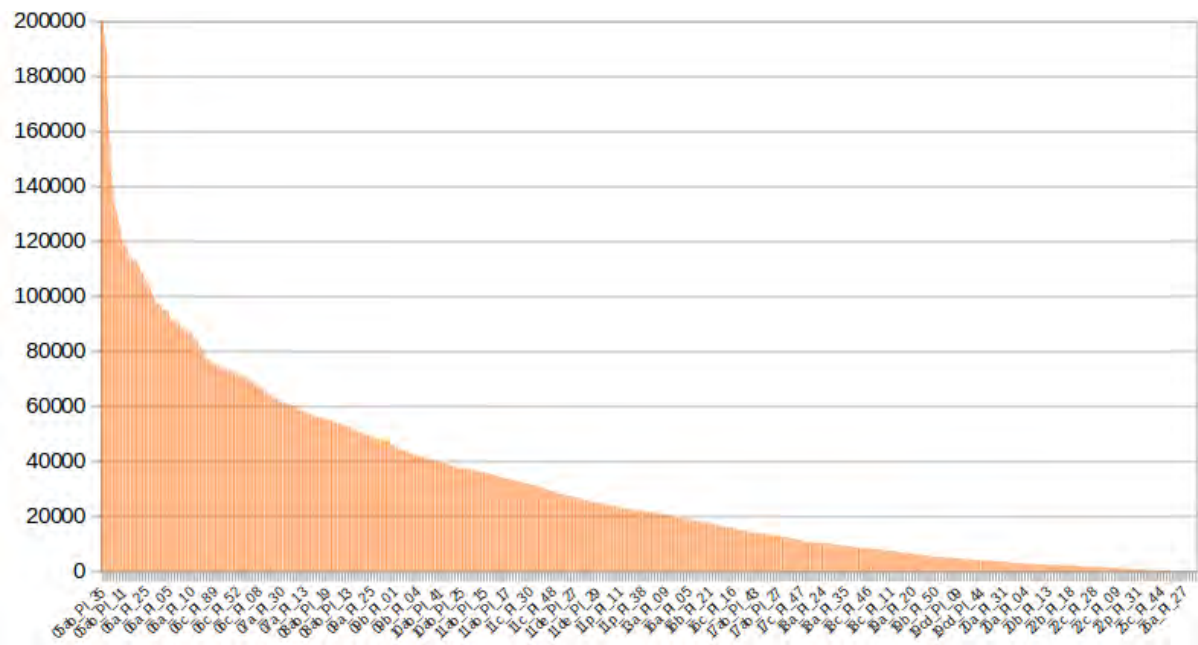


A)

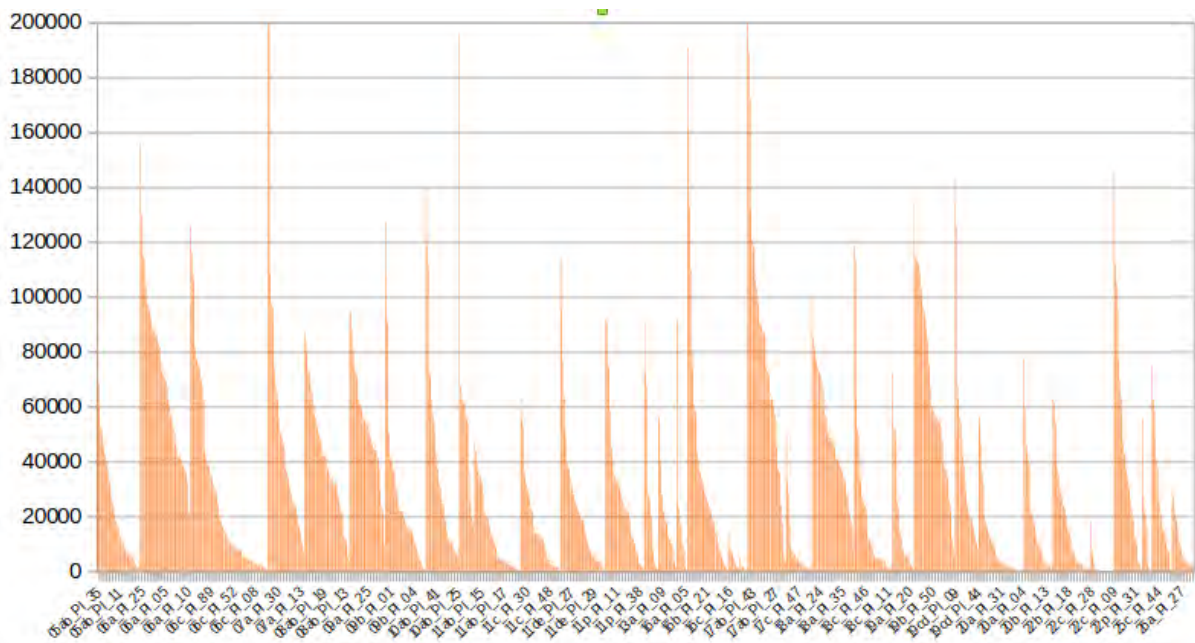


B)

Figure 1.5.2.11: Retained reads after demultiplexing for (A) the whole *P. longirostris* dataset, and (B) each population separately



A)



B)

Fig. 1.5.2.12: ustacks loci for the (A) whole *P. longirostris* dataset, and (B) each population separately

Stacks' module populations and vcftools were used to filter the initial SNP panel. Only the individuals genotyped at least at 10% of the initial loci were retained. This first filtering step eliminated all the individuals from GSA06a and most of those of GSA06c. Despite having a reasonable number of raw reads (in line with other GSAs) these samples shared almost no tag with the others, and most of the catalog tags added by these samples could not be found in the rest of the GSAs. Similar patterns are found when individuals from different species are included in the same library and analyzed together. We could exclude that these samples were confused with samples of *Aristeus antennatus* by comparing the tags with *Aristeus's* catalog, but we eventually couldn't understand the origin of the potential contamination. After this filter, for the remaining 782 samples, only SNPs shared by at least 50% of the individuals and with at least 10X coverage were kept. A total of 1225 SNPs were retained for 782 samples.

Observed heterozygosities were similar among localities and very low, ranging from 0.01 to 0.05, and smaller than expected heterozygosities, which varied between 0.03 and 0.09 (table 1.5.2.13).

Table 1.5.2.13: Summary of genetic variability estimates across sampling locations in *P. longirostris*. Reported are the average observed (H_o) and average unbiased expected heterozygosity (H_e) for GSA represented by at least 10 samples.

GSA	H_e	H_o
05ab	0.0741	0.0329
06c	0.0646	0.0114

07a	0.0739	0.04
08ab	0.0731	0.0393
09a	0.0651	0.0482
09b	0.0707	0.027
10ab	0.0762	0.0362
10c	0.0648	0.0469
11ab	0.0784	0.0247
11c	0.072	0.0269
11de	0.068	0.0252
11p	0.0677	0.0256
12a	0.0686	0.0328
13a	0.0729	0.0308
16b	0.0676	0.0334
17ab	0.0681	0.0457
17c	0.0766	0.0225
18a	0.0709	0.0396
18c	0.0782	0.0278
19a	0.0729	0.024
19b	0.0684	0.0438
19cd	0.0739	0.0315
20a	0.0783	0.0258
20b	0.0777	0.0255
22b	0.0788	0.0262
22p	0.0705	0.0403
22p	0.0705	0.0403
25c	0.085	0.0349
26a	0.076	0.0205

Pairwise F_{st} showed 41 significant pairwise differentiation involving 22 GSAs (range of significant pairwise F_{st} values from 0.01223 to 0.08997), mostly including comparisons between most Western and the most Eastern populations (Figure 1.5.2.13). Global F_{st} value was small but statistically significant (0.05573, F_{st} P-value < 0.001).

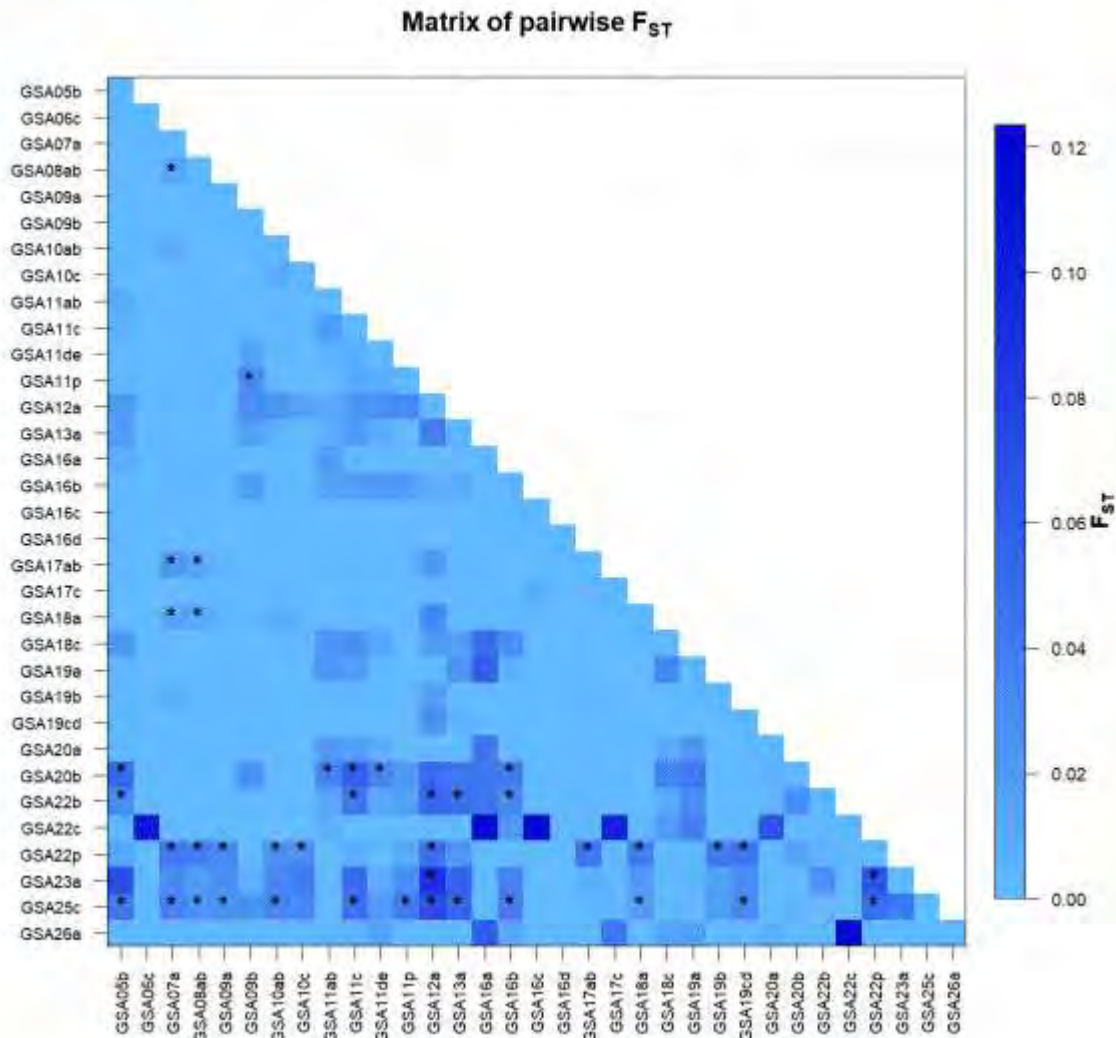


Fig. 1.5.2.13: Matrix of pairwise F_{ST} for *P. longirostris* based on 1225 polymorphic SNPs present in >50% of the whole sampling set. Significant pairwise F_{ST} values, after Benjamini–Hochberg correction, are indicated with asterisks.

Hierarchical AMOVA pointed out a significant differentiation, as indicated by the existence of a significant percentage of genetic variation (measured by the differentiation metrics F_{ct}), across groups of populations. According to this analysis, the best scenario explaining population structure, maximizing F_{ct} value, corresponded to three groups, the first including samples from the Western and Central Mediterranean Sea up to the Strait of Sicily (GSA05ab-GSA16d), the second including the remaining samples from the Central Mediterranean Sea (GSA17ab-GSA20a) except, intriguingly, the sample GSA20b that clustered with the third group including the samples from the Eastern Mediterranean Sea (GSA22b-GSA26a). This group subdivision explained about 0.9% of the total genetic variation ($F_{ct}=0.00931$, $P<0.001$; Table 1.5.2.14). Within this structure, the strongest

differentiation is between populations divided by the Strait of Sicily. Indeed, about 0.6% of the overall genetic variation is attributable to this subdivision (Table 1.5.2.14)

Table 1.5.2.14: AMOVA results for several scenarios of grouping *P. longirostris* samples.

	Scenario 1	Scenario 2	Scenario 3
Group 1	05ab,,06c,07a,08a b,09a,09b,10ab,1 0c,11ab,11c,11de, 11p,12a,13a,16a, 16b,16c,16d	05ab,06c,07a,08a b,09a,09b,10ab,1 0c,11ab,11c,11de, 11p,12a,13a,16a, 16b,16c,16d	05ab,06c,07a,08a b,09a,09b,10ab,1 0c,11ab,11c,11de, 11p,12a
Group 2	17ab,17c,18a,18c, 19a,19b,19cd,20a ,20b,22b,22c,22p, 23a,25c,26a	17ab,17c,18a,18c, 19a,19b,19cd,20a	13a,16a,16b,16c,1 6d,17ab,17c,18a, 18c,19a,19b,19cd, 20a,20b
Group 3		20b,22b,22c,22p, 23a,25c,26a	22b,22c,22p,23a, 25c,26a
Variation %			
Among groups	0.60	0.93	0.51
Among populations within groups	5.25	5.02	5.25
Within populations	94.15	94.05	94.24
Fixation Indices			
FST	0.05852	0.05947	0.05761
P-value	0.000	0.000	0.000
FSC	0.05285	0.05064	0.05280
P-value	0.000	0.000	0.000
FCT	0.00598	0.00931	0.00508
P-value	0.016	0.000	0.012

These subdivisions were not evident from DAPC plots (Figure 1.5.2.14), which only pointed out the distinctiveness of two of the East Mediterranean samples (23a and 25c). STRUCTURE analysis (Figure 1.5.2.15), instead, indicated K=2 as the best grouping and a slightly different cluster composition for samples from West and East from the Strait of Sicily. Similar results were obtained during the pilot study, with a different size selection procedure leading to a completely distinct set of SNPs.

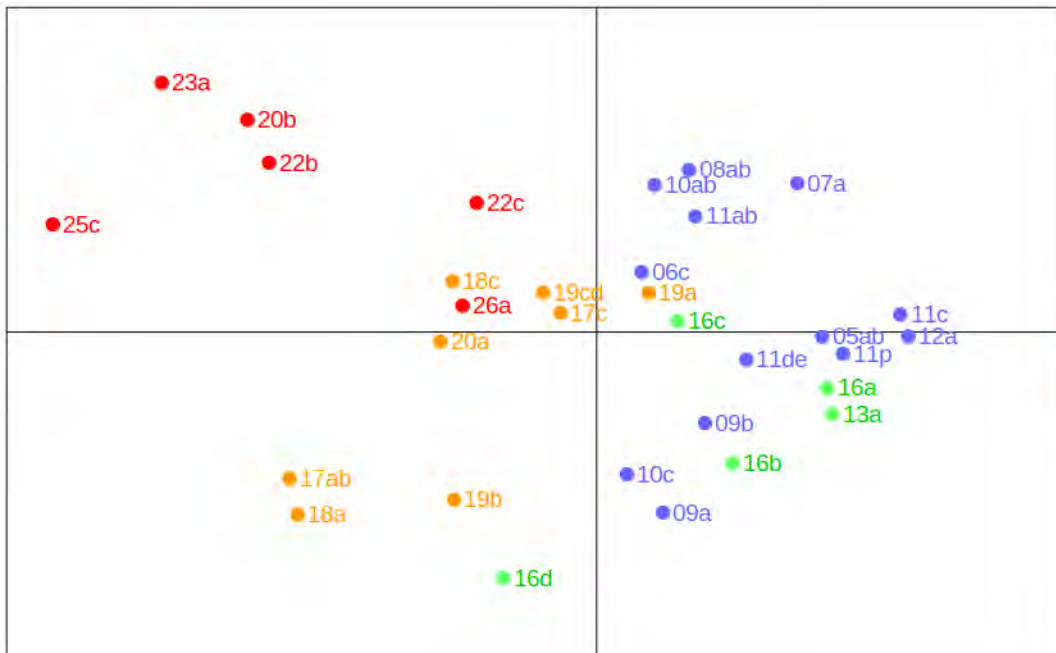
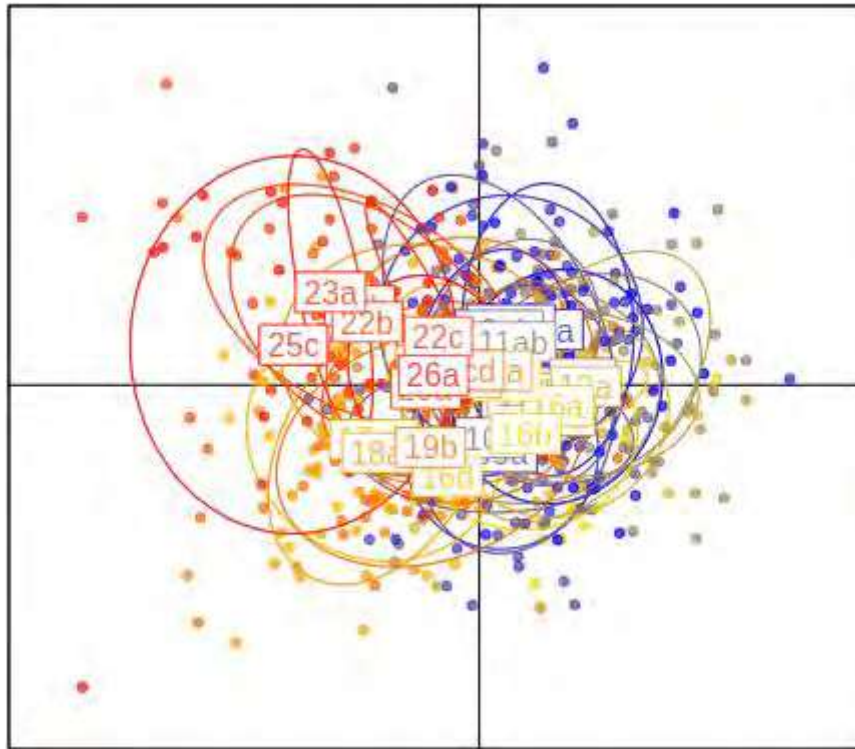


Fig. 1.5.2.14: DAPC plot for *P. longirostris* based on 1,225 polymorphic SNP present in >50% of the whole sampling set. In the bottom figure, each point represents the average coordinates of the individuals of a single GSA

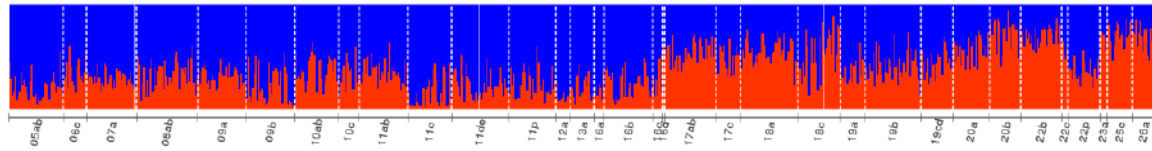


Fig. 1.5.2.15: Proportions of membership to each K inferred clusters for *P. longirostris*. Best K=2, consensus plot over 10 replicates.

Only two studies, to our knowledge, report on populations genetics of *P. longirostris*, the first one performed using mtDNA sequencing on samples collected along the Atlantic–Mediterranean transition (Garcia-Merchan et al. 2012) and the second one using mtDNA sequencing and AFLP on samples on 8 populations samples from Tyrrhenian Sea, Strait of Sicily, Adriatic and Aegean Sea (Lo Brutto et al. 2013). The results of the final study confirm the existence of significant differentiation among Mediterranean samples, previously reported by Lo Brutto and colleagues (2013) that detected the deepest differentiation when comparing Aegean samples with other Mediterranean sites. Our results suggest that the major genetic breakpoint is located in the Strait of Sicily, whereas no stronger differentiation was detected when comparing samples from the extremes of the sampling range and, indeed, a possible further subdivision is present at the boundary between Central and Eastern Mediterranean (between GSA20a and GSA20b). From a technical point of view, the optimization of the library preparation and the reduction of multiplexing (144 individuals per library) provided more robust results in terms of average coverage per sample, but the level of missing data in the final dataset is still particularly high, due to the poor sharing of tags across samples.

Conclusions

The comprehensive genetic study allowed to obtain ddRAD sequencing data for the last three out of the six species considered. The ddRAD approach evidenced different levels of differentiation for the three species: higher in the Norway lobster (*Nephrops norvegicus*) which is something novel for the species, medium-high in deep-water rose shrimp (*Parapenaeus longirostris*), and very low for the red mullet (*Mullus barbatus*).

In the red mullet (*Mullus barbatus*), the complete study had high-coverage sampling locations and included 2,133 specimens from 38 localities from the W. Mediterranean (GSA1b, N. Alboran Sea) to the South-Eastern Mediterranean Sea (GSA25c, 26 and 27). Unlike the hake study, the red mullet samples collected resulted in low DNA quality which led to discard approximately one third of them. We ended up using 1,373 specimens for the ddRAD library preparation, and unfortunately samples from seven GSAs (like 4, 11, 16 and 22) were not used at all. But again, after sequencing and demultiplexing there was a high underrepresentation of certain samples and we opted to continue with samples with more than 3,000 stack loci, and for SNPs present in at least 70% of the samples; in this way, the dataset composed of 771 samples (from 30 localities) and 853 high-quality SNPs. Our results are in agreement with most genetic studies conducted so far in the species and point

out an evident lack of genetic differentiation in the Mediterranean Sea. Therefore, we measured very low pairwise F_{st} values (divergence between sites) and the absence of clear population structure in the Mediterranean; moreover, the greatest part of the identified genetic variation was attributed to differences among individuals in the populations, and much less among groups. However, when trying to test for different differentiation scenarios we found very small but still significant indications for three major pools that group samples from the Western (from the west GSA1 till N.Tunisia GSA12), the Central (till the S. Ionian GSA20) and the Eastern Mediterranean and less for the West (till GSA12) to East (all the rest) Mediterranean differentiation; however, this differentiation explains only 1.8 and 1.3%, respectively, of the divergence between groups for the two above scenarios.

For the Norway lobster (*Nephrops norvegicus*), the complete genetic study included 1,537 specimens from 30 localities sampled from the N. Alboran Sea (GSA1b) in W. Mediterranean (GSA1b) to the Aegean Sea (GSA22b) in the Eastern Mediterranean Sea and samples in the Central Mediterranean were quite dense. The Norway lobster was the shrimp species with the best DNA quality after the extractions and most samples were included in the library preparations. From the 1,152 specimens we finally included in the molecular work, only two GSAs were not represented (18a and 22b). Filtering out samples with low performance after sequencing and demultiplexing (<3,000 stack loci) and including only SNPs present in at least 70% of the samples, we had a powerful dataset composed of 890 samples (for 27 localities) and 730 high quality SNPs. Our results reaffirm the evidence we had from the pilot study and showed a significant differentiation of the samples eastern of the Adriatic Sea (GSA17) against the others from the central and Western Mediterranean Sea. Additionally, when testing for alternative scenarios relatively high and significant values were also encountered for the separation of the Adriatic Sea (GSA17 to 19) from the neighboring basins to the west (GSA1 to 11) and the east (GSA22). Current results come to affirm some indications that previous genetic studies had put forward arguing for some level of differentiation in the Mediterranean Sea, besides the discrepancy from the Atlantic ones.

For the deep water rose shrimp (*Parapenaeus longirostris*), the full study was conducted with samples from 35 localities (GSA06a to GSA26). Quality of DNA extracts was poor for many samples and a total of 1008 individuals from 34 locations were used for ddRAD. Differently from the pilot study (performed on DNA extracted from GSA06a to 09b, 11de, 11p, 16b, 17ab, 18a and 19b), where most of the samples provided high or medium quality DNA extracts, quality of DNA extracts was poor for many of the additional samples. Considering that during this study we made a strong effort to standardize protocols for specimen manipulation after collection and for preservation, and, when we were able to control all the procedure, good quality DNA was obtained, this result indicated that further care is needed on this aspect. Accordingly, to DNA extraction results, the sample size was reduced and a total of 1008 individuals (of which 772 with high quality DNA) from 34 locations were used for ddRAD. Still the obtained dataset contained individuals with a large number of missing loci, and after further quality filters, a total of 782 individuals genotyped at 1225 SNPs were retained. Observed heterozygosities were similar among localities and very low. Pairwise F_{st} , a metric of population differentiation among pairs of samples, showed many significant pairwise differentiation (ranging from 0.01223 to 0.08997), and the global F_{st} value was highly significant (F_{st} = 0.05573, P-value < 0.001) and much higher than what recorded for *Aristeus antennatus* (F_{st} = 0.01044, P<0.001), reported in deliverable 1.5.1. Genetic clustering methods confirmed the existence of differentiation, indicating potential

subdivisions of East Mediterranean samples using DAPC, and two groups with a slightly different cluster composition for samples from West and East of the Strait of Sicily using STRUCTURE. As for other species analyzed in this project, much of the genetic variation was distributed among individuals in the populations, but a support for three groups was found: a “western-central” one including samples from Western and Central Mediterranean Sea up to the Strait of Sicily, a “central” one including the remaining samples from the Central Mediterranean Sea except the easternmost Ionian sample, and an “eastern” group that included the samples from Eastern Mediterranean Sea and the Ionian leftover. The pilot study for this species, due to the change in the size selection procedure that was adopted to optimize the ddRAD protocol based on preliminary results, provided a non-overlapping set of SNPs. Interestingly also in that case global F_{st} were statistically significant, and AMOVA and DAPC pointed out the genetic differentiation between specimens from GSA17 from the others, thus indicating that the Sicily Strait could represent an area of discontinuity in *P. longirostris* genetic pool. The results of the final study are also in line with the results of the only previous study investigating populations genetics at the Mediterranean scale, which detected using mitochondrial DNA a deep differentiation between Aegean samples and Tyrrhenian Adriatic ones (Lo Brutto et al. 2013). Our results suggest that the major genetic breakpoint is indeed located in the Strait of Sicily and that a possible further subdivision is present at the boundary between Central and Eastern Mediterranean, though not coinciding with the boundary between Ionian and Aegean Sea. The existence of significant genetic differentiation should be taken into account in the future and additional investigations are needed particularly in the Sicily Strait area, where *Parapenaeus longirostris* is an important resource.

General conclusion

As a concluding remark for this Deliverable and the Deliverable 1.5.1, the application of ddRADseq to this comprehensive genetic study evidenced a different level of differentiation of the six species that is in line with previous studies. We reconfirm that there are two main critical steps in the application of the ddRADseq, the first related to DNA quality and the second to the optimization of the protocol for library preparation and sequencing. For DNA quality, a particular care should be taken during sample collection and preservation, to reduce the number of DNA extract not suitable to ddRAD sequencing. We found that the time passed from collection to processing was critical and good quality DNA was obtained when samples were processed fresh, a few hours after the collection and placed directly on board in an appropriate volume of ethanol. Yet, the quality of DNA was still variable for specimens collected at different sites and, to a lesser extent, for specimens collected at the same site, implying that further optimization is needed.

For the optimization of the protocol for library preparation and sequencing we found that the critical step was the selection of restriction enzyme pairs for the ddRADseq digestion. We also found that, with the available enzyme pairs for ddRADseq and due to the big size of crustacean genomes, a very high number of loci is obtained, which required further optimization of size-selection of fragments for high throughput sequencing, level of multiplexing, and/or filtering for bioinformatic analysis. Based on these experiences, we believe that future applications of ddRADseq in new species, particularly crustaceans, can be facilitated by performing a preliminary optimization study, on a very small scale, to address these issues in advance.

References

Apostolidis, A. P., K. A. Moutou, C. Stamatis & Z. Mamuris, 2009. Genetic structure of three marine fishes from the Gulf of Pagasitikos (Greece) based on allozymes, RAPD, and mtDNA RFLP markers. *Biologia* 64(5): 1005-1010 doi:10.2478/s11756-009-0161-0.

Arculeo, M., S. Lo Brutto, M. Cammarata, M. Scalisi & N. Parrinello, 1999. Genetic variability of the Mediterranean Sea red mullet, *Mullus barbatus* (Pisces, Mullidae). *Russian Journal of Genetics* 35(3): 292-296.

Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Ann. Stat.*, 29, 1165–1188.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores & W. A. Cresko, 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol* 22(11):3124-3140 doi:10.1111/mec.12354.

Excoffier, L. & Lischer, H. E. L. Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10, 564–567 (2010).

Falush, D., M. Stephens & J. K. Pritchard, 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567-1587.

Falush, D., M. Stephens & J. K. Pritchard, 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7(4):574-578 doi:10.1111/j.1471-8286.2007.01758.x.

Felix-Hackradt, F. C., C. W. Hackradt, A. Perez-Ruzafa & J. A. Garcia-Charton, 2013. Discordant patterns of genetic connectivity between two sympatric species, *Mullus barbatus* (Linnaeus, 1758) and *Mullus surmuletus* (Linnaeus, 1758), in south-western Mediterranean Sea. *Marine Environmental Research* 92:23- 34. doi:10.1016/j.marenvres.2013.08.008

Galarza JA, Turner GF, Macpherson E, Rico C. 2009. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). *Can J Fish Aquat Sci*, 66(9): 1478-1490.

Gallagher, Jeanne, John A. Finarelli, Jónas P. Jonasson, and Jens Carlsson. 2019. "Mitochondrial D-Loop DNA Analyses of Norway Lobster (*Nephrops Norvegicus*) Reveals Genetic Isolation between Atlantic and Mediterranean Populations. *Journal of the Marine Biological Association of the United Kingdom* 99(4):933–40. doi: <https://doi.org/10.1017/S0025315418000929>.

Garoia F, Guarniero I, Piccinetti C, Tinti F. 2004. First microsatellite loci of red mullet (*Mullus barbatus*) and their application to genetic structure analysis of Adriatic shared stock. *Marine Biotechnology*, 6(5): 446-452.

Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi: 10.1093/bioinformatics/btn129.

Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr52.

Lo Brutto, S., T. Maggio & M. Arculeo, 2013. Isolation by distance (IBD) signals in the deep-water rose shrimp *Parapenaeus longirostris* (Lucas, 1846) (Decapoda, Panaeidae) in the Mediterranean sea. *Marine Environmental Research* 90:1-8 doi:10.1016/j.marenvres.2013.05.006.

Maggio T, Lo Brutto S, Garoia F, Tinti F, Arculeo M. 2009. Microsatellite analysis of red mullet *Mullus barbatus* (Perciformes, Mullidae) reveals the isolation of the Adriatic Basin in the Mediterranean Sea. *ICES Journal of Marine Science*, 66(9): 1883-1891.

Maltagliati F, Camilli L, Biagi F, Abbiati M: Genetic structure of Norway lobster, *Nephrops norvegicus* (L.) (Crustacea : Nephropidae), from the Mediterranean Sea. *Scientia Marina* 1998, 62:91-99.

Mamuris Z, Apostolidis AP, Theodorou AJ, Triantaphyllidis C. 1998a. Application of random amplified polymorphic DNA (RAPD) markers to evaluate intraspecific genetic variation in red mullet (*Mullus barbatus*). *Marine Biology*, 132(2): 171-178.

Mamuris Z, Apostolidis AP, Triantaphyllidis C. 1998c. Genetic Protein Variation in Red Mullet (*Mullus barbatus*) and Striped Red Mullet (*M. surmuletus*) Populations from the Mediterranean Sea. *Marine Biology*, 130, no. 3: 353-60.

Mamuris, Z., C. Stamatis, K. A. Moutou, A. P. Apostolidis & C. Triantaphyllidis, 2001. RFLP analysis of mitochondrial DNA to evaluate genetic variation in striped red mullet (*Mullus surmuletus* L.) and red mullet (*Mullus barbatus* L.) populations. *Marine Biotechnology* 3(3): 264-274 doi:10.1007/s101260000075.

Matic-Skoko, S., T. Segvic-Bubic, I. Mandic, D. Izquierdo-Gomez, E. Arneri, P. Carbonara, F. Grati, Z. Ikica, J. Kolutari, N. Milone, P. Sartor, G. Scarcella, A. Tokac & E. Tzanatos, 2018. Evidence of subtle genetic structure in the sympatric species *Mullus barbatus* and *Mullus surmuletus* (Linnaeus, 1758) in the Mediterranean Sea. *Sci Rep* 8: 14 doi:10.1038/s41598-017-18503-7.

Mora-Márquez F, García-Olivares V, Emerson BC, López de Heredia U. ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. *Mol Ecol Resour.* 2017;17(2):230-246. doi:10.1111/1755-0998.12550

Passamonti, M., B. Mantovani, V. Scali & C. Frogli, 1997. Allozymic Characterization of Scottish and Aegean Populations of *Nephrops norvegicus*. *Journal of the Marine Biological Association of the United Kingdom* 77(3):727-735 doi:10.1017/S0025315400036158.

Pritchard, J. K., M. Stephens & P. Donnelly, 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155(2):945.

Stamatis, C., A. Triantafyllidis, K. A. Moutou & Z. Mamuris, 2004. Mitochondrial DNA variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. *Molecular Ecology* 13(6):1377-1390 doi:10.1111/j.1365-294X.2004.02165.x.

Stamatis C, Triantafyllidis A, Moutou KA, Mamuris Z: Allozymic variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. *ICES J. Mar. Sci.* 2006, 63:875-882.

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP1 - Population genetics and phylogeographic studies for identification of biological units of priority species

Deliverable 1.6 – Detailed protocol for routine sampling and genetic monitoring

LEAD: RITA CANNAS (CONISMA)

PARTICIPANTS: LUCA BARGELLONI, ALESSIA CARIANI, LAURA CARUGATI, LEONARDO CONGIU, RACHELE CORTI, ALICE FERRARI, RAFAELLA FRANCH, SILVIA IORI, FRANCESCO MAROSO, ILARIA ANNA MARIA MARINO, RICCARDO MELIS, CHIARA PAPETTI, FEDERICA PIATTONI, ELISABETTA PIAZZA, FAUSTO TINTI, LORENZO ZANE (CONISMA)

COSTAS TSIGENOPOULOS, DIMITRIS TSAPARIS, VASSO TERZOGLU, TEREZA MANOUSAKI, KATERINA EKONOMAKI, PETROULA BOTSIDOU, XENIA SARROPOULOU (HCMR)

Date: June 14, 2021

Contents

EXECUTIVE SUMMARY	3
OBJECTIVE	6
OUTCOMES FROM MED_UNITS WP1	7
OVERALL EVALUATION	7
<i>Sampling: temporal and spatial coverage</i>	7
<i>Tissue types and DNA extraction</i>	15
<i>Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses</i>	15
SPECIES BY SPECIES EVALUATION	18
<i>Aristaeomorpha foliacea</i>	18
Sampling	18
Tissue types and DNA extraction	18
Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses	19
<i>Aristeus antennatus</i>	21
Sampling	21
Tissue types and DNA extraction	21
Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses	22
<i>Merluccius merluccius</i>	24
Sampling	24
Tissue types and DNA extraction	24
Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses	25
<i>Mullus barbatus</i>	27
Sampling	27
Tissue types and DNA extraction	27
Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses	28
<i>Nephrops norvegicus</i>	30
Sampling	30
Tissue types and DNA extraction	30
Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses	31
<i>Parapeneus longirostris</i>	33
Sampling	33
Tissue types and DNA extraction	33
Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses	34
LESSONS LEARNED FROM MED_UNITS WP1	36
<i>Sampling: spatial and temporal coverage</i>	36
<i>Tissue types, DNA extraction and DNA quality</i>	36
<i>Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses</i>	37
CONCLUSIONS	38
FINAL UPDATED FIELD PROTOCOLS SAMPLING TO BE USED IN GENOMIC STUDIES	42
REFERENCES	43

Executive Summary

The present document summarizes and evaluates the procedures and methodologies adopted within the MED_UNITS genetics Work Package (WP1).

The recommended protocols include full details of the different phases, with special attention devoted to the revised sampling procedures. The text is accompanied by illustrations with clear and concise schemes, that could eventually be printed as stand-alone cards and used as supports in the field and laboratory work.

Deliverable 1.6 starts with the detailed description of the main outcomes of WP1 (Tasks 1.3-1.4-1.5) in terms of samples analysed and data obtained. The discussion on the number of stocks identified per species, and their biological meaning is out of the scope of the present document, but it is fully addressed in previous documents (D1.4, D1.5.1 and D1.5.2).

In overall, the MED_UNITS sampling was satisfactory for the spatial coverage and number of specimens collected with a total of 10,670 specimens for all six species. Contrarily to the initial expectations, the timely sampling has finally proved to be the most problematic aspect.

According to the sampling protocol for genetic analyses, tissues were collected both from freshly caught (28.7%) and frozen specimen (52%). Unfortunately, for a fraction of tissues no indication was given on the sampling conditions and tissue type (19.3%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium quality or even unusable DNAs.

All DNAs classified as ‘unusable’ were discarded and were not included in the following step (library preparation for the genotyping). A total of 7,544 individuals were included in libraries, most of them (69.8%) having a ‘good’ DNA quality rating while the remaining nearly one third (30.2%) were ‘medium-quality’ samples.

A total of 5,690 individuals were successfully genotyped (75.4% of those included in the libraries). The genotype success of the ‘good DNA quality’ tissues was high (82%) while only 61% of the ‘medium quality’ DNAs included in the libraries ended up producing genotype data.

However, the quality of DNAs obtained by tissue type as well as the genotype success varied considerably by species; therefore, results are described and discussed in detail at the species level.

In brief, for the fish species:

- *M. merluccius* gave the best results. A total of 1,664 individuals were successfully genotyped (97.6% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 98.5% versus 92.9% of ‘medium’ DNAs.
- *M. barbatus* gave the worst results: a total of 771 individuals were successfully genotyped (56.1% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 57.8% versus 54% of ‘medium’ DNAs

On overall, for the crustacean species:

Similar results were obtained for *A. antennatus* and *P. longirostris* with about 79% of individuals included in libraries successfully genotyped, followed by *N. norvegicus* (72.2%) and *A. foliacea* (60.8%). In detail:

- *A. foliacea*: A total of 770 individuals were successfully genotyped. The genotype success of the ‘good DNA quality’ tissues was 68.4% versus only 53.1% of ‘medium’ DNAs.
- *A. antennatus*: A total of 825 individuals were successfully genotyped. The genotype success of ‘good DNA quality’ tissues was 86.6% versus only 60% of ‘medium quality’ DNAs.
- *N. norvegicus*: A total of 890 individuals were successfully genotyped. The genotype success of ‘good DNA quality’ tissues was 84.2% versus only 50% of ‘medium’ DNA.
- *P. longirostris*: A total of 770 individuals were successfully genotyped. The genotype success of ‘good DNA quality’ tissues was 78.1.% versus 71.9% of ‘medium’ DNAs.

Our results confirm what it is known in the literature: the ‘good DNA quality’ tissues produced better results measured as number of genotyped individuals. As the proportion of medium-quality DNA in libraries increased, the whole performance of the sequencing output decreased on overall for all the samples included. This is explained by the fact that the level of DNA degradation affects the efficiency of reduced representation sequencing, increasing the number of missing data/loci, reducing the number of total reads and SNPs when using a given threshold to finally obtain a solid dataset for measuring population differentiation and stock identity.

Among the main lessons learned in this project we can list:

- A timely sampling is the crucial step for having reliable genomic data in a short period project.
- A clear indication is to rely preferentially on fresh tissues and speed up the sampling and processing procedures.
- Larger quantities of samples should be collected in order to have enough samples per area. Extra samples will allow to overcome the unavoidable decrease in numbers due to low DNA quality samples or failures during the experimental steps.

In more general terms we can conclude that:

- The sampling is very important. It should be designed in order to fulfil the needs of ‘genomics’ requirements (in terms of quality of samples, procedures, storage, timing) with special indications for shorter preservation times, or the need to include additional steps that may increase cost and logistical constraints.
- The scientific surveys at sea (as MEDITS or other similar surveys) are a good opportunity to implement proper sampling design for genomic analyses, because of the wide geographical coverage, the relatively homogeneous temporal sampling and the sampling locations. However, the sampling for genomic analyses cannot be a collateral activity or side-project, opportunistically realized during the standard surveys at sea, and a supplementary task in addition to the multiple activities performed on board by the scientific staff; thus, adequate resources should be allocated for this activity.
- If the sampling is performed by multiple groups of scientists operating in different areas, not necessarily involved in the laboratory genetic analyses, it is highly recommended to set up a central Hub in charge of the coordination of the sampling activities between the samplers in the field and the technicians in the laboratory.
- The timeframe for collecting tissues should be carefully defined to address the specific biological problem (stock identification), taking into consideration the biology (life history cycle) of the species under investigation and the necessity to make meaningful comparisons among populations (areas).
- In the planning of any project, the timing for the experimental phases cannot be too strict but extensive pilot experiments at the beginning of a project are recommended to be performed for testing different experimental conditions.

- The collection of tissues should preferentially be realized from alive/freshly caught individuals and the sampling realized as soon as possible within 1-2 hours from the death of the animal, always kept in optimal conditions (cold temperature).
- Crustaceans can be particularly problematic. In this case, it is highly advisable to use only fresh and well-preserved tissues and to allocate the proper time to optimize the molecular protocols for each species under study. This is especially true for shrimps because their muscle tissues consist of numerous fiber proteins, majorly composed of myofibrillar and sarcoplasmic proteins. Thus, finding the best conditions for digesting these fibrous proteins from the initial step of the DNA extraction method applied may result in higher DNA yields and quality for downstream analytical approaches.

Objective

The present document summarizes and evaluates the procedures and methodologies adopted within the MED_UNITS genetics Work Package (WP1). It highlights the main steps necessary to develop a detailed protocol based on genetic data to be used in a routine basis in the future to help stock identification.

A step-by-step procedure is described on the way to use genetic methods and allocate samples into different stocks, while minimizing effort and costs.

The recommended protocols include full details of the different phases, with special attention devoted to the revised sampling procedures. The text is accompanied by illustrations and clear/concise schemes, that could be eventually printed as stand-alone cards and used as supports in the field and laboratory work.

The sampling is aimed at satisfying the high standard requirement of the genomic analyses, in particular based on the ddRAD methodology which at present represents one of the most effective scientific tools for the genetic differentiation of stocks.

It is worth pointing out that while standard sampling procedures can be defined and applied to any species, standard and robust ddRAD protocols need to be optimized and tailored for each species separately, and also frequently updated based on the progress of technology and methodologies. Therefore, the descriptions of the ddRAD protocols adopted (i.e., DNA extraction, DNA quality check, library preparation, NGS sequencing, the bioinformatic pipeline and the population genetic analyses) are not included in the present document but are provided and fully discussed in previous separate documents, namely Deliverables D.1.3, D1.4, D1.5.1, and D1.5.2.

Deliverable 1.6 is composed of three main parts:

- Description of the main outcomes of WP1 (Tasks 1.3-1.4-1.5) in terms of samples analysed and data obtained. The discussion on the number of stocks identified per species, is fully addressed in the Deliverables D1.4, D1.5.1 and D1.5.2.
- Strengths and weaknesses of the adopted procedures.
- Final protocols that will eventually amend/integrate the original ones, proposed at the starting of the project.

It was originally listed as an additional objective of Task 1.6 to indicate the best types of markers/genetic analyses at the species level as well as well to describe the most effective standard scientific methods/tools for the genetic differentiation of stocks. It was included as a sound objective, considering that we were testing for the first time the SNP markers in five species out of six (all except *M. merluccius*), leaving the door open to using other types of markers (i.e., microsatellites) in case of failure of the chosen methodology and because in this species there are plenty of microsatellite loci available. In case we ended up using different markers for the different species, the methods could eventually be tuned/adjusted and compared case by case. Given that we succeeded in obtaining SNPs for all the six species, the markers and the pipeline for the population genetic analyses and stock identification methods/tools used were the same for all the species. They were fully described in D1.4, D1.5.1 and D1.5.2 and hence they are no further commented here.

We consider the markers and methods of analysis as the most valid for the studied species at present for obtaining population genomic data.

Outcomes from MED_UNITS WP1

OVERALL EVALUATION

The following paragraphs will evaluate the full results of WP1 (sampling design, pilot and full study: Tasks 1.3-1.4-1.5) in the different steps of the experimental procedure, in order to evaluate their accomplishment. The final goal is to identify the main drawbacks and propose modifications to the protocols used and described in previous documents (Deliverables and Reports). The protocols are here revised to be ready for routinely future monitoring.

Sampling: temporal and spatial coverage

The strategy and rationale of the sampling scheme are here briefly described. The project included a first set of tissues samplings for the pilot study (Task 1.4), and a second set for the full study (Task 1.5).

Samplings for the pilot studies should have been accomplished by month 5 while those for the full study by month 12.

In particular, the pilot study was a pre-screening phase that aimed at testing:

1. the results of previous studies
2. the markers to be used (type and number)
3. the sampling size for the full and comprehensive studies

In brief, a preliminary phase, called test phase, was also performed within Task 1.4 at the very beginning of the project (months 2-5) as described in D0.3. These preliminary tests aimed at optimizing the ddRADseq protocol in each species and to give indication on the best procedure to realize the whole sampling activities.

In particular, 18 sampling sites from five GSAs (GSA9, GSA11, GSA18, GSA19 and GSA23) have been collected in February-March by Conisma-UNICA, CIBM, HCMR and COISPA; tissues for the six species were made available to Conisma-UNIPD and HCMR for the preliminary laboratory analyses.

Total genomic DNA has been isolated from these tissues using commercial kits, following manufacturer's protocol. The amount and quality of DNA has been quantified using a spectrophotometer and resulted in general low-medium in terms of quantity and quality. Based on these provisional results, indications were obtained for both the extraction protocols and sampling activities of the following tasks.

In particular, in order to have the best quality/quantity of DNA

- sampling protocols have been revised, stressing the need to collect the tissues immediately after the capture of the animals (within 1-2 hours from the haul), and to strictly avoid the use of any preservative (chemicals) especially in crustaceans. Fin and legs (pereiopods) were eliminated from the sampling protocol as possible alternative for sampling, given the better results in terms of DNA obtained in the preliminary tests using muscle tissues.
- Laboratory protocols have been optimized, leaving aside the commercial kits and selecting a more demanding (in terms of time and staff) but satisfactory method of DNA extraction. Standard DNA salt extraction protocols by Miller et al., 1988 and Cruz et al. 2017 were used in HCMR and ConISMA, respectively. Full details are given in D1.4.

Once defined the best protocols and received the samples, we started working on eight sites and 50 individuals from each location. The choice of the pilot sites was determined by the availability of samples collected by July 2019. An effort was made to include as many as possible sites corresponding to the different genetic stocks identified in the previous studies, but this was not always possible (Table 1).

In the full study, as many as possible additional locations were added in the analyses, extending the geographical coverage to include the whole Mediterranean areas where the different species exist.

The sampling was a critical step because the success of the entire WP was much dependent on the availability of tissues with precise requirements in terms of:

- spatial and temporal coverage
- number (quantity of different individuals per species).

In overall, the sampling was satisfactory for the first and third aspects (spatial coverage number of specimens). while the timely sampling was the more problematic aspect.

Sampling started in January 2019 and ended in July 2020 mostly to accommodate for the collection of samples from non-European areas, as requested by the Contracting Authority and delayed given the COVID_19 pandemic restrictions and consequences. Sampling included a total of 210 different ‘sampling sites’ in 22 GSA + 1 Atlantic sample. About 38% of tissues were collected within the first 6 months, reaching 61% by July, and 84% of samples by the end of 2019. In particular, most of samples were collected during surveys at sea by July 2019 (Figure 1).

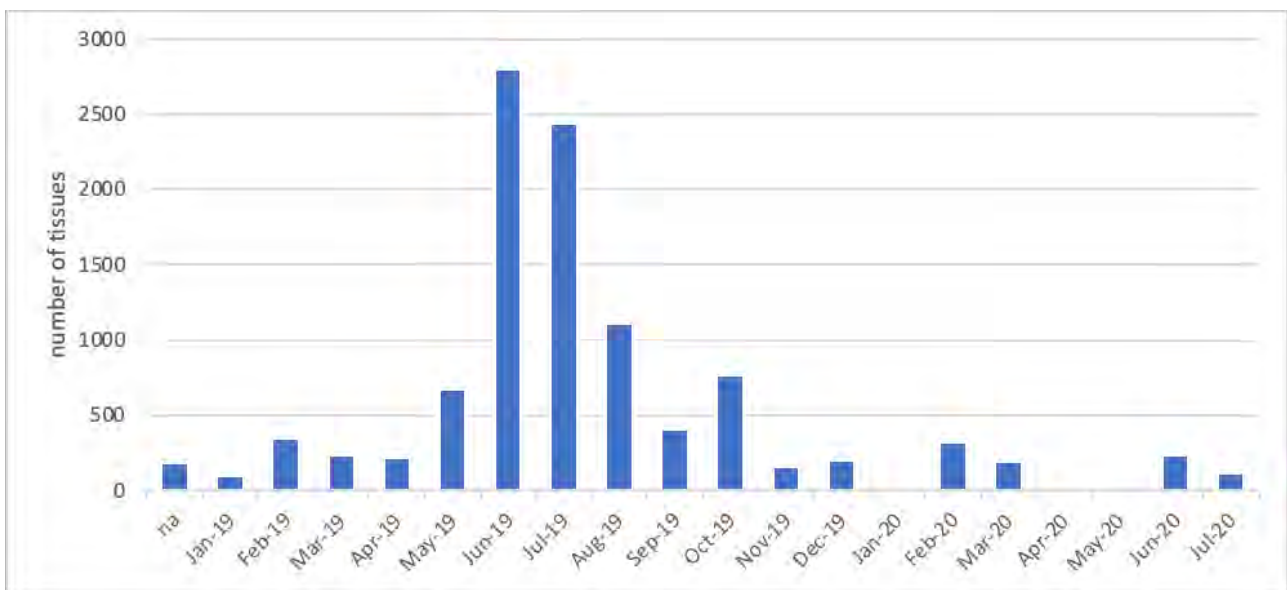


Figure 1 Monthly temporal distribution of tissues collection for the six species under investigation. The term ‘na’ refers to samples for which no indication of a collection date is given by the samplers.

The samples available by July 2019 were not all coincident with the ones selected for the pilot phase in terms of geographical distribution (Table 1).

MED_UNITS Deliverable 1.6 – Detailed protocol for routine sampling and genetic monitoring

Table 1 Pilot sampling sites. In the upper rows are pilot sites as originally proposed (based on previous studies where genetic discontinuities were recorded or suspected), while in the lower rows are the pilot sites as realized (based on the samples available at the end of July 2019). The text and cells in green indicate the sites not foreseen or in excess with respect to the original list, while the text and cells in red indicate the sites foreseen but finally not available due to the failure of sampling (GSA3) or the postponing of the samples collection/delivery.

Original Pilot sites																		
Species	GSA1	GSA3	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	Total
<i>A. foliacea</i>			1				1	1	1	1			1	1	1			8
<i>A. antennatus</i>	1		1	1			1		1	1			1	1				8
<i>M. merluccius</i>	1	1							1	1		1	1		1		1	8
<i>M. barbatus</i>			1				1				2	2			1		1	8
<i>N. norvegicus</i>	1		1	1			1		1	1	1				1			8
<i>P. longirostris</i>							1	1		1	1	1	1		1	1		8
total	3	1	4	2			5	2	4	5	4	4	4	2	5	1	2	48
Realized Pilot sites																		
Species	GSA1	GSA3	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA16	GSA17	GSA18	GSA19	GSA20	GSA22	GSA23	GSA25	Total
<i>A. foliacea</i>			0				1	1	2	1		1	2	0	0			8
<i>A. antennatus</i>	0		0	2	1		1		2	0		1	1	0				8
<i>M. merluccius</i>	0	0		2			1		2	1		1	1		0		0	8
<i>M. barbatus</i>			0	1			1		2		2	2			0		0	8
<i>N. norvegicus</i>	0		0	2		1	1		2	0	1	1			0			8
<i>P. longirostris</i>				1			1	0	2	1	1	1	1		0	0		8
total	0	0	0	8	1	1	6	1	12	3	4	7	5	0	0	0	0	48

Considering the whole sampling, specimens were totally missing for four GSAs (GSA3, GSA15, GSA21 and GSA24)(Figure 2, Figure 3, Figure 4). In some GSAs, all six species were sampled while in others only one species or few of them were obtained (Table 2, Figure 2).

A total of 10,670 specimens were recorded as sampled in the MED_UNITS dataset. Almost the totality of tissues was processed for DNA extraction (n=10,550), excluding a few that were recorded in the database but were finally unavailable being possibly lost in the route from the field to the HUB/laboratory or (n=17), and some not included in the planned analyses being duplicates or supernumerary in a given area (n= 60,

Table 3).

The mean number of individuals analyzed per species was 1,758.3 (range 1,470-2,173), in most cases well above the maximum number of specimens originally foreseen (1,450 per species, adding the 1,050 specimens of the full study to the 400 included in the first pilot phase). The mean number of individuals analyzed per GSA was 458.7 (range 50-1,353). The highest numbers were reached in large GSAs, where multiple sites in different subareas were sampled, and/or tissues were collected in the same GSA (subarea) twice (in the pilot phase and in the full study). In general, the re-sampling was realized to overcome quality problems of the first tissue batches (Figure 2). On average, 50.3 individuals per species per sampling site (subarea) were processed (range 1-100; Figure 3) for the DNA extraction, in general well above the number of 35 individuals per site foreseen for the full study sites.

MED_UNITS Deliverable 1.6 – Detailed protocol for routine sampling and genetic monitoring

Table 2 Number of sampling sites by species and GSA as originally foreseen (Milestone 1.2; row 'PLANNED') and actually accomplished (row 'REALIZED'). If no indication is provided in the cell, the sample is composed of >35 specimens, * indicates that the sample has <35 individuals but >20; ** indicates that one of the samples has < 20 individuals but > 10; *** indicates that one of the samples has < 10 individuals. The column Atlantic refers to tissues collected for European hake specimens, not originally foreseen in the planning design.

		GSA1	GSA3	GSA4	GSA5	GSA6	GSA7	GSA8	GSA9	GSA10	GSA11	GSA12	GSA13	GSA14	GSA15	GSA16	GSA17	GSA18	GSA19	GSA20	GSA21	GSA22	GSA23	GSA24	GSA25	GSA26	GSA27	ATLAN TIC	TOTAL
PLANNED	<i>A. foliacea</i>			1	2	3		2	3	3	3	1	1	1	1	3		2	3	2	1	2	1	1	1	1			38
REALIZED	<i>A. foliacea</i>				2*			2*	3	3*	5	1	1			3		2	3	2		2	1		1	1			32
PLANNED	<i>A. antennatus</i>	1	1	1	2	3	2	1	3	2	3	1	1	1	1	2		2	3	2	1	1	1	1	1	1			38
REALIZED	<i>A. antennatus</i>	1			2	5	1		2*	2	5	1	1		1		2*	4	2*		1**	1*		1***	1				33
PLANNED	<i>M. merluccius</i>	1	1	1	2	3	1	1	2	2	3	1	1	1	1	2	2	2	2	1	1	2	1	1	1	1	1		38
REALIZED	<i>M. merluccius</i>	1		2*	2*	3*	1	2**	2	2	4	2	1*			3*	3	2	3	1		3	1		1	1	1	1	42
PLANNED	<i>M. barbatus</i>	1	1	1	1	2	2	1	2	2	2	1	1	1	1	2	3	3	2	1	1	2	1	1	1	1	1		38
REALIZED	<i>M. barbatus</i>	1		2	1	2	1	1	2	2	3	2		1		3	3	3	2	1		2	2		1	2	2		38
PLANNED	<i>N. norvegicus</i>	1	1	1	1	3	1	1	2	3	4	1	1	1	1	3	2	2	2	2	1	3	1						38
REALIZED	<i>N. norvegicus</i>	1			1	3*	1	1	2	2*	5**	1	1			3	2	3	1	1		2*							30
PLANNED	<i>P. longirostris</i>	1	1	1	1	2	2	1	2	2	2	1	1	1	1	3	2	2	2	2	1	2	1	1	1	1	1		38
REALIZED	<i>P. longirostris</i>				1	2	1	1	2	2	4	1	1			4	2	2	3	2		3	1		2**	1			35
PLANNED	TOTAL	5	5	6	9	16	8	7	14	14	17	6	6	6	6	15	9	13	14	10	6	12	6	5	5	5	3		228
REALIZED	TOTAL	4		4	9	15	5	7	13	13	26	8	5	1		17	10	14	16	9		13	6		5	6	3	1	210

Table 3 Number of DNA extractions per species. The column 'NA' refers to specimens included in the dataset for which tissues/tubes were unavailable; the column 'no' refers to tissues for which DNA extraction was not performed being duplicates or supernumerary and hence non included in the planned analyses; the column 'yes' refers to the tissues processed for DNA extraction.

Species	DNA extractions			Total tissues
	NA	no	yes	
<i>A. foliacea</i>		1	1,692	1,693
<i>A. antennatus</i>			1,470	1,470
<i>M. merluccius</i>	14	42	1,928	1,984
<i>M. barbatus</i>	3	16	2,173	2,192
<i>N. norvegicus</i>			1,537	1,537
<i>P. longirostris</i>		1	1,750	1,751
Total	17	60	10,550	10,627

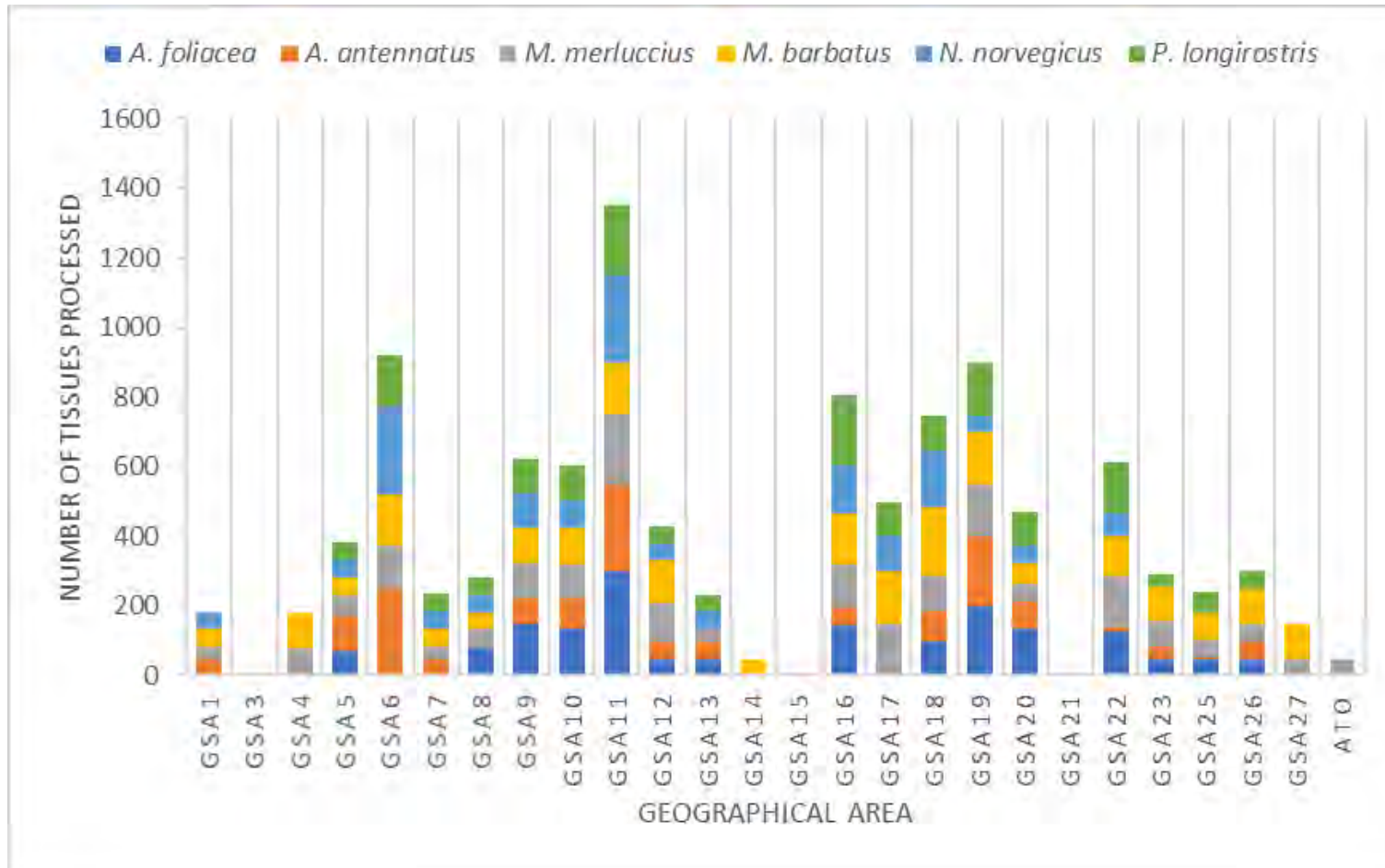


Figure 2 Number of individual tissues processed in the genetic laboratories by species and GSA. ATO refers to European hake samples collected from the Atlantic Ocean.

MED_UNITS Deliverable 1.6 – Detailed protocol for routine sampling and genetic monitoring

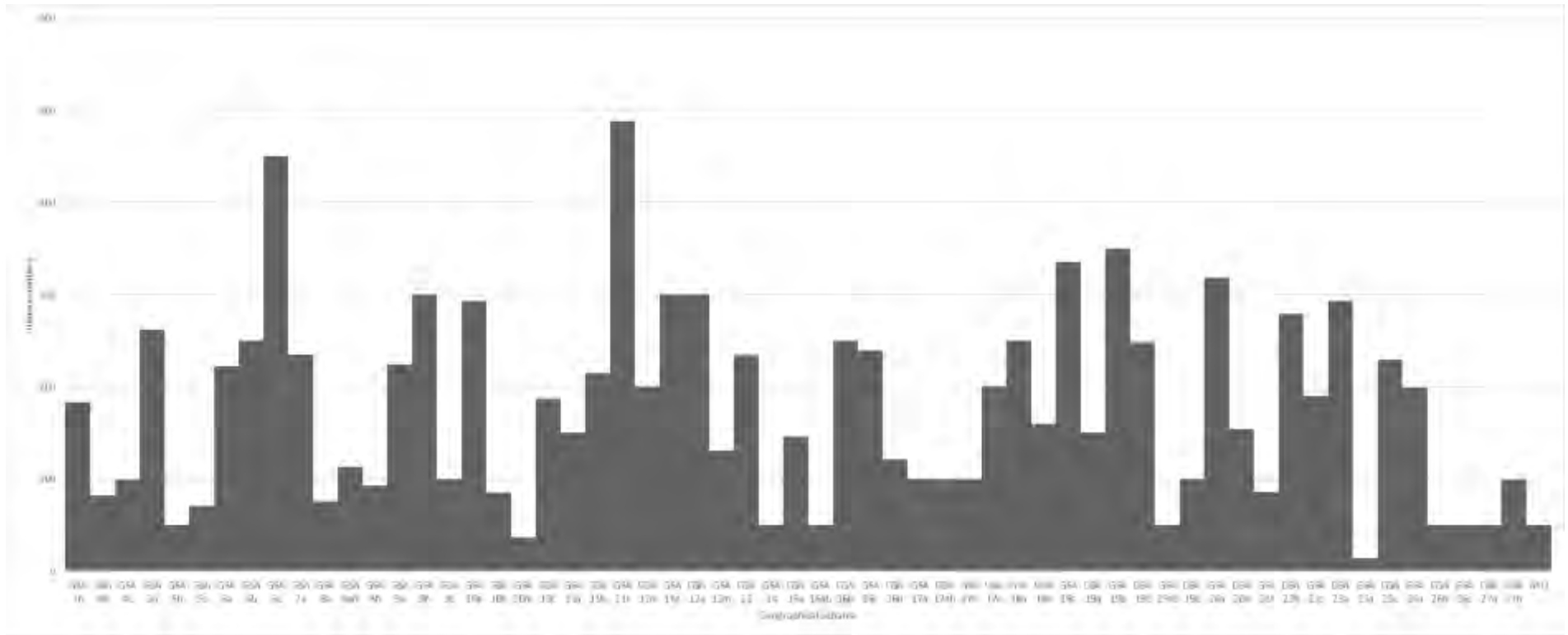


Figure 3 Number of individual tissues processed in the genetic laboratories by subareas. ATO refers to European hake samples collected from the Atlantic Ocean.

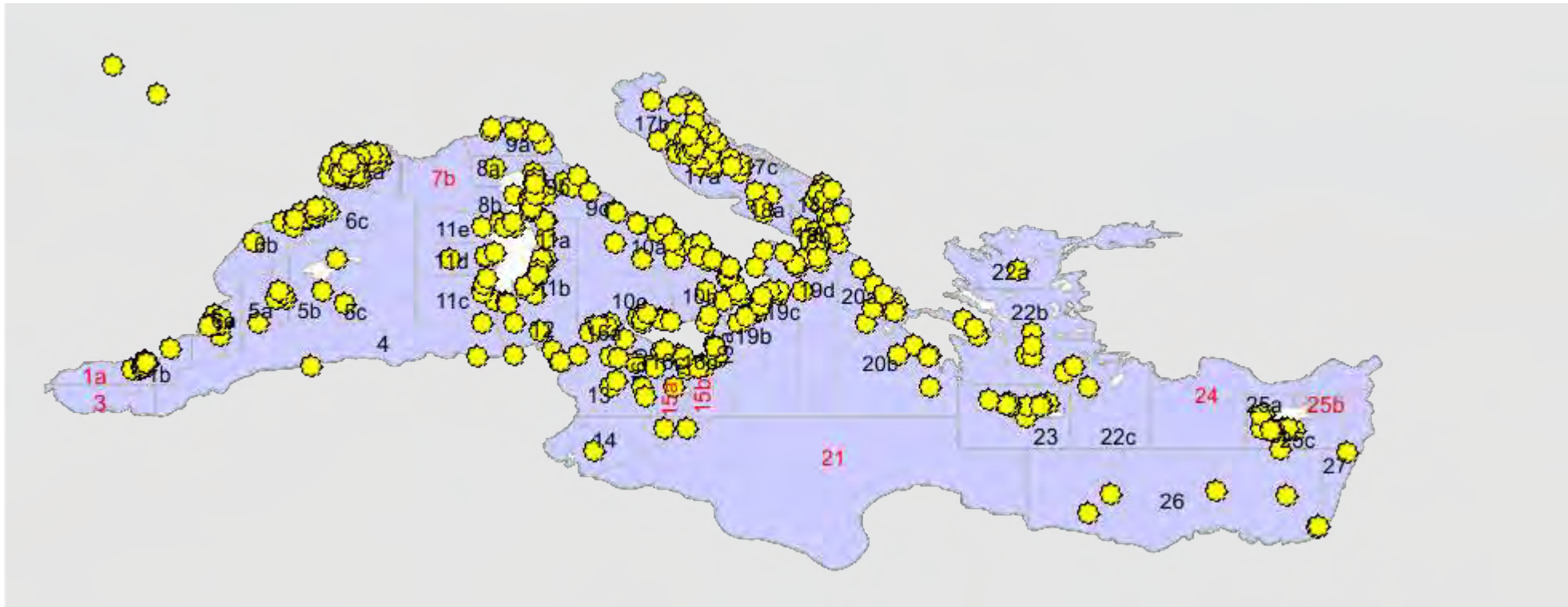


Figure 4 Spatial distribution of tissues for the six species under investigation. In red the GFCM areas or the subareas for which no tissues were available for the genetic analyses.

Tissue types and DNA extraction

A very important and critical aspect of sampling was the quality of tissues, from which high molecular weight DNA must be obtained as required for the ddRAD protocols.

According to the sampling protocol for genetic analyses (Annex 7 in D0.3), tissues were collected both from freshly caught (28.7%) and frozen specimen (52%). Unfortunately, for a fraction of tissues no indication was given on the sampling conditions and tissue type (19.3%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium quality or even unusable DNAs (Figure 5).

The definition of DNA quality categories (good, medium, unusable) is provided in Deliverable D1.4. In brief:

- Category 1 (good quality) refers to DNA extracts where only a sharp band of HMW DNA is present (no smear).
- Category 2 (medium quality) refers to DNA extracts where both a band of HMW DNA and some smears of degraded DNA are present, or to cases that the band of HMW is absent.
- Category 3 where the DNA is degraded or absent and cannot be finally used in the study.



Figure 5 Overall outcome of DNA extractions by type of tissues (see definition of DNA quality categories in the main text)

Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses

All DNAs classified as 'unusable' were discarded and not included in the following step (library preparation for the genotyping, Figure 6).

A total of 7,544 individuals were included in libraries, most of them (69.8%) having a 'good' DNA quality, the remaining (30.2%) were 'medium-quality' samples. Samples with good or medium DNAs were not used in libraries preparations when they were in excess, having met the requirements of the minimum

number of specimens per species/subarea/GSA (≥ 35 individuals) already included the analyses (Figure 6).

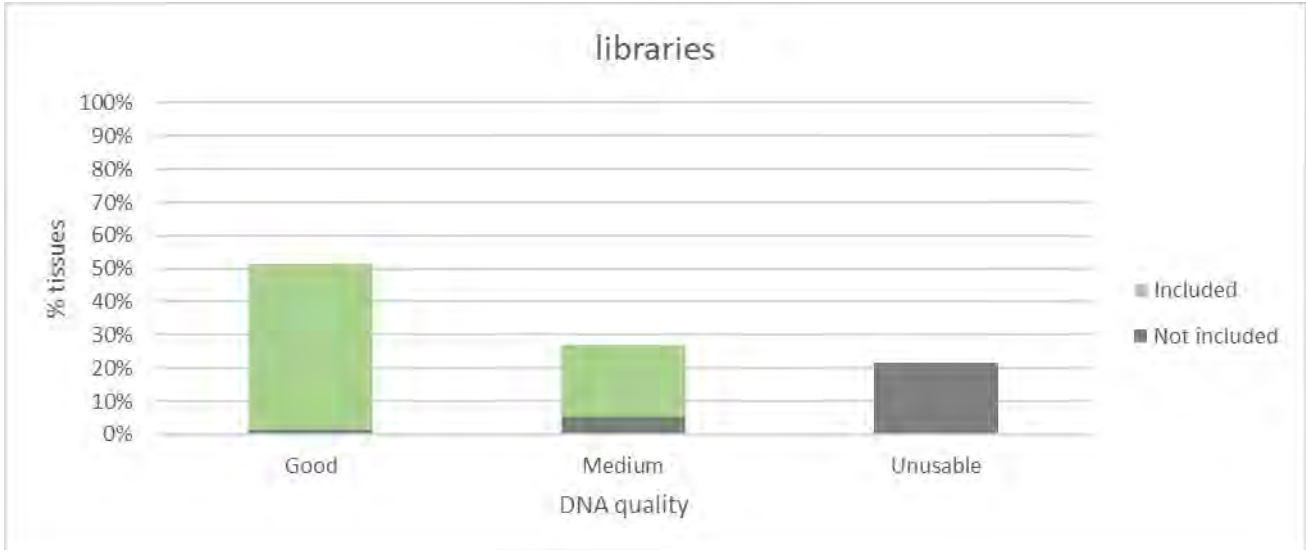


Figure 6 Overall percentage of tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 5,690 individuals were successfully genotyped (75.4% of those included in the libraries). The genotype success of the 'good DNA quality' tissues was high (82%) while only 61% of the 'medium quality' DNAs included in the libraries ended up producing genotype data (Figure 7).

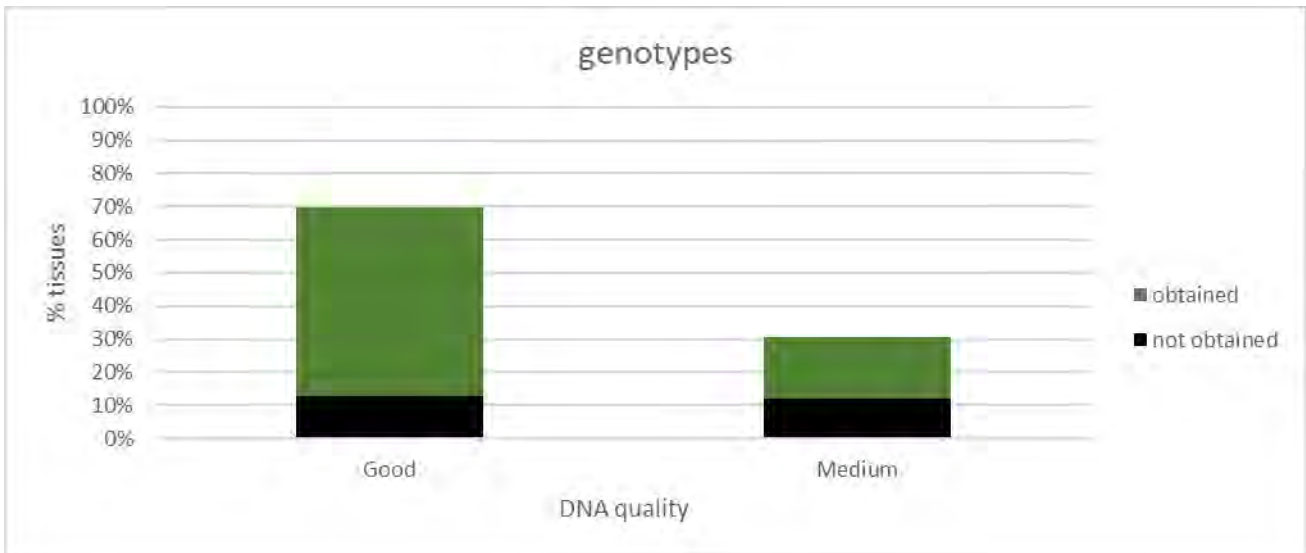


Figure 7 Overall percentage of tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Our results confirm what has already been reported in the literature. In general, 'good DNA quality' tissues gave better results measured as number of genotyped individuals. As the proportion of medium-quality DNA in libraries increased, the whole performance of the sequencing output decreased on overall for all the samples included. This is explained by the fact that the level of DNA degradation affects

the efficiency of reduced representation sequencing, increasing the number of missing loci. For instance, Guo, Yang, Chen, Li, and Guo (2018) showed that degraded samples resulted in the reduction of total reads using ddRAD. Similar results were observed by Graham et al. (2015), where incubation at room temperature of samples up to 96 hr induced DNA degradation reduced the final numbers of SNPs. These studies showed that RADseq performs best when using high-molecular-weight DNA to generate population data.

Since the quality of DNAs obtained by tissues as well as the genotype success varied considerably, detailed results by species are reported in the following paragraphs.

As concerns the bioinformatic pipeline and the population genetic analyses and stock identification methods, the same procedures (software and tools) were for all the species, fully described in D1.4, D1.5.1 and D1.5.2. No further details are included in this document on these aspects.

We consider the markers (SNPs) and methods of analysis as the most valid (suitable) for the studied species at present in producing population genetic data.

SPECIES BY SPECIES EVALUATION

Aristaeomorpha foliacea

SAMPLING

The detailed spatial distribution of samples analysed is depicted in Figure 8.

The vast majority of *A. foliacea* samples (89%) were collected within the first 12 months (by the end of 2019). Most of the samples from the western (GSA5) and eastern areas (GSA 20 and 22) were available after July 2019 so they were not included in the pilot phase.

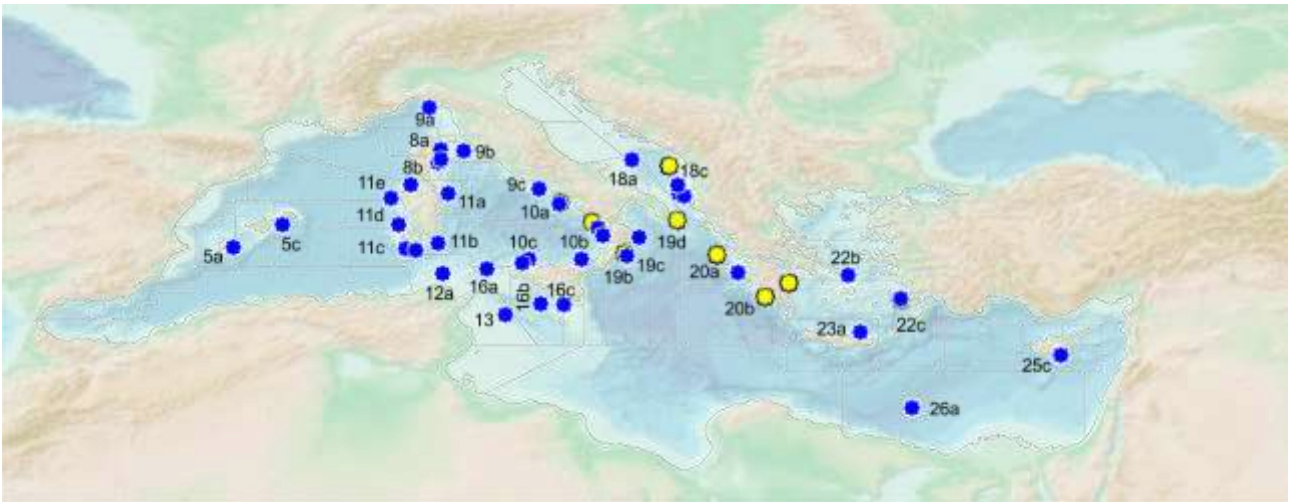


Figure 8 Spatial distribution of tissues for *A. foliacea*. The blue dots indicate samples that were successfully analysed and for which population genetic data were obtained. The yellow dots indicate samples for which tissues were available, or DNAs were extracted but population genetics data were not obtained.

TISSUE TYPES AND DNA EXTRACTION

Tissues were collected both from freshly caught (37.8%) and frozen specimen (41.4%). Unfortunately, for a noticeable fraction of tissues no indication was given on the sampling conditions and tissue type (20.8%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium or even unusable DNAs (Figure 9).

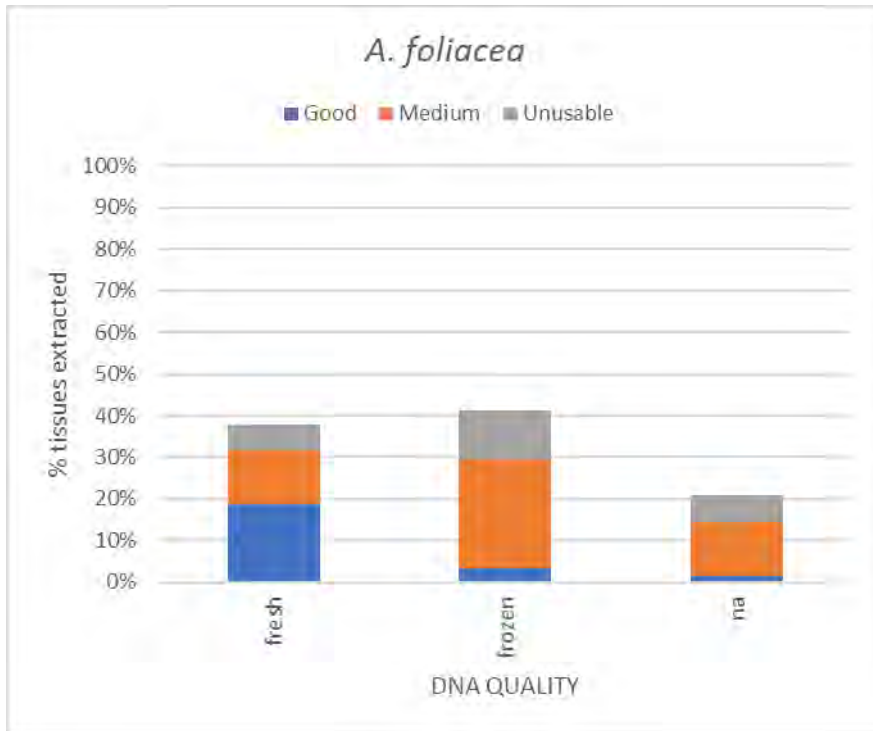


Figure 9 Overall outcome of *A. foliacea* DNA extractions by type of tissues (see definition of DNA quality categories in the main text).

LIBRARY PREPARATIONS, GENOTYPING-BY-SEQUENCE (GBS), BIOINFORMATIC ANALYSIS & POPULATION GENETIC ANALYSES

All DNAs classified as ‘unusable’ were discarded and not included in the following step (library preparation). A total of 1,266 individuals were included in libraries, having a ‘good’ or “medium” DNA quality (Figure 10). While all samples with good DNAs were used in libraries preparations, some medium DNAs were excluded when they were in excess, having met the requirements of the minimum number of specimens per species/subarea/GSA (≥ 35 individuals) already included the analyses.



Figure 10 Overall percentage of *A. foliacea* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 770 individuals were successfully genotyped (60.8% of those included in the libraries). The genotype success of the 'good DNA quality' tissues was 68.4% versus only 53.1% of 'medium' DNAs (Figure 11).

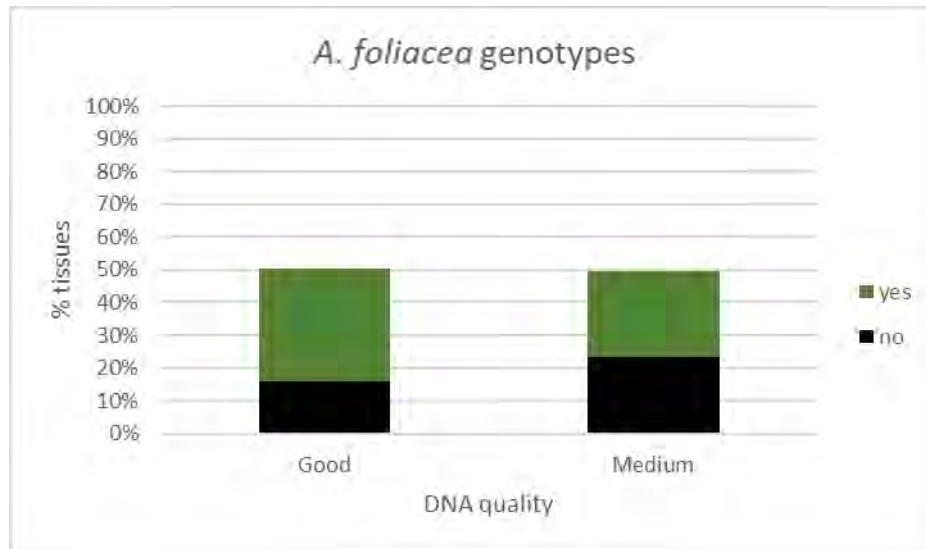


Figure 11 Overall percentage of *A. foliacea* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Aristeus antennatus

SAMPLING

The detailed spatial distribution of samples analysed is depicted in Figure 12.

The vast majority of *A. antennatus* samples (84%) were collected within the first 12 months (by the end of 2019). Some samples from the western areas (GSA1 and 5) as well as eastern areas (GSA 20-26) were available only after July 2019, so they were not included in the pilot phase.

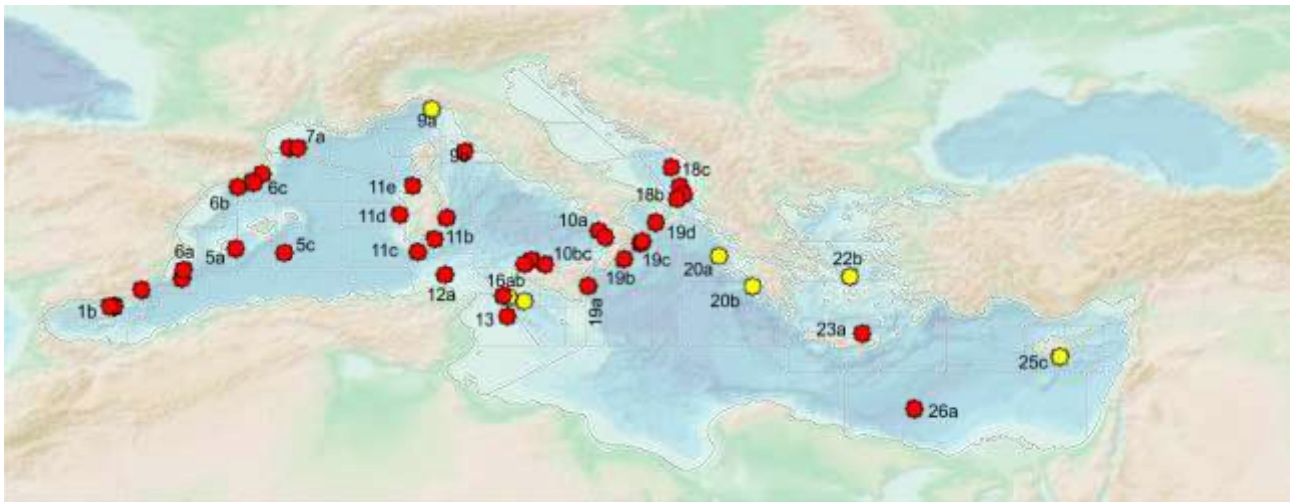


Figure 12 Spatial distribution of tissues for *A. antennatus*. The red dots indicate samples that were successfully analysed and for which population genetic data were obtained. The yellow dots indicate samples for which tissues were available, DNAs were extracted but population genetics data were not obtained.

TISSUE TYPES AND DNA EXTRACTION

Tissues were collected both from freshly caught (52.1%) and frozen specimen (25.1%). Unfortunately, for a conspicuous fraction of tissues no indication was given on the sampling conditions and tissue type (22.8%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium or even unusable DNAs (Figure 13).

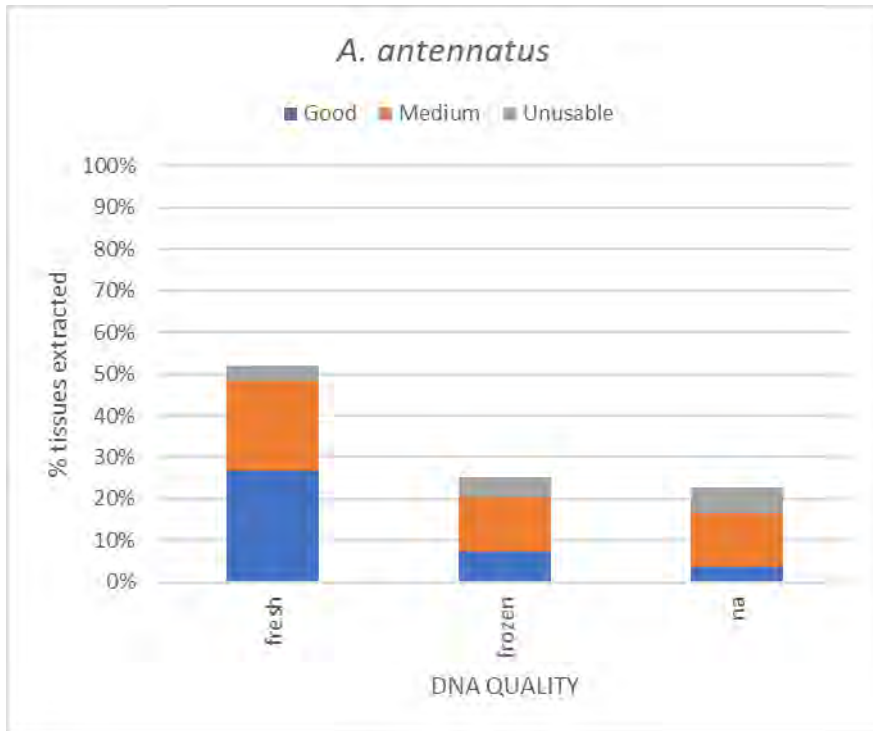


Figure 13 Overall outcome of *A. antennatus* DNA extractions by type of tissues (see definition of DNA quality categories in the main text).

LIBRARY PREPARATIONS, GENOTYPING-BY-SEQUENCE (GBS), BIOINFORMATIC ANALYSIS & POPULATION GENETIC ANALYSES

All DNAs classified as ‘unusable’ were discarded and not included in the following step (library preparation). A total of 1,040 individuals were included in libraries, having a ‘good’ or medium DNA quality (Figure 14). A few samples with good or medium DNAs were excluded from libraries when they were in excess, having met the requirements of the minimum number of specimens per species/subarea/GSA (≥ 35 individuals) already included in the analyses.

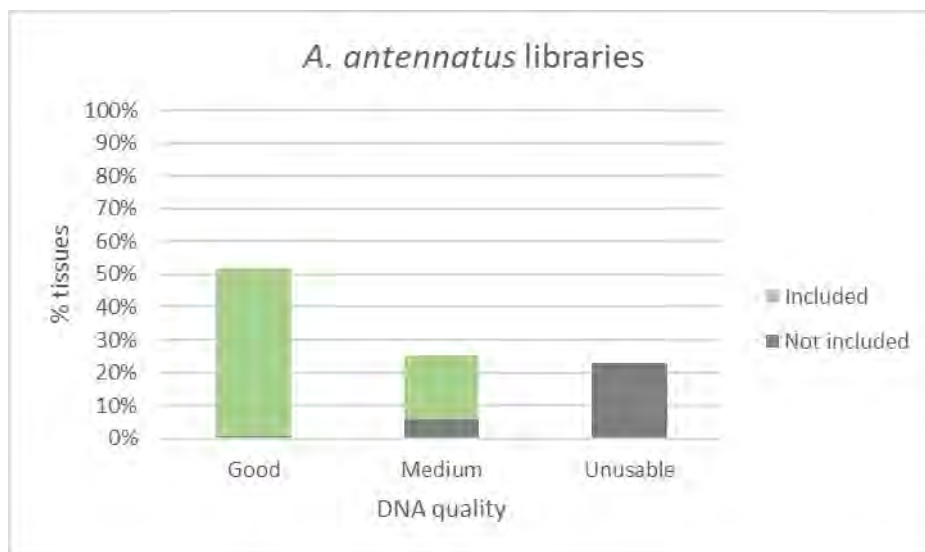


Figure 14 Overall percentage of *A. antennatus* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 825 individuals were successfully genotyped (79.3% of those included in the libraries). The genotype success of 'good DNA quality' tissues was 86.6% versus only 60% of 'medium quality' DNAs (Figure 15).

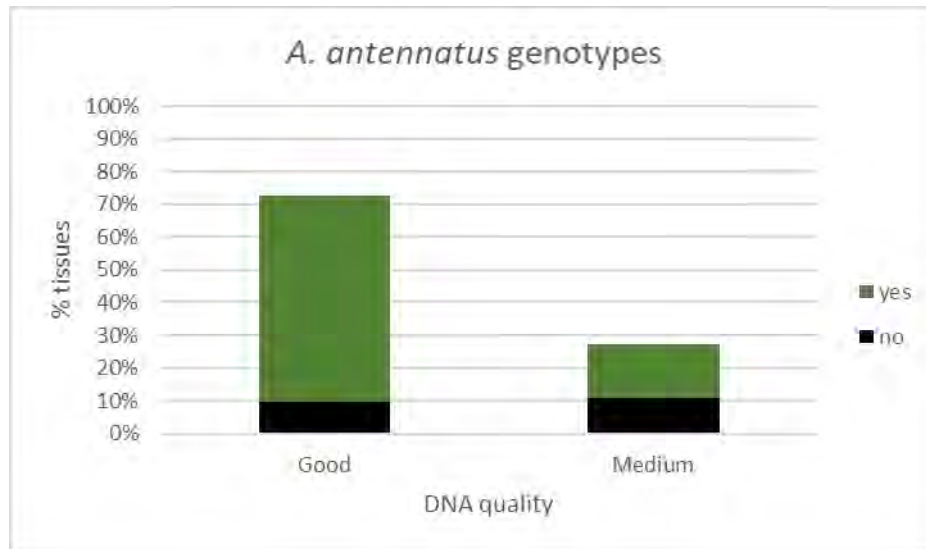


Figure 15 Overall percentage of *A. antennatus* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Merluccius merluccius

SAMPLING

The detailed spatial distribution of samples analysed is depicted in 19.

The vast majority of *M. merluccius* samples (87%) were collected within the first 12 months (by the end of 2019). Most of the samples from the westernmost and eastern areas were available after July 2019, so they were not included for the pilot phase (Table 1).

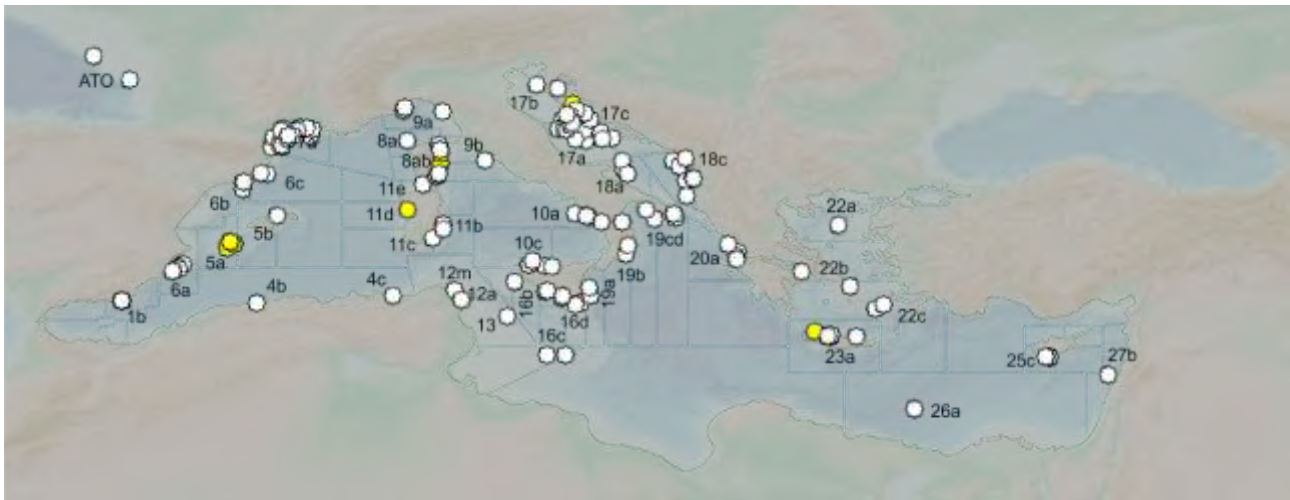


Figure 16 Spatial distribution of tissues for *M. merluccius*. The white dots indicate samples that were successfully analysed and for which population genetic data were obtained. The yellow dots indicate samples for which tissues were available, DNAs were extracted but population genetics data were not obtained (see main text for details).

TISSUE TYPES AND DNA EXTRACTION

Tissues were collected both from freshly caught (79.7%) and frozen specimen (16.5%). For a small fraction of tissues, no indication was given on the sampling conditions and tissue type (3.8%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium or even unusable DNAs (Figure 17).

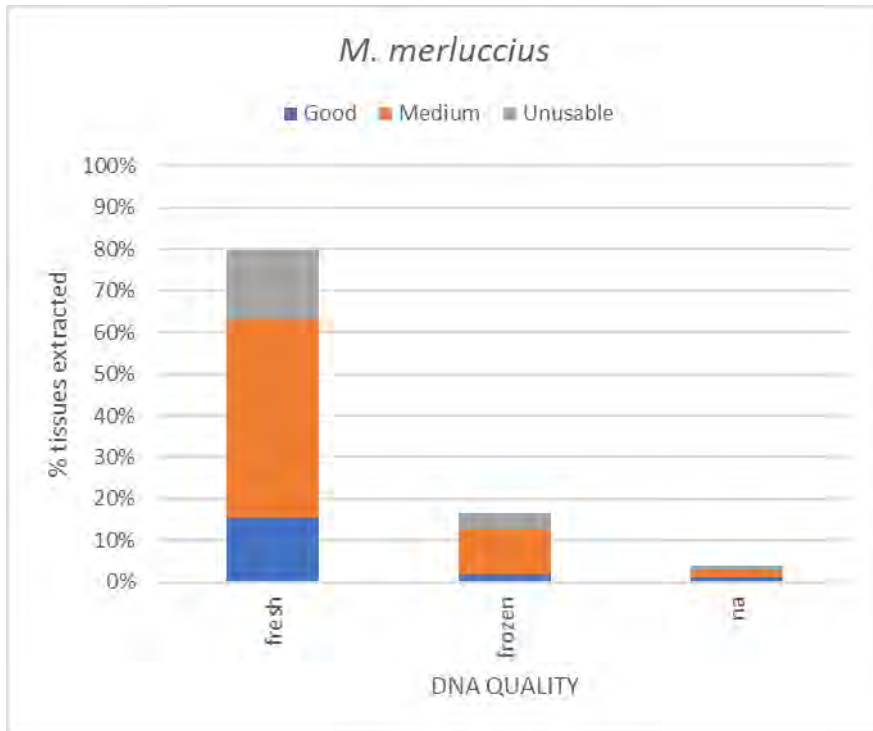


Figure 17 Overall outcome of *M. merluccius* DNA extractions by type of tissues (see definition of DNA quality categories in the main text).

LIBRARY PREPARATIONS, GENOTYPING-BY-SEQUENCE (GBS), BIOINFORMATIC ANALYSIS & POPULATION GENETIC ANALYSES

All DNAs classified as ‘unusable’ were discarded and not included in the following step (library preparation). A total of 1,705 individuals were included in libraries, having a ‘good’ or medium DNA quality (Figure 18). A few samples with good or medium DNAs were excluded from libraries when they were in excess, having met the requirements of the minimum number of specimens per species/subarea/GSA (≥ 35 individuals) already included in the analyses.

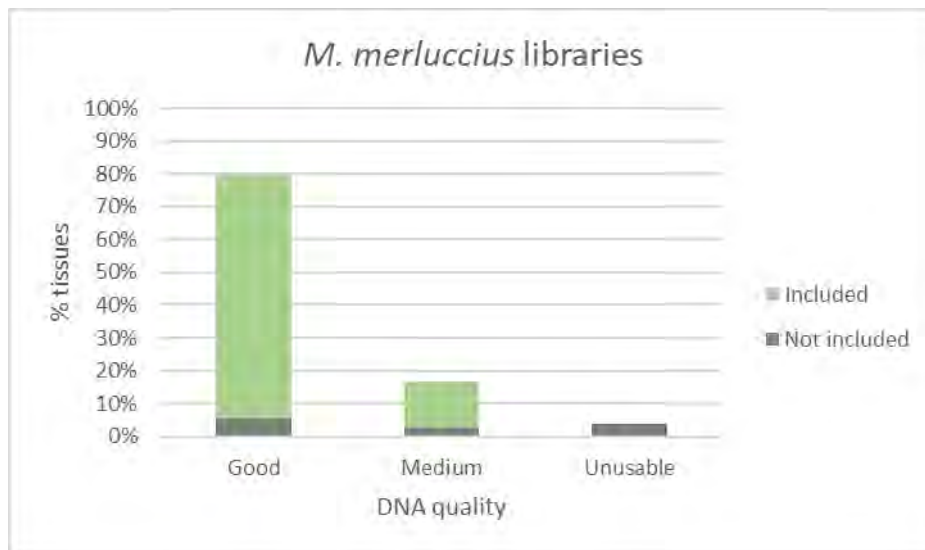


Figure 18 Overall percentage of *M. merluccius* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 1,664 individuals were successfully genotyped (97.6% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 98.5% versus 92.9% of ‘medium’ DNAs (Figure 19).

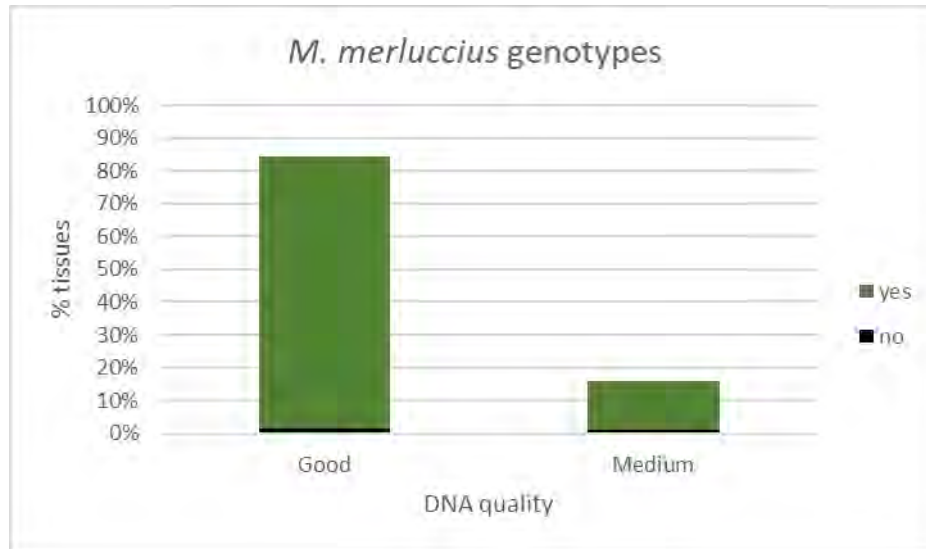


Figure 19 Overall percentage of *M. merluccius* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Mullus barbatus

SAMPLING

The detailed spatial distribution of samples analysed is depicted in Figure 20.

The vast majority of *M. barbatus* samples (83%) were collected within the first 12 months (by the end of 2019). Most of the samples from the westernmost and eastern areas were available after July 2019, so they were not included in the pilot phase (Table 1).

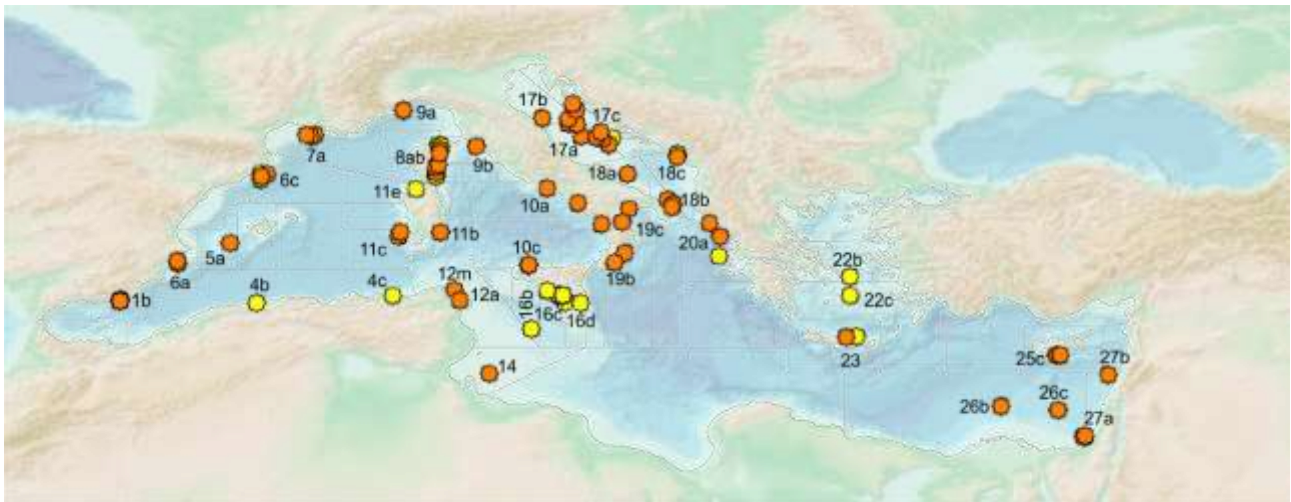


Figure 20 Spatial distribution of tissues for *M. barbatus*. The orange dots indicate samples that were successfully analysed and for which population genetic data were obtained. The yellow dots indicate samples for which tissues were available, DNAs were extracted but population genetics data were not obtained (see main text for details).

TISSUE TYPES AND DNA EXTRACTION

Tissues were collected both from freshly caught (35.2%) and frozen specimen (32.1%). Unfortunately, for a very conspicuous fraction of tissues no indication was given on the sampling conditions and tissue type (32.7%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium or even unusable DNAs (Figure 21).

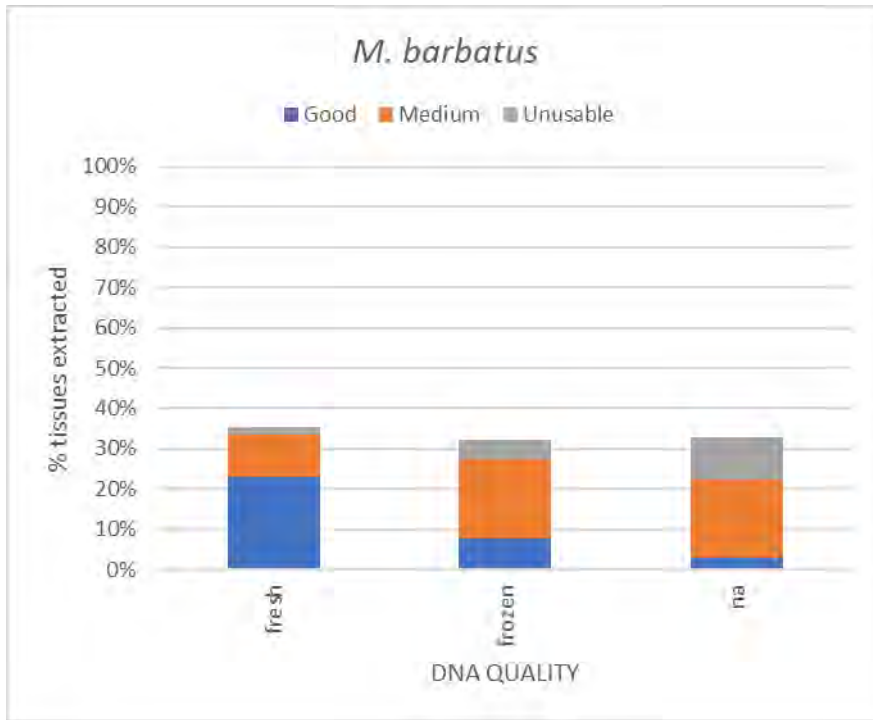


Figure 21 Overall outcome of *M. barbatus* DNA extractions by type of tissues (see definition of DNA quality categories in the main text).

LIBRARY PREPARATIONS, GENOTYPING-BY-SEQUENCE (GBS), BIOINFORMATIC ANALYSIS & POPULATION GENETIC ANALYSES

All DNAs classified as ‘unusable’ were discarded and not included in the following step (library preparation). A total of 1,373 individuals were included in libraries, having a ‘good’ or medium DNA quality (Figure 22). A few samples with medium DNAs were excluded from libraries when they were in excess, having met the requirements of the minimum number of specimens per species/subarea/GSA (≥ 35 individuals) already included the analyses.

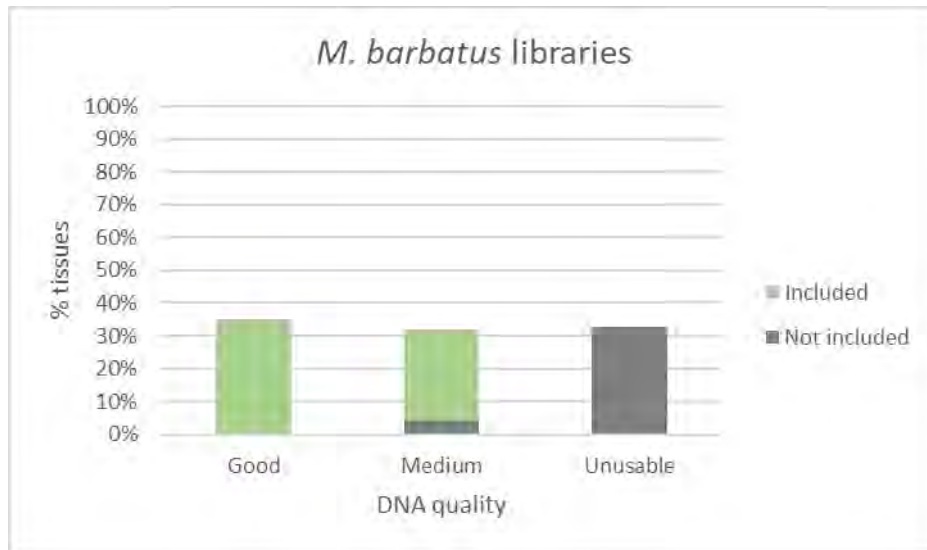


Figure 22 Overall percentage of *M. barbatus* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 771 individuals were successfully genotyped (56.1% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 57.8% versus 54% of ‘medium’ DNAs (Figure 23).

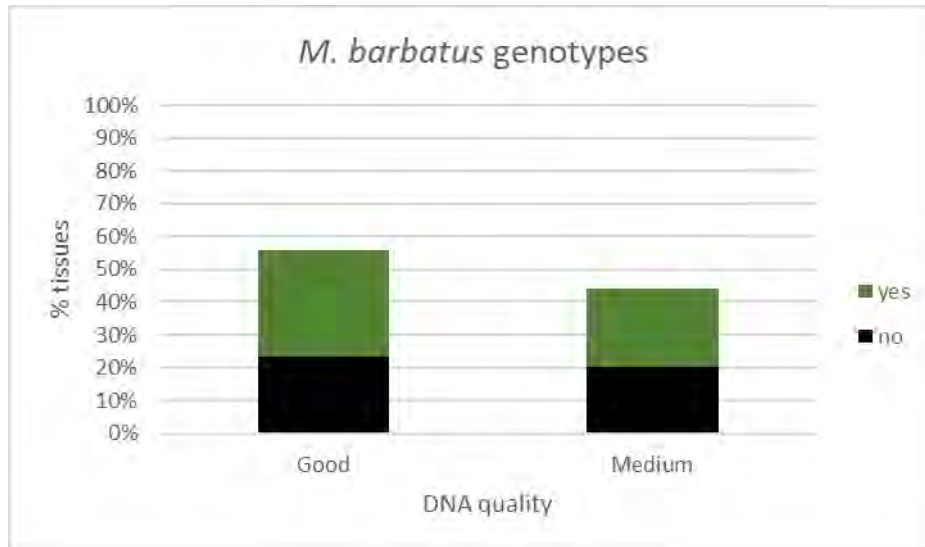


Figure 23 Overall percentage of *M. barbatus* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Nephrops norvegicus

SAMPLING

The detailed spatial distribution of samples analysed is depicted in Figure 24.

The vast majority of *N. norvegicus* samples (84%) were collected within the first 12 months (by the end of 2011). Most of the samples from the westernmost and eastern areas were available after July 2019, so they were not included in the pilot phase (Table 1).

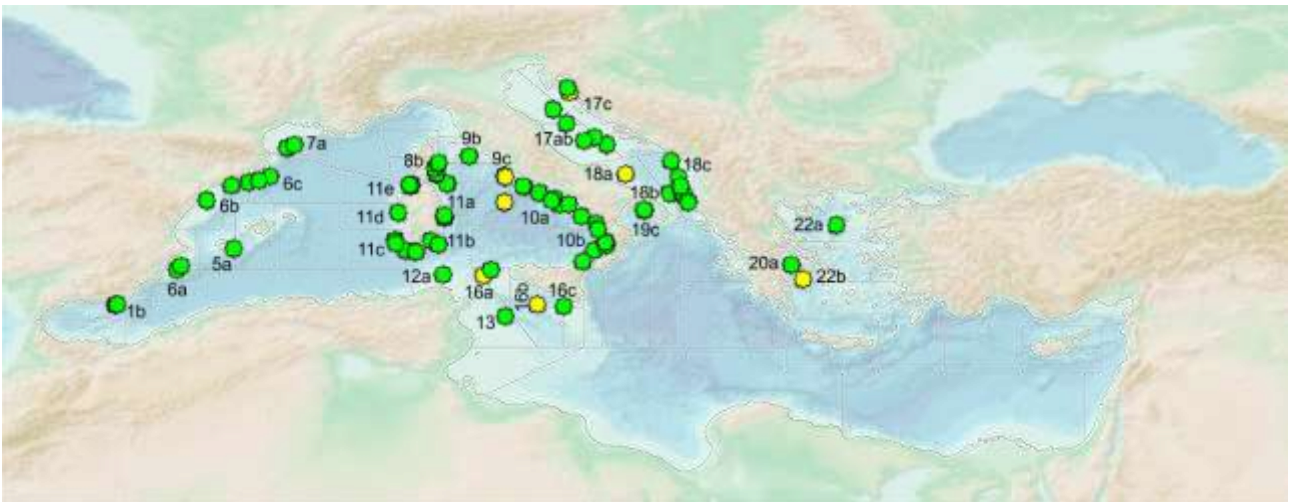


Figure 24 Spatial distribution of tissues for *N. norvegicus*. The green dots indicate samples that were successfully analysed and for which population genetic data were obtained. The yellow dots indicate samples for which tissues were available, DNAs were extracted but population genetics data were not obtained (see main text for details).

TISSUE TYPES AND DNA EXTRACTION

Tissues were collected both from freshly caught (61.2%) and frozen specimen (20.2%). Unfortunately, for a noticeable fraction of tissues no indication was given on the sampling conditions and tissue type (18.6%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium or even unusable DNAs (Figure 25).

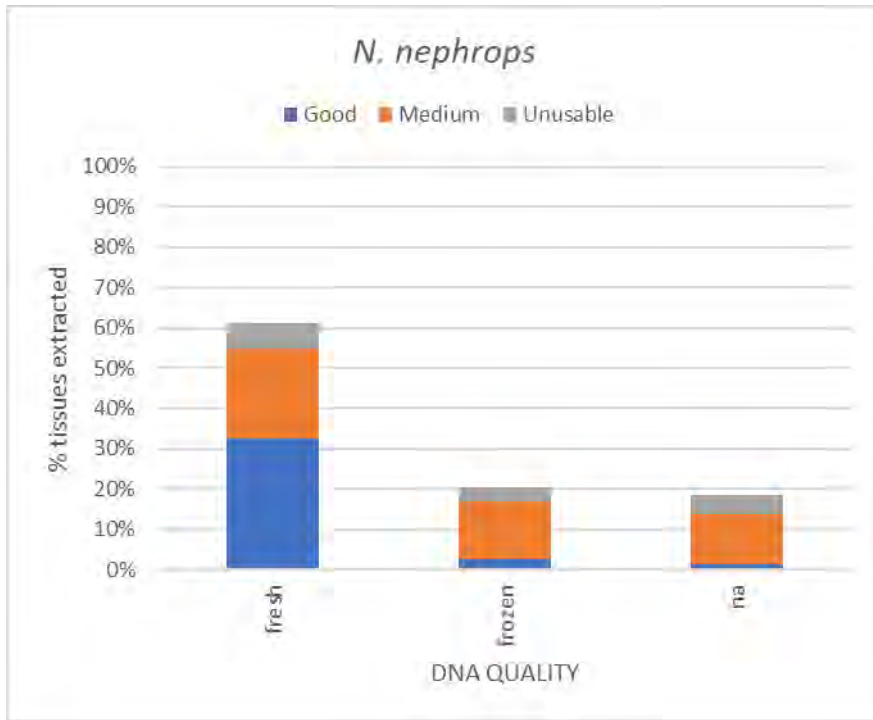


Figure 25 Overall outcome of *N. norvegicus* DNA extractions by type of tissues (see definition of DNA quality categories in the main text).

LIBRARY PREPARATIONS, GENOTYPING-BY-SEQUENCE (GBS), BIOINFORMATIC ANALYSIS & POPULATION GENETIC ANALYSES

All DNAs classified as ‘unusable’ were discarded and not included in the following step (library preparation). A total of 1,152 individuals were included in libraries, having a ‘good’ or medium DNA quality (Figure 26). A few samples with good or medium DNAs were excluded from libraries when they were in excess, having met the requirements of the minimum number of specimens per species/subarea/GSA (≥ 35 individuals) already included in the analyses.

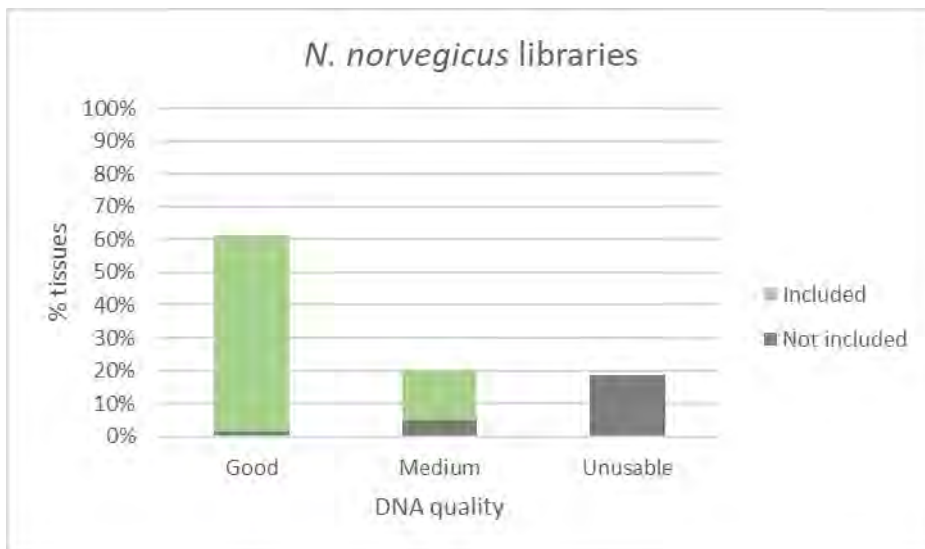


Figure 26 Overall percentage of *N. norvegicus* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 890 individuals were successfully genotyped (72.2% of those included in the libraries). The genotype success of ‘good DNA quality’ tissues was 84.2% versus only 50% of ‘medium’ DNAs (Figure 27).

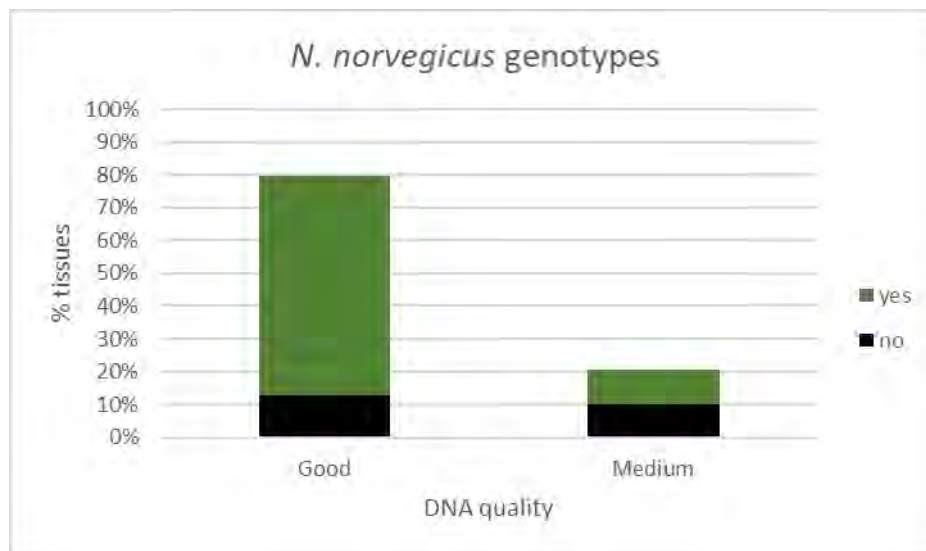


Figure 27 Overall percentage of *N. norvegicus* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Parapeneus longirostris

SAMPLING

The detailed spatial distribution of samples analysed is depicted in Figure 28.

The vast majority of *P. longirostris* samples (88%) were collected within the first 12 months (by the end of 2019). Most of the samples from the easternmost areas were available after July 2019, so they were not included in the pilot phase (Table 1).

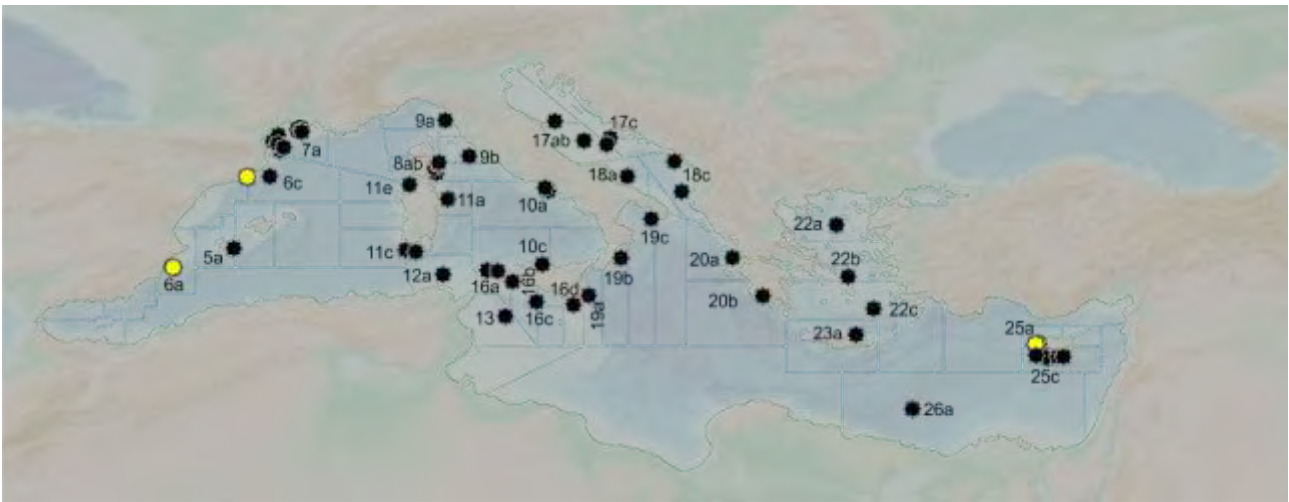


Figure 28 Spatial distribution of tissues for *P. longirostris*. The black dots indicate samples that were successfully analysed and for which population genetic data were obtained. The yellow dots indicate samples for which tissues were available, DNAs were extracted but population genetics data were not obtained (see main text for details).

TISSUE TYPES AND DNA EXTRACTION

Tissues were collected both from freshly caught (44.1%) and frozen specimen (26.0%). Unfortunately, for a conspicuous fraction of tissues no indication was given on the sampling conditions and tissue type (29.9%). In general, DNA of good quality was easily obtained from fresh tissues, while frozen tissues generated larger fraction of medium or even unusable DNAs (Figure 29).

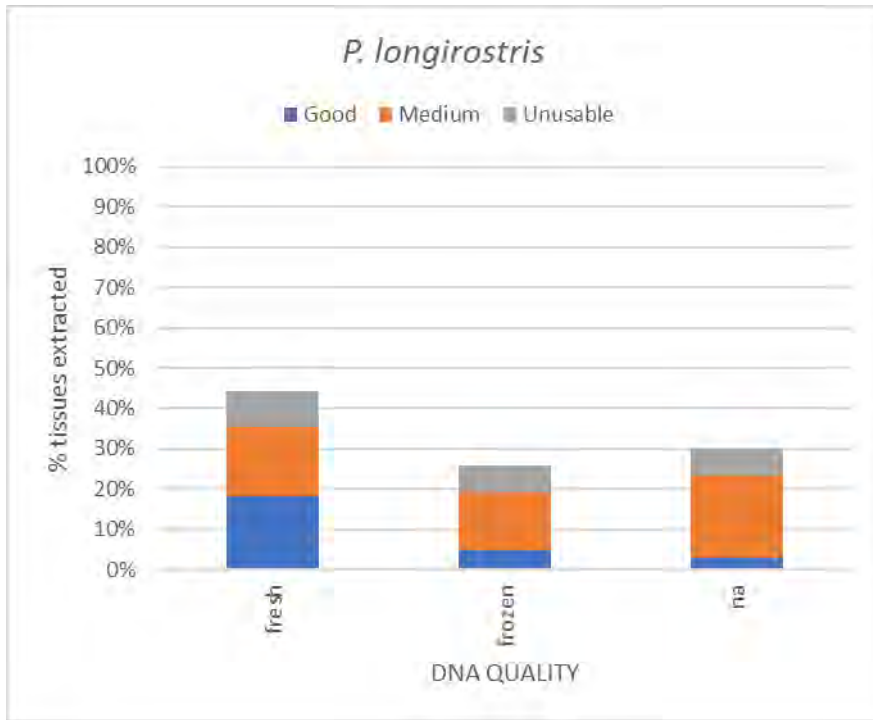


Figure 29 Overall outcome of *P. longirostris* DNA extractions by type of tissues (see definition of DNA quality categories in the main text).

LIBRARY PREPARATIONS, GENOTYPING-BY-SEQUENCE (GBS), BIOINFORMATIC ANALYSIS & POPULATION GENETIC ANALYSES

All DNAs classified as ‘unusable’ were discarded and not included in the following step (library preparation). A total of 1008 individuals were included in libraries, having a ‘good’ or medium DNA quality (Figure 30). A few samples with good or medium DNAs were excluded from libraries when they were in excess, having met the requirements of the minimum number of specimens per species/subarea/GSA (≥ 35 individuals) already included the analyses.

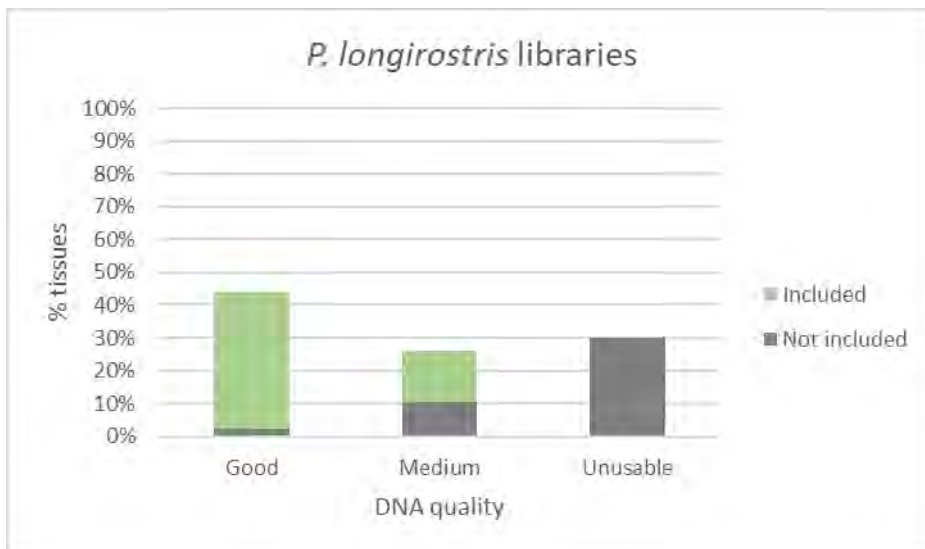


Figure 30 Overall percentage of *P. longirostris* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text)

A total of 770 individuals were successfully genotyped (79.4% of those included in the libraries). The genotype success of 'good DNA quality' tissues was 78.1% versus 71.9% of 'medium' DNAs (Figure 31).

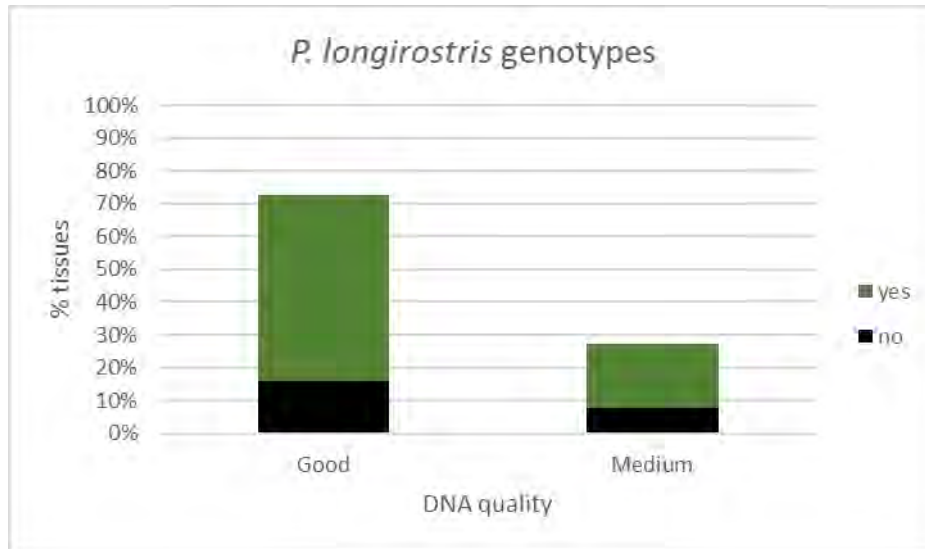


Figure 31 Overall percentage of *P. longirostris* tissues included in libraries by their DNA quality (see definition of DNA quality categories in the main text) for which genotypes were obtained

Lessons learned from MED_UNITS WP1

For the different aspects analyzed in the previous sections the main strong points (SP), weak points (WP) and suggestions (S) are summarized below.

Sampling: spatial and temporal coverage

SP: the accurate and wide- ranging sampling design with a clear subdivision in subareas allowed to obtain good geographical coverage, encompassing most of the geographical distribution range of the different species. A very efficient strategy proved to be setting up of a Genetic Hub in charge of selecting and sending suitable material for sampling to samplers, recollect the tissues from samplers, check and temporarily store tissues, and distribute tissues to final genetic labs performing the analyses.

WP: the collection of tissues was spread across 18 months (January 2019 – July 2020). A large fraction of tissues was collected during surveys at sea realized in the 2019 summer months. Contrarily to what was initially foreseen, the sampling for the full study was not tuned based on the results from the pilot study that were finally available some months after the end of the sampling campaigns at sea (Figure 1). Moreover, the samples available by July 2019 for the pilot phase were not all coincident with ones selected at the beginning. Therefore, Task 1.4 was inevitably modified based on the availability of samples present in the laboratories in August 2019 (Table 1). In some areas, especially in extra-EU countries, some bureaucratic obstacles made the shipping of tissues very problematic, contributing to exacerbate the delays.

S: Timely sampling is the crucial step for having reliable genomic data in a short-term project. A proper budget, time, storage facilities and staff for the sampling should be mandatorily allocated to each project. The timeframe for collecting tissues should be carefully defined to address the specific biological problem (stock identification), but also taking into consideration the biology (life history cycle) of the species under investigation and the necessity to make meaningful comparisons among populations (areas).

Tissue types, DNA extraction and DNA quality

SP: large numbers of individual tissues were collected by species, an unprecedented huge effort, leading to amounts not comparable with respect to previous studies. For the fish species, they were in general accompanied by the sampling of otoliths from the same individuals.

WP: a large fraction of tissues was collected from frozen individuals, that in general lead to medium-low DNA quality extracts and higher failure rates in the genotyping process; for microsatellites or mtDNA analyses, DNA samples could be more or less suitable but not for the latest ddRAD methodology. This was especially evident in *M. barbatus* and the three shrimp species where we were obliged to use many medium-quality DNAs in order to have all the areas (GSA) adequately represented.

S: A clear suggestion should be given to collect fresh tissues and speed up the sampling and processing procedures. Stronger specifications should be given of what “fresh” means with clear indication of the duration of the sampling and processing procedure in order to avoid animals to be kept for long periods in unsuitable conditions during the sampling. This hold particularly true for many crustacean species, where post-mortem DNA degradation is known to occur very quickly.

Library Preparations, Genotyping-By-Sequence (GBS), Bioinformatic Analysis & Population genetic analyses

SP: optimization of species-specific protocols, especially challenging for the crustacean species.

WP: very stringent time constraints that limited the possibility to further refine the protocols based on additional trials. The need to analyse samples from important (strategic areas) even when their quality was suboptimal. The unknown, and probably large and complex genomes of crustaceans considerably complicated the analyses.

S: Larger number of samples should be collected in order to overcome the unavoidable decrease in numbers due to low quality samples or failures during the various steps (library preparation, sequencing, bioinformatics). Crucial to have adequate time for the optimization of species-specific protocols with the possibility to make finer rearrangements on species basis.

Conclusions

- **The sampling is the crucial step**

Sample design and collection is the first step in any research project, but it can be very challenging when genomic analyses are foreseen. It often involves a tremendous effort from many people. However, field conditions, limited funding and time constraints often hinder the accurate collection and sample preservation.

When qualified scientific personnel perform the sampling onboard, the conditions could be not ideal (Rodriguez-Ezpeleta et al. 2013) because of

- ✓ the lack of clean and sterile surfaces as well as weather conditions that can prevent proper manipulation of the samples;
- ✓ the large number of specimens generally caught at the same time, allowing sampling to be performed soon after death only in some individuals;
- ✓ the sampling is the result of collaborative efforts among different institutions, not directly involved in DNA analysis and, therefore, are not aware of the strictness required during the sampling process.

When sampling onboard is not possible, samples are collected at landing or once the specimens arrived at the laboratories, but the conditions in which they have been stored in the boat or during the transport to the laboratories and after (till tissue dissection) are not known or not optimal (Rodriguez-Ezpeleta et al. 2013).

Nevertheless, the proper sampling and preservation of samples prior to DNA extraction is very important, because depending on the storage condition, DNA degrades over time and becomes unsuitable for new generation molecular studies.

Therefore, we highly recommend that:

- A proper budget, time and manpower for the sampling should be mandatorily allocated in each project.
- The sampling should be designed in order to fulfil the needs of ‘genomics’ requirements (in terms of quality of samples, procedures, storage, timing) with special indications for shorter preservation times, or the need to include additional steps that may increase cost and logistical constraints. The opposite is not valid: the genomic analyses cannot accommodate or rely on any sample taken whatever conditions only because they are the usual or simplest ways to work in the field.
- The sampling for genomic analyses cannot be a collateral activity, opportunistically realized during the standard surveys at sea, as a supplementary task in addition to the multiple activities performed on board by the scientific staff.
- It is important to collect more samples than actually needed in a given area with a surplus of at least 15-20%, to compensate for the unavoidable decrease in numbers due to the presence of low-quality tissues that are discarded immediately or the subsequent failures in the different steps leading to the genotyping.

- **The pilot phase for optimizing protocols by species cannot be skipped**

It is always worth evaluating the best method of preservation to ensure that samples will yield the best data possible. In particular, preservation of tissues for DNA extraction is a really important procedure to make samples kept at the original best quality (Oosting et al. 2020).

It is not easy to define a standard high performing protocol with a universal validity since it is highly dependent on the organism, tissue type, and study design i.e., time spent in the field, available resources/staff for sampling and storage facilities (Oosting et al. 2020).

Therefore, we highly recommend that:

- ✓ in the planning of any project, the timing for the experimental phases cannot be too strict but extensive pilot experiments at the beginning of a project are always performed for testing different preservation solutions under a range of storage conditions over different periods of time, to provide practical guidelines for DNA preservation (Oosting et al. 2020), library preparation, sequencing conditions, bioinformatic, population genomic analyses and stock identification.

- **The fresher the better**

Recently, several documents have defined and revised the procedure for collecting and analysing biological material for genetic identification of the species, for informational purposes only, or as an expert evidence for all possible presentations of fishery products (Bandarra et al. 2020a; Bandarra et al. 2020b; ICES 2021).

However, it is widely recognized that the requirements of DNA preservation for population genomic research are more demanding and can be difficult to meet when sampling field conditions are suboptimal (Oosting et al. 2020; Rodriguez-Ezpeleta et al. 2013). High rates of DNA degradation can potentially render samples unusable for most genomic applications. In fact, Next-generation sequencing (NGS) applications typically require high-molecular-weight (HMW) DNA (>20 kbp) for library preparation and sequencing to obtain high-quality genomic datasets. Such requirements are more demanding than those for traditional PCR-based approaches. As a consequence, traditional methods for sampling and preserving samples for genetic analysis may perform sub-optimally or fail to meet requirements of DNA preservation for genomic research (Oosting et al. 2020). Above all, it is well known that DNA degradation affects the efficiency of reduced representation library preparation and sequencing, increasing the number of missing loci (Guo et al. 2018) and/or reducing the final number of SNPs (Graham et al. 2015).

DNA degradation and DNA damage occur through enzymatic processes, oxidative damage, UV radiation, and hydrolysis. DNA degradation starts within minutes or hours after sampling from a live specimen and will continue to operate regardless of how the DNA has been preserved. DNA should be ideally extracted immediately after tissue sampling or stored at subzero temperatures and extracted shortly after (e.g., within 24 h). Keeping samples at very low temperatures will slow down the enzymatic degradation of DNA (Oosting et al. 2020).

Measuring the effect of tissue type, storage time, preserving conditions and post-mortem interval in fish DNA integrity showed that post-mortem interval is the one that most drastically affects DNA integrity. Fish tissues sampled 24 hours after death yield partial or totally degraded DNA (Rodriguez-Ezpeleta et al. 2013). Similarly, a recent study detected early signs of DNA degradation in crustacean (crab) tissues and hence confirmed that crustacean tissue starts to degrade immediately after death. Besides, it proved that freezers with constant temperature cycling also cause DNA shearing when tissues are left in there for long periods (days or weeks). The possible reason for such rapid degradation is believed to be hemocyte degranulation in crustaceans (Mulcahy et al. 2016).

We highly recommend that:

- The collection of tissues should preferentially be realized from alive/freshly caught individuals; additionally, we suggest the immediate addition of a preservation solution and the storage in proper conditions until the DNA extraction is performed.
- It should be clearly understood that the collection of fresh tissues means not only that the specimens will not get frozen but also that the sampling is realized as soon as possible within 1-2 hours from the death of the animal, always kept in optimal conditions (cold temperature).

- **Crustaceans can be particularly tricky**

Crustaceans as well as other marine invertebrates are reported to have genomes with high density of repetitive sequences, which hinder most genomic approaches. To overcome this challenge, starting DNA should be of high quality and integrity to fully exploit the capacity of the long-read sequencing platform (Angthong et al. 2020).

Obtaining such high-quality genomic DNA from crustaceans faces several challenges because crustacean DNA is easily degraded due to endonuclease enzymes (Rahman et al. 2017) and the purity can be jeopardized by high amounts of polysaccharides and polyphenolic proteins, which can inhibit molecular applications such as DNA library preparations (Angthong et al. 2020; Panova et al. 2016).

While several DNA extraction methods have been evaluated for routine PCR applications, only a few have already been tested for long-read sequencing. For instance, Angthong et al. (2020) compared the effectiveness of five DNA extraction protocols in obtaining shrimp high molecular weight (HMW) DNA for a long-read sequencing platform and identified a specific kit which yielded the highest quality genomic DNA. In a different study, high quality DNA was produced only through spin column extraction method from tissue samples preserved in 100% ethanol and RNAlater (Rahman et al. 2017). Finally, different tissue preservation methods were found to influence the production of genome-quality DNA (gDNA) for crustacean (crab) suggesting that 95% ethanol was a better preservative than RNAlater (Mulcahy et al. 2016).

Based on the literature and our experience in this project, we highly recommend, especially for crustaceans:

- to use only fresh and well-preserved tissues
- to allocate the proper time to optimize the DNA extraction method for each crustacean species under study. This is especially true for shrimps because their muscles consist of

numerous fiber proteins, majorly composed of myofibrillar and sarcoplasmic proteins. Thus, finding the best conditions for digesting these fibrous proteins from the beginning may result in higher DNA yields (Angthong et al. 2020).

Final updated field protocols sampling to be used in genomic studies

Sampling protocol available as **ANNEX 1**

Visual sampling protocol available as **ANNEX 2**

Excel sheet for recording sampling data available as **ANNEX 3**

References

- Angthong, P., T. Uengwetwanit, W. Pootakham, K. Sittikankaew, C. Sonthirod, D. Sangsrakru, T. Yoocha, I. Nookaew, T. Wongsurawat, P. Jenjaroenpun, W. Rungrassamee & N. Karoonuthaisiri, 2020. Optimization of high molecular weight DNA extraction methods in shrimp for a long-read sequencing platform. *Peerj* 8 doi:10.7717/peerj.10340.
- Bandarra, N., A. R., C. Cardoso, G.-C. F., A. H. Rojas, N. Rodriguez-Ezpeleta & H. Arrizabalaga, 2020a. Fish products sampling for DNA testing. *International Manual of Procedures*.
- Bandarra, N., A. R., C. Cardoso, G.-C. F., A. H. Rojas, N. Rodriguez-Ezpeleta & H. Arrizabalaga, 2020b. Study to produce an International Manual of Procedures (IMP) to be used in the NAFO Regulatory Area to guide the collection of samples from fisheries products for genetic analysis. Final Report. Study to produce an International Manual of Procedures (IMP) to be used in the NAFO Regulatory Area to guide the collection of samples from fisheries products for genetic analysis Final Report
- Graham, C. F., T. C. Glenn, A. G. McArthur, D. R. Boreham, T. Kieran, S. Lance, R. G. Manzon, J. A. Martino, T. Pierson, S. M. Rogers, J. Y. Wilson & C. M. Somers, 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular ecology resources* 15(6):1304-15 doi:10.1111/1755-0998.12404.
- Guo, Y., G. Yang, Y. Chen, D. Li & Z. Guo, 2018. A comparison of different methods for preserving plant molecular materials and the effect of degraded DNA on ddRAD sequencing. *Plant Diversity* 40(3):106-116 doi:https://doi.org/10.1016/j.pld.2018.04.001.
- ICES, 2021. Workshop on the manual for genetic sampling from fisheries products in the NAFO area (WKGenMan). ICES.
- Mulcahy, D. G., K. S. Macdonald, III, S. G. Brady, C. Meyer, K. B. Barker & J. Coddington, 2016. Greater than X kb: a quantitative assessment of preservation conditions on genomic DNA quality, and a proposed standard for genome-quality DNA. *PeerJ* 4:e2528 doi:10.7717/peerj.2528.
- Oosting, T., E. Hilario, M. Wellenreuther & P. A. Ritchie, 2020. DNA degradation in fish: Practical solutions and guidelines to improve DNA preservation for genomic research. *Ecology and Evolution* 10(16):8643-8651 doi:10.1002/ece3.6558.
- Panova, M., H. Aronsson, R. A. Cameron, P. Dahl, A. Godhe, U. Lind, O. Ortega-Martinez, R. Pereyra, S. V. Tesson, A. L. Wrangé, A. Blomberg & K. Johannesson, 2016. DNA Extraction Protocols for Whole-Genome Sequencing in Marine Organisms. *Methods in molecular biology* (Clifton, NJ) 1452:13-44 doi:10.1007/978-1-4939-3774-5_2.
- Rahman, S., D. Schmidt & J. Hughes, 2017. Development of an Efficient Method for Producing High Quality Genomic DNA in Crustaceans. *Asian Journal of Animal Sciences* 11(5):214-220 doi:10.3923/ajas.2017.214.220.
- Rodriguez-Ezpeleta, N., I. Mendibil, P. Álvarez & U. Cotano, 2013. Effect of fish sampling and tissue storage conditions in DNA quality: considerations for genomics studies. *Revista de Investigación Marina, AZTI-Tecnalia* 20(6):77-87.

MED_UNITS



“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP2 - OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

D.2.1

Section I - Report on the literature review on the otolith shape

RESPONSIBLE: ANDREA MASSARO (CIBM)

Section II - Report on the literature review on the otolith microchemistry

RESPONSIBLE: IGNACIO CATALAN (IMEDEA CSIC/UIB)

March, 30 2019

Index

SECTION I.....	3
I.Executive summary	3
I.1. INTRODUCTION.....	3
I.2. MATERIAL AND METHODS.....	4
I.3. RESULT AND DISCUSSION	5
Shape analysis methodologies	6
Variation of otolith shape according to potential effects.....	10
Case study of <i>Merluccius</i> spp.....	13
Case study of <i>Mullus</i> spp.....	15
I.4. REFERENCES	18
SECTION II	29
II.Executive summary	29
II.1. INTRODUCTION	30
II.2. MATERIAL AND METHODS	30
II.3. RESULTS	31
Literature review results.....	31
General history	31
Microchemistry analysis methodologies	31
Otolith preparation for LA-ICPMS.....	32
Otolith preparation for Carbonate and Oxygen isotopes analysis.....	32
Otolith analysis with LA-ICPMS	33
Otolith analysis for Cand O isotopes	34
II.4. DISCUSSION	34
II.5. CONCLUSION.....	35
II.6. REFERENCES.....	36

SECTION I

I.Executive summary

A review of more than 600 papers related with otolith shape analysis was performed to obtain a general overview on applications and methodologies of this analytical approach and its potentiality. Particular attention was addressed to stock identification studies focusing on *Merluccius spp.* and *Mullus spp.*

Otolith shape analysis have applications in different research fields, most of the publications concern shape analysis as tool to discriminate different stocks.

Different approaches have been used to describe and compare the morphology of otoliths, and two main groups can be recognized to describe and compare otoliths shape: univariate methods, using linear measurement, and multivariate methods describing the whole otolith shape from a mathematical point of view. Fourier, Wavelet Transform, Geomorphometric Analysis, Geodesic approach and Curvature Scale Space represent the most used methods to describe the whole otolith shape; while Shape Indices (Roundness, Circularity, Form Factor, Rectangularity, Ellipticity, Aspect Ratio) represents an option to analyze otolith shape differences.

Ten studies were carried out to discriminate *Merluccius spp.* stocks in different areas. Mediterranean population of *M. merluccius* was evaluated in one study combining a multiple approach analysis (chemical and shape). With reference to *Mullus barbatus*, five studies were performed.

In general, otolith shape analysis represents an useful tool in stock identification; moreover it seems a very low-cost method with high accuracy in the results making this analyses a feasible tool to evaluate the structure of populations at different levels.

I.1. INTRODUCTION

Otoliths, calcified structures overlying the sensory epithelia in the inner ears (Popper and Lu 2000; Campana 2004), are used firstly to produce the ageing data which are necessary to management and assessment of stocks. In the world, each year, 800 000 otoliths are collected to follow the fish stocks (Campana and Thorrold, 2001). Moreover, the otoliths are used for other applications in fisheries sciences. They grow throughout the life of the fish and, unlike scales and bones, are metabolically inert (i.e. once deposited, otolith material is unlikely to be resorbed or altered, Casselman, 1987). Consequently, otolith shape remains unaffected by short-term changes in fish condition (Campana and Casselman, 1993) or environmental variations (Campana, 1999). Consequently, they represent efficient tool for fish stock identification (Campana and Neilson, 1985; Campana and Casselman, 1993; Ihssen *et al.*, 1981; Pawson and Jennings, 1996; Cardinale *et al.*, 2004), based on phenotypic differences

(Casselman et al., 1981; Begg and Waldman, 1999). A wide number of techniques were developed and applied to identify and discriminate stocks, such as tagging experiments or analyses of spatial variation in arrange of markers including genetic markers, morphological traits, life-history traits at various life-stages, parasite load or infracommunity structure, or contaminant concentration (e.g. Pawson and Jennings, 1996; Garcia et al., 2011; Cadrin et al., 2014; ICES, 2016, Pita et al., 2016). However, the otolith shape with the rapid development of all steps of images analysis and statistical tools, has become a very used low-cost method with high accuracy in the results (Begg and Brown, 2000).

Otolith's shape is species-specific, but can also show intra-specific geographic differences within a species (L'Abée-Lund, 1988; Vignon, 2012; Lombarte and Lleonart, 1993). Several factors could affect otoliths shape: behaviour (food or spatial-temporal differences) (Aguirre and Lombarte, 1999; Cardinale *et al.*, 2004; Mérigot *et al.*, 2007; Gauldie and Crampton, 2002; Lombarte and Cruz, 2007; Sadighzadeh et al., 2014; Lombarte et al., 2010), abiotic factors (temperature, salinity) (Campana and Casselman, 1993; Gallego *et al.*, 1996; Hüsey, 2008), variations in diet habits (Gagliano and McCormick, 2004), somatic growth rate (Campana and Neilson, 1985; Campana and Casselman, 1993; Begg and Brown, 2000; Simoneau *et al.*, 2000; Cardinale *et al.*, 2004) and genetic (Vignon and Morat, 2010; Smith *et al.*, 2002).

Within stock, sex, age, and year class variation could lead to bias in discrimination (Castonguay *et al.*, 1991; Begg and Brown, 2000).). Moreover, otolith shape is also known to vary potentially intra-individually as left otolith shape may not be perfectly symmetrical to right otolith shape one and vice-versa (Díaz-Gill et al., 2015, Mahé et al., 2018).

This work aims at performing an overview of otolith shape analysis applications and methodologies, with particular attention on stock identification studies focusing on *Merluccius* spp. and *Mullus* spp.

I.2. MATERIAL AND METHODS

The search of the existing literature was performed using Web of Science selecting Web of Science Core Collection as database representing the world's leading scholarly journals, books, and proceedings in the sciences.

A general search process was carried out using specific topics to identify peer-reviewed articles related to otolith shape analysis used to discriminate fish stocks. In particular, research was carried out using as topic words

- "OTOLITH" and "SHAPE"
- "OTOLITH" and "SHAPE" and "STOCK"

In order to focus on literature regarding any otolith shape studies carried out on *Merluccius* spp. and *Mullus* spp. "STOCK" word was not considered.

- "OTOLITH" and "SHAPE" and "MERLUCCIUS"
- "OTOLITH" and "SHAPE" and "MULLUS"

Moreover, due to limited research regarding shape analysis on *Merluccius* spp. and *Mullus* spp. in the Mediterranean Sea, "MEDITERRANEAN" was not used in the research.

All abstracts resulting from systematic review were analyzed: any duplicity was deleted and only studies closely related to otolith shape topic were read entirely.

Results of the articles with the most relevant information to understand the use and methodologies in otolith shape analysis were summarized.

I.3. RESULT AND DISCUSSION

Preliminary research combined "OTOLITH" and "SHAPE" and resulted in 607 papers; 415 out of them were dealing with topics related to the shape of otoliths (FIG. 1). Subsequently, 127 papers emerged from the research carried out using "OTOLITH" and "SHAPE" and "STOCK". Regarding research with "OTOLITH" and "SHAPE" and "MERLUCCIUS" as topic words, 38 papers were found, but only 10 dealt with the research; as concerns "OTOLITH" and "SHAPE" and "MULLUS", 8 papers were identified and 5 were used.

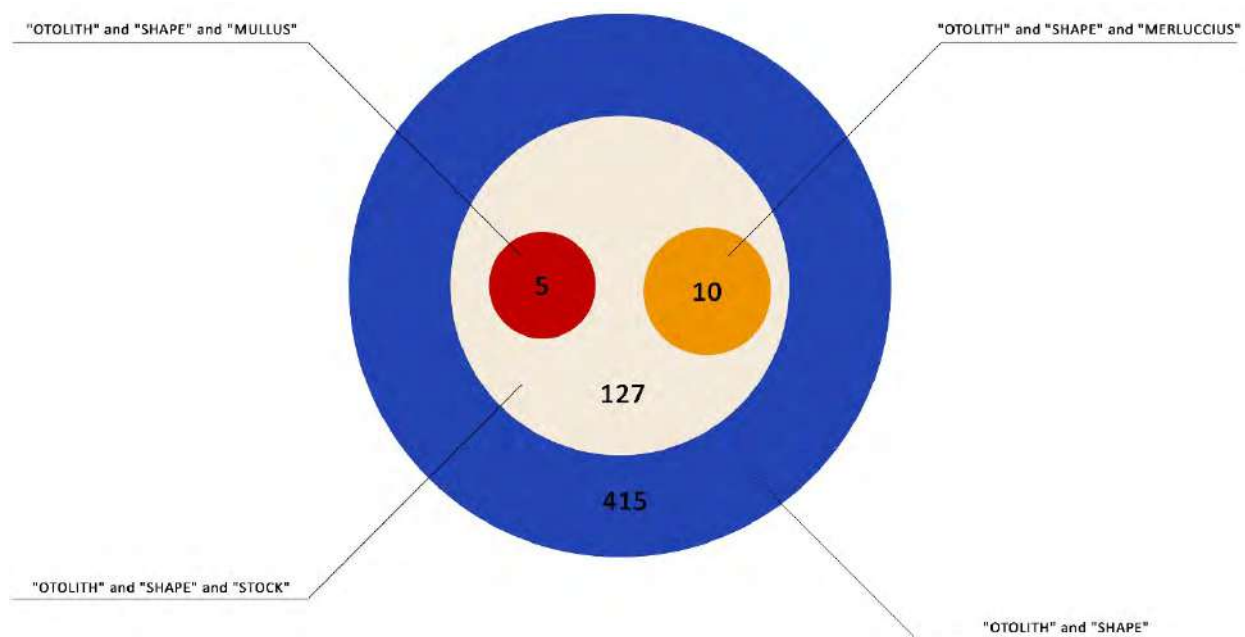


Figure 1 – General scheme of literature search used in the review (source Web of Science)..

General history

The first scientific articles related with otolith shape date back to the 1950s, when otolith characteristics were used to identify herring groups in the northwest Atlantic (Einarsson 1951; Ziklstra 1958; Parrish and Sharman 1958, 1959; Raitt 1961). Fifty years later, 22% of otolith-oriented papers were dealing with otolith shape (Campana,

2005). These studies were focusing on otolith allometry, species identification and tracer application (Campana, 2005). Studies on cod highlighted that environmental factors, such as feeding history (Gagliano *et al.*, 2004), and genetics play a fundamental role in otolith shape (Cardinale, *et al.*, 2004). Atlases of otoliths (Assis, 2003; Campana 2004, Tuset *et al.*, 2008) are useful tools in identification of species in stomach contents, i.e. prey-predator studies (Škeljo *et al.*, 2012), and to identify Lessepsian fish species (Tuset *et al.*, 2012). From 2006 onwards, AFORO website, <http://isis.cmima.csic.es/aforo/> (Lombarte *et al.*, 2006), offers an online catalogue of otolith images which allows to identify species or anatomical structures of species of different taxonomical groups (Parisi-Baradad *et al.*, 2010).

At present, publications involving otolith shape research reach over 600 publication, with a significant increase in the last 10 years (FIG. 2). Publications deal with different objectives, i.e. stock identification, methodologies or otolith description.

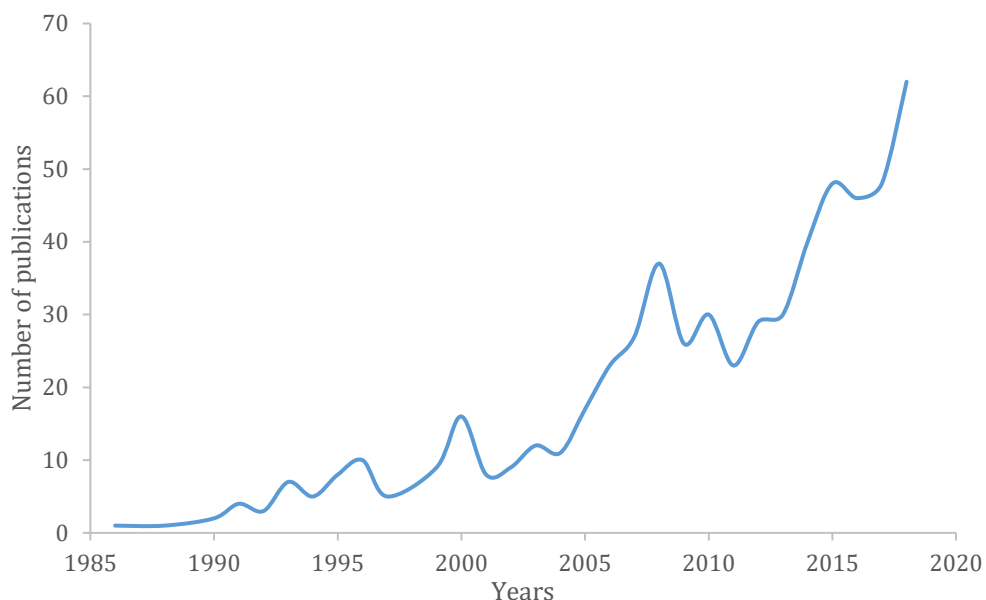


Figure 2 – Evolution of the number of publications related to “OTOLITH” and “SHAPE” topics (source Web of Science)

Shape analysis methodologies

Different approaches have been used to describe and compare the morphology of otoliths (Kuhl and Giardina 1982; Ponton, 2006; Nasreddine *et al.*, 2009). There are 2 mainly groups of extracted data from the otolith outline with univariate and multivariate data (FIG. 3 and 4).

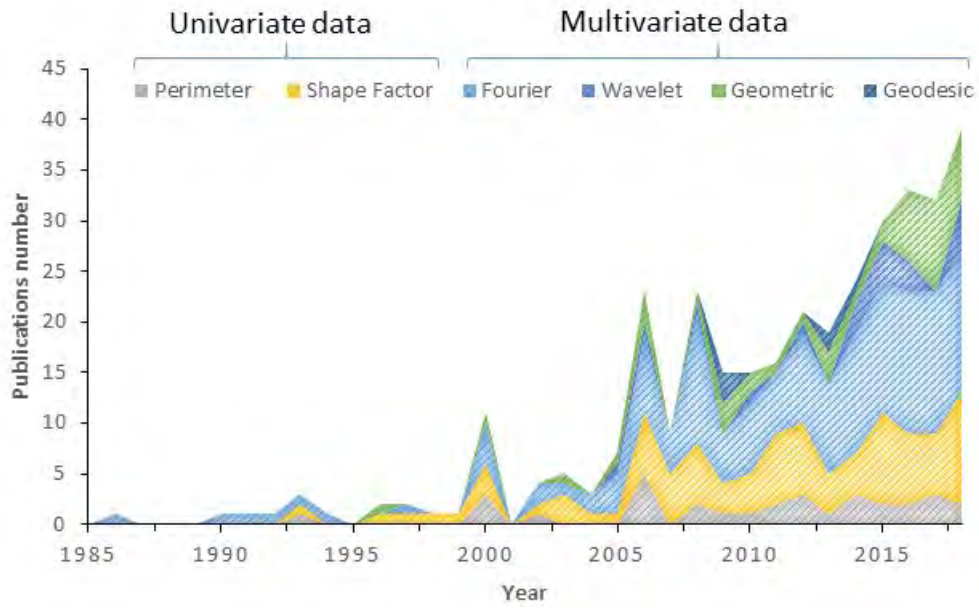


Figure 3 – Evolution of the number of publications with prior selected “OTOLITH” and “SHAPE”, by type of data extracted from the otolith outline (source Web of Science).

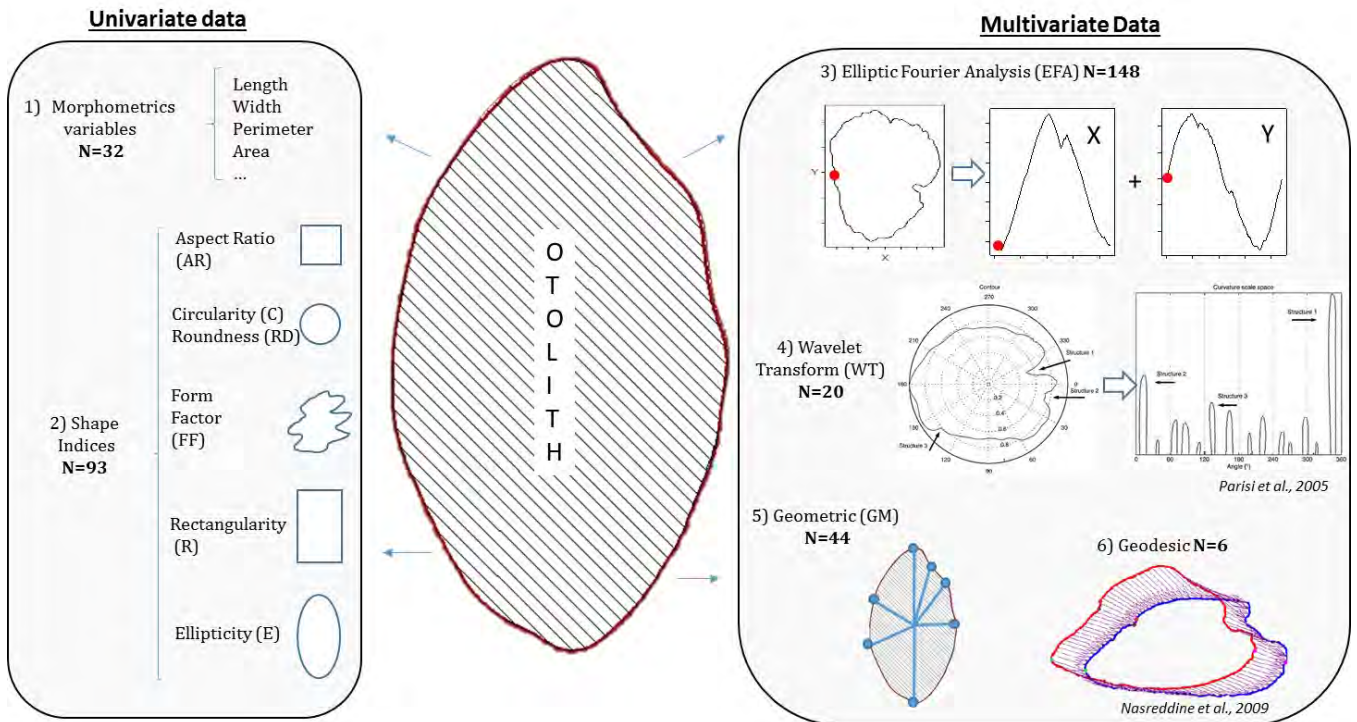


Figure 4 –Type of data extracted from the otolith outline with the number of publications with prior selected “OTOLITH” and “SHAPE” (source K. Mahé).

The comparison between shape otoliths is based on the linear measurement (univariate data) or on the multivariate data which described mathematical the whole otolith shape (Messieh 1972, Rojo 1977, King 1985, Payne 1985, Grygiel 1987, lapp 1990, Dawson 1991, Smith 1992;

Bird *et al.*, 1986, Castonguay *et al.*, 1991, Smith 1991, Campana and Casselman 1993)(Afanasyev *et al.*, 2017) (Yu *et al.*, 2014; Bookstein *et al.*, 1985; Rohlf and Bookstein, 1990; Rohlf and Marcus, 1993; Marcus *et al.*, 1996; Stransky, 2014). Among the multivariate approaches, there are mainly Fourier, wavelet transform and geomorphometric analysis. The others method are little developed as geodesic (Nasreddine *et al.*, 2009 ; Benzinou *et al.*, 2013), fractal dimension (Duarte-Neto *et al.*, 2008) and Curvature Scale Space (Mapp *et al.*, 2013). Multivariate approaches are more precise than the univariate data but the biological interpretation more complex than interpretation of the results obtained from linear morphometry (Stransky and MacLellan, 2005 ; Cadrin and Friedland, 1999). In fact, the multivariate and univariate data are often coupled in the study to increase the discrimination power and its biological interpretation (Galley *et al.*, 2006 ; Canas *et al.*, 2012).

ELLIPTIC FOURIER ANALYSIS (EFA)

At present, the most common outline methods involve Fourier analysis (Campana and Casselman, 1993). It represents the most popular method (Cardinale *et al.* 2004; Duarte-Neto *et al.* 2008; Torres *et al.*, 2000; Kristoffersen and Magoulas 2008; Mériqot *et al.*, 2007; Smith *et al.* 2002; Stransky *et al.* 2008) used in shape analysis and describes information about the outlines of enclosed shapes in a quantifiable manner (Kuhl and Giardina, 1982; Lestrel 1997; Tracey *et al.*, 2006). As reported by Cadrin and Friedland (1999), this technique is an efficient method for describing the variation in the shape contours and the individual differences in the scale of the otolith contour (Campana and Casselman, 1993).

Benzinou *et al.* (2013) showed how shape can be described using complex Fourier descriptors (Granlund, 1972) or using elliptic Fourier descriptors (Kuhl and Giardina, 1982): both techniques have been extensively used and proved to be efficient (Cardinale *et al.*, 2004; Duarte-Neto *et al.*, 2008; Galley *et al.*, 2006; Kristoffersen and Magoulas 2008; Mériqot *et al.*, 2007; Schulz-Mirbach *et al.*, 2008; Smith *et al.*, 2002; Stransky *et al.*, 2008; Torres *et al.*, 2000; Robertson and Talman, 2002). Although these descriptors represent a rapid, objective and semi-automated means to obtain information on fish stock, age distributions and environmental history (Bird *et al.*, 1986), the ordinary Fourier coefficients are difficult to use (Granlund, 1972).

Elliptic Fourier Analysis describes the shape through the use of specific components, called harmonics, formed by four coefficients, or elliptical Fourier Descriptor (EEFAs) resulting from the linearization of each outline point on the x and y axes. The amplitude of each harmonic represents the contribution to the empirical shape of an object (Bird *et al.*, 1986) and the accuracy of the outline descriptor is directly proportional to the number of harmonics (Kuhl and Giardina, 1982).

Although EFA represents the most widely used method of discriminating stocks, there are different kinds of this model based on different mathematical approaches. An example is represented by the Fast Fourier Transform (FFT) that works on the tangent angle as a function of arc-length connecting the coordinates (Zahn *et al.*, 1972). As reported by several authors,

EFA descriptors are often considered as more powerful than those derived from FFT for shape analysis (Gonzalez-Salas and Lenfant 2007; Lord *et al.*, 2012; Mériqot, *et al.*, 2007; Stransky, *et al.*, 2005; Tracey *et al.*, 2006), due to their relatively high efficiency to describe outline information (Stransky and MacLellan 2005; Tracey *et al.*, 2006; Yu *et al.*, 2014). In some cases, EFA are more appropriate than complex Fourier descriptor when otolith contours are composed of series of ellipse arcs, as reported by Benzinou *et al.* (2013) in their study on *Mullus surmuletus*.

GEOMETRIC METHOD (GM)

The geometric method is based on comparison of homologous points named landmarks to describe the shape. However, these landmarks are very difficult to identify on the otolith outline and consequently this method for the otolith used the semi-landmarks which are defined by grid (Pavlinov and Mikeshina, 2002 ; Ponton, 2006 ; Afanasyev *et al.*, 2017).

WAVELET TRANSFORM (WT)

Another contour method is represented by the Wavelet transform (WT): it provides a powerful alternative to the more commonly known Fourier method (Libungan *et al.*, 2015; Parisi-Baradad *et al.*, 2005). WT is a multiscale analysis of local points of a contour (Lombarte and Tuset, 2015) and is more accurate when more detailed information about shape is required, i.e. which area of the otolith shape is most involved in the separation between populations (Libungan *et al.*, 2015).

SHAPE INDICES

The shape indices represent the pattern of otolith growth in a bidimensional plane (Tuset *et al.* 2003). Measurements taken on otoliths have the advantage of being unaffected by short-term changes in fish condition or by preservation, as long as acidic preservatives are avoided (Campana and Casselman, 1993).

The most widely used size-dependent variables are anteroposterior length (OL), dorso-ventral width (OW), otolith perimeter (OP), sulcus perimeter (SP), otolith surface (OA), sulcus surface (SS) and distances from various points between the margin and center (Ponton, 2006). From these linear measurements, shape indices (TAB. 1) have been used efficiently to identify stock from different regions (Castonguay *et al.*, 1991; Torres *et al.*, 2000).

TABLE 1 – Shape indices applied on the otolith shape with formula.

SHAPE INDICES	FORMULA
Form-Factor (FF)	$(4\pi OA) / OP^2$
Aspect Ratio (AR)	OL / OW
Roundness (RD)	$(4OA) / (\pi OL^2)$
Circularity (C)	OP^2 / OA
Ellipticity (E)	$(OL-OW) / (OL+OW)$
Rectangularity (R)	$OA / (OL*OW)$

Russo (1990) defined geometrically the shape indices as follow:

- Roundness (RD) and Circularity (C) provide indication on the similarity to a perfect circle;
- Form Factor (FF) is a mean to estimate the irregularity of a surface area;
- Rectangularity (R) describes the variations of length and width with respect to the area;
- Ellipticity (E) indicates the proportional change of the short and long axes;
- Aspect Ratio (AR) is the result of the division of otolith length by otolith width.

In comparison with multivariate approaches, this approach have the advantage of being easier to calculate and applicate (Tuset *et al.*, 2003). Messieh (1972), Smith (1992) and Tuset *et al.*, (2003) indicated shape indices as a natural tags useful for identifying and differentiating species such as herring, red snapper (*Etelius carbunculus*) and *Serranus* spp.

Although shape indices are not the best option to analyze otolith shape differences (Lombarte and Tuset, 2015) and size descriptor can provide some understanding of the morphological differences among stocks (Cadrin and Friedland, 1999), some authors confirmed that the species identification and stock differentiation can be made by means of the analysis of meristic characteristic of the otoliths (Gaga 1993; Tuset *et al.*, 2003).

Variation of otolith shape according to potential effects

Many studies were carried out using shape analysis approach with different purpose.

Studies involving shape analysis as tool to discriminate fish stock were 30.6% (TAB. 2). 67.6% of the reviewed studies were performed applying a Fourier descriptor to analyze the otolith shape, and in 34.2% Fourier analysis was the only analysis used. Papers using only one descriptor were 42.3% (Wavelet transform=4.5%; OM/OSI=3.6%). Several papers were carried out using Fourier and Wavelet (17.2%) and Fourier and OM or OSI (16.2%), while 4.5% combined Wavelet and OM or OSI methods. Finally, 11.7% applied more than two methods.

Regarding investigated areas, 9.4% of papers were conducted considering only the Mediterranean Sea, (Arechavala-Lopez *et al.*, 2012; Barhoumi *et al.*, 2017; Bourehail *et al.*, 2015; Callicó Fortunato *et al.*, 2016; Cresson *et al.*, 2015; Ider *et al.*, 2017; Khemiri *et al.*, 2018; Marengo *et al.*, 2017; Mérigot *et al.*, 2007; Montanini *et al.* 2017; Trojette *et al.*, 2014, Zorica *et al.*, 2010) while 8.8% compared Mediterranean species with other areas (Agüera and Brophy 2011; Bacha *et al.*, 2016; Jemaa *et al.*, 2015; Lombarte and Castellón 1991; Lombarte and Lleonart 1993; Lombarte and Morales-Nin 2000; Stransky *et al.*, 2008; Torres *et al.*, 2000; Tuset *et al.*, 2003; Mahé *et al.*, 2018).

Most of the investigated species were marine species, while freshwater species have been studied in few cases (Avigliano *et al.*, 2014, 2015; Lord *et al.*, 2012; Radhakrishnan *et al.*, 2012; Schulz-Mirbach and Plath 2012; Watkinson and Gillis, 2014; Yedier *et al.*, 2016).

Concerning structure, otoliths represented the most used structure. Schulz-Mirbach and Plath (2012) compared contour outlines of all three otolith pairs by applying Fourier analysis and they demonstrated, on a quantitative basis, that shape information not only of *sagittae*, but also of *lapilli* and especially *asterisci* significantly contributes to species discrimination. This hypothesis was confirmed by Yedier *et al.* (2016) using *asterisci*, *lapilli* and scale shape to identify *Garra rufa* species. Statoliths were used to distinguish three cubozoan species in Australian waters (Mooney and Kingsford, 2016b).

Watkinson and Gillis (2005, 2015) used scales in their studies. Comparing otolith and scale shape to identify different stock of *Mugil curema* in Gulf of Mexico, Ibáñez *et al.* (2017) suggested that scale shape offers a straightforward, non-destructive, accessible, quick and inexpensive method to trace fish phenotypic stocks. Arechavala-Lopez *et al.* (2012) distinguished farmed European sea bass (*Dicentrarchus labrax*) from wild stocks using scales and otoliths shape. The same approach was used by Katayama and Isshiki (2007) to discriminate between wild and released Japanese flounder (*Paralichthys olivaceus*) individuals and attributing the morphological differences to biotic and abiotic factor between wild and hatchery environment. Experiment to test the influence of biotic and abiotic factors, such as food availability, on otoliths shape was carried out by Gagliano and McCormick (2004) and Hüsey (2008). Environmental factors are major determinants of otolith growth (Cardinale *et al.*, 2004). The interaction between biotic factors, i.e temperature and food availability, and otolith shape is related to the effects of these factors on growth rate (Campana and Casselman, 1993; Gallego *et al.*, 1996; Hüsey, 2008). Variation in growth rate produces corresponding variation in otolith microstructure and shape (Gauldie and Nelson, 1990), due to the proportional relationship between otolith growth and somatic growth (Campana and Neilson, 1985; Campana and Casselman, 1993; Begg and Brown, 2000; Simoneau *et al.*, 2000; Cardinale *et al.*, 2004). In Cardinale *et al.* (2004), individuals belonging to the same stock but growing in different temperature and feeding regimes showed otoliths with different morphometric characters. Otolith shape differences appears to coincide with geographical differences in temperature (Bolles and Begg, 2000), water depth (Lombarte and Leonart, 1993), salinity (Capoccioni *et al.*, 2011) and substrate type (Mérigot *et al.*, 2007). From this assumption, measurements of growth increments (lengths from nucleus to annual rings) and shape analysis of growth rings can be a powerful method for stock discrimination (Burke *et al.*, 2008a, 2008b; Torres *et al.*, 1996; Zhang *et al.*, 2014).

Genetic heterogeneity is decisive in the otolith shape (Campana and Casselman, 1993; Cadrin and Friedland, 1999; Cardinale *et al.*, 2004; Vignon and Morat, 2010): it regulates otolith shape at interspecific level, while the environmental factors affect otolith shape at intraspecific level (Vignon and Morat, 2010; Vignon, 2012). For this reasons, genetic and environmental influences should be taken into account in otolith shape analysis. In some studies, otolith shape differences were observed between groups of fish living in the same geographic area (Galley *et al.*, 2006; Pothin *et al.*, 2006) and for stocks that could not be separated using genetic techniques (DeVries *et al.*, 2002).

As reported by many authors, a number of confounding effects must be considered to ensure that classification is based on shape alone (Burke *et al.*, 2008a). Individual parameters such as sex, age (Castonguay *et al.*, 1991), size (Hüssy, 2008), and sexual maturity (Mérigot *et al.*, 2007) may influence otolith shape (Cardinale *et al.*, 2004), leading to misinterpretation of spatial variation in shape parameters (Burke *et al.*, 2008a). The recent study identified the head side as other potential effect which influence the misinterpretation of the stocks structure (Mahé *et al.*, 2018).

In studies where sex effect on otolith shape have been tested, no significant differences in Atlantic mackerel *Scomber scombrus* (Castonguay *et al.*, 1991), Atlantic cod *Gadus morhua* (Cardinale *et al.*, 2004), blue whiting *Micromesistius poutassou* (Mahé *et al.*, 2013) and Atlantic herring *Clupea harengus* (Bird *et al.*, 1986; Mille *et al.* 2015) were found. In silver hake *Merluccius bilinearis* (Bolles and Begg 2000), *Porichthys notatus* (Bose *et al.*, 2017) and the forkbeard *Phycis phycis* (Vieira *et al.*, 2014), sex effect was significant: in these cases other factors should be taken into account (Campana and Casselman, 1993).

Age and size are a potential source of variability in morphometric measures because are linked to individual ontogeny (Lleonart *et al.*, 2000). Differences in shape analysis due to the increased size of otoliths in large individuals (Castonguay *et al.*, 1991, Campana and Casselman, 1993; Begg and Brown, 2000) can be avoided using a sampling strategy focusing on a specific year-class or length (Vasconcelos *et al.*, 2018) or with defined numbers of sampled individuals for each year-class (Paul *et al.*, 2013). Begg and Brown (2000) also suggested that for stock discrimination from otolith shape variables should be recalculated each year for each age group. Furthermore, some statistical procedures have been proposed, i.e. linear model or allometric growth model (Lleonart *et al.*, 2010), in order to eliminate size and age effects.

Sexual maturity can change otolith shape (Cardinale *et al.*, 2004). During maturation, metabolism shifts in favor of reproduction with respect to growth, thus affecting the otolith morphology (Härkönen, 1986). To minimize the effect of sexual maturity, only mature individuals (Mahé *et al.*, 2016 Mahé *et al.*, 2013) or immature of indeterminate sex (Burke *et al.*, 2008) should be taken into account in the sampling.

Recent reviews point out that a multidisciplinary approach is useful for understanding the structure of a population (Cadrin *et al.*, 2013). In this way, several studies have been carried out using shape analysis along other analysis.

Smith *et al.* (2002) applied multiple techniques such as life history traits (age, spawning time and L_{50}), length frequency distributions (LEFA), shape analysis and genetic for determining stock relationships in *Hoplostethus atlanticus* in Tasman Sea; Marengo *et al.* (2017) combined otolith shape, microsatellite and parasite community to assess population structure of *Dentex dentex*. Parasite community were used also by Vignon *et al.* (2008). Through shape and genetic analysis, Stransky *et al.*, (2008) studied differences in costal and offshore Norwegian cod and highlighted that differences in otolith morphology cannot directly be linked to genetic structure, while environmental condition (temperature) plays a fundamental role in otolith shape development. Villegas-Hernández *et al.* (2014) obtained a high level of discrimination in *Haemulon plumieri* with shape and genetic analysis, suggesting these analyses can be feasible

tool to evaluate structuration of this species at a regional level. In contrast, some studies have demonstrated that genetic approaches are less efficient for stock identification at smaller scales when relatively low levels of exchange between stocks occur (Jemaa *et al.*, 2015).

Interdisciplinary approach was carried out in *Gadus morhua* in U.S. waters combining genetic, tagging, life histories parameters, larval dispersal, chemistry otolith and body shape analysis (Zemeckis *et al.*, 2018), *Sebastes mentella* and *S. norvegicus* using morphological characteristics, genetic and otolith shape analysis (Christensen *et al.*, 2018), and *Scomberomorus* spp.. studying variation in DNA and otolith morphometric (Ramírez-Pérez *et al.*, 2015).

In all cases, approach based on otolith shape and/or morphometric represents the cheapest methodology, and more efficient than several others (e.g. genetics, parasites, isotope and micro-chemical discrimination techniques) (Neves *et al.*, 2011).

Case study of *Merluccius* spp.

Alados *et al.* (1993) carried out a study to quantify Developmental Instability (DI) using asymmetry of otolith length, width, growth rate, and weight, as well as right-left otolith shape differences to test the influence of El Niño (1982-1983) as a possible cause of dislocations of North Pacific hake stock (*Merluccius productus*) on the summer fishing grounds (Francis and Hollowed, 1984).

Bolles and Begg (2000) used otolith morphometric parameters (length, width, area and perimeter), and two shape indices (circularity and rectangularity) to confirm that two stocks of silver hake (*Merluccius bilinearis*) exist in the Northwest Atlantic.

Differences between silver hake stocks are probably the result of both environmental and genetic influences. They found a correlation between growth rate and morphometric: northern stock grows at a slower rate and shows longer, wider and greater otolith than the southern stock, that grows at a faster rate. This correlation was already observed by several authors (Templeman and Squires, 1956; Reznick *et al.*, 1989; Fowler and Short, 1996). Finally, they underlined how otolith morphometric, in combination with an image analysis system, represents a relatively inexpensive, objective, and efficient tool to distinguish fish stocks (Messieh *et al.*, 1989).

Lombarte and Castellón (1991) compared morphological variation in the sagittal otoliths of six species of the genus *Merluccius*: *M. merluccius*, *M. capensis*, *M. paradoxus*, *M. bilinearis*, *M. productus*, and *M. gayi*. Two geographic and phylogenetic groups based on the otoliths from individuals longer than 20 cm were identified: an Euro-African group (*M. capensis*, *M. merluccius*, and *M. paradoxus*), which exhibited greater development of the transverse axis than of the longitudinal axis of the *sagitta*; an American group (*M. bilinearis*, *M. gayi*, and *M. products*), where otolith were longer along the antero-posterior axis in relation to their width. In order to eliminate variation in shape due to the growth of the *sagittae*, shape analysis was carried out using four size classes of fish: less than 20 cm total length; 21-40 cm; 41-60 cm; and greater than 60 cm. Differences in otolith shape in the two *Merluccius* groups thus increased with size

of fish: in particular, the otoliths from small individuals of various species exhibit plesiomorphic characters and are quite similar to one another.

Differences in *Merluccius* spp. were investigated also by Torres *et al.* (2000), analyzing the morphology of the *sulcus acusticus* of the *sagitta*.

As reported in the paper, the *sulcus acusticus* has a species-specific character and it can be a good taxonomic tool (Schwarzahns, 1980; Gaemers, 1984; Nolf, 1985); in fact also in this case it was possible to divide *Merluccius* spp. into an American group (*M. gayi* and *M. hubbsi*) and an Euro-African group (*M. capensis*, *M. merluccius* and *M. senegalensis*)

Another study on the morphology (size and shape) of the saccular otolith (*sagitta*) was carried out on *Merluccius gayi*, *M. hubbsi* and *M. merluccius* (Lombarte and Morales-Nin, 2000). Morphometry and outline analysis conducted on otolith and *sulcus acusticus* showed spatial differences of predicted groups confirming shape analysis as a very appropriate method for determining the origin of the otoliths. Moreover for *M. merluccius*, a gradient in the positioning of otoliths was observed between samples from NE Atlantic and Mediterranean Sea. PCA results exhibited high correlation between all morphometric variables and first principal component: these results indicated that the differences between the two groups were related to the greater relative size of the morphometric measurements of the otoliths from the Mediterranean. Differences in otolith size can be explained by oceanographic differences between NE Atlantic and Mediterranean areas (Casey and Pereiro, 1995; Morales-Nin *et al.*, 1998b). Genetics differences between the NE Atlantic and the Mediterranean populations were found by Roldan *et al.* (1998).

Lombarte and Leonart (1993) carried out a study to investigate whether there is an environmental effect on otolith growth related to depth using carbonate concentration, and understand which factor (genetic vs environmental) acts on otolith morphology. *Merluccius* spp. was selected to investigate relation between otolith size and environmental factors, while *Coelorhynchus* spp. to evaluate the relationship between depth and otolith size. For each otolith, maximum length, maximum height, areas of the *sagitta* and the *sulcus acusticus*, perimeters of the *sagitta* and *sulcus acusticus* and *sagitta* weight were measured. They demonstrated clearly how the temperature affects the otolith size. In this way, *M. merluccius* from NW Mediterranean Sea has largest relative size, living in the warmest waters (13 to 14°C) (Salat and Font 1987), *M. guyi*, *M. bilinearis* and *M. capensis* species with intermediate otolith relative sizes live in waters with temperatures of 7°C-12°C, and *M. paradoxus* and *M. productus*, which live in the coldest waters (4°-10°C) (Botha, 1971; Inada, 1981), have the smallest otoliths. Finally, they concluded that otolith development occurs under dual regulation: genetic conditions regulate the form of the otolith, while environmental conditions, mainly temperature in carbonate-saturated waters, regulate the quantity of material deposited during the formation of the otolith.

In Vaz-Dos-Santos *et al.* (2017), morphometric measurements (length, height, thickness, length and height of the *sulcus acusticus*, length of the posterior region, area and perimeter) and otolith

shape index (ellipticity, rectangularity, form factor, roundness and circularity) were used to describe *M. hubbsi* population from four areas in SW Atlantic Ocean.

Variation in the *M. hubbsi* and *M. australis* population structure in SW Atlantic was investigated by Pierce *et al.* (2002) through multivariate analysis of external and skeletal morphometric data, counts of fin rays and teeth, and measurements on scales. Two groups of *M. hubbsi* within the study area and differences between the two species were recognized.

In order to discriminate Argentine hake (*M. hubbsi*) populations, Torres *et al.* (1996) proposed another approach: they analyzed measures of pelagic, demersal and first years rings of otolith belonging to two length group (TL 25 cm and TL 40 cm) identifying the presence of different stocks within the area of study.

The only study on *M. merluccius* in Mediterranean Sea was conducted by Cresson *et al.* (2015): they studied European hake population inhabiting French Mediterranean waters, combining a multiple approach analysis (chemical and shape) in order to discriminate between fish populations at large and small spatial scale. Isotope ratios, contamination levels and otolith shape analysis allowed to identify two groups of European hake in the Gulf of Lions and Corsica.

Case study of *Mullus* spp.

Morat *et al.* (2012) were the only one using shape analysis to discriminate *Mullus barbatus* in the Mediterranean Sea. In particular, they described population of *M. barbatus barbatus*, *M. barbatus ponticus* and *M. surmuletus* from NW Mediterranean Sea, Black Sea and Aegean Sea. The shape of each otolith was assessed with the elliptic Fourier analysis and Canonical Discriminant Analysis was performed. With respect to large spatial scale, they identified *M. barbatus* population belonging to the three different investigated areas. Over regional scale, discrimination between individuals living in deep and shallow sites was recognized. Finally, they highlighted the importance of using both otoliths in the analysis because left and right otoliths contain complementary information useful to identify populations.

In Soria *et al.* (2014), otolith shape of *M. barbatus* was used to test fish stock identification, using geodesic approach and Density-based Library Local Discriminant Bases method, DLLBD (Soria and Parisi-Baradad, 2011), in Bay of Biscay, Celtic Sea/Western English Channel and Eastern English Channel/North Sea. They demonstrated that both methods are useful to differentiate *M. barbatus* populations and different living conditions, and that different environments can affect otolith shape.

Benzinou *et al.* (2013) investigated stocks of striped red mullet using three automatic methods of stock identification based on otolith shape and growth marks. In their study, they compared method based on Fourier descriptor, PCA and a recently proposed method based on shape geodesics (Nasreddine *et al.*, 2009). Population of striped red mullet in Northwest European seas can be divided in three geographical zones: the Bay of Biscay, a mixing zone composed of the Celtic Sea and the Western English Channel, and a northern zone composed of the Eastern English Channel and the North Sea.

Ecomorphological comparisons of *sagittae* in *M. barbatus* and *M. surmuletus* in the North-Western Mediterranean Sea was carried out by Aguirre and Lombarte (1999). Relationships between several morphological otolith parameters were analyzed in order to determine correlation with fish growth. For *M. barbatus*, proportional relation between otolith area and *sulcus acusticus* area was found; moreover, decrement in the size of *sagitta* related to body growth seems to not affect the *sulcus acusticus* area.

Jaramillo *et al.* (2014) worked on *M. surmuletus* otolith to understand how different species interacting with substrate. Morphology and morphometry were expressed in terms of E (maximum width of the otolith (OW) / maximum length of the otolith (OL) %) and S (*sulcus* area (SS)/otolith area (OS) %). Results show eco-morphological differences related with habitat use and the type of substrate where the fish inhabit.

TABLE 2 - Studies involving shape analysis as tool to discriminate fish stock.

Structure: O=Otolith; S=Scale; St=Statolith

Method: EFA Elliptic Fourier Analysis; WT= Wavelet transform; OM: Otolith Morphometric; OSI=Otolith Shape Indices; MoG: measurement of growth

Autor	Species	Area	Structure	Method
Afanasyev <i>et al.</i> , 2017	<i>Sebastes</i> spp.	Russia	O	EFA
Agüera and Brophy, 2011	<i>Scorpaenopsis scorpaenoides</i>	NE Atlantic Ocean / Mediterranean Sea	O	EFA
Alados <i>et al.</i> , 1993	<i>Merluccius productus</i>	NW Atlantic Ocean	O	OM
Arechavala-Lopez <i>et al.</i> , 2012	<i>Sparus aurata</i> , <i>Dicentrarchus labrax</i>	Mediterranean	S - O	EFA - WT - OM
Avigliano <i>et al.</i> , 2014	<i>Odontesthes bonariensis</i>	Uruguay	O	WT - OM
Avigliano <i>et al.</i> , 2015	<i>Plagioscion ternetzi</i>	Paraguay	O	
Bacha <i>et al.</i> , 2016	<i>Sardinella aurita</i>	NW Africa	O	EFA - OM
Bacha <i>et al.</i> , 2016	<i>Engraulis encrasicolus</i>	SW Mediterranean Sea / Atlantic Ocean	O	EFA - WT - OM
Banaru <i>et al.</i> , 2017	<i>Neogobius melanostomus</i>	Black Sea	O	EFA
Barhoumi <i>et al.</i> , 2017	<i>Oblada melanura</i>	SW Mediterranean Sea	O	EFA
Begg and Brown 2000	<i>Melanogrammus aeglefinus</i>	NW Atlantic Ocean	O	EFA - WT - OM
Benzinou <i>et al.</i> , 2013	<i>Mullus surmuletus</i>	NW European Seas	O	EFA
Bergenius <i>et al.</i> , 2002	<i>Plectropomus leopardus</i>	Australia	O	EFA - WT - OM
Bird <i>et al.</i> , 1986	<i>Cleupea</i> spp.	NW Atlantic Ocean	O	EFA
Bolles and Begg 2000	<i>Merluccius bilinearis</i>	NW Atlantic Ocean	O	OM - OSI
Bose <i>et al.</i> , 2017	<i>Porichthys notatus</i>	NW Atlantic Ocean	O	EFA - WT - OM
Bostanci and Yedier, 2018	<i>Atherina boyeri</i>	Turkey	O	OM
Bourehail <i>et al.</i> , 2015	<i>Sphyræna sphyraena</i>	SW Mediterranean Sea	O	EFA - OM
Burke <i>et al.</i> , 2008a	<i>Clupea harengus</i>	Celtic Sea	O	EFA - OM - OSI
(Burke <i>et al.</i> , 2008b)	<i>Clupea harengus</i>	Irish Sea	O	EFA - WT - OM
Callicó Fortunato <i>et al.</i> , 2017	<i>Mugil cephalus</i>	NW Mediterranean Sea	O	OM
Callicó Fortunato <i>et al.</i> , 2016	<i>Mugil liza</i>	SW Atlantic Ocean	O	WT - OM
Cardinale <i>et al.</i> , 2004	<i>Gadus morhua</i>	NE Atlantic (Faroe Islands)	O	EFA - WT
Castronguay <i>et al.</i> , 1991	<i>Scorpaenopsis scorpaenoides</i>	NW Atlantic	O	EFA
Christensen <i>et al.</i> , 2018	<i>Sebastes mentella</i> , <i>Sebastes norvegicus</i>	Greenland	O	OSI - MoG
Clardy <i>et al.</i> , 2008	<i>Scorpaenopsis scorpaenoides</i>	South Florida	O	EFA - WT
Cresson <i>et al.</i> , 2015	<i>Merluccius merluccius</i>	French Mediterranean	O	EFA
da Silva Santos <i>et al.</i> , 2017	<i>Micropogonias furnieri</i>	SW Atlantic Ocean	O	EFA
de Carvalho <i>et al.</i> , 2015	<i>Anchoa tricolor</i>	Brazil	O	
DeVries, Grimes <i>et al.</i> , 2002	<i>Scorpaenopsis scorpaenoides</i>	Gulf of Mexico	O	EFA - WT
Doering-Arjes <i>et al.</i> , 2008	<i>Gadus morhua</i>	Faroe Bank	O	EFA - WT
Doustdar <i>et al.</i> , 2018	<i>Acanthopagrus arabicus</i>	Northern Persian Gulf and Oman Sea	O	OM

Duarte-Neto <i>et al.</i> , 2008	<i>Coryphaena hippurus</i>	NE Brazil	0	EFA - WT - OM
Duncan <i>et al.</i> , 2018	<i>Thunnus alalunga</i>	NE Atlantic Ocean	0	EFA
Eggers <i>et al.</i> , 2014	<i>Clupea harengus</i>	Norway	0	MoG
Farias <i>et al.</i> , 2009	<i>Aphanopus carbo</i>	Portugal coast	0	EFA
Félix-Uraga <i>et al.</i> , 2005	<i>Sardinops sagax</i>	California	0	WT
Ferguson <i>et al.</i> , 2011	<i>Argyrosomus japonicus</i>	Australia	0	EFA - OM
Gagliano and McCormick, 2004	<i>Amphiprion akindynos</i> , <i>Pomacentrus amboinensis</i>	Australia	0	EFA
Galley <i>et al.</i> , 2006	<i>Gadus morhua</i>	North Sea	0	EFA - OM
Gonzalez-Salas and Lenfant, 2007	<i>Engraulis encrasicolus</i>	Bay of Biscay	0	EFA
Hamer <i>et al.</i> , 2012	<i>Macruronus novaezelandiae</i>	Australia	0	EFA
He <i>et al.</i> , 2018	<i>Scomber</i> spp.	South China Sea, Norway and Japan Sagami Bay	0	EFA - WT - OM
Higgins <i>et al.</i> , 2010	<i>Gadus morhua</i>	NE Atlantic Ocean	0	WT
Hüssy <i>et al.</i> , 2015	<i>Gadus morhua</i>	Baltic Sea	0	EFA
Hüssy <i>et al.</i> , 2016	<i>Gadus morhua</i>	Baltic Sea	0	EFA
Ider <i>et al.</i> , 2017	<i>Boops boops</i>	SW Mediterranean Sea	0	EFA
Ibáñez <i>et al.</i> , 2017	<i>Mugil curema</i>	Gulf of Mexico		
Javor <i>et al.</i> , 2011	<i>Sardinops sagax</i>	N America	0	WT - OM
Jemaa <i>et al.</i> , 2015	<i>Engraulis encrasicolus</i>	NE Atlantic Ocean / Mediterranean Sea	0	EFA - WT
Jemaa <i>et al.</i> , 2015	<i>Sardina pilchardus</i>	Atlantic/Mediterranean Sea	0	EFA - OM
Katayama and Isshiki, 2007	<i>Paralichthys olivaceus</i>	Sagamy Bay	0	
Keating <i>et al.</i> , 2014	<i>Micromesistius poutassou</i>	NE Atlantic	0	EFA - WT - OM
Khemiri <i>et al.</i> , 2018	<i>Engraulis encrasicolus</i>	SW Mediterranean Sea	0	EFA
Leguá <i>et al.</i> , 2013	<i>Micromesistius australis</i>	SW Atlantic Ocean	0	EFA - WT - OM
Libungan <i>et al.</i> , 2015	<i>Clupea harengus</i>	N Atlantic Ocean	0	WT
Libungan <i>et al.</i> , 2015	<i>Clupea harengus</i>	Norway	0	WT
Libungan <i>et al.</i> , 2016	<i>Clupea pallasii</i>	Alaska, Norway, Russia	0	WT
Lombarte and Castellón, 1990	<i>Merluccius</i> spp.	Mediterranean Sea, NW Atlantic Ocean, SE Atlantic Ocean, E Pacific Ocean	0	OM
Lombarte and Leonart, 1993	<i>Merluccius</i> spp. and <i>Coelorhynchus</i> spp.	Mediterranean Sea, SE Atlantic Ocean, NE Pacific Ocean	0	OM
Lombarte and Morales-Nin, 2000	<i>M. gayi</i> , <i>M. hubbsi</i> and <i>M. merluccius</i>	Mediterranean Sea, NE Atlantic Ocean, E Pacific Ocean	0	EFA - OM - OSI
Longmore <i>et al.</i> , 2010	<i>Coryphaenoides rupestris</i>	N Atlantic Ocean	0	EFA - OM
Lord <i>et al.</i> , 2012	<i>Sicyopterus</i> spp.	New Caledonia	0	EFA
Lorenzo <i>et al.</i> , 2011	<i>Merluccius hubbsi</i>	SE Atlantic Ocean	0	WT
Mahé <i>et al.</i> , 2013	<i>Micromesistius poutassou</i>	NE Atlantic Ocean	0	FD - OM
Mahé <i>et al.</i> , 2016	<i>Xiphias gladius</i>	W Indian Ocean	0	FD - OM - OSI
Mahé <i>et al.</i> , 2018	<i>Boops boops</i>	Mediterranean Sea, NE Atlantic Ocean	0	FD - OM - OSI
Marengo <i>et al.</i> , 2017	<i>Dentex dentex</i>	Mediterranean Sea	0	EFA - OM
Mérigot <i>et al.</i> , 2007	<i>Solea solea</i>	NW Mediterranean Sea	0	EFA - OM
Midway <i>et al.</i> , 2014	<i>Paralichthys lethostigma</i>	Atlantic Ocean and Gulf of Mexico	0	EFA - OM
Montanini <i>et al.</i> , 2015	<i>Aspitrigla cuculus</i> , <i>Eutrigla gurnardus</i>	Adriatic Sea	0	WT
Mooney and Kingsford, 2016	<i>Cubomedusae</i>	Australia	0	EFA
Mooney and Kingsford, 2016b	<i>Cubomedusae</i>	Australia	St	EFA
Moreira <i>et al.</i> , 2019	<i>Trachurus picturatus</i>	NE Atlantic Ocean	0	OM
Neves <i>et al.</i> , 2011	<i>Helicolenus dactylopterus</i>	NE Atlantic Ocean	0	EFA
Paul <i>et al.</i> , 2013	<i>Gadus morhua</i>	Baltic Sea	0	EFA
Pavlov, 2018	<i>Upeneus tragula</i>	Vietnam	0	EFA - WT - OM
Pierce <i>et al.</i> , 2002	<i>Merluccius hubbsi</i> and <i>Merluccius australis</i>	SW Atlantic Ocean	0	OM
Pothin <i>et al.</i> , 2006	<i>Mulloidichthys flavolineatus</i>	SW Indian Ocean	0	EFA - OM
Radhakrishnan <i>et al.</i> , 2012	<i>Coilia ectenes</i>	China	0	EFA - WT - OM

Ramírez-Pérez <i>et al.</i> , 2015	<i>Scomberomorus</i> spp.	California	0	WT
Renán <i>et al.</i> , 2016	<i>Mycteroperca microlepis</i>	Mexico	0	EFA - WT - OM
Schulz-Mirbach and Plath, 2012	<i>Poecilia</i> spp.	San Antonio River (Texas)	0	EFA
Serrano Gordo <i>et al.</i> , 2009	<i>Aphanopus carbo</i>	NE Atlantic Ocean	0	EFA
Smith <i>et al.</i> , 2002	<i>Hoplostethus atlanticus</i>	Tasman Sea	0	OM
Stransky <i>et al.</i> , 2008	<i>Trachurus trachurus</i>	NE Atlantic Ocean / Mediterranean Sea	0	EFA
Stransky <i>et al.</i> , 2008	<i>Gadus morhua</i>	NE Artic / Norway	0	EFA - WT
Stransky, 2005	<i>Sebastes marinus</i> , <i>S. mentella</i>	N Atlantic Ocean	0	EFA
Torres <i>et al.</i> , 1996	<i>Merluccius hubbsi</i>	SW Atlantic Ocean	0	WT - MoG
Torres <i>et al.</i> , 2000	<i>Merluccius</i> spp.	Mediterranean Sea, NW Atlantic Ocean, SE Atlantic Ocean, E Pacific Ocean	0	EFA - OM - OSI
Treinen-Crespo <i>et al.</i> , 2012	<i>Haemulon plumieri</i>	Yucatan Peninsula	0	EFA
Trojette <i>et al.</i> , 2014	<i>Scorpaena porcus</i>	SW Mediterranean Sea	0	EFA
Tuset <i>et al.</i> , 2003	<i>Serranus cabrilla</i>	Mediterranean Sea / Atlantic Ocean	0	OM
Tuset <i>et al.</i> , 2014	<i>Sebastes</i> spp.	NE Pacific Ocean	0	MoG
Vasconcelos <i>et al.</i> , 2018	<i>Trachurus picturatus</i>	NE Atlantic Ocean	0	EFA - WT
Vaz-Dos-Santos <i>et al.</i> , 2017	<i>Merluccius hubbsi</i>	SW Atlantic Ocean	0	OM - OSI
Vieira <i>et al.</i> , 2014	<i>Phycis phycis</i>	NE Atlantic Ocean	0	EFA
Vignon <i>et al.</i> , 2008	<i>Lutjanus kasmira</i>	French Polynesia	0	EFA
Villegas-Hernández <i>et al.</i> , 2014	<i>Haemulon plumieri</i>	Yucatan	0	EFA
Watkinson and Gillis, 2005	<i>Sander vitreus</i>	Lake Winnipeg	S	EFA - OM
Watkinson and Gillis, 2014	<i>Stizostedion vitreum</i>	Manitoba lakes	S	EFA
Yedier <i>et al.</i> , 2016	<i>Garra rufa</i>	Turkey	S - 0	OM
Yu <i>et al.</i> , 2014	<i>Gobies</i> spp.	Chinese coastal waters	0	EFA
Zemeckis <i>et al.</i> , 2018	<i>Gadus morhua</i>	NW Atlantic Ocean	0	OM
Zengin <i>et al.</i> , 2015	<i>Engraulis encrasicolus</i>	Marmara Seas	0	OM
Zhang <i>et al.</i> , 2013	<i>Scomberomorus niphonius</i>	Yellow Sea and Bohai Sea	0	EFA - OM
Zhang <i>et al.</i> , 2014	<i>Larimichthys polyactis</i>	Chinese coast	0	EFA - WT - OM - OSI
Zhang <i>et al.</i> , 2017	<i>Pampus</i> spp.	Chinese coast	0	EFA - OM
Zhao <i>et al.</i> , 2017	<i>Larimichthys polyactis</i>	Chinese coast	0	EFA
Zhao <i>et al.</i> , 2017a	<i>Collichthys</i> spp.	Chinese coast	0	EFA
Zischke <i>et al.</i> , 2016	<i>Scomberomorus</i> spp.	Australia	0	WT - OM
Zorica <i>et al.</i> , 2010	<i>Engraulis encrasicolus</i> , <i>Sardina pilchardus</i> , <i>Scombrus scombrus</i> , <i>Scomber japonicus</i> , <i>Belone belone</i>	Adriatic Sea	0	WT - OM

I.4. REFERENCES

- Afanasyev, P. K., Orlov, A. M., and Rolsky, A. Y. (2017). Otolith Shape Analysis as a Tool for Species Identification and Studying the Population Structure of Different Fish Species. *Biology Bulletin*, 44(8), 952–959.
- Agüera, A., and Brophy, D. (2011). Use of saggital otolith shape analysis to discriminate Northeast Atlantic and Western Mediterranean stocks of Atlantic saury, *Scomberesox saurus saurus* (Walbaum). *Fisheries Research*, 110(3), 465–471.
- Aguirre, H., and Lombarte, A. (1999). Ecomorphological comparisons of sagittae in *Mullus barbatus* and *M. surmuletus*. *Journal of Fish Biology*, 55(1), 105–114.
- Alados, C. L., Escos, J., and Emlen, J. M. (1993). Developmental instability as an indicator of environmental stress in the Pacific hake (*Merluccius productus*). *Fishery Bulletin*, 91(4), 587–593.
- Arechavala-Lopez, P., Sanchez-Jerez, P., Bayle-Sempere, J. T., Sfakianakis, D. G., and Somarakis, S. (2012). Discriminating farmed gilthead sea bream *Sparus aurata* and European sea bass *Dicentrarchus labrax* from wild stocks through scales and otoliths. *Journal of Fish Biology*, 80(6), 2159–2175.
- Assis, C. A. (2003). The lagenar otoliths of teleosts: their morphology and its application in species

- identification, phylogeny and systematics. *Journal of Fish Biology*, 62(6), 1268-1295.
- Avigliano, E., Martinez Riaños, C. F., Volpedo, A. V., Martinez Riaños, C. F., and Volpedo, A. V. (2014). Combined use of otolith microchemistry and morphometry as indicators of the habitat of the silverside (*Odontesthes bonariensis*) in a freshwater-estuarine environment. *Fisheries Research*, 149, 55-60.
- Avigliano, E., Comte, G., Rosso, J. J., Mabrugaña, E., Della Rosa, P., Sanchez, S., ... Schenone, N. F. (2015). Identificación de stocks pesqueros de la corvina de río (*Plagioscion ternetzi*) de los ríos Paraguay y Paraná, mediante el análisis morfométrico de sus otolitos. *Latin America Journal of aquatic research*, 43(4), 718-725.
- Bacha, M., Jeyid, A. M., Jaafour, S., Yahyaoui, A., Diop, M., and Amara, R. (2016). Insights on stock structure of round sardinella *Sardinella aurita* off north-west Africa based on otolith shape analysis. *Journal of Fish Biology*, 89(4), 2153-2166.
- Banaru, D., Morat, F., and Creteanu, M. (2017). Otolith shape analysis of three gobiid species of the northwestern black sea and characterization of local populations of *Neogobius melanostomus*. *Cybium*, 41(4), 325-333.
- Barhoumi, M., Khoufi, W., Kalai, S., Ouerhani, A., Essayed, S., Zaier, G., ... Fehri-Bedoui, R. (2017). The use of Fourier analysis as a tool for *Oblada melanura* (Linnaeus, 1758) stock unit separation in the south central Mediterranean Sea. *Journal of the Marine Biological Association of the United Kingdom*, 98(07), 1725-1732.
- Begg, G. A., and Brown, R. (2000). Stock identification of haddock *Melanogrammus aeglefinus* on Georges Bank based on otolith shape analysis. *Transactions of the American Fisheries Society*, 129(4), 935-945.
- Begg, G. A., and Waldman, J. R. (1999). An holistic approach to fish stock identification. *Fisheries research*, 43 (1-3), 35-44.
- Benzinou, A., Carbini, S., Nasreddine, K., Elleboode, R., and Mahé, K. (2013). Discriminating stocks of striped red mullet (*Mullus surmuletus*) in the Northwest European seas using three automatic shape classification methods. *Fisheries Research*, 143, 153-160.
- Bergenius, M. . A. J., Begg, G. A., and Mapstone, B. D. (2002). The use of otolith morphology to indicate the stock structure of common coral trout on the Great Barrier Reef, Australia. *FISHERY BULLETIN*, 104(4), 498-511.
- Bird, J. L., Eppler, D. T., and Checkley, D. M. (1986). Comparisons of Herring Otoliths Using Fourier Series Shape Analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 43(6), 1228-1234.
- Bolles, K. L., and Begg, G. A. (2000). Distinction between silver hake (*Merluccius bilinearis*) stocks in U.S. waters of the northwest Atlantic based on whole otolith morphometrics. *Fishery Bulletin*, 98(3), 451-462.
- Bookstein, F. L., Chernoff, B., Elder, R., Humphries, J., Smith, G., and Strauss, R. (1985). Morphometrics in evolutionary biology. Special Publication 15. *Academy of Natural Sciences, Philadelphia*, 277.
- Bose, A. P. H., Adragna, J. B., and Balshine, S. (2017). Otolith morphology varies between populations, sexes and male alternative reproductive tactics in a vocal toadfish *Porichthys notatus*. *Journal of Fish Biology*, 90(1), 311-325.
- Bostanci, D., and Yedier, S. (2018). Discrimination of invasive fish *Atherina boyeri* (Pisces: Atherinidae) populations by evaluating the performance of otolith morphometrics in several lentic habitats. *Fresenius Environmental Bulletin*, 27(6), 4493-4501.
- Botha, L. (1970). Growth and otolith morphology of the Cape hakes *Merluccius capensis* (Cast.) and *M. paradoxus* (Franca) (*Doctoral dissertation, Stellenbosch: Stellenbosch University*).
- Bourehail, N., Morat, F., Lecomte-Finiger, R., and Kara, M. H. (2015). Using otolith shape analysis to distinguish barracudas *Sphyraena sphyraena* and *Sphyraena viridensis* from the Algerian coast *Sphyraena sphyraena* and *Sphyraena viridensis*. *Cybium*, 39(4), 271-278.
- Burke, N., Brophy, D., and King, P. A. (2008a). Otolith shape analysis: Its application for discriminating between stocks of Irish Sea and Celtic Sea herring (*Clupea harengus*) in the Irish Sea. *ICES Journal of Marine Science*, 65(9), 1670-1675.
- Burke, N., Brophy, D., and King, P. A. (2008b). Shape analysis of otolith annuli in Atlantic herring (*Clupea harengus*); a new method for tracking fish populations. *Fisheries Research*, 91(2-3), 133-143.
- Cadrin, S. X., and Friedland, K. D. (1999). The utility of image processing techniques for morphometric analysis and stock identification. *Fisheries Research*, 43(1-3), 129-139.

- Cadrin, S. X., Kerr, L. A., and Mariani, S. (Eds.). (2013). Stock identification methods: applications in fishery science. *Academic Press*.
- Cadrin S. X., Kerr, L. A., and Mariani, S. (2014). Stock Identification Methods: Applications in Fishery Science. 2nd edn, Elsevier Academic Press, Amsterdam. 566 pp.
- Callicó Fortunato, R., Benedito Durà, V., and Volpedo, A. (2016). Otolith morphometry and microchemistry as habitat markers for juvenile *Mugil cephalus* Linnaeus 1758 in nursery grounds in the Valencian community, Spain. *Journal of Applied Ichthyology*, 33(2), 163–167.
- Callicó Fortunato, R., Benedito Durà, V., González-Castro, M., and Volpedo, A. (2017). Morphological and morphometric changes of sagittae otoliths related to fish growth in three Mugilidae species. *Journal of Applied Ichthyology*, 33(6), 1137–1145.
- Callicó Fortunato, R., González-Castro, M., Reguera Galán, A., García Alonso, I., Kunert, C., Benedito Durà, V., ... García Alonso, I. (2017). Identification of potential fish stocks and lifetime movement patterns of *Mugil liza* Valenciennes 1836 in the Southwestern Atlantic Ocean. *Fisheries Research*, 193(July 2016), 164–172.
- Campana, S. E. (1999). Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecology Progress Series*, 188: 263–297.
- Campana, S. E. (2004). Photographic atlas of fish otoliths of the Northwest Atlantic Ocean Canadian special publication of fisheries and aquatic sciences No. 133. *NRC Research press*.
- Campana, S. E. (2005). Otolith science entering the 21st century. *Marine and Freshwater Research*, 56(5), 485. <https://doi.org/10.1071/mf04147>
- Campana, S. E., and Neilson, J. D. (1985). Microstructure of fish otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*, 42(5), 1014–1032.
- Campana, S. E., and Casselman, J. M. (1993). Stock Discrimination Using Otolith Shape Analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 50 (5), 1062–1083.
- Campana, S. E., and Thorrold, S.R. (2001). Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? *Can. J. Fish. Aquat. Sci.* 58(1): 30-38.
- Cañás, L., Stransky, C., Schlickeisen, J., Sampedro, M. P., Fariña, A. C. (2012) Use of the otolith shape analysis in stock identification of anglerfish (*Lophius piscatorius*) in the Northeast Atlantic. *ICES J Mar Sci J Cons* 69:250–256
- Capoccioni, F., Costa, C., Aguzzi, J., Menesatti, P., Lombarte, A., and Ciccotti, E. (2011). Ontogenetic and environmental effects on otolith shape variability in three Mediterranean European eel (*Anguilla anguilla*, L.) local stocks. *Journal of Experimental Marine Biology and Ecology*, 397(1), 1–7. <https://doi.org/10.1016/j.jembe.2010.11.011>
- Cardinale, M., Doering-Arjes, P., Kastowsky, M., and Mosegaard, H. (2004). Effects of sex, stock, and environment on the shape of known-age Atlantic cod (*Gadus morhua*) otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(2), 158–167.
- Casey, J. and Pereiro, J. (1995). European hake (*Merluccius merluccius*) in the north-east Atlantic. In *Hake: Fisheries, ecology and markets* (ed. J. Alheit and T.J. Pitcher), pp. 125-147. London: Chapman and Hall Book Series
- Casselman, J. M. (1987). Determination of age and growth. In *The Biology of Fish Growth*, pp. 209-242. Ed. by A. H. Weatherley and H. S. Gill. Academic Press, New York. 443 pp
- Casselman, J. M., Collins, J. J., Grossman, E. J., Ihssen, P. E., and Spangler, G. R. (1981). Lake whitefish (*Coregonus clupeaformis*) stocks of the Ontario waters of Lake Huron. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(12), 1772-1789.
- Castonguay, M., Simard, P., and Gagnon, P. (1991). Usefulness of Fourier Analysis of Otolith Shape for Atlantic Mackerel (*Scomber scombrus*) Stock Discrimination. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(2), 296–302.
- Christensen, H. T., Rigét, F., Backe, M. B., Saha, A., Johansen, T., and Hedeholm, R. B. (2018). Comparison of three methods for identification of redfish (*Sebastes mentella* and *S. norvegicus*) from the Greenland east coast. *Fisheries Research*, 201(May), 11–17.
- Clardy, T. R., Patterson, W. F., DeVries, D. A., and Palmer, C. (2008). Spatial and temporal variability in the relative contribution of king mackerel (*Scomberomorus cavalla*) stocks to winter mixed fisheries off South Florida. *Fishery Bulletin*, 106(2), 152–160.
- Cresson, P., Bouchoucha, M., Morat, F., Miralles, F., Chavanon, F., Loizeau, V., and Cossa, D. (2015). A multitracer approach to assess the spatial contamination pattern of hake (*Merluccius merluccius*)

- in the French Mediterranean. *Science of the Total Environment*, 532, 184–194.
- da Silva Santos, R., de Azevedo, M. C. C., de Albuquerque, C. Q., and Araújo, F. G. (2017). Different sagitta otolith morphotypes for the whitemouth croaker *Micropogonias furnieri* in the Southwestern Atlantic coast. *Fisheries Research*, 195(July), 222–229.
- de Carvalho, B. M., Vaz-dos-santos, A. M., Spach, H. L., and Volpedo, A. (2015). Ontogenetic development of the sagittal otolith of the anchovy, *Anchoa tricolor*, in a subtropical estuary. *Scientia Marina*, 79(4), 409–418.
- Dawson, W. A. (1991). Otolith measurement as a method of identifying factors affecting first-year growth and stock separation of mackerel (*Scomber scombrus* L.). *ICES Journal of Marine Science*, 47(3), 303–317.
- DeVries, D. A., Grimes, C. B., and Prager, M. H. (2002). Using otolith shape analysis to distinguish eastern Gulf of Mexico and Atlantic Ocean stocks of king mackerel. *Fisheries Research*, 57(1), 51–62.
- Díaz-Gil, C., Palmer, M., Catalán, I. A., Alós, J., Fuiman, L. A., García, E., del Mar Gil, M. Grau, A., Kang, A., Maneja, R. H., Mohan, J. A., Morro, B., Schaffler, J. J., Buttay, L., Inmaculada, R. B. Borja, T., and Morales-Nin, B. (2015). Otolith fluctuating asymmetry: a misconception of its biological relevance? *ICES Journal of Marine Science*, 72(7): 2079–2089.
- Doering-Arjes, P., Cardinale, M., and Mosegaard, H. (2008). Estimating population age structure using otolith morphometrics: a test with known-age Atlantic cod (*Gadus morhua*) individuals. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(11), 2342–2350.
- Duarte-Neto, P., Lessa, R., Stosic, B., and Morize, E. (2008). The use of sagittal otoliths in discriminating stocks of common dolphinfish (*Coryphaena hippurus*) off northeastern Brazil using multishape descriptors. *ICES Journal of Marine Science*, 65(7), 1144–1152.
- Duncan, R., Brophy, D., and Arrizabalaga, H. (2018). Otolith shape analysis as a tool for stock separation of albacore tuna feeding in the Northeast Atlantic. *Fisheries Research*, 200(December 2017), 68–74.
- Eggers, F., Slotte, A., Libungan, L. A., Johannessen, A., Kvamme, C., Moland, E., ... Nash, R. D. M. (2014). Seasonal dynamics of atlantic herring (*Clupea harengus* L.) populations spawning in the vicinity of marginal habitats. *PLoS ONE*, 9(11).
- Einarsson, H. (1951). Racial analyses of Icelandic herring by means of the otoliths. *Rapp. P.-v. Reun. Cons. Perm. Int. Explor. Mer* 128, 55e74.
- Farias, I., Vieira, A. R., Serrano Gordo, L., and Figueiredo, I. (2009). Otolith shape analysis as a tool for stock discrimination of the black scabbardfish, *Aphanopus carbo* Lowe, 1839 (Pisces: Trichiuridae), in Portuguese waters. *Scientia Marina*, 73(S2), 47–53.
- Félix-Uraga, R., Quiñonez-Velázquez, C., Hill, T. K., Gomez-Munoz, V. M., Melo-Barrera, F. N., and Garcia-Franco, W. (2005). Pacific sardine (*Sardinops sagax*) stock discrimination off the west coast of Baja California and southern California using otolith morphometry. *California Cooperative Oceanic Fisheries Investigations Reports*, 46(2004), 113–121.
- Ferguson, G. J., Ward, T. M., and Gillanders, B. M. (2011). Otolith shape and elemental composition: Complementary tools for stock discrimination of mullet (*Argyrosomus japonicus*) in southern Australia. *Fisheries Research*, 110(1), 75–83.
- Fowler, A. J., and Short, D. A. (1996). Temporal variation in the early life-history characteristics of the King George whiting (*Sillaginodes punctata*) from analysis of otolith microstructure. *Marine and Freshwater Research*, 47(6), 809–818.
- Gaemers, P. A. (1983). Taxonomic position of the Cichlidae (Pisces, Perciformes) as demonstrated by the morphology of their otoliths. *Netherlands Journal of Zoology*, 34(4), 566–595.
- Gaga, J. F. (1993). Morphology of the saccular otoliths of six species of lanternfishes of the genus *Symbolophorus* (Pisces: Myctophidae). *Bulletin of marine science*, 52(3), 949–960.
- Gagliano, M., and McCormick, M. I. (2004). Feeding history influences otolith shape in tropical fish. *Marine Ecology Progress Series*, 278, 291–296.
- Gallego, A., Heath, M. R., McKenzie, E., and Cargill, L. H. (1996). Environmentally induced short-term variability in the growth rates of larval herring. *Marine Ecology Progress Series*, 137(1–3), 11–23.
- Galley, E. A., Wright, P. J., and Gibb, F. M. (2006). Combined methods of otolith shape analysis improve identification of spawning areas of Atlantic cod. *ICES Journal of Marine Science*, 63(9), 1710–1717.
- Garcia, A., Mattiucci, S., Damiano, S., Santos, M. N., and Nascetti, G. (2011). Metazoan parasites of swordfish, *Xiphias gladius* (Pisces: Xiphiidae) from the Atlantic Ocean: implications for host stock identification. *ICES Journal of Marine Science*, 68: 175–182.

- Gauldie, R. W., and Nelson, D. G. A. (1990). Interactions between crystal ultrastructure and microincrement layers in fish otoliths. *Comparative Biochemistry and Physiology Part A: Physiology*, 97(4), 449-459.
- Gauldie, R. W., and Crampton, J. S. (2002). An eco-morphological explanation of individual variability in the shape of the fish otolith: comparison of the otolith of *Hoplostethus atlanticus* with other species by depth. *Journal of Fish Biology*, 60(5), 1204-1221.
- Gonzalez-Salas, C., and Lenfant, P. (2007). Interannual variability and intraannual stability of the otolith shape in European anchovy *Engraulis encrasicolus* (L.) in the Bay of Biscay. *Journal of Fish Biology*, 70(1), 35-49.
- Granlund, G. H. (1972). Fourier Preprocessing for Hand Print Character Recognition. *IEEE Transactions on Computers*, C-21(2), 195-201.
- Grygiel, W. (1987). An attempt at a differentiation of herring geographical populations in the southern Baltic on the basis of otolith morphological structure in juvenile spring spawners. *Biul. Morsk. Inst. Rybackiego* 18(3-4): 3-13 (in Polish).
- Hamer, P. A., Kemp, J., Robertson, S. G., and Hindell, J. S. (2012). Multiple otolith techniques aid stock discrimination of a broadly distributed deepwater fishery species, blue grenadier, *Macruronus novaezelandiae*. *Fisheries Research*, 113(1), 21-34.
- Härkönen, T. (1986). Guide to the otoliths of the bony fishes of the Northeast Atlantic. *Danbiu ApS. biological consultants*.
- He, T., Cheng, J., Qin, J. G., Li, Y., and Gao, T. X. (2018). Comparative analysis of otolith morphology in three species of *Scomber*. *Ichthyological Research*, 65(2), 192-201.
- Higgins, R. M., Danilowicz, B. S., Balbuena, J. A., Daníelsdóttir, A. K., Geffen, A. J., Meijer, W. G., ... Wilson, B. (2010). Multi-disciplinary fingerprints reveal the harvest location of cod *Gadus morhua* in the Northeast Atlantic. *Marine Ecology Progress Series*, 404, 197-206.
- Hüssy, K. (2008). Otolith shape in juvenile cod (*Gadus morhua*): Ontogenetic and environmental effects. *Journal of Experimental Marine Biology and Ecology*, 364(1), 35-41.
- Hüssy, K., Hinrichsen, H. H., Eero, M., Mosegaard, H., Hemmer-Hansen, J., Lehmann, A., and Lundgaard, L. S. (2015). Spatio-temporal trends in stock mixing of eastern and western Baltic cod in the Arkona Basin and the implications for recruitment. *ICES Journal of Marine Science*, 69(1), 1-11.
- Hüssy, K., Mosegaard, H., Albertsen, C. M., Nielsen, E. E., Hemmer-Hansen, J., and Eero, M. (2016). Evaluation of otolith shape as a tool for stock discrimination in marine fishes using Baltic Sea cod as a case study. *Fisheries Research*, 174, 210-218.
- Ibáñez, A. L., Hernández-Fraga, K., and Alvarez-Hernández, S. (2017). Discrimination analysis of phenotypic stocks comparing fish otolith and scale shapes. *Fisheries Research*, 185, 6-13.
- ICES (2016). Report of the Stock Identification Methods Working Group (SIMWG). ICES CM 2016/SSGEPI:16. 47 pp.
- Ider, D., Ramdane, Z., Mahé, K., Duffour, J. L., Bacha, M., and Amara, R. (2017). Use of otolith-shape analysis for stock discrimination of *Boops boops* along the Algerian coast (Southwestern mediterranean sea). *African Journal of Marine Science*, 39(3), 251-258.
- Ihssen, P. E., Booke, H. E., Casselman, J. M., McGlade, J. M., Payne, N. R., and Utter, F. M. (1981). Stock identification: materials and methods. *Canadian journal of fisheries and aquatic sciences*, 38(12), 1838-1855.
- Inada, T. (1981). Studies on the Merluccidae fishes. *Bulletin-Far Seas Fisheries Research Laboratory (Japan)*.
- Japp, D. W. (1990). A new study on age and growth of kingklip *Genypterus capensis* off the south and west coasts of South Africa, with comments on its use for stock identification. *South African journal of marine science*, 9(1), 223-237.
- Jaramillo, A. M., Tombari, A. D., Benedetto Durà, V., Eugeni Rodrigo, M., and Volpedo, A. (2014). Otolith Eco-Morphological Patterns of Benthic Fishes From the Coast of Valencia (Spain). *Thalassas*, 30(1), 57-66.
- Javor, B., Lo, N., and Vetter, R. (2011). Otolith morphometrics and population structure of Pacific Sardine (*Sardinops sagax*) along the west coast of North America. *Fishery Bulletin*, 109(4), 402-415.
- Jemaa, S., Bacha, M., Khalaf, G., and Amara, R. (2015). Evidence for population complexity of the European anchovy (*Engraulis encrasicolus*) along its distributional range. *Fisheries Research*, 168, 109-116.

- Jemaa, S., Bacha, M., Khalaf, G., Dessailly, D., Rabhi, K., and Amara, R. (2015). What can otolith shape analysis tell us about population structure of the European sardine, *Sardina pilchardus*, from Atlantic and Mediterranean waters? *Journal of Sea Research*, 96, 11–17.
- Katayama, S., and Isshiki, T. (2007). Variation in otolith macrostructure of Japanese flounder (*Paralichthys olivaceus*): A method to discriminate between wild and released fish. *Journal of Sea Research*, 57(2–3 SPEC. ISS.), 180–186.
- Keating, J., Brophy, D., Officer, R. A., and Mullins, E. (2014). Otolith shape analysis of blue whiting suggests a complex stock structure at their spawning grounds in the Northeast Atlantic. *Fisheries Research*, 157, 1–6.
- Khemiri, S., Gaamour, A., Ben Abdallah, L., and Fezzani, S. (2018). The use of otolith shape to determine stock structure of *Engraulis encrasicolus* along the Tunisian coast. *Hydrobiologia*, 821(1), 73–82.
- King, D. P. F. (1985). Morphological and meristic differences among spawning aggregations of north-east Atlantic herring, *Clupea harengus* L. *Journal of fish biology*, 26(5), 591–607.
- Kristoffersen, J. B., and Magoulas, A. (2008). Population structure of anchovy *Engraulis encrasicolus* L. in the Mediterranean Sea inferred from multiple methods. *Fisheries Research*, 91(2–3), 187–195.
- Kuhl, F. P., and Giardina, C. R. (1982). Elliptic Fourier Features of a Closed Contour. *Computer Graphics and Image Processing*, 18, 236–258.
- L'Abée-Lund, J. H. (1988). Otolith shape discriminates between juvenile Atlantic salmon, *Salmo salar* L., and brown trout, *Salmo trutta* L. *Journal of Fish Biology*, 33(6), 899–903.
- Leguá, J., Plaza, G., Pérez, D., and Arkhipkin, A. (2013). Otolith shape analysis as a tool for stock identification of the southern blue whiting, *Micromesistius australis*. *Lat. Am. J. Aquat. Res.*, 41(3), 479. h
- Lestrel, P. E. (Ed.). (1997). *Fourier descriptors and their applications in biology*. Cambridge University Press.
- Libungan, L. A., Óskarsson, G. J., Slotte, A., Jacobsen, J. A., and Pálsson, S. (2015). Otolith shape: A population marker for Atlantic herring *Clupea harengus*. *Journal of Fish Biology*, 86(4), 1377–1395.
- Libungan, L. A., Slotte, A., Husebø, Å., Godiksen, J. A., and Pálsson, S. (2015). Latitudinal gradient in otolith shape among local populations of Atlantic herring (*Clupea harengus* L.) in Norway. *PLoS ONE*, 10(6).
- Libungan, L. A., Slotte, A., Otis, E. O., and Pálsson, S. (2016). Otolith variation in Pacific herring (*Clupea pallasii*) reflects mitogenomic variation rather than the subspecies classification. *Polar Biology*, 39(9), 1571–1579.
- Lleonart, J., Salat, J., and Torres, G. J. (2000). Removing allometric effects of body size in morphological analysis. *Journal of theoretical Biology*, 205(1), 85–93.
- Lombarte, A., and Castellón, A. (1991). Interspecific and intraspecific otolith variability in the genus *Merluccius* as determined by image analysis. *Canadian Journal of Zoology*, 69(9), 2442–2449.
- Lombarte, A., Chic, O., Parisi-Baradad, V., Olivella, R., Piera, J., and García-Ladona, E. (2006). A web-based environment for shape analysis of fish otoliths. The AFORO database. *Scientia Marina*, 70(1), 147–152.
- Lombarte, A., and Lleonart, J. (1993). Otolith size changes related with body growth, habitat depth and temperature. *Environmental Biology of Fishes*, 37(3), 297–306.
- Lombarte, A., and Morales-Nin, B. (2000). Sagittal otolith size and shape variability to identify geographical intraspecific differences in three species of the genus *Merluccius*. *Journal of the Marine Biological Association of the United Kingdom*, 80(2), 333–342.
- Lombarte, A., and Cruz, A. (2007). Otolith size trends in marine fish communities from different depth strata. *Journal of Fish Biology*, 71(1), 53–76.
- Lombarte, A., Palmer, M., Matallanas, J., Gómez-Zurita, J., and Morales-Nin, B. (2010). Ecomorphological trends and phylogenetic inertia of otolith sagittae in Nototheniidae. *Environmental Biology of Fishes*, 89(3–4), 607–618.
- Lombarte, A., and Tuset, V. (2015). Morfometria de otólitos. *Métodos de estudo com otólitos: princípios e aplicações*. Buenos Aires: CAFP-BA-PIESCI, 269–302.
- Longmore, C., Fogarty, K., Neat, F., Brophy, D., Trueman, C., Milton, A., and Mariani, S. (2010). A comparison of otolith microchemistry and otolith shape analysis for the study of spatial variation in a deep-sea teleost, *Coryphaenoides rupestris*. *Environmental Biology of Fishes*, 89(3), 591–605.
- Lord, C., Morat, F., Lecomte-Finiger, R., and Keith, P. (2012). Otolith shape analysis for three Sicyopterus (Teleostei: Gobioidae: Sicydiinae) species from New Caledonia and Vanuatu. *Environmental Biology*

of Fishes, 93(2), 209–222.

- Lorenzo, M. I., Vaz-dos-Santos, A. M., and Rossi-Wongtschowski, C. L. D. B. (2011). Growth pattern of the young of the year Argentine hake *Merluccius hubbsi* Marini, 1933 (Gadiformes Merlucciidae) along the Brazilian and Uruguayan coasts. *Environmental Biology of Fishes*, 91(2), 155–164. <https://doi.org/10.1007/s10641-011-9768-4>
- Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Worsoe Clausen, L., Petursdottir, G., Rasmussen, H., Meland, M., Mullins, E., Pinnegar, J. K., Hoines, A., Trenkel, V. M. (2013). Identifying blue whiting (*Micromesistius poutassou*) stock structure in the Northeast Atlantic by otolith shape analysis. *Canadian Journal of Zoology*, 73(9), 1363–1371.
- Mahé, K., Evano, H., Mille, T., Muths, D., and Bourjea, J. (2016). Otolith shape as a valuable tool to evaluate the stock structure of swordfish *Xiphias gladius* in the Indian Ocean. *African Journal of Marine Science*, 38(4), 457–464.
- Mahé, K., Ider, D., Massaro, A., Hamed, O., Jurado-Ruzafa, A., Gonçalves, P., Anastasopoulou, A., Jadaud, A., Mytilineou, C., Elleboode, R., Ramdane, Z., Bacha, M., Amara, R., de Pontual, H., Ernande, B. (2018). Directional bilateral asymmetry in otolith morphology may affect fish stock discrimination based on otolith shape analysis. *CES Journal of Marine Science*, 76 (1), 232-234.
- Mapp, J., Hunter, E., Van Der Kooijc, J., Songer, S., and Fisher, M. 2017. Otolith shape and size: The importance of age when determining indices for fish-stock separation. *Fisheries Research*, 190: 43–52.
- Marcus, L. F., Corti, M., Loy, A., Naylor, G. J., and Slice, D. E. (Eds.). (2013). *Advances in morphometrics* (Vol. 284). Springer Science and Business Media.
- Marengo, M., Baudouin, M., Viret, A., Laporte, M., Berrebi, P., Vignon, M., ... Durieux, E. D. H. (2017). Combining microsatellite, otolith shape and parasites community analyses as a holistic approach to assess population structure of *Dentex dentex*. *Journal of Sea Research*, 128, 1–14.
- Mérigot, B., Letourneur, Y., and Lecomte-Finiger, R. (2007). Characterization of local populations of the common sole *Solea solea* (Pisces, Soleidae) in the NW Mediterranean through otolith morphometrics and shape analysis. *Marine Biology*, 151(3), 997–1008.
- Messieh, S. N., MacDougall, C., and Claytor, R. (1989). Separation of Atlantic herring (*Clupea harengus*) stocks in the southern Gulf of St. Lawrence using digitized otolith morphometrics and discriminant function analysis. *Biological Sciences Branch, Scotia-Fundy Region, Department of Fisheries and Oceans, Bedford Institute of Oceanography*.
- Messieh, S. N. (1972). Use of otoliths in identifying herring stocks in the southern Gulf of St. Lawrence and adjacent waters. *Journal of the Fisheries Board of Canada*, 29(8), 1113-1118.
- Midway, S. R., Cadrin, S. X., and Scharf, F. S. (2014). Southern flounder (*Paralichthys lethostigma*) stock structure inferred from otolith shape analysis. *Fishery Bulletin*, 112(4), 326–338.
- Mille, T., Mahe, K., Villanueva, M. C., De Pontual, H., and Ernande, B. (2015). Sagittal otolith morphogenesis asymmetry in marine fishes. *Journal of fish biology*, 87(3), 646-663.
- Montanini, S., Stagioni, M., Benni, E., and Vallisneri, M. (2017). Ontogenetic changes in otolith morphology and shape analyses in *Chelidonichthys cuculus* (Linnaeus, 1758) and *Chelidonichthys lucerna* (L., 1758). *Journal of Applied Ichthyology*, 33(2), 217–220.
- Montanini, S., Stagioni, M., Valdrè, G., Tommasini, S., and Vallisneri, M. (2015). Intra-specific and inter-specific variability of the sulcus acusticus of sagittal otoliths in two gurnard species (Scorpaeniformes, Triglidae). *Fisheries Research*, 161, 93–101.
- Mooney, C. J., and Kingsford, M. J. (2016a). Statolith morphometrics as a tool to distinguish among populations of three cubozoan species. *Hydrobiologia*, 787(1), 111–121.
- Mooney, C. J., and Kingsford, M. J. (2016b). Statolith morphometrics can discriminate among taxa of cubozoan jellyfishes. *PLoS ONE*, 11(5).
- Morales-Nin, B., Torres, G.J., Lombarte, A. and Recasens, L. (1998b). Otolith growth and age estimation on European hake *Merluccius merluccius*. *Journal of Fish Biology*, 53, 1155-1168.
- Morat, F., Letourneur, Y., Nérini, D., Banaru, D., and Batjakas, I. E. (2012). Discrimination of red mullet populations (Teleostean, Mullidae) along multi-spatial and ontogenetic scales within the Mediterranean basin on the basis of otolith shape analysis. *Aquatic Living Resources*, 25(1), 27–39.
- Moreira, C., Froufe, E., Vaz-Pires, P., and Correia, A. T. (2019). Otolith shape analysis as a tool to infer the population structure of the blue jack mackerel, *Trachurus picturatus*, in the NE Atlantic. *Fisheries Research*, 209(April 2018), 40–48.

- Nasreddine, K., Benzinou, A., and Fablet, R. (2009). Shape geodesics for the classification of calcified structures: Beyond Fourier shape descriptors. *Fisheries Research*, 98(1-3), 8-15.
- Neves, A., Sequeira, V., Farias, I., Vieira, A. R., Paiva, R. B., and Serrano Gordo, L. (2011). Discriminating bluemouth, *Helicolenus dactylopterus* (Pisces: Sebastidae), stocks in Portuguese waters by means of otolith shape analysis. *Journal of the Marine Biological Association of the United Kingdom*, 91(6), 1237-1242.
- Nolf, D. (1989). Evidence from otoliths for establishing relationships within gadiforms. *Papers on the systematics of gadiform fishes*.
- Parisi-Baradad, V., Lombarte, A., Garcia-Ladona, E., Cabestany, J., Piera, J., and Chic, O. (2005). Otolith shape contour analysis using affine transformation invariant wavelet transforms and curvature scale space representation. *Marine and Freshwater Research*, 56(5), 795-804.
- Parisi-Baradad, V., Manjabacas, A., Lombarte, A., Olivella, R., Chic, Ò., Piera, J., and García-Ladona, E. (2010). Automated Taxon Identification of Teleost fishes using an otolith online database-AFORO. *Fisheries Research*, 105(1), 13-20.
- Parrish, B.B., Sharman, D.P., (1958). Some remarks on methods used in herring 'racial' investigations with special reference to otolith studies. *Rapp. P.-v. Reun. Cons. Perm. Int. Explor. Mer* 143, 66e80.
- Parrish, B.B., Sharman, D.P., (1959). Otolith Types amongst Summer-Autumn Spawning Herring in the Northern North Sea, *ICES Journal of Marine Science*, Volume 25, Issue 1, 81-92,
- Paul, K., Oeberst, R., and Hammer, C. (2013). Evaluation of otolith shape analysis as a tool for discriminating adults of Baltic cod stocks. *Journal of Applied Ichthyology*, 29(4), 743-750.
- Pavlinov, I., and Mikesheina, N. G. (2002). Principles and methods of geometric morphometrics. *Zhurnal obshchei biologii*, 63(6), 473-493.
- Pavlov, D. A. (2018). Differentiation of Freckled Goatfish *Upeneus tragula* Richardson, 1846 (Mullidae) in the Coastal Zone of Vietnam Based on Otolith Shape Analysis. *Russian Journal of Marine Biology*, 44(5), 404-414.
- Pawson, M. G., and Jennings, S. (1996). A critique of methods for stock identification in marine capture fisheries. *Fisheries Research*, 25(3-4), 203-217.
- Pierce, G. J., Santos, M. B., Bishop, A. J., Bellido, J. M., Rasero, M., and Portela, J. M. (2002). Morphometric and meristic variation in Argentine hake (*Merluccius hubbsi*) and southern hake (*Merluccius australis*) from the southwest Atlantic. *Filanmanifmgeomarde*, 20, 1-20.
- Pita, A., Casey, J., Hawkins, S. J., Villarreal, M., and Gutiérrez, M. J. (2016). Conceptual and practical advances in fish stock delineation. *Fisheries Research*, 173(3): 185-193.
- Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith shape of different fish species? *Journal of Morphology*, 267(6), 750-757.
- Popper, A. N., and Lu, Z. (2000). Structure-function relationships in fish otolith organs. *Fisheries research*, 46(1-3), 15-25.
- Pothin, K., Gonzalez-Salas, C., Chabanet, P., and Lecomte-Finiger, R. (2006). Distinction between *Mulloidichthys flavolineatus* juveniles from Reunion Island and Mauritius Island (south-west Indian Ocean) based on otolith morphometrics. *Journal of Fish Biology*, 69(1), 38-53.
- Radhakrishnan, K. V., Yuxuan, L., Jayalakshmy, K. V., Liu, M., Murphy, B. R., and Xie, S. (2012). Application of otolith shape analysis in identifying different ecotypes of *Coilia ectenes* in the Yangtze Basin, China. *Fisheries Research*, 125-126, 156-160.
- Roldán, M. I., García-Marín, J. L., Utter, F. M., and Pla, C. (1998). Population genetic structure of European hake, *Merluccius merluccius*. *Heredity*, 81(3), 327.
- Rohlf, F. J., and Bookstein, F. L. (1990). *Proceedings of the Michigan morphometrics workshop*. University of Michigan Museum of Zoology.
- Rohlf, F. J., and Marcus, L. F. (1993). A revolution morphometrics. *Trends in ecology and evolution*, 8(4), 129-132.
- Raitt, D. F. S. (1961). Otolith Studies of Southern North Sea Herring, *ICES Journal of Marine Science*, 26(3), 312-328
- Ramírez-Pérez, J. S., García-Rodríguez, F. J., Quiñonez-Velázquez, C., and Rodríguez-Domínguez, G. (2015). Recognizing species from commercial catches: Molecular and morphometric analyses of *Scomberomorus* spp. off the Mexican Pacific Coast. *Turkish Journal of Fisheries and Aquatic Sciences*, 15(2), 201-211.
- Renán, X., Montero-Muñoz, J., Garza-Pérez, J. R., and Brulé, T. (2016). Age and Stock Analysis Using

- Otolith Shape in Gags from the Southern Gulf of Mexico. *Transactions of the American Fisheries Society*, 145(6), 1252–1265.
- Reznick, D., Lindbeck, E., and Bryga, H. (1989). Slower growth results in larger otoliths: an experimental test with guppies (*Poecilia reticulata*). *Canadian Journal of Fisheries and Aquatic Sciences*, 46(1), 108-112.
- Rojó, A. L. (1977). El crecimiento relativo del otolito como criterio identificador de poblaciones de bacalao del Atlántico noroeste. *Inv. Pesq*, 41(2), 239-261.
- Robertson, S., and Talman, S. (2002). Shape analysis and ageing of orange roughy otoliths from the south Tasman Rise. *Marine and freshwater Resources Institute, Department of Natural Resources and the Environment, Victoria, Australia*.
- Salat, J., and Font, J. (1987). Water mass structure near and offshore the Catalan coast during winters of 1982 and 1983. *Ann. Geoph., B*, 5(1), 49-54.
- Schulz-Mirbach, T., and Plath, M. (2012). Corrigendum to: All good things come in threes – species delimitation through shape analysis of saccular, lagenar and utricular otoliths. *Marine and Freshwater Research*, 66(8), 757.
- Sadighzadeh, Z., Valinassab, T., Vosugi, G., Motallebi, A. A., Fatemi, M. R., Lombarte, A., and Tuset, V. M. (2014). Use of otolith shape for stock identification of John's snapper, *Lutjanus johnii* (Pisces: Lutjanidae), from the Persian Gulf and the Oman Sea. *Fisheries Research*, 155, 59-63.
- Schulz-Mirbach, T., Stransky, C., Schlickeisen, J., and Reichenbacher, B. (2008). Differences in otolith morphologies between surface- and cave-dwelling populations of *Poecilia mexicana* (Teleostei, Poeciliidae) reflect adaptations to life in an extreme habitat. *Evolutionary Ecology Research*, 10(4), 537–558.
- Serrano Gordo, L., Lourenço, H., Bordalo-Machado, P., Reis, S., Nunes, M. L., Baptista, I., ... Cruz, C. (2009). Stock structure of black scabbardfish (*Aphanopus carbo* Lowe, 1839) in the southern northeast Atlantic. *Scientia Marina*, 73(S2), 89–101.
- Simoneau, M., Casselman, J. M., and Fortin, R. (2000). Determining the effect of negative allometry (length/height relationship) on variation in otolith shape in lake trout (*Salvelinus namaycush*), using Fourier-series analysis. *Canadian Journal of Zoology*, 78(9), 1597-1603.
- Škeljo, F., and Ferri, J. (2012). The use of otolith shape and morphometry for identification and size-estimation of five wrasse species in predator-prey studies. *Journal of Applied Ichthyology*, 28(4), 524–530. <https://doi.org/10.1111/j.1439-0426.2011.01925.x>
- Smith, J. N., Nelson, R., and Campana, S. E. (1991). The use of Pb-210/Ra-226 and Th-228/Ra-228 disequilibria in the ageing of otoliths of marine fish. In *Radionuclides in the study of marine processes* (pp. 350-359). Springer, Dordrecht.
- Smith, M. K. (1992). Regional differences in otolith morphology of the deep slope red snapper *Etelis carbunculus*. *Canadian Journal of Fisheries and Aquatic Sciences*, 49(4), 795-804.
- Smith, P. J., Robertson, S. G., Horn, P. L., Bull, B., Anderson, O. F., Stanton, B. R., and Oke, C. S. (2002). Multiple techniques for determining stock relationships between orange roughy, *Hoplostethus atlanticus*, fisheries in the eastern Tasman Sea. *Fisheries Research*, 58(2), 119–140.
- Soria, J. A., and Parisi-Baradad, V. (2011). Best-basis development towards the automatical detection of otolith irregularities in fishes. In *2011 7th International Conference on Next Generation Web Services Practices* (pp. 332-337). IEEE.
- Soria, J. A., Nasreddine, K., Parisi-Baradad, V., Ferrer-Arnau, L., and Benzinou, A. (2014). Otolith shape classification for fish stock discrimination. *International Image Processing, Applications and Systems Conference, IPAS 2014*, 1–5.
- Stransky, C. (2005). Geographic variation of golden redfish (*Sebastes marinus*) and deep-sea redfish (*S. mentella*) in the North Atlantic based on otolith shape analysis. *ICES Journal of Marine Science*, 62(8), 1691–1698.
- Stransky, C. (2014). Morphometric outlines. In *Stock Identification Methods* (pp. 129-140). Academic Press.
- Stransky, C., Baumann, H., Fevolden, S. E., Harbitz, A., Høie, H., Nedreaas, K., ... Skarstein, T. H. (2008). Separation of Norwegian coastal cod and Northeast Arctic cod by outer otolith shape analysis. *Fisheries Research*, 90(1–3), 26–35.
- Stransky, C., and MacLellan, S. E. (2005). Species separation and zoogeography of redfish and rockfish (genus *Sebastes*) by otolith shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*,

62(10), 2265–2276.

- Stransky, C., Murta, A. G., Schlickeisen, J., and Zimmermann, C. (2008). Otolith shape analysis as a tool for stock separation of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean. *Fisheries Research*, 89(2), 159–166.
- Schwarzthans, W., (1980). Die tertiäre Teleosteer-Fauna Neuseelands, rekonstruiert anhand von otolithen. *Berliner Geowissenschaftliche Abhandlungen A* 26, 1±211.
- Templeman, W., and Squires, H. J. (1956). Relationship of otolith lengths and weights in the haddock *Melanogrammus aeglefinus* (L.) to the rate of growth of the fish. *Journal of the Fisheries Board of Canada*, 13(4), 467-487.
- Torres, G. J., Lombarte, A., and Morales-Nin, B. (2000). Variability of the sulcus acusticus in the sagittal otolith of the genus *Merluccius* (Merlucciidae). *Fisheries Research*, 46(1–3), 5–13.
- Torres, G. J., Norbis, W., and Lorenzo, M. I. (1996). Variations in the measures of argentine hake (*Merluccius hubbsi*) rings otoliths during their first-year: evidence for stocks separation? *Scientia Marina*, 60(2–3), 331–338.
- Tracey, S. R., Lyle, J. M., and Duhamel, G. (2006). Application of elliptical Fourier analysis of otolith form as a tool for stock identification. *Fisheries Research*, 77(2), 138–147.
- Treinen-Crespo, C., Villegas-Hernández, H., Guillén-Hernández, S., Ruiz-Zárate, M. Á. A., and González-Salas, C. (2012). Otolith shape analysis as a tool for population discrimination of the white grunt (*Haemulon plumieri*) stock in the northern coast of the Yucatan Peninsula, Mexico. *Rev. Mar. Cost.*, 4(1), 157–168.
- Trojette, M., Fatnassi, M., Ben Alaya, H., Mahouachi, N. E. H., Chalh, A., Quignard, J. P., and Trabelsi, M. (2014). Applying Sagitta otolith shape in the discrimination of fish populations *Scorpaena porcus* (Linnaeus, 1758) (Scorpaenidae) in the Tunisian coasts. *Cahiers de Biologie Marine*, 55(4), 499–506.
- Tuset, V. M., Azzurro, E., and Lombarte, A. (2012). Identification of Lessepsian fish species using the sagittal otolith. *Scientia Marina*, 76(2), 289–299.
- Tuset, V. M., Imondi, R., Aguado, G., Otero-Ferrer, J. L., Santschi, L., Lombarte, A., and Love, M. S. (2014). Otolith patterns of rockfishes from the northeastern pacific. *Journal of Morphology*, 276(4), 458–469.
- Tuset, V. M., Lombarte, A., and Assis, C. A. (2008). Otolith atlas for the western Mediterranean, north and central eastern Atlantic. *Scientia Marina*, 72(S1), 7–198.
- Tuset, V. M., Lombarte, A., González, J. A., Pertusa, J. F., and Lorente, M. J. (2003). Comparative morphology of the sagittal otolith in *Serranus* spp. *Journal of Fish Biology*, 63(6), 1491–1504.
- Tuset, V. M., Lozano, I. E., González, J. A., Pertusa, J. F., and García-Díaz, M. M. (2003). Shape indices to identify regional differences in otolith morphology of comber, *Serranus cabrilla* (L., 1758). *Journal of Applied Ichthyology*, 19(2), 88–93.
- Vasconcelos, J., Vieira, A. R., Sequeira, V., González, J. A., Kaufmann, M., and Serrano Gordo, L. (2018). Identifying populations of the blue jack mackerel (*Trachurus picturatus*) in the Northeast Atlantic by using geometric morphometrics and otolith shape analysis. *Fishery Bulletin*, 116(1), 81–92.
- Vaz-Dos-Santos, A. M., Dos Santos-Cruz, N. N., De Souza, D., Giombelli Da Silva, A., Gris, B., and Rossi-Wongtschowski, C. L. D. B. (2017). Otoliths sagittae of *Merluccius hubbsi*: An efficient tool for the differentiation of stocks in the southwestern atlantic. *Brazilian Journal of Oceanography*, 65(3), 520–525.
- Vieira, A. R., Neves, A., Sequeira, V., Paiva, R. B., and Serrano Gordo, L. (2014). Otolith shape analysis as a tool for stock discrimination of forkbeard (*Phycis phycis*) in the Northeast Atlantic. *Hydrobiologia*, 728(1), 103–110. <https://doi.org/10.1007/s10750-014-1809-5>
- Vignon, M. (2012). Ontogenetic trajectories of otolith shape during shift in habitat use: Interaction between otolith growth and environment. *Journal of Experimental Marine Biology and Ecology*, 420, 26–32.
- Vignon, M., Morat, F., Galzin, R., and Sasal, P. (2008). Evidence for spatial limitation of the bluestripe snapper *Lutjanus kasmira* in French Polynesia from parasite and otolith shape analysis. *Journal of Fish Biology*, 73(10), 2305–2320.
- Vignon, M., and Morat, F. (2010). Environmental and genetic determinant of otolith shape revealed by a non-indigenous tropical fish. *Marine Ecology Progress Series*, 411, 231–241.
- Villegas-Hernández, H., Rodríguez-Canul, R., Guillén-Hernández, S., Zamora-Bustillos, R., and González-

- Salas, C. (2014). Population differentiation in *Haemulon plumieri* juveniles across the northern coast of the Yucatan Peninsula. *Aquatic Biology*, 20(2), 129–137.
- Watkinson, D. A., and Gillis, D. M. (2005). Stock discrimination of Lake Winnipeg walleye based on Fourier and wavelet description of scale outline signals. *Fisheries Research*, 72(2–3), 193–203.
- Watkinson, D. A., and Gillis, D. M. (2014). Stock differentiation of walleye based on the Fourier approximation of averaged scale outline signals. *North American Journal of Fisheries Management*, 23(1), 91–99.
- Yedier, S., Konaş, S., Bostanci, D., and Polat, N. (2016). Otolith and scale morphologies of doctor fish *Garra rufa* inhabiting Kangal Balkl Çermik thermal spring. *Iranian Journal of Fisheries Sciences*, 15(4), 1593–1608.
- Yu, X., Cao, L., Liu, J., Zhao, B. B., Shan, X., and Dou, S. (2014). Application of otolith shape analysis for stock discrimination and species identification of five goby species (Perciformes: Gobiidae) in the northern Chinese coastal waters. *Chinese Journal of Oceanology and Limnology*, 32(5), 1060–1073.
- Zahn, C. T., and Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on computers*, 100(3), 269–281.
- Zemeckis, D. R., Martins, D., Kerr, L. A., and Cadrin, S. X. (2018). Stock identification of Atlantic cod (*Gadus morhua*) in US waters: an interdisciplinary approach. *ICES Journal of Marine Science*, 73, 1494–1502.
- Zengin, M., Saygin, S., and Polat, N. (2015). Otolith shape analyses and dimensions of the anchovy *Engraulis encrasicolus* L. in the Black and Marmara Seas. *Sains Malaysiana*, 44(5), 657–662.
- Zhang, C., Ye, Z., Panhwar, S. K., and Shen, W. (2013). Stock discrimination of the Japanese Spanish mackerel (*Scomberomorus niphonius*) based on the otolith shape analysis in the Yellow Sea and Bohai Sea. *Journal of Applied Ichthyology*, 29(2), 368–373.
- Zhang, C., Ye, Z., Wan, R., Ma, Q., and Li, Z. (2014). Investigating the population structure of small yellow croaker (*Larimichthys polyactis*) using internal and external features of otoliths. *Fisheries Research*, 153, 41–47.
- Zhang, C., Fan, Y., Ye, Z., Li, Z., Yu, H., Zhenjiang, Y., ... Hongliang, Y. (2017). Identification of five Pampus species from the coast of China based on sagittal otolith morphology analysis. *ACTA OCEANOLOGICA SINICA*, 36(2), 51–56.
- Zhao, B. B., Liu, J., Song, J., Cao, L., and Dou, S. (2017a). Evaluation of removal of the size effect using data scaling and elliptic Fourier descriptors in otolith shape analysis, exemplified by the discrimination of two yellow croaker stocks along the Chinese coast. *Chinese Journal of Oceanology and Limnology*, 35(6), 1482–1492.
- Zhao, B. B., Liu, J., Song, J., Cao, L., and Dou, S. (2017b). Otolith shape analysis for stock discrimination of two *Collichthys* genus croaker (Pisces: Sciaenidae) from the northern Chinese coast. *Journal of Oceanology and Limnology*, 36(3), 981–989.
- Zischke, M. T., Litherland, L., Tilyard, B. R., Stratford, N. J., Jones, E. L., and Wang, Y. G. (2016). Otolith morphology of four mackerel species (*Scomberomorus* spp.) in Australia: Species differentiation and prediction for fisheries monitoring and assessment. *Fisheries Research*, 176, 39–47. <https://doi.org/10.1016/j.fishres.2015.12.003>
- Zorica, B., Sinovčić, G., and Keč, V. Č. (2010). Preliminary data on the study of otolith morphology of five pelagic fish species from the Adriatic Sea (Croatia). *Acta Adriatica*, 51(1), 89–96.

SECTION II

II.Executive summary

Section II of D.2.1 aims at performing an overview of otolith microchemistry analysis applications and methodologies, with particular attention on stock identification studies focusing on *Merluccius merluccius* and *Mullus* spp. The aim is to provide a background for the analysis to be performed in the framework of Med_Units project and to contribute to a better understanding of the methodologies to be applied.

The overarching idea behind the review lies in that otolith microchemistry from individual fish, and particularly the dynamics of seven commonly analysed elements (Li, Mg, Mn, Cu, Zn, Sr and Ba) is linked to their concentration in the surrounding water. Ba and Sr are particularly linked to water properties (i.e. salinity, chemical composition) or productivity, and, when analysed in otoliths and coupled to age information, can indicate the environmental history of individual fish. IN addition, stable isotope analysis in otoliths (C and O) can be concurrently used to derive additional environmental history records of fish (e.g., temperature).

A search on Web of Science was conducted for data extraction on the published information on otolith chemistry for European hake and red mullets. Information on sampling characteristics, individual fish variables, instrumental procedure, type of analyses, statistical treatment, environmental parameters and reference details were collated. Results were discussed in terms of reliability and usability within the project.

A total of 82 references were found for hake. From these, a filter based on otoliths and the Mediterranean resulted in 11 manuscripts containing relevant information for the data review. On the other hand, a total of 30 references were found for red mullet albeit none for otoliths microchemistry; therefore, all microchemistry analyses generated in Med_Units will be novel for this species. The conclusions of the review highlight that, in order to maximize the success of the activities within the project, we must 1) use otoliths from immature fish, 2) follow the procedures established by Morales-Nin et al. (2014) for European hake otoliths, 3) test the procedures more suitable for the red mullet otoliths, 4) analyze otolith cores and otolith edges to obtain the natal and sampling location signatures, 5) cover the widest geographical area inside the Mediterranean to ensure enough differentiation of the populations considered.

II.1. INTRODUCTION

The analysis of otolith microchemistry from individual fish has become a widely used technique in exploring relevant aspects of fish ecology, such as the association of individuals with a particular habitat through their life-cycle or in disentangling migrations and population structures (Campana 1999, Elsdon et al. 2008, Phillis et al. 2011, Sturrock et al. 2012; 2014). Seven elements (Li, Mg, Mn, Cu, Zn, Sr and Ba) are routinely used to infer past location in fishes, and strontium (Sr) and barium (Ba) are the most-studied elements due to the positive relationship between their ambient concentrations and their concentrations in otoliths (Elsdon & Gillanders 2005, Elsdon et al. 2008, Tabouret et al. 2010). Typically, marine fishes derive between 83% and 98% of their otolith Sr and Ba from the surrounding water, and the changes produced by marked variations in water properties (i.e. salinity, chemical composition) or productivity are indicative of a particular habitat (Chen & Jones 2006, Walther & Thorrold 2006).

This work aims at performing an overview of otolith microchemistry analysis applications and methodologies, with particular attention on stock identification studies focusing on *Merluccius merluccius* and *Mullus* spp. The aim is to provide a background for the analysis to be performed in the framework of Med_Units project and to contribute to a better understanding of the methodologies to be applied.

II.2. MATERIAL AND METHODS

Web of Science (WoS V.5.29) was used for the data-extraction. Grey literature was not considered. A search in the WoS for (“*Merluccius*“ OR “hake”) AND (otolith* OR geotags OR microchemistry) AND (stocks) was performed. The outcome screen from WoS was saved for a quality check and an initial scan of results. Any paper potentially containing results that matched the objective was kept. Results were listed in a separate spreadsheet. The structure of the spreadsheet was agreed between the partners and contained 18 fields collating all pertinent information (Table 1).

In parallel a search in the WoS for (“*Mullus surmuletus*“ OR “red mullet”) AND (otolith* OR geotags OR microchemistry) was performed.

Table 1. Summary of the spreadsheet to collate the otoliths microchemistry information.

Sampling information	Sampling point, Sampling dates
Fish information	Fish length, Age, Sex, Maturity, number of analysed otoliths
Instrumental procedure	Analytical method, reference materials
Sample analysis	Isotopes, zone analyzed
Statistical analysis	Data treatment
Ambient parameters	Provided, no provided
Author	Author, reference

Results of the articles with the most relevant information to understand the use and methodologies in otolith microanalysis were summarized.

II.3. RESULTS

Literature review results

A total of 82 references were found for hake. From these, a filter based on otoliths and Mediterranean resulted in 11 manuscripts containing relevant information for the data review.

A total of 30 references were found for red mullet albeit none for otoliths microchemistry. Therefore all work performed in this species will be novel.

General history

The first publication dealing with European hake was published in 2005. Most work has been centered in immature fish with total length less than 30 cm (Figure 1). This length threshold has been adopted in order to avoid changes in the otolith's composition due to the possible modification of endolymph due to element mobilization towards gametogenesis (Kalish 1989).

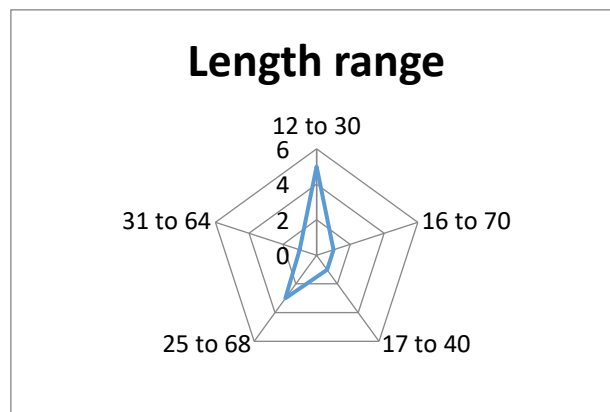


Figure 1. Number of publications on European hake otolith's microchemistry related to the fish length.

Most work has been centered in elucidating the differences between Atlantic and Mediterranean stocks (54.5%) covering a wide range of sites in the Atlantic and W Mediterranean. Only one work was centered in the Atlantic and another in the E Mediterranean, whilst three were dealing with the W Mediterranean. The geographical location of the samples considered reflected clearly the location of the research groups involved in the studies.

Microchemistry analysis methodologies

Two main methodologies are used for otolith microanalysis depending of the isotopes to be analyzed. Each method requires a specific methodology for preparing the otoliths and for their analysis. Methods to analyze trace elements include 1) Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICPMS) (most papers), 2) two papers analyzed light isotopes (carbon and Oxygen) by microdrilling discrete otolith areas and posteriorly analyzing the powder with a mass-spectrometer, and 3) two papers used a combination of both methods (Figure 2).

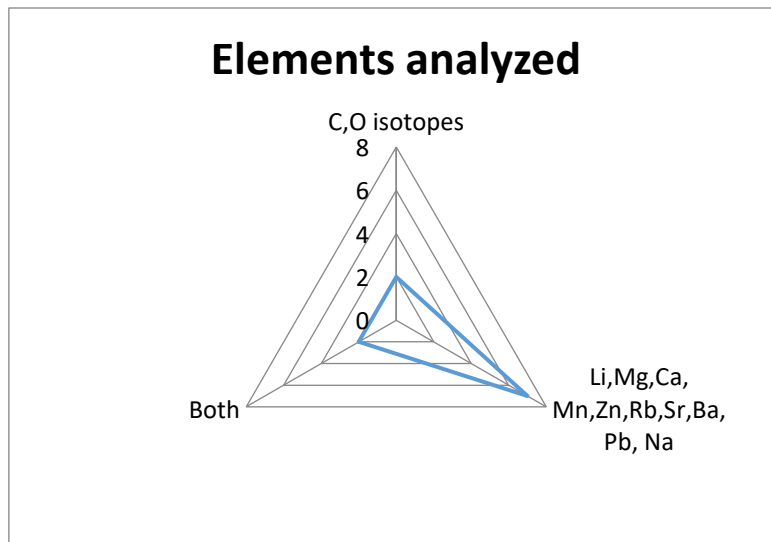


Figure 2. Number of published works on European hake otolith composition depending on the elements considered. Symbols refer to standard chemical notation for the analyzed chemical elements.

Otolith preparation for LA-ICPMS

The analytical procedure requires thoroughly polished otolith sections that could be transversal or sagittal and through the otolith core. Here we describe the sagittal section mostly used due to the major surface available for the analysis (Morales-Nin et al. 2005). For this, each otolith (right or left) was adhered to a piece of cover glass using a thermoplastic resin (Crystalbond) with its sulcal groove side (medial) up, the cover glass with the otolith on it was glued with Crystalbond to a microscope glass slide. Increasingly fine-grain lapping papers (P1200-P2400-P4000) and MilliQ water were used to grinding the otoliths until the nucleus was exposed and the otolith surface was completely flat. Diamond suspensions of 3 μm and 1 μm were used to polish the otolith surface, followed by a final rinse of the otolith in MilliQ water and drying in a laminar flow hood overnight. The cover glass with its otolith on it was released from the microscope glass slide and several otoliths, randomly assigned, were glued to a petrographic glass slide for microchemical analysis.

Calibrated digital images were taken of each otolith, these images were used to measure the distance from otolith nucleus to the area to be analyzed using ImageJ software, so that this feature could be located by the laser ablation system camera. The otolith preparation process could contribute to its surface contamination, thus each petrographic glass slide underwent a final cleaning. This consisted of an ultrasound bath in MilliQ water for 1 min, followed by a 15 sec bath in acetic acid 5% (Emsure quality, Merck), triple rinse with MilliQ water, and a second ultrasound bath in MilliQ water for 1 min. Finally the slides were dried overnight in a laminar flow hood and were individually stored in double zip plastic bags.

Otolith preparation for Carbonate and Oxygen isotopes analysis

The published papers used transversal (Hidalgo et al. 2008a, b) or sagittal (Tanner et al. 2012) sections. For sagittal sections the procedure was as in the previous text. For transversal sections otoliths were embedded in epoxy resin (EpoThin Buehler®) and cut into 500 μm thick sections through the primordium with a slow speed saw using a double-diamond blade. These sections were affixed to glass

slides with thermoplastic glue *Crystalbond*, and ground and polished consecutively with P-400, P-1000 and P-4000 grid paper (FEPA standards).

A high-resolution sampling technique consisting on a computer-controlled micromill was used to microdrill the carbonate in the hake otoliths. Sampling was designed to meet the requirements of the Mass Spectrometer used, whose minimum sample size requirements range between 40 and 50 μg of carbonate mass. Assuming an average otolith density of $2.7 \text{ gr}\cdot\text{cm}^{-3}$ (Hoff and Fuiman 1993) the necessary volume ranged between 0.014 mm^3 and 0.018 mm^3 . About 0.025 mm^3 was extracted from each sample to avoid invalid results. Otolith carbonate was sampled in each otolith at the edge, along the ventro-proximal side, and inside the primordium (Hidalgo et al. 2008) or in the core and otolith edge (Tanner et al. 2012).

Prior to otolith sampling, an external and internal calibration of the Micromill and the software were carried out. From a digital image of the otolith section the region to be sampled was outlined and that line interpolated towards the selected area. Interpolated lines must be separated having in consideration the diameter of the drill.

Otolith analysis with LA-ICPMS

Otolith elemental composition was determined using a Nd:YAG UP-213 laser ablation system (NewWave Research) coupled to an ElementXR plasma mass spectrometer (Thermo-Finnigan and Thermo-Finnigan Element2). This method allowed to measure several elements (Mg^{24} , Ca^{43} , Ca^{44} , Mn^{55} , Sr^{88} and Ba^{138}). Otoliths were sampled with laser line scans, that can measure $200 \mu\text{m}$ in length, $55 \mu\text{m}$ in width, with a scan speed of $3 \mu\text{m/s}$ (Morales-Nin et al. 2014) or $80\times 150 \mu\text{m}$ in the core and $390 \mu\text{m}$ in the dorsal otolith edge (Tanner et al. 2012). Line scans generally were made in the core region (natal signature), just after the settlement mark (~ 60 d old signature), and on the ventral edge of the otolith (capture area signature). Certified Reference Materials (CRMs) NIST610, NIST612, NIES-22 (Yoshinaga et al., 2000) and FEBS-1 (Sturgeon et al. 2005) were analyzed before and after the analysis of the otolith sections on each petrographic glass slide.

Output data from LA-ICPMS was obtained as element intensities that were processed off-line using the software Glitter (GEMOC, Macquarie University) or a R-routine (belonging to the schlerochronology service at IMEDEA <http://imedeaiuib-csic.es/departament.php?d=30>) to obtain element concentrations ($\mu\text{g}_{\text{Me}}/\text{g}_{\text{otolith}}$) using Ca^{43} as internal standard. The four CRMs analyzed were used to calibrate element concentration and the election of one specific CRM for one element in a given session (day) was determined in terms of accuracy and precision (Geffen et al., 2013).

However, a key trade-off for the Laser ablation ICPMS is selecting the size of the ablation area to maximize the resolution for discrete temporal intervals during the life of an individual fish, vs the amount of otolith material required to produce reliable data. Three different widths of ablation lines were used to analyze the otoliths of European hake (*Merluccius merluccius*) (Chang et al. 2012). The best temporal resolution was produced by ablation lines of $10 \mu\text{m}$ width, corresponding to less than 2 weeks in the fish's life, but the data quality from this configuration were variable, often below the detection limit for many elements. Ablation lines of 20 and $30 \mu\text{m}$ width produced accurate and precise data corresponding to approximately 20 and 30 days in terms of temporal resolution. When tested on hake otoliths, the measured element concentrations differed significantly between the 20 and $30 \mu\text{m}$ lines.

The 30 µm ablation line resulted in a better multivariate model for discrimination between populations, with higher classification success and higher probability of individual assignment to source location, which is a key information point to be used in MED_UNITS.

Otolith analysis for Cand O isotopes

Isotope analyses were performed on a Finnegan MAT (M-251) or a Finnigan MAT253 mass spectrometer. Measurements were taken of CO₂ gas from the reaction of the otolith carbonate with orthophosphoric acid at 70°C in an automated on-line system with acid added to the sample in individual reaction chambers. The method used for the spectrometer was CARBO L1/L2 3. Seawater samples were equilibrated with CO₂ at 20°C in an automated Finnigan preparation line, and measured for δ¹⁸O with a Finnigan DELTA-E mass spectrometer with an analytical precision of 0.1‰. Carbonate sample results were reported relative to the VPDB (Vienna Pee Dee Belemnite) standard through calibration against CM03 (Carrara Morbal 03) standards. Water sample results were reported relative to VSMOW (Vienna Standard Mean Ocean Water). All values were reported using standard δ notation:

$$\delta = \left(\frac{R_{\text{sample}} - R_{\text{standard}}}{R_{\text{standard}}} \right) \times 1000 (\text{‰})$$

where R is the ratio ¹⁸O:¹⁶O or ¹³C:¹²C in the sample or standard. The reproducibility of the system was ±0.07‰ for δ¹⁸O and ±0.06‰ for δ¹³C and was based on replicate measurements of an internal carbonate standard.

II.4. DISCUSSION

Fish incorporate different chemical elements from the environment into their otoliths as they grow, either directly via the gills, indirectly through the diet, or from mobilized body stores (Campana, 1999; Gronkaer 2016; Kalish, 1989). Otolith growth and composition is a combined result of physiological processes such as growth (Hoff and Fuiman, 1993; Sadovy and Severin, 1994), metabolism (Borelli et al.2001; Clarke and Friedland 2004), osmoregulation (Kalish, 1989; 1991) and factors such as habitat use (Chang and Geffen 2013). Otoliths are assumed to be metabolically inert, so otolith composition should reflect if populations live and grow in a discrete area, or follow migration routes (Campana 1999).

Otolith chemical composition has been a particularly valuable tool for marine and fishery ecologists in understanding the spatial ecology of marine fish (Edmonds et al., 1991, 1992;Threster 1999; Elston andGillanders, 2002; Kraus and Secor, 2005; Secor and Zdanowicz, 1998; Thorrold et al., 2001). A principal application of otolith microchemistry is to assign widely distributed adults to their natal nursery habitats using otoliths as birth certificates (Secor, 2004) or natural tags (Campana et al., 2000; Ruttenberg 2005). Classification accuracy using otolith microchemistry is moderate and often ranges between 60% and 80% (Arslan and Secor, 2005; Forrester and Swearer, 2002; Rooker et al., 2001; Secor et al., 2002; Swan et al., 2006).

The otolith geotags as indicators of stock structure (Campana et al., 2000; Sturrock et al. 2015; Tanner et al. 2016; Izzo et al. 2018) supported the differentiation of Atlantic and Mediterranean populations. Swan et al. (2006) identified the capture locations in the Northeast Atlantic Ocean and Mediterranean

Sea based on the otolith edge composition with an accuracy of 79% and 65.5%, respectively. Tanner et al. (2012) used the otolith geochemistry, both of the otolith core and edge, showing significant differences among the collection locations in the Atlantic Ocean and the Mediterranean Sea. Individuals were assigned to three different groups in both areas, and the results suggested movement of individuals among local populations within each basin, but little or no movement of hake through the Strait of Gibraltar.

However, Morales-Nin et al. (2015) found that the otolith edge results showed uncertainty in the assignment of fish to its capture site. This edge heterogeneity and lack of geographical differences in the otolith edge composition might be related to limitations on the discriminate power of the few elements analyzed, to low residence time in the location (implying fish movements) or to similarities in the water mass characteristics. European hake is a mobile species undertaking ontogenetic and seasonal migrations to different depths and areas being adults less restricted in depth and location (Recasens et al., 1998). Recruits are located in spatially differentiated high productivity zones that act as nurseries (Maynou et al., 2003). Daily migrations between Merluccidae have been reported related to their feeding rhythms (Orsi-Relini et al., 1989; Bozzano et al., 2005; Cartes et al., 2009), moreover existing data from marking studies evidenced important individual vertical migrations in the water column and along the slope (de Pontual et al., 2003, 2012; 2013). Importantly was the different behavior of individuals, whilst most marked hake remained close to the marking location, some traveled long distances suggesting that some exchange at population level was possible (de Pontual et al., 2013). During the vertical migrations, hake experienced rapid temperature changes of $>7^{\circ}\text{C}$ as they moved frequently across highly stratified water layers albeit there was a high individual variability (Fig.2, de Pontual et al., 2012). This high mobility and heterogeneous behavior might also be reflected in the lack of differentiation in the otolith signal.

Another potential confounding process is the existence of fish endogenous effects. All hake otolith geochemical studies have reported ontogenetic variation in otolith composition. Morales-Nin et al. (2005) found age-related trends in the otolith composition, but were masked by variability between the widely separated fishing locations in the Atlantic versus Mediterranean. Tomás et al. (2006) analyzed Sr:Ca across the otoliths of hake from three different environments suggesting the existence of an endogenous pattern of variation. Tanner et al. (2012) found that some of the element:Ca and stable isotope ratios had consistent pattern between the otolith core and edge zones in all the locations sampled. The ontogenetic habitat changes are observed in the otolith isotope composition (Hidalgo et al., 2008b). Strontium and Ba were more abundant and with less variability in the settlement mark, suggesting association with more warm saline waters and higher productivity.

II.5. CONCLUSION

Therefore, we can conclude that to maximize the success of the proposed activity within MED_UNITS, we have to follow the following recommendations:

- Use otoliths from immature fish.
- Follow the procedures established by Morales-Nin et al. (2014) for European hake otoliths.
- Test the procedures more suitable for the red mullet otoliths.
- Analyze otolith cores and otolith edges to obtain the natal and sampling location signatures.
- Cover the widest geographical area inside the Mediterranean to ensure enough differentiation of the populations considered.

II.6. REFERENCES

- Borelli G, Mayer-Gostan N, Pontual H De, Boeuf G, Payan P (2001) Biochemical relationships between endolymph and otolith matrix in the trout (*Oncorhynchus mykiss*) and turbot (*Psetta maxima*). *Calcif Tissue Int* 69:356–64
- Bozzano, A., Recasens, L., Sartor, P,(1997) Diet of the European hake *Merluccius merluccius* (Pisces:Merluciidae) in theWestern Mediterranean (Gulf of Lions). *Sci.Mar.* 61, 1–8.
- Campana SE (1999) Chemistry and composition of fish otoliths: pathways, mechanisms and application. *Mar Ecol Prog Ser* 188:263–297
- Campana, S.E., Chouinard, G.A., Hanson, J.M., Fréchet, A. and Bratney, J. (2000) Otolith elemental fingerprints as biological tracers of fish stocks.*Fisheries Research* 46(1-3): 343-357
- Cartes, J.; Maynou, F.; Moranta, J.; Massutí, E.; Lloris, D.; Morales-Nin, B. (2004) Patterns of bathymetric distribution among deep-sea fauna at local spatial scale: comparison of mainland vs. insular areas. *Progress in Oceanography* 60:29-45
- Chang, M.Y., Geffen, A.J., Kosler, J., Dundas, S.H. and Maes, G.E.(2012) The effect of ablation pattern on LA-ICPMS analysis of otolith element composition in hake, *Merluccius merluccius*. *Environmental Biology of Fishes* 95(4): 509-520
- Chang MY, Geffen AJ (2013) Taxonomic and geographic influences on fish otolith microchemistry. *Fish Fish* 14:458–492
- Chen Z, Jones CM (2006) Simultaneous determination of 33 major, minor, and trace elements in juvenile and larval fish otoliths by high resolution double focusing sector Weld inductively coupled plasma mass spectrometry. Presented at the. In: 2006 Winter Conferenc eon Plasma Spectrochemistry, Tucson, Arizona, p 8–14
- Clarke LM, Friedland KD (2004) Influence of growth and temperature on strontium deposition in the otoliths of Atlantic salmon. *J Fish Biol* 65:744–759
- de Pontual, H., Bertignac, M., Battaglia, A., Bavouzet, G., Mogueudet, P., Groison, A.-L. (2003) A pilot tagging experiment on European hake (*Merluccius merluccius*): methodology and preliminary results. *ICES J. Mar. Sci.* 60, 1318–1327.
- de Pontual, H., Jolivet, A., Bertignac, M., Fablet, R. (2012) Diel vertical migration of European hake *Merluccius merluccius* and associated temperature histories: insights from a pilot data-storage tagging (DST) experiment. *J. Fish Biol.* 81, 728–734.
- de Pontual, H., Jolivet, A., Garren, F., Bertognac, M. (2013.) New insights on European hake biology and population dynamics from a sustained tagging effort in the Bay of Biscay.*ICES J. Mar. Sci.* 70 (7), 1416–1428.
- Elsdon, T.S. and Gillanders, B.M.(2002) Interactive effects of temperature and salinity on otolith chemistry: challenges for determining environmental histories of fish.*Canadian Journal of Fisheries and Aquatic Sciences* 59(11): 1796-1808
- Elsdon TS, Gillanders BM (2005) Consistency of patterns between laboratory experiments and field collected fish in otolith chemistry: An example and applications for salinity reconstructions. *Mar Freshw Res* 56:609–617
- Elsdon TS, Wells BK, Campana SE, Gillanders BM, Jones CM, Limburg KE, Secor DH, Thorrold SR, Walther BD (2008) Otolith chemistry to describe movements and life-history parameters of fishes: hypotheses, assumptions, limitations and inferences. *Ocean Mar Biol Ann Rev* 46:297–330
- Geffen AJ, Morales-Nin B, Pérez-Mayol S, Cantarero-Roldán AM, Skadal J, Tovar-Sánchez A (2013) Chemical analysis of otoliths: Cross validation between techniques and laboratories. *Fish Res* 143:67–80
- Gronkjaer P (2016) Otoliths as individual indicators: a reappraisal of the link between fish physiology and otolith characteristics. *Mar Freshw Res* 67:881–888
- Hamer PA, Jenkins GP, Coutin P (2006) Barium variation in *Pagrus auratus* (Sparidae) otoliths: A potential indicator of migration between an embayment and ocean waters in south-eastern Australia. *Estuar Coast Shelf Sci* 68:686–702
- Hidalgo, M.; Massutí, E.; Moranta, J.; Cartes, J.L.; Lloret, J.; Oliver, P.; Morales-Nin, B. (2008a) Seasonal and short spatial patterns in European hake (*Merluccius merluccius*, L.) recruitment process

- at the Balearic Sea (NW Mediterranean): the role of environment on distribution and condition. *Journal of Marine Systems* 71 (3-4): 367- 384
- Hidalgo, M.; Tomás, J.; Hole, H.; Morales-Nin, B.; Ninnemann, B.(2008b) Environmental influences on the recruitment process inferred from otolith stable isotopes in *Merluccius merluccius* off the Balearic Islands. *Aquatic Biology* 3(3):195-207
- Hoff GR, Fuiman LA (1995) Environmentally induced variation in elemental composition of red drum (*Sciaenops ocellatus*) otoliths. *Bull Mar Sci* 56:578–591
- Izzo, C., Reis-Santos, P. and Gillanders, B.M. (2018) Otolith chemistry does not just reflect environmental conditions: A meta-analytic evaluation. *Fish and Fisheries* 19(3): 441-454
- Kalish JM (1989) Otolith microchemistry: validation of the effects of physiology, age and environment on otolith composition. *Journal of Experimental Marine Biology and Ecology* 132:151-178
- Kalish JM (1991) Determinants of otolith chemistry: seasonal variation in the composition of blood plasma, endolymph and otoliths of bearded rock cod *Pseudophycis barbatus*. *Mar Ecol Prog Ser* 74:137–159
- Limburg KE (1995) Otolith strontium traces environmental history of subyearling American shad
- Maynou, F., Leonart, J., Cartes, J.E. (2003) Seasonal and spatial variability of hake (*Merlucciusmerluccius* L.) recruitment in the NW Mediterranean. *Fish. Res.* 60, 65–78.
- Mercier L, Darnaude AM, Bruguier O, Vasconcelos RP, Cabral HN, Costa MJ, Lara M, Jones DL, Mouillot D (2011) Selecting statistical models and variable combinations for optimal classification using otolith microchemistry. *Ecol Appl* 21:1352–1364
- Miller JA (2011) Effects of water temperature and barium concentration on otolith composition along a salinity gradient: Implications for migratory reconstructions. *J Exp Mar Bio Ecol* 405:42–52
- Miller JA, Kent AJR (2009) The determination of maternal run time in juvenile Chinook salmon (*Oncorhynchus tshawytscha*) based on Sr/Ca and 87Sr/86Sr within otolith cores. *Fish Res* 95:373–378
- Morales-Nin, B., Swan, S.C., Gordon, J.D.M., Palmer, M., Geffen, A.J., Shillfield, T., Sawyer, T. 2005. Age related trends on otolith microchemistry of *Merluccius merluccius* from the north-eastern Atlantic Ocean and the western Mediterranean Sea. *Marine Freshwater Research* 56: 599-607.
- Morales-Nin B, Pérez-Mayol S, Palmer M, Geffen AJ (2014) Coping with connectivity between populations of *Merluccius merluccius*: An elusive topic. *J Mar Syst* 132:211–219
- Orsi-Relini, L., Fiorentino, F., Cappanera, M., (1986) The timing of recruitment of *Merluccius merluccius* in the Ligurian Sea. *Rapp. Comm. Int. Mer Médit.* 30 (2), 224.
- Payan P, Pontual H De, Bœuf G, Mayer-Gostan N (2010) Endolymph chemistry and otolith growth in fish. *Comptes Rendus Palevol* 3: 535–547
- Phillis CC, Ostrach DJ, Ingram BL, Weber PK (2011) Evaluating otolith Sr/Ca as a tool for reconstructing estuarine habitat use. *Can J Fish Aquat Sci* 68:360–373
- Recasens, L.; Lombarte, A.; Morales-Nin, B.; Torres, G. (1998) Spatiotemporal variation in the population structure of the European hake in the NW Mediterranean. *Journal of Fish Biology* 53: 387-401
- Rooker, J.R., Zdanowicz, V.S. and Secor, D.H. (2001) Chemistry of tuna otoliths: Assessment of base composition and postmortem handling effects. *Marine Biology* 139(1): 35-43
- Ruttenberg, B.I., Hamilton, S.L., Hickford, M.J.H., Paradis, G.L., Sheehy, M.S., Standish, J.D., Ben-Tzvi, O. and Warner, R.R.(2005) Elevated levels of trace elements in cores of otoliths and their potential for use as natural tags. *Marine Ecology Progress Series* 297: 273-281
- Sadovy Y, Severin KP (1994) Elemental patterns in red hind (*Epinephelus guttatus*) otoliths from Bermuda and Puerto Rico reflect growth rate, not temperature. *Can J Fish Aquat Sci* 51:133–141
- Sturgeon RE, Willie SN, Yang L, Greenberg R, Spatz RO, Chen Z, Scriver C, Clancy V, Lam JW, Thorrold S (2005) Certification of a fish otolith reference material in support of quality assurance for trace element analysis. *J Anal At Spectrom* 20:1067–1071
- Sturrock AM, Hunter E, Milton JA, Johnson RC, Waring CP, Trueman CN (2015) Quantifying physiological influences on otolith microchemistry (E Leder, Ed.). *Methods Ecol Evol* 6:806–816

- Sturrock AM, Trueman CN, Darnaude AM, Hunter E (2012) Can otolith elemental chemistry retrospectively track migrations in fully marine fishes? *J Fish Biol* 81:766–795
- Sturrock AM, Trueman CN, Milton JA, Waring CP, Cooper MJ, Hunter E (2014) Physiological influences can outweigh environmental signals in otolith microchemistry research. *Mar Ecol Prog Ser* 500:245–264
- Sturrock, A.M., Hunter, E., Milton, J.A., Johnson, R.C., Waring, C.P., Trueman, C.N. and EIMF (2015) Quantifying physiological influences on otolith microchemistry. *Methods in Ecology and Evolution* 6(7): 806-816
- Swan, S.C., Geffen, A.J., Gordon, J.D.M., Morales-Nin, B. and Shimmield, T. (2006a) Effects of handling and storage methods on the concentrations of elements in deep-water fish otoliths. *Journal of Fish Biology* 68(3): 891-904
- Swan, N., Geffen, A., Morales_Nin, B., Gordon, J., Shimmield, T., Sawyer, T. and Massuti, E. (2006b) Otolith chemistry: an aid to stock separation of *Helicolenus dactylopterus* (bluemouth) and *Merluccius merluccius* (European hake) in the Northeast Atlantic and Mediterranean. *ICES Journal of Marine Science* 63(3): 504-513
- Tabouret H, Bareille G, Claverie F, Pécheyran C, Prouzet P, Donard OFX (2010) Simultaneous use of strontium:calcium and barium:calcium ratios in otoliths as markers of habitat: Application to the European eel (*Anguilla anguilla*) in the Adour basin, South West France. *Mar Environ Res* 70:35–45
- Tanner, S.E.; Vasconcelos, R.P.; Cabral, H.N.; Thorrold, S.R. (2012) Testing an otolith geochemistry approach to determine population structure and movements of European hake in the northeast Atlantic Ocean and Mediterranean Sea. *Fisheries Research* 125-126: 198-205
- Tanner, S.E., Reis-Santos, P. and Cabral, H.N. (2016) Otolith chemistry in stock delineation: A brief overview, current challenges and future prospects. *Fisheries Research* 173: 206-213
- Thresher, R.E. (1999) Elemental composition of otoliths as a stock delineator in fishes. *Fisheries Research* 43(1-3): 165-204
- Thorrold SR, Jones CM, Campana SE (1997) Response of otolith microchemistry to environmental variations experienced by larval and juvenile Atlantic croaker (*Micropogonias undulatus*). *Limnol Oceanogr* 42:102–111
- Tomás, J., Geffen, A.J., Millner, R.S., Piñeiro, C.G. and Tserpes, G. (2006) Elemental composition of otolith growth marks in three geographically separated populations of European hake (*Merluccius merluccius*). *Marine Biology* 148(6): 1399-1413
- Torres, G.; Lombarte, A.; Morales-Nin, B. (2000) Sagittal otolith size and shape variability to identify geographical intraspecific differences in three species of the genus *Merluccius*. *Journal of the Marine Biological Association of the United Kingdom* 80:333-342
- Warburton, M.L., Reid, M.R., Stirling, C.H. and Closs, G. (2017) Validation of depth-profiling LA-ICP-MS in otolith applications. *Canadian Journal of Fisheries and Aquatic Sciences* 74(4): 572-581
- Walther BD, Thorrold SR (2006) Water, not food, contributes the majority of strontium and barium deposited in the otoliths of a marine fish. *Mar Ecol Prog Ser* 311:125–130
- Webb SD, Woodcock SH, Gillanders BM (2012) Sources of otolith barium and strontium in estuarine fish and the influence of salinity and temperature. *Mar Ecol Prog Ser* 453:189–199
- Yoshinaga J, Nakama A, Morita M, Edmonds JS (2000) Fish otolith reference material for quality assurance of chemical analyses. *Mar Chem* 69:91–97
- Zimmerman CE (2005) Relationship of otolith strontium-to-calcium ratios and salinity: experimental validation for juvenile salmonids *Can J Fish Aquat Sci* 62: 88–97

Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-
Project MED_UNITS

*“Study on Advancing fisheries assessment and management
advice in the Mediterranean by aligning biological and
management units of priority species”*

WP2- OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

Milestone 2.1. – Protocol to extract and store otoliths for shape and microchemistry analyses

Compiled by: Sílvia Pérez Mayol (IMEDEA), Kélig Mahé (IFREMER) and Marco
Stagioni (CoNISMa)

February 2019

1. INTRODUCTION

The following protocol explains the procedures to extract and store fish otoliths for elemental quantification by laser ablation or solution-based inductively coupled plasma-mass spectrometry (LA-ICPMS and SO-ICPMS) and otolith shape analysis.

This protocol will be applied for the *Merluccius merluccius* and *Mullus barbatus* otoliths that will be studied in the framework of the MED-UNITS Project. Therefore, the same protocol will be applied for the otoliths of both Task 2.2 (shape) and Task 2.3 (microchemistry) to ease the work.

Despite for microchemistry purposes it is recommended to extract and store the otoliths in very clean conditions, in this case, since the otoliths will be primarily used for shape analysis in IFREMER facilities, we will apply a protocol to facilitate manipulation tasks to scientist from IFREMER and after for scientist of IMEDEA and CoNISMa.

Afterwards, prior to microchemistry analyses, in order to remove any contamination and ensure reliable results, *M. merluccius* otoliths analysed in IMEDEA will suffer a stricter cleaning protocol and *M. barbatus* otoliths analysed in CoNISMa will be pre-ablated.

2. PROTOCOL

2.1. SAMPLING PROCEDURE

According to the project proposal, samples will be obtained during the pilot genetic study (in May 2019) from commercial samplings in different Mediterranean locations, and during the main phase (full genetic study) of the project (in December 2019) from the MEDITS surveys. If MEDITS campaigns do not obtain enough fish, the rest of the samples will be provided by the fish markets of different Mediterranean locations.

- **a minimum of 50 juvenile fish in the case of hake** (<28 cm TL for females and <16 cm TL for males) and **a minimum of 50 adults fish in the case of Mullus barbatus** (>of 12 cm TL) **of both sexes** and site will be sampled at the different locations defined in WP1 for the genetic study. These will be used to extract otoliths (both sides) for otolith shape analysis.
- different information will be provided using the D1.1_sample_sheet provided in

WP1 for the collection of genetic and otolith samples:

- fish identification (ID) (the same ID used for the genetic sampling)
 - date
 - location
 - fishing information (depth, haul, bottom temperature, etc)
 - fish total length (TL, cm)
 - fish weight (W, g)
 - sex and maturity staging
- among the 50 otoliths taken for shape analysis and following these analyses, a **minimum of 30 otolith of juvenile females for hake and a minimum of 30 adult females in the case of mullus** will be put aside to be used for microchemistry analysis. Of these 30 pairs of otoliths only 25 left sagittae will finally be analysed, but more are needed in case of otolith damage during processing.
 - samples for the microchemistry study on hake will all be provided by the Pilot study which include the following 8 sites: GSA1, GSA9, GSA11, GSA16, GSA18, GSA19, GSA22 and GSA25. If some samples can be provided by North African countries (i.e. GSA3) they will also be analysed. In the case of red mullet microchemistry analyses a total of 8 sites will be sampled: GSA5, GSA9, GSA17 (two sites), GSA18 (two sites) (as an alternative one site can be replaced by GSA19), GSA22, GSA25.

2.2. MATERIAL

Any of this material will be provided by partners of the WP2.

- knives
- gloves
- non-metallic tweezers (polypropylene (PP), polytetrafluoroethylene (PTFE))
- 2x glass Petri dish 60-80mm Ø
- deionized water type I (i.e. MilliQ water)
- stock solution of hydrogen peroxide (H₂O₂) 30% Suprapur quality. A final solution of 5% H₂O₂ will be finally used
- 500mL glass bottle
- 500 mL graduated cylinder
- dispenser bottle for deionized water type I
- lab paper
- 2mL PP vials for *M. barbatus*
- 5mL PP vials for *M. merluccius*
- labels for vials identification

- double zip plastic bags
- permanent marker pen.

2.2.1. Preparation of the 5% H₂O₂ solution

The objective of this procedure is to prepare 500mL of 5% H₂O₂ from the stock solution of hydrogen peroxide:

- using the graduated cylinder measure a volume of 416.70mL of MilliQ water and transfer it to the 500mL glass bottle
- using the graduated cylinder measure a volume of 83.30mL of 30% H₂O₂ Suprapur (stock solution) and transfer it to the 500mL glass bottle
- place the plug to the bottle and gently mix it

2.3. OTOLITH EXPOSURE

The otolith extraction necessitates the opening of the cranium as they are located here. In order to access the cavities which enclose the otoliths there are several slicing methods possible. The frontal section slice is the most common and may be used successfully for all types of fish (whichever the species, the individual size or the cranial morphology). However, a given method is used for a given species further to a phase of adaptation and of technical adjustment. The slicing utensils vary according to the size of the cranium but in general a knife is perfectly suitable. The slice must be made carefully in order to avoid severing the internal ear and the otoliths. Having made the appropriate slice, the otoliths are normally removed with tweezers.

2.3.1. Frontal head section

The fish is held by the eyes between the thumb and the index finger, a slice at 45° is made on the forehead (Fig. 1).

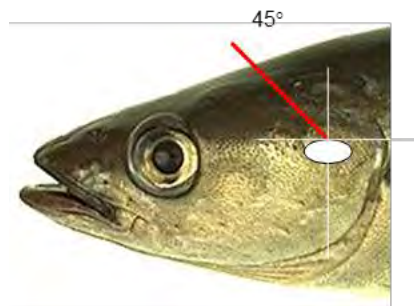


Figure 1: Position of the cranial frontal slice (red line).

Another technique used for the striped red mullet and the red mullet is an inversed frontal slice (Fig. 2).



Figure 2: Position of the inversed cranial frontal slice (white line).

Having opened the cranium and moved the encephalon by turning over the anterior part of the fish head, the two biggest otoliths (the sagittae) are easily detected. They are removed with tweezers (Fig. 3).

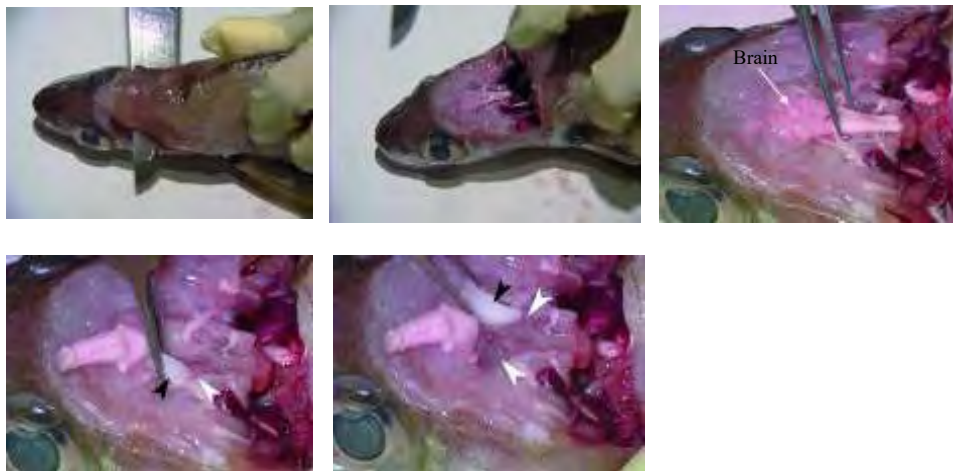


Figure 3. Removal of otoliths via frontal slice on a round fish (*In Panfili et al., 2002*).

2.3.2. Transverse head section

The transversal slice is carried out by separating the body from the head of the fish. This slice is made from the dorsal part towards the ventral part (Fig. 4 & 5). For red mullet a transverse section cut is also used (see figure 5). Removal of otoliths is carried out as in figures 6 and 7.



Figure 4: Position of the transversal slice of the cranium.



Figure 5 Position of transverse section cut (blue line), relative to the otolith position (white circle)

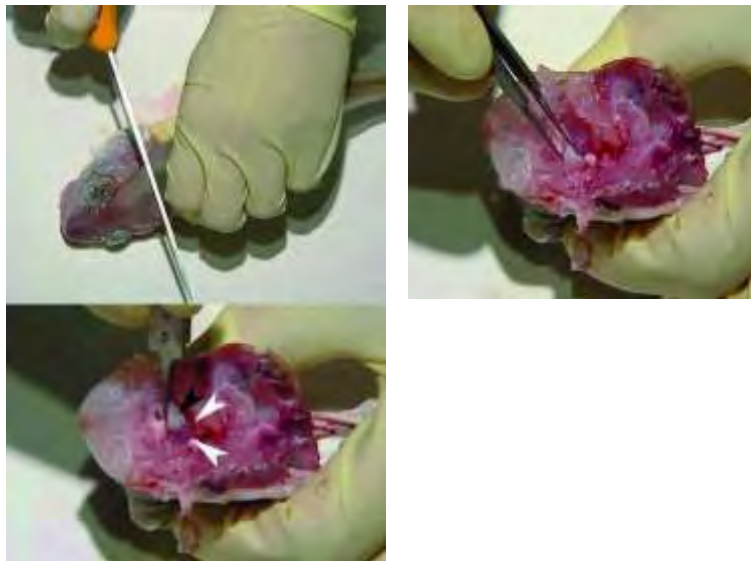


Figure 6. Removal of otoliths via a transversal slice on a round fish (*In* Panfili et al., 2002).



Figure 7 Extraction of otolith via transversal head section

2.3.3. Posterior head section

For the cut of *Merluccius merluccius* also the technique of a posterior section is usually used (Fig. 8)



Figure 8 Posterior section cut (blue line) in European hake, *M. merluccius*, relative to the otolith position (white ellipse).

The otolith extraction via posterior section is reported in figure 9.



Figure 9: Otolith extraction via posterior section in *Merluccius merluccius*.

2.4. OTOLITH EXTRACTION AND STORAGE

From the three pairs of otoliths present in fish's head (*sagitta*, *lapilli* and *asteriscus*), only left and right *sagittae* will be preserved (Fig. 10).

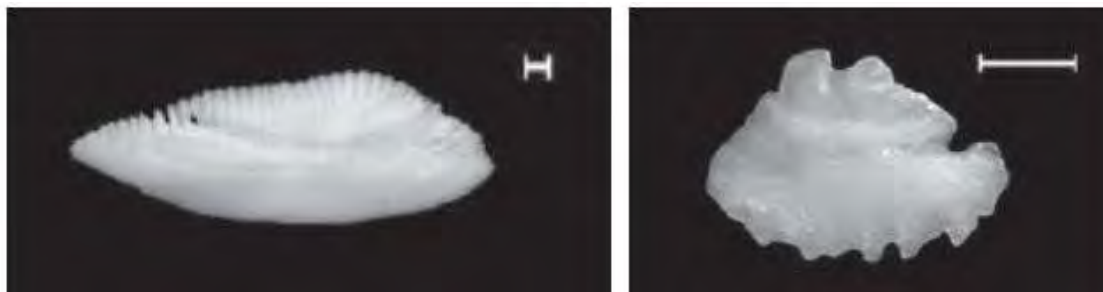


Figure 10. Left) left *sagitta* otolith of a European hake 53 cm TL; Right) left *sagitta* otolith of a red mullet 17.9 cm TL. Scale bars 1 mm. (In Tuset et al, 2008.).

- transfer a small part of the 5% H₂O₂ solution to one Petri dish
- fill the other Petri dish with MilliQ water using the dispenser bottle
- follow the preferred technique explained in section 2.3. to expose the otoliths
- once the *sagitta* pair of otoliths are exposed, extract them using the non-metallic tweezers
- put the pair of otoliths in the Petri dish filled with 5% H₂O₂ and clean them from any biological debris using the non-metallic tweezers
- transfer the otoliths to the second Petri dish filled with MilliQ water and clean the otoliths again, be sure otoliths remain free of any adhered material
- put the otoliths on a piece of paper and allow them to dry for some minutes
- use a marker pen to label a PP vial using the same individual code as for the genetic sampling
- store the pair of otoliths in the labelled PP vial
- empty the two Petri dishes and rinse them with MilliQ water
- rinse the tweezers accordingly
- fill again the two Petri dishes as explained above
- start with another fish
- once all the otoliths have been extracted store all the vials from one species in a double zip plastic bag and label it (species, location) with the marker pen
- contact to Pierluigi Carbonara from COISPA who is the person in charge to gather the otoliths and all their related informations

REFERENCES

Panfili, J.; Pontual, H. (de); Troadec, H. and Wright, P.J. (eds), 2002. Manual of fish sclerochronology. Brest, France: Ifremer-IRD coedition, 4664 p.

Tuset, V.M.; Lombarte, A. and Assis, C.A. 2008. Otolith atlas for the western Mediterranean, north and central eastern Atlantic. *Scientia Marina*, 72(S1): 7-198.

For questions and comments please contact to:

- Sílvia Pérez (IMEDEA): silvia@imedea.uib-csic.es
- Marco Stagioni (CoNISMa): marco.stagioni3@unibo.it
- Kélig Mahé (IFREMER): kelig.mahe@ifremer.fr

MED_UNITS

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP2 - OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

D.2.2 - Protocol for otolith shape analysis and data-treatment of otoliths of European hake and red mullet

RESPONSIBLE: KÉLIG MAHE (IFREMER)

Date: May 28, 2019

Summary

Sagittal otoliths (left and right otoliths) will be extracted from the cranial cavity. After cleaning, to minimize distortion errors within the normalization process during image analysis, images of the whole left and right sagittal otoliths will be scanned under reflected light and stored with high resolution (3200 dpi). Image acquisition process will be standardized and automatized to limit the user bias. External Outline information will be extracted from TNPC software. With these data, three methods will be used to describe otolith contours: the size parameters (Length; Width; Perimeter; Area); shape indices and the Elliptic Fourier Analysis. To test Elliptical Fourier Descriptors (EFDs) of otolith outline, firstly, Principal Components Analysis (PCA) will be applied on EFDs matrix and a subset of the resulting Principal Components (PCs) will be then selected as otolith shape descriptors according to the broken stick model. After the pre-processing of EFDs, several mixed-effects models (multivariate for EFDs and univariate for size parameters or shape indices) will be fitted with potential factors (or explanatory variables) on the otolith shape (response variable described by size parameters, or shape indices or the PCs matrix for EFDs). These analysis may provide a better understanding of the drivers (and their interaction) which control the otolith shape as directional asymmetry, ontogenic effect, environmental effects, and genetic difference and finally the geographical effect. By extracting that the share of variance due to geographical effect, it will be possible to study the stock limits from Linear Discriminant Analysis (LDA) with Jackknifed prediction and cluster analysis according to Ward's hierarchical agglomerative algorithm based on squared Euclidean distances for red mullet (*Mullus barbatus*) and hake (*Merluccius merluccius*).

1. Choice of used shape data

Different approaches have been used to describe and compare the morphology of otoliths which identified 2 mainly groups of extracted data from the otolith outline with univariate and multivariate data (Fig. 1).

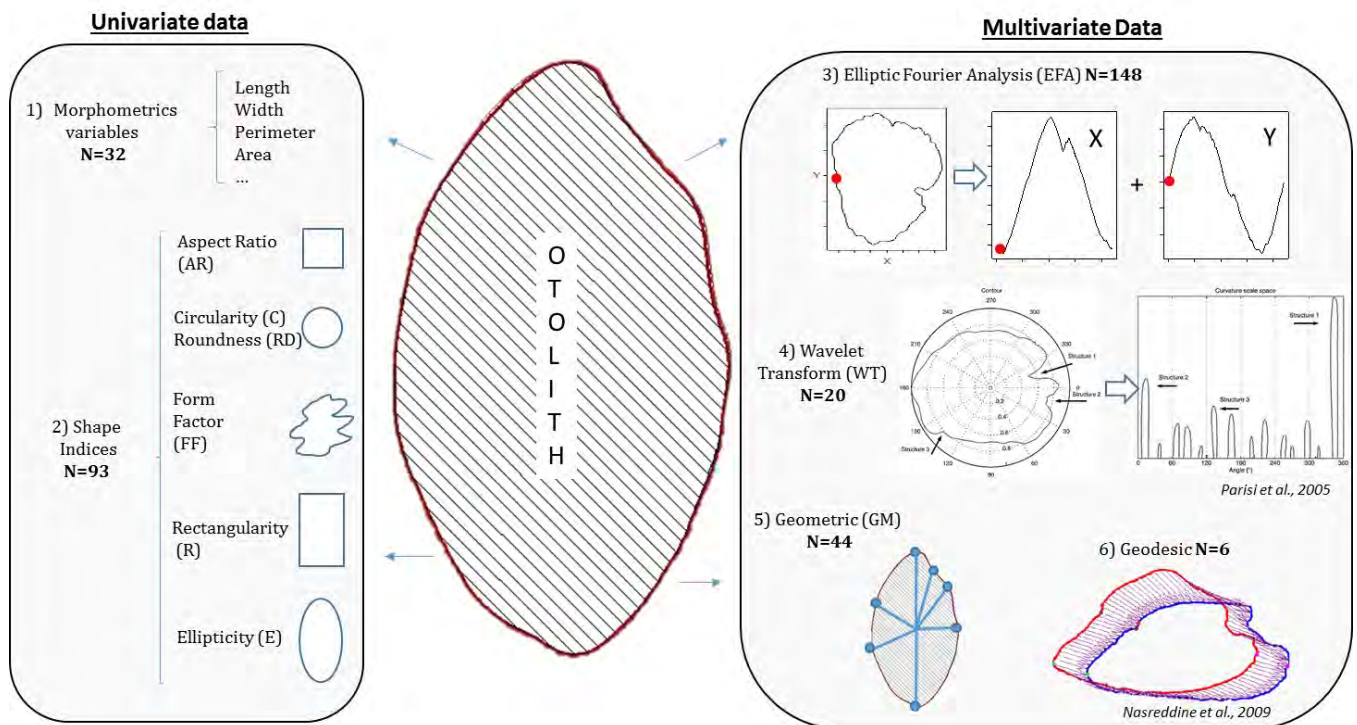


Figure 1 –Type of data extracted from the otolith outline with the number of publications with prior selected “OTOLITH” and “SHAPE” (source K. Mahé).

The comparison between shape otoliths is based on the linear measurement (univariate data) or on the multivariate data which described mathematical the whole otolith. Among the multivariate approaches, Fourier analysis is the most used method (Deliverable 2.1 of this project).

In this project, three methods will be used to describe otolith contours: the size parameters (Length; Width; Perimeter; Area); shape indices and the Elliptic Fourier Analysis (e.g. Lestrel, 2008). The used shape indices are described the Table 1.

TABLE 1 – Shape indices used to describe the otolith shape with formula.

SHAPE INDICES	FORMULA
Form-Factor (FF)	$(4\pi OA) / OP^2$
Aspect Ratio (AR)	OL / OW
Roundness (RD)	$(4OA) / (\pi OL^2)$
Circularity (C)	OP^2 / OA
Ellipticity (E)	$(OL-OW) / (OL+OW)$
Rectangularity (R)	$OA / (OL*OW)$

2. Image acquisition process

Sagittal otoliths (left and right otoliths) will be extracted from the cranial cavity. After cleaning, to minimize distortion errors within the normalization process during image analysis, images of the whole left and right sagittal otoliths will be scanned under reflected light and stored with high resolution (3200 dpi). Scanner will be used to have a standardized acquisition of calibrated images (Elleboode & Mahé, 2016). During this process, a fixed single magnification will be used to ensure as high a resolution as possible. Image processing will be performed using the image analysis system TNPC (Digital processing for calcified structures, version 7; Mahé *et al.*, 2011) with the *sulcus acusticus* facing up (Fig. 2).

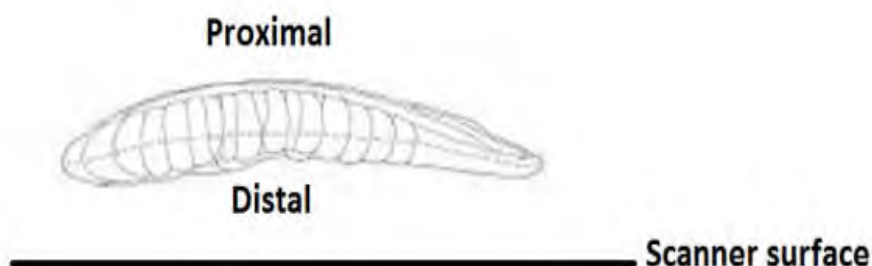


Figure 2 - A lateral image of the right otolith, showing the otolith position with the distal surface (concave side) in the scanner surface (In Gonçalves *et al.*, 2017).

After multiple automatic acquisition step, each otolith image will be individualized. Following this step, a normalization procedure will be applied to these raw images to be invariant with respect to translation, rotation and scaling, so that the normalized shape is the result of the fish history, independently of acquisition settings. The most difficult part of the normalization step is rotation normalization. A simple way to do that would be to normalize in rotation according to the main axis of the shape. Finally, before to extract automatically the outline information, all images are binarized (Figure 3). In order to compare left and right otolith shapes, mirror images of left otoliths will be used.

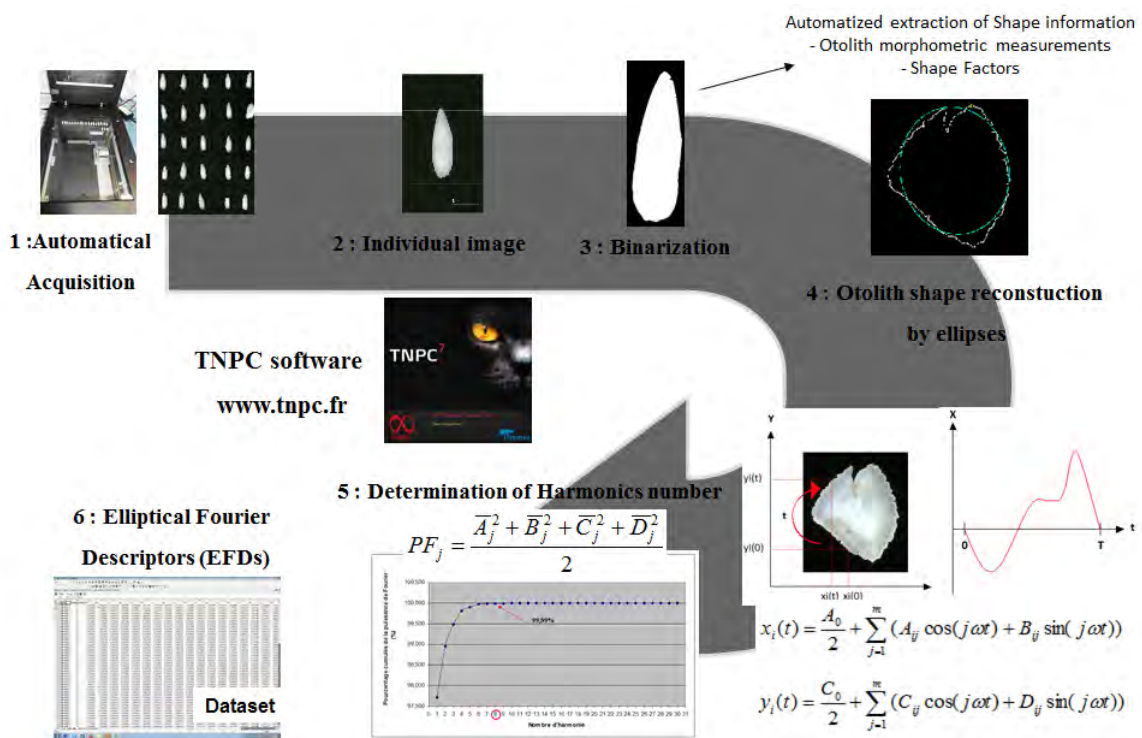


Figure 3 - Protocol to extract the otolith shape information from TNPC software.

The size parameters (Length; Width; Perimeter; Area), the shape indices and the Elliptic Fourier descriptors (EFD) will be automatically extracted from TNPC 7 software.

3. Pre-processing of Elliptic Fourier Descriptors

Elliptic Fourier Descriptors (EFD's) (e.g. Lestrel, 2008) will be carried out on each otolith contour delineated and extracted after image binarization. For each otolith, the first 99 elliptical Fourier

harmonics (H) will be extracted and normalised with respect to the first harmonic and will be thus invariant to otolith size, rotation and starting point of contour description (Kuhl and Giardina, 1982). To determine the number of harmonics required to reconstruct the otolith outline, the cumulated Fourier Power (F) will be calculated for each individual otolith as a measure of the precision of contour reconstruction obtained with n_k harmonics (i.e., the proportion of variance in contour coordinates accounted for by the n_k harmonics):

$$F_{(n_k)} = \sum_{i=1}^{n_k} \frac{A_i^2 + B_i^2 + C_i^2 + D_i^2}{2}$$

where A_i , B_i , C_i and D_i are the coefficients of the H_i harmonic. $F_{(n_k)}$ and n_k will be calculated for each individual otolith k in order to ensure that each individual otolith in the sample will be reconstructed with a precision of 99.99% (Lestrel, 2008). The maximum number of harmonics $n = \max(n_k)$ across all otoliths will be then used to reconstruct each individual otolith.

To test Elliptical Fourier Descriptors (EFDs) of otolith outline, firstly, Principal Components Analysis (PCA) will be applied on EFDs matrix (Rohlf and Archie, 1984) and a subset of the resulting Principal Components (PCs) will be then selected as otolith shape descriptors according to the broken stick model (Legendre and Legendre, 1998). This procedure allowed us to decrease the number of variables used to describe otolith shape variability while ensuring that the main sources of shape variation will be kept, as well as to avoid co-linearity between shape descriptors (Rohlf and Archie, 1984).

4. Potential Directional asymmetry of the otolith shape

Directional Asymmetry (DA) is defined by lateralization process fixed on the same side and depends on its genotype, on the influence of environmental (biotic and abiotic) factors during its life (Fig. 4; Mahé et al., 2019a).

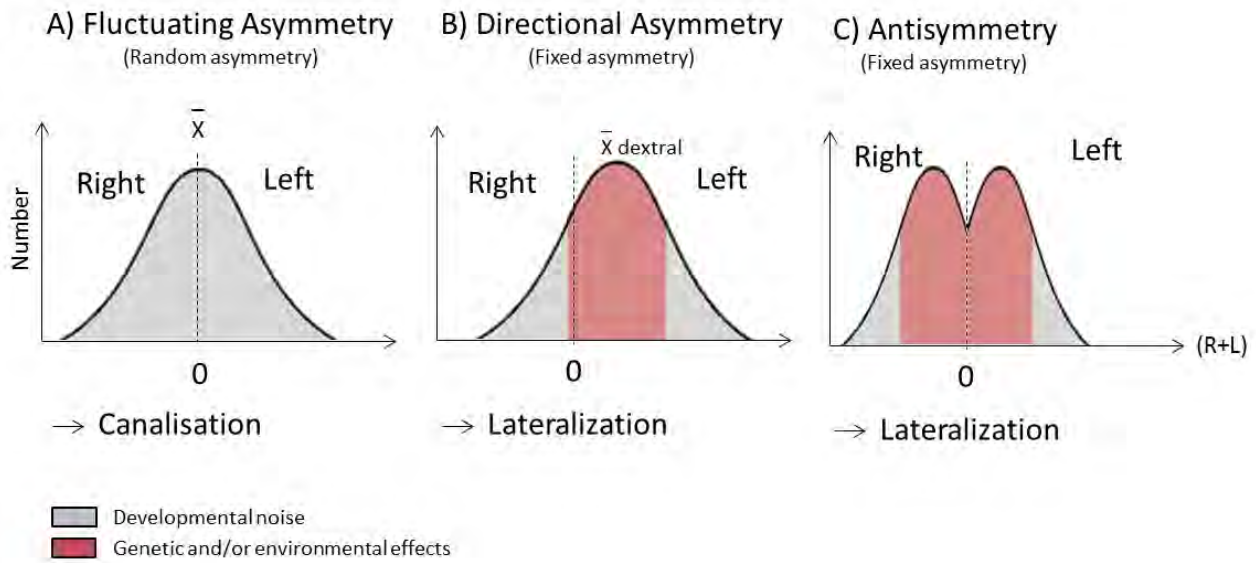


Figure 4 - Categories of asymmetry : A) Fluctuating Asymmetry represented by minor non-directional deviations from perfect symmetry limited by canalization process and the result of developmental noise, B) Directional Asymmetry or lateralization process fixed on the same side and depends on its genotype, on the influence of environmental (biotic and abiotic) factors during its life, C) Antisymmetry or lateralization process fixed on the side which varies randomly among individuals (In Mahé et al., 2019a).

Directional Asymmetry (DA) could affected the results of stock identification from the shape otolith (Mahé et al., 2019b). Consequently, DA in otolith shape of red mullet and hake will be analysed as the effect of otolith’s location side, i.e., left versus right inner ear (side *SI*, thereafter) on otolith shape. The matrix of selected PCs, with PCs from the PCA as columns and otoliths as lines, is referred to as the shape matrix *S* hereafter.

The shape matrix (*S*) will be analyzed using the following multivariate mixed-effects model:

$$S \sim L + SI + LO + SI:LO + i$$

Where otolith shape variations due to side (*SI*), sampling location (*LO*) and their interaction (*SI:LO*) are represented by fixed effects. More precisely, *SI* measures DA at the global scale, *LO* assesses shape variation across sampling locations affecting both otoliths and *SI:LO* represents variation in DA across sampling locations. Individuals’ total length *L* will be also included as a covariate to remove some potentially confounding ontogenetic effect on otolith shape. Finally, a random intercept (*i*) will be used to account for variability due to individuals (or some of their characteristics, such as total length for instance) and autocorrelation between left and right otolith shape within individuals.

The model will be fitted with a different variance for each PC of the shape matrix S . Normality of the residuals and the random effects as well as homoscedasticity of the residuals will be assessed by visual inspection of diagnostic plots. The significance of explanatory variables at 5% will be tested by likelihood ratio tests between nested models while respecting marginality of the effects (type 2 tests; Fox and Weisberg 2011) that are supposed to follow a χ^2 distribution under the null hypothesis. To visualise differences in otolith shape between right and left sides, an average otolith shape will be rebuilt for each side based on EFD's. Moreover, the direction and amplitude of DA at the global scale and at each sampling location will be extracted from the multivariate mixed-effects model as the estimators of the side effect SI and of the interaction between head side and sampling location $SI:LO$, respectively. To ease interpretation, it will be also evaluated as the average percentage of non-overlapping surface between the right and left otoliths' shapes reconstructed on the basis of the EFD's at the individual level. The percentage will be computed relative to the total area.

4. Potential drivers which control the otolith shape

Before to determine the boundaries of stocks for red mullet and hake, the potential drivers which control the otolith shape will be tested from the same process with DA. In fact, mixed-effects models (multivariate for EFDs and univariate for size parameters or shape indices) will be fitted with potential factors (or explanatory variables) on the otolith shape (response variable described by size parameters, or shape indices or the PCs matrix for EFDs).

There will be different potential factors:

- Ontogenic effect with total length and age effects
- Environmental effects with temperature, salinity, depth water, chlorophyll-a...
- Genetic difference from WP1 of this project
- Geographical effect with location effect from GSA areas.

5. Identification of boundaries of stocks for red mullet and hake

To discriminate fish from the all sampled locations based on otolith shape, from PCs, a Linear Discriminant Analysis (LDA) with Jackknifed prediction will be applied to the residuals R_S of a

redundancy analysis (RDA) $S \sim L$ of the shape matrix S explained by individuals' total length L (Fig. 5). The use of the residual matrix R_S instead of the shape matrix S is meant to avoid potential confounding effects due to otolith shape variation across sampling locations related to variations in individuals' size originating from different size-selectivity of the capture procedure/gear at different sampling sites (Rencher and Christensen, 2012). To evaluate the resulting discriminant functions, the percentage of correct classification of individuals into sampling areas will be calculated using jackknife cross-validation (Klecka, 1980) and compared to those obtained from random distribution. Moreover, the performance of the discriminant analyses will be assessed using the Wilks' λ . This value is the ratio between the intra-group variance and the total variance, and provides an objective way of calculating the percentage of agreement between real and predicted groups' membership. Wilks' λ values range from 0 to 1 and the closer to 0, the better the discriminating power of the RDA. To complete the stock identification procedure, a cluster analysis according to Ward's hierarchical agglomerative algorithm based on squared Euclidean distances will be performed on the residual shape matrix R_S to group individuals with similar otolith shapes. These analyses will be carried out three times: on left otoliths only, on right otoliths only and on both.

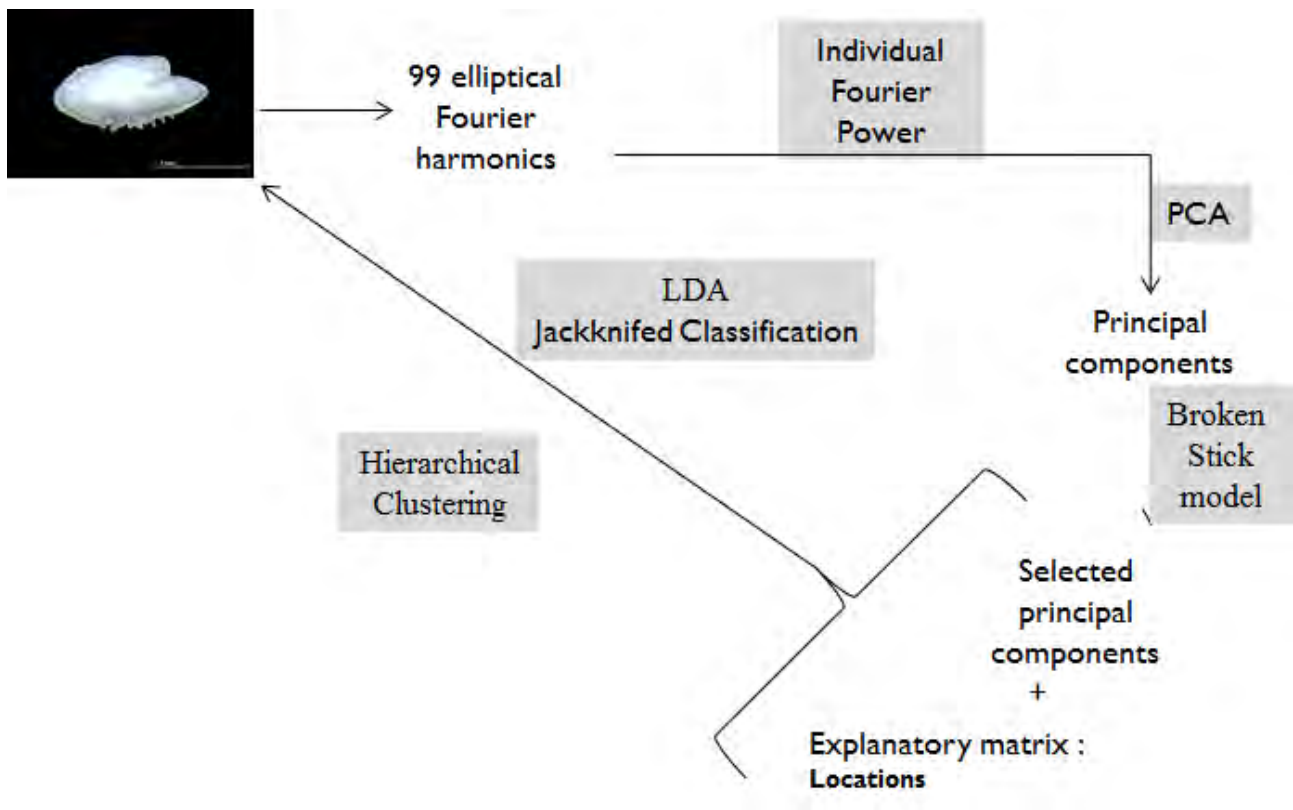


Figure 5 – Analysis of EFD to discriminate the stocks of red mullet and hake.

5. R packages

Statistical analyses will be performed using the following packages in the statistical environment R (R Development Core Team, 2016): ‘nlme’ (Pinheiro *et al.*, 2016), ‘Effects’ (Fox, 2003), ‘Vegan’ (Oksanen *et al.*, 2013), ‘SP’ (Bivand *et al.*, 2013), ‘ggplot2’ (Wickham, 2016), ‘RGEOS’ (Bivand *et al.*, 2013), ‘MASS’ (Venables and Ripley, 2002) and ‘RRCOV’ (Todorov and Filzmoser, 2009).

6. REFERENCES

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. 2013. Applied spatial data analysis with R, Second edition. Springer, New York. 405 pp.

Elleboode R, and Mahe K. 2016. Utilisation du logiciel TNPC (Traitement Numérique des Pièces Calcifiées): réaliser des acquisitions automatiques d’images d’otolithes en lumière réfléchie à l’aide d’un Scanner . SEANOE . <https://doi.org/10.17882/43114>

Fox, J. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8(15): 1-18.

Fox, J., and Weisberg, S. 2011. An {R} Companion to Applied Regression, 2nd edn, Thousand Oaks CA. 472 pp.

Gonçalves P., Mahe K., Elleboode R. Chantre C., Murta A., Avila De Melo A., and Cabral H., 2017. Blue whiting otoliths pair’s symmetry side effect . *International Journal of Fisheries and Aquatic Studies* , 5(3 Part A), 06-09 .

Klecka, W. R. 1980. Discriminant analysis. Sage Publications, Beverly Hills.

Kuhl, F., and Giardina, C. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 18: 236–258.

Legendre, P., and Legendre, L. F. J. 1998. Numerical Ecology. 2nd edn, Elsevier Science. 853 pp.

Lestrel, P. E. 2008. Fourier Descriptors and their Applications in Biology. Cambridge University Press, Cambridge. 460 pp.

Mahé K., Ider D., Massaro A., Hamed O., Jurado-Ruzafa A., Gonçalves P., Anastasopoulou A., Jadaud A., Mytilineou C., Randon M., Elleboode R., Morell A., Ramdane Z., Smith J., Bekaert K., Amara R., de Pontual H., and Ernande B., 2019a. Directional bilateral asymmetry in fish otolith: is it a potential tool to evaluate the stock boundaries? *In* : Lestrel P.E. (Ed.) Biological Shape Analysis: Proceedings of the 5th International Symposium of Biological Shape Analysis. World Scientific, Pub Singapore and New Jersey, USA 5: In press

Mahé K., Ider D., Massaro A., Hamed O., Jurado-Ruzafa A., Gonçalves P., Anastasopoulou A., Jadaud A., Mytilineou C., Elleboode R., Ramdane Z., Bacha M., Amara R., de Pontual H., and Ernande B. 2019b. Directional bilateral asymmetry in otolith morphology may affect fish stock discrimination based on otolith shape analysis. *ICES Journal of Marine Science*, 76(1), 232-243.

Mahe K., Fave S., and Couteau J. 2011. TNPC User guide. <https://archimer.ifremer.fr/doc/00032/14288/>

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, H. M. H., and Wagner, H. 2013. *Vegan: Community Ecology Package*. R package version 2.0–10. 292 pp.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. 2016. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-128.

Rencher, A. C., and Christensen, W. F. 2012. *Methods of Multivariate Analysis*. 3rd edn, Wiley, New York. 800 pp.

Rohlf, F. J., and Archie, J. W. 1984. A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitoes (Diptera: Culicidae). *Systematic Biology*, 33: 302-317.

Todorov, V. and Filzmoser P. 2009. An Object-Oriented Framework for Robust Multivariate Analysis. *Journal of Statistical Software*, 32(3): 1-47.

Venables, W. N., and Ripley, B. D. 2002. *Modern Applied Statistics with S*, 4th edn, Springer, New York. 446 pp.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. 182 pp.

MED_UNITS

Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species

D.2.4

Report on the results of otolith shape analysis
and multivariate analysis of European hake and
red mullet

Responsible: Kélig MAHE (IFREMER)

October 2020

1. Summary 4

2. Introduction 5

3. Sample collection 6

4. Pre-processing of Elliptic Fourier Descriptors 7

5. Potential Directional asymmetry of the otolith shape 10

6. Potential drivers which control the otolith shape 12

7. Geographical effect: compared average shapes 13

8. Identification of boundaries of stocks 16

8.1 Identification of boundaries of stocks for red mullet 16

8.2 Identification of boundaries of stocks for hake 23

9. New approach: 3D analysis of otolith shape 30

10. References 32

1. Summary

Before to identify the boundaries of stock from otolith shape, an analysis was performed to test the potential drivers, which control the otolith shape. Directional Asymmetry (DA; the effect of otolith's location side, i.e., left *versus* right inner ear) was not significant and there was no significant effect of the DA on the relationship between otolith shape and the sampling location ($P > 0.05$). Ontogenic effect (total length) was tested on the otolith shape too. It was significant for both species and so it has been taken into account in the analyses. Moreover, among the potential drivers, sexual dimorphism was not significant on the otolith shape for both species. Consequently, the females and males were compiled in the same analysis. The last tested potential effect was the sampling location. With these data sets, there was significant geographical effect on the otolith shape for red mullet and for hake.

Red mullet population in the Mediterranean Sea analyzed from the otolith shape of 1845 adults from 37 subunits of geographical subareas showed 3 potentials stocks : Western Mediterranean Sea (from GSA1 to GSA16), Adriatic sea with Central Mediterranean Sea (from GSA17 to GSA20) and Eastern Mediterranean Sea (from GSA22 to GSA27).

For hake, 1868 juveniles from 39 subunits of GSAs in the Mediterranean Sea and from 4 ICES areas in the Atlantic Ocean were compared. The results showed that the hake population could be divided in 4 stocks: Atlantic Ocean (from ICES IV to ICES VIII), Western Mediterranean Sea (from GSA1 to GSA13), Adriatic sea with Central Mediterranean Sea (from GSA16 to GSA20) and Eastern Mediterranean Sea (from GSA22 to GSA27).

The new method for the analysis of otolith shape variation in 3D constitutes a high methodological advance, but also new insights in the characterisation and analysis of otoliths. In the future, this type of 3D analysis should be developed to increase potentially the discrimination among fish stocks.

2. Introduction

Otolith shape remains unaffected by the short-term changes in fish condition (Campana and Casselman, 1993) or environmental variations (Campana, 1999). Accordingly, the shape of the otolith has been used as a tool to identify the species, to reconstruct the composition of the diet of predators (fish, seabird, seal, etc.) and to discriminate fish stocks. Since Campana and Casselman (1993), many fishery scientists have developed this type of analysis for stock discrimination studies, as a base for understanding fish population dynamics and achieving reliable assessments for fishery management (Reiss et al., 2009). As a result, more than 90 papers were published from 1993 to 2017 on the identification of marine fish populations or stock structure using otolith shape.

Otolith shape analysis of European was realized on hake and red mullet in the Mediterranean Sea and Atlantic Ocean. This analysis was carried out in several stages:

- Protocol for sampling and conservation of otoliths
- Synthesis of the relevant scientific literature on otolith shape
- Protocol for otolith shape analysis
- Otolith shape analysis and multivariate analysis of European hake and red mullet (this document present the results of these analysis)

3. Sample collection

All the otoliths were collected and extracted by several laboratories. After cleaning the otoliths, an image of each otolith was taken in IFREMER institute and only the whole otoliths were kept for shape analysis (Fig. 1).

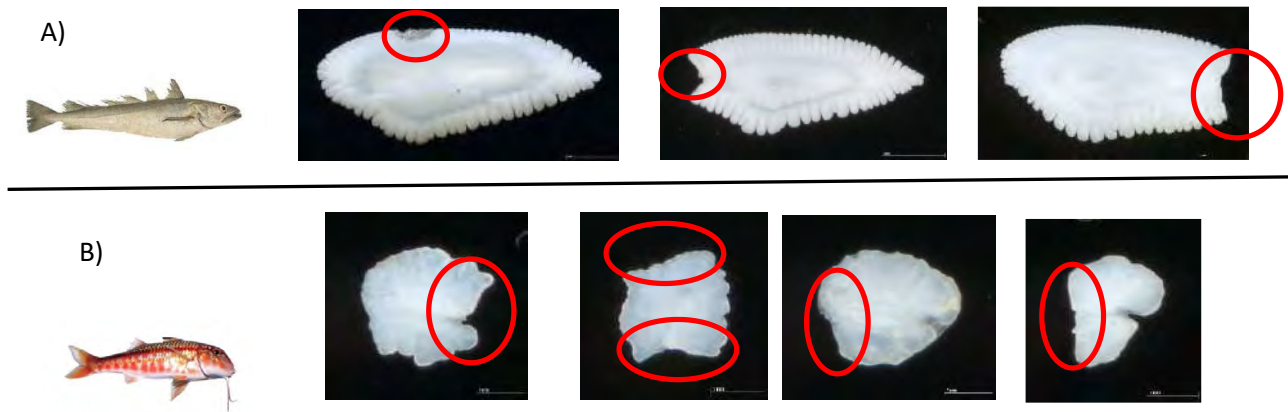


Figure 1 : Several examples of unusable otoliths for otolith shape analysis of European hake (A.) and red mullet (B.).

The used data sets were composed by 1898 adults of red mullets from 37 subunits of geographical subareas and by 1883 juveniles of European hake from 39 subunits of geographical subareas and from 4 ICES areas (Fig 2).

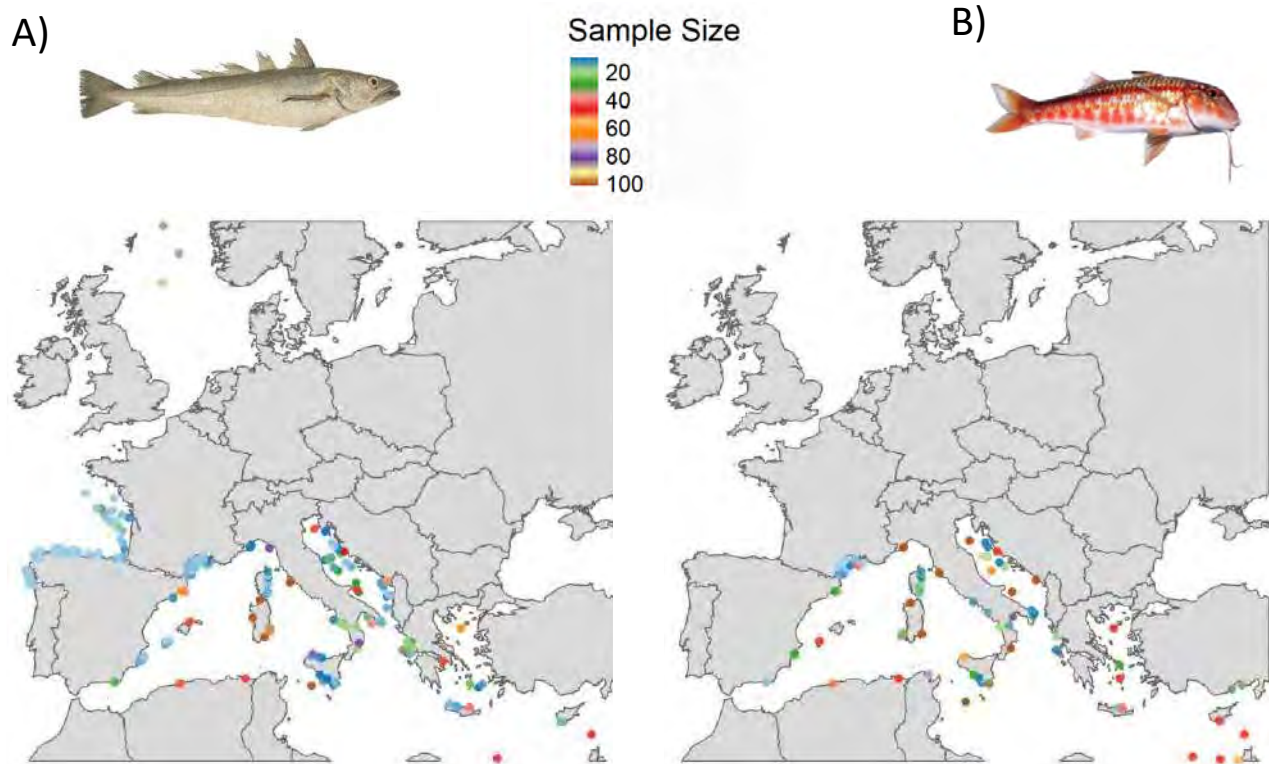


Figure 2 : Available otoliths number for otolith shape analysis of European hake (A.) and red mullet (B.).

4. Pre-processing of Elliptic Fourier Descriptors

For each otolith, the first 99 elliptical Fourier harmonics (H) were extracted corresponding to 396 parameters to describe each external shape. Therefore, the first step was to reduce the number of descriptors (Elliptical Fourier Descriptors; EFDs) using the cumulated Fourier Power (F) for each individual otolith and applying Principal Components Analysis (PCA) on EFDs. The maximum number of harmonics (n_k) to reconstruct the otolith shape with a precision of 99.99% was $n_k=29$ for red mullet and $n_k=50$ for hake (Fig. 3).

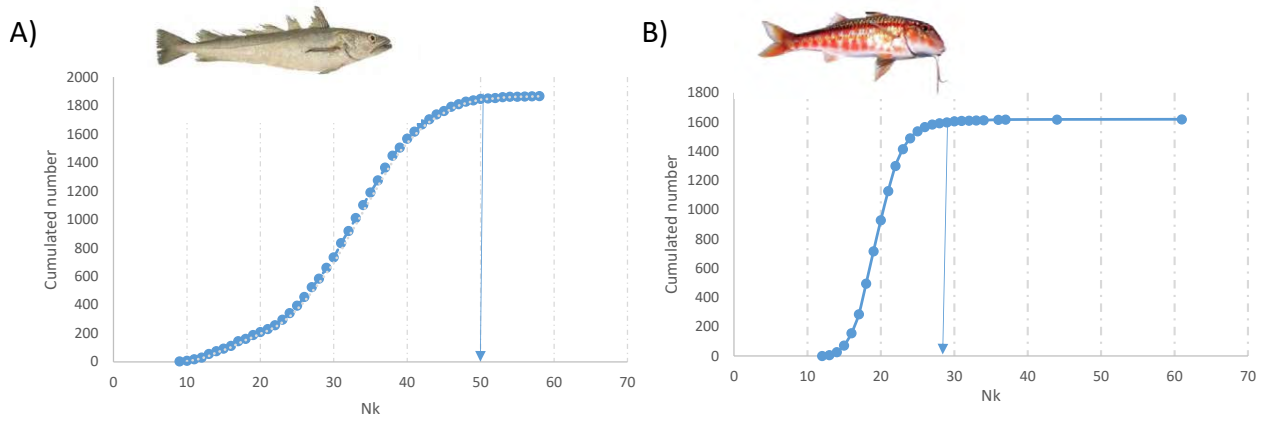


Figure 3 – Cumulated otolith number according to the individual number of harmonics (n_k) needed to reconstruct the otolith shape with a precision of 99.99%.

The maximum number of harmonics $n = \max(n_k)$ across all otoliths was higher for hake than for red mullet because the hake otolith as all gadoid species present the lobes that increase the complexity of the external shape. During this step, the outliers corresponding to very characteristic shapes were removed from the analysis. Consequently, the final number of individuals analysed was 1845 for red mullet and 1868 for hake (Tab. 1).

Table 1 – Available otoliths number by subGSA/subICES areas for otolith shape analysis of European hake and red mullet.

GSAsub	Hake	Red mullet
IVa	67	
VIIIa	39	
VIIIb	41	
VIIIC	49	
1b	25	40
4b	26	47
4c	45	46
5a	9	40
5b	50	
6a	24	42
6b	45	
6c	50	48
7a	28	88
8a	39	91
8b	15	
9a	47	48
9b	47	
10a	50	44
10b	48	49
11b	50	47
11c	48	46
11d	50	
11e	47	46
12	98	106
13	26	
14		46
16b	45	38
16c		43
16d	20	45
17a	50	49
17b	32	48
17c	44	
18a	47	50
18b		49
18c	50	50
19b	44	49
19c	17	48
19d	33	
20a	76	55
22a	57	31
22b	40	31
22c	40	46
23a	65	50
24a		46
25c	57	72
26a	40	
26b		41
26c		36
27a		52
27b	48	42
Total	1868	1845

Principal Components Analysis (PCA) was applied on EFDs matrix to continue to decrease the number of variables used to describe otolith shape variability while ensuring that the main sources of shape variation. The number of significant Principal Components selected from the broken stick model was 6 for both species. Principal Components Analysis of the first 29 Fourier harmonics for red mullet showed that the first and the second PCs accounted for 33% and 17% of the total variance respectively. For hake, PCA of the first 50 Fourier harmonics showed that the first and the second PCs accounted for 45% and 15% of the total variance respectively (Fig. 4).

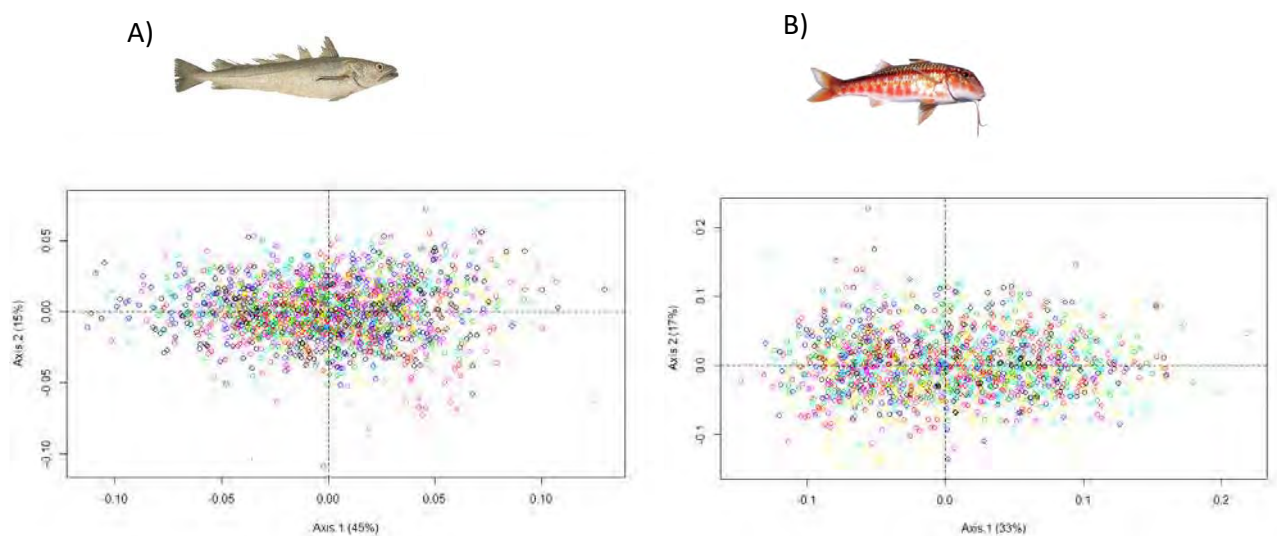


Figure 4 – Distribution of individuals along the axes of the first 2 PCs of PCA applied on the otolith shape dataset of hake (A.) and red mullet (B.).

5. Potential Directional asymmetry of the otolith shape

Before to identify the boundaries of stock from otolith shape, an analysis was performed to test the potential drivers, which control the otolith shape. Directional Asymmetry (the effect of otolith's location side, i.e., left *versus* right inner ear) was tested from subset of dataset (11 GSA subunits; red mullet n= 952; hake n=930) using the following multivariate mixed-effects model:

$$S \sim SI + LO + SI:LO + i$$

The results showed that there was no significant side effect for hake (Tab. 2). Consequently, for this species, it was possible to use left otolith or right otolith when the left otolith was broken. For red mullet, the side effect was significant on the otolith shape. Consequently, we used the left otolith only because this side presented more available data than the right side.

Table 2 – Multivariate mixed-effects models applied to directional asymmetry effect on the otolith shape of European hake and red mullet (shape (S); side (SI); sampling location (LO). Significant effect were in red case.

	Hake	Red mullet
SI	0,79209	0,00001
LO	0,00001	0,00001
SI:LO	0,08838	0,4926

For both species, the side effect was not significant on the relationship between otolith shape and the sampling location (S:SI:LO ; Tab2; $P > 0.05$). In fact, we analyzed the left side only. To visualize differences in average shapes between sides for each species, the reconstructed outlines of the mean Fourier harmonics of the left and right side were plotted as overlay picture (Fig. 5).

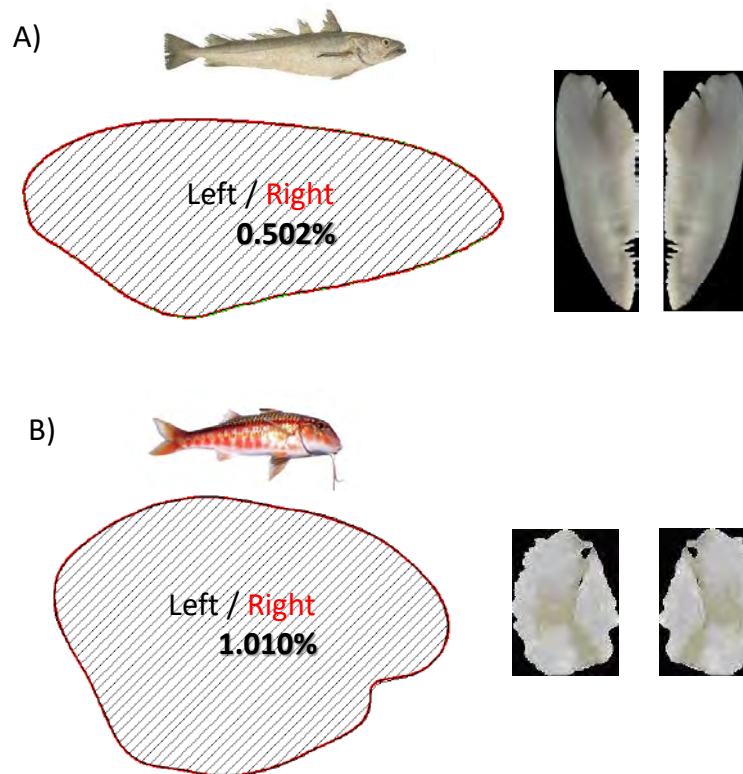


Figure 5 – Mean otolith outline shapes formed with reverse Fourier transform of the outline showing the overlap and variations between two sides. Percentages are the average percentages of non-overlapping surface between the two reconstructed otolith shapes at the individual level.

6. Potential drivers which control the otolith shape

After DA effect, other potential drivers, which control the otolith shape (S), were tested from the same process using the following multivariate mixed-effects model:

$$S \sim TL + Sex + GSA + SubGSA + GSA:SubGSA + i$$

In this model, several effects (or explanatory variables) on the otolith shape (response variable described by the PCs matrix for EFDs) were tested with Sexual dimorphism (Sex) and the geographical effect at two different levels (GSA/ICES area level and GSA/ICES subarea level). Ontogenic effect (total length : TL) is integrated to this model as random effect.

Table 3 – Multivariate mixed-effects models applied to sexual dimorphism and geographical effects on the otolith shape of European hake and red mullet. Significant effect were in red case.

	Hake	Red mullet
Sex	0,107	0,063
GSA	0,004	0,001
SubGSA	0,001	0,001
GSA:SubGSA	0,078	0,017

Sexual dimorphism was not significant effect on the otolith shape for both species. Consequently, all individuals were integrated to the final analysis for both species. The geographical effect was significant on the otolith shape for both species and at two geographical scales (at GSA level and at subGSA level). However, the relationship between GSA and SubGSA was significant for red mullet only. Consequently, for red mullet, the optimisation of boundaries between stocks could be obtained at subGSA level.

7. Geographical effect: compared average shapes

To visualize geographical differences, average shape by reconstructed outlines of the mean Fourier harmonics of all geographical locations were compared (Fig. 6 for Hake and Fig. 7 for red mullet). This type of information does not allow us to identify the stocks composition but it shows the difference level among geographic locations. For both species, the geographical level was higher in the Western Mediterrean Sea than in the Adriactic Sea and in Central Mediterrean Sea and in the eastern Mediterrean Sea.

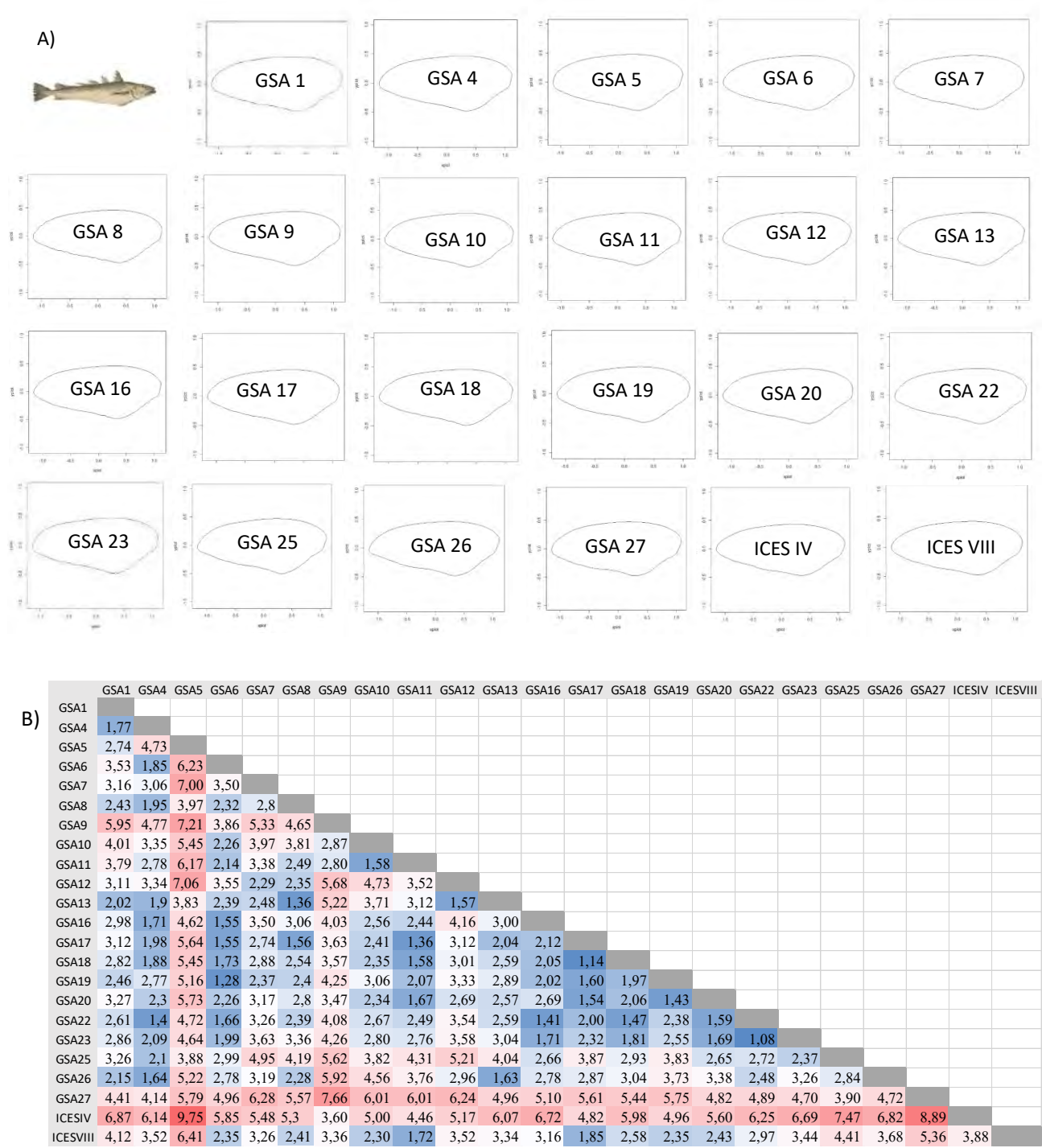


Figure 6 – Mean otolith outline shapes formed with reverse Fourier transform of the outline showing the difference among geographical location from hake (A.) and the average percentages of non-overlapping surface between the two reconstructed otolith shapes between two geographical locations (from low difference in blue case to high difference in red case).



B)

	1	4	5	6	7	8	9	10	11	12	14	16	17	18	19	20	22	23	24	25	26	27	
1																							
4	2,7																						
5	3,42	5,35																					
6	2,59	4,7	3,24																				
7	6,20	4,64	5,62	7,53																			
8	3,74	6,51	3,22	3,94	0,23																		
9	4,51	4,27	4,46	3,87	3,71	6,59																	
10	3,61	3,48	3,92	3,12	2,86	5,85	2,40																
11	4,10	4,46	3,58	4,27	1,79	4,42	2,79	2,61															
12	3,86	2,6	4,02	3,56	3,32	5,96	3,93	2,68	3,13														
14	3,23	3,22	4,77	3,73	3,78	5,90	4,78	5,11	3,82	2,04													
16	5,67	2,55	5,19	5,35	3,64	5,23	2,99	3,02	2,80	2,86	2,84												
17	4,35	2,96	3,44	3,72	2,87	5,52	3,26	1,88	2,95	2,13	4,29	2,22											
18	6,30	2,45	5,90	5,12	5,07	6,73	4,39	3,14	4,72	3,13	3,81	2,63	3,34										
19	3,98	2,91	4,09	3,81	3,31	6,11	2,99	1,90	3,20	2,51	3,01	2,28	2,06	1,77									
20	2,86	4,56	4,13	2,96	3,22	6,32	4,04	3,01	3,27	3,23	4,47	4,46	2,85	4,00	2,85								
22	4,05	3,33	3,99	3,14	3,31	6,05	3,88	1,64	3,28	2,16	3,09	2,96	1,72	2,21	1,76	2,58							
23	4,27	4,69	4,95	3,93	3,70	5,41	4,27	3,14	3,18	3,76	4,67	3,98	3,74	3,88	3,15	3,61	3,08						
24	3,62	4,83	4,55	3,94	3,59	6,58	3,75	4,24	4,06	3,96	5,57	4,98	3,70	4,75	3,42	4,92	3,29	4,22					
25	4,11	4,43	3,90	3,02	3,16	5,49	3,06	2,60	3,02	3,19	4,65	3,81	2,14	2,98	2,54	2,09	2,17	3,54	4,94				
26	6,61	3,46	7,05	6,68	6,50	7,56	7,10	5,63	6,91	4,35	3,65	4,88	5,48	5,09	5,28	6,67	5,77	6,91	6,41	6,40			
27	4,96	2,67	4,72	3,81	3,84	6,49	4,64	2,68	4,40	2,24	3,66	3,04	2,72	2,61	2,55	3,63	2,95	4,28	5,50	3,29	3,81		

Figure 7 – Mean otolith outline shapes formed with reverse Fourier transform of the outline showing the difference among geographical location from red mullet (A.) and the average percentages of non-overlapping surface between the two reconstructed otolith shapes between two GSA areas (B, from low difference in blue case to high difference in red case).

8. Identification of boundaries of stocks

To discriminate fish from the all sampled locations based on otolith shape, from PCs, two complementary analysis were performed with the hierarchical clustering (Unsupervised Machine Learning) and the Linear Discriminant Analysis (LDA) with Jackknifed prediction (Supervised Machine Learning). The Unsupervised Machine Learning identifies the number of groups and their composition. With this information, the Supervised Machine Learning allows to optimize the composition of groups according to administrative location (GSA/ICES areas).

8.1 Identification of boundaries of stocks for red mullet

The results of the hierarchical clustering (Unsupervised Machine Learning) showed an optimized distribution of individuals in 3 clusters (Fig. 8). The distribution of individuals in the 3 clusters does not seem to indicate geographical patterns (Tab. 4).

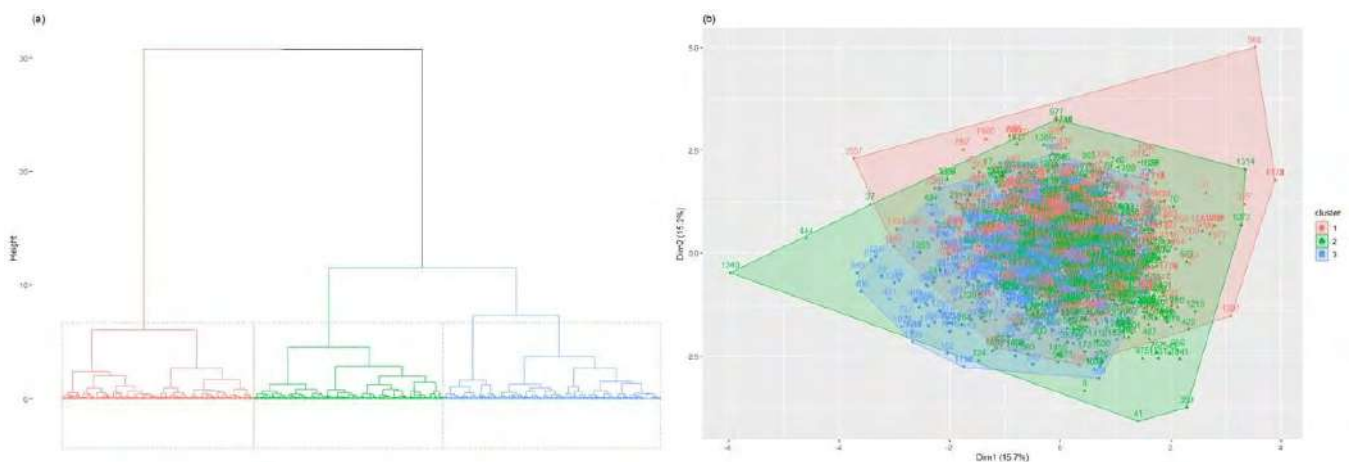


Figure 8 – Hierarchical clustering analysis applied on the otolith shape of the red mullet dataset.

Table 4 – Classification matrix of red mullet from each sampling location (by GSA : A. ; by Subunits of GSA : B) among the different clusters (grey case showing the cluster with the highest number of individuals).

A)

GSA	cluster 1	cluster 2	cluster 3
1	9	19	12
4	28	39	26
5	11	18	11
6	30	24	36
7	32	20	36
8	33	21	37
9	13	21	14
10	31	32	30
11	28	54	57
12	38	47	21
14	15	19	12
16	42	40	44
17	28	33	36
18	46	47	56
19	31	40	26
20	14	19	22
22	25	43	40
23	18	15	17
24	23	16	7
25	21	17	34
26	38	33	6
27	29	54	11

B)

GSAsub	cluster 1	cluster 2	cluster 3
1a	9	19	12
4b	15	18	14
4c	13	21	12
5a	11	18	11
6a	18	13	11
6c	12	11	25
7a	32	20	36
8ab	33	21	37
9a	13	21	14
10a	8	21	15
10b	23	11	15
11b	10	19	18
11c	11	16	19
11e	7	19	20
12	38	47	21
14	15	19	12
16b	11	12	15
16c	16	14	13
16d	15	14	16
17a	15	16	18
17b	13	17	18
18a	12	16	22
18b	22	15	12
18c	12	16	22
19b	17	19	13
19c	14	21	13
20A	14	19	22
22A	10	13	8
22B	6	14	11
22c	9	16	21
23a	18	15	17
24a	23	16	7
25c	21	17	34
26b	15	21	5
26c	23	12	1
27a	12	30	10
27b	17	24	1

The second analysis (LDA) is supervised to predict the position of individuals according to their origin geographical area and thus evaluate the percentage of correct classification (predicted area=actual area). Sampling area was used as an explanatory variable in the subsequent LDA. The overall jackknifed classification success was 9.91% and 6.39% at the geographical scale of the GSA (Tab. 5) and Subunits of GSA (Tab. 6) respectively. The analysis showed significant differences among groups of red mullet sampled in different areas of Mediterranean Sea (at GSA scale : Wilks'λ = 0.7398; χ²=551.26; p=0.001 ; at Subunits of GSA scale : Wilks'λ = 0.6421; χ²=789.69; p=0.001).

Table 5 – Jackknifed correct classification matrix of the linear function discriminant analysis for red mullet (N=1845) between GSA areas based on the selected Principal Components (PC) matrix. The percentages in each row represent the classification into the sampling area given in columns (correct classification in grey square). Overall classification success: 9.91%.

		Predicted GSA																										Correct classification rate
		1	4	5	6	7	8	9	10	11	12	14	16	17	18	19	20	22	23	24	25	26	27					
Actual GSA	1	3	1	5	3	0	1	4	3	1	1	0	0	0	0	0	1	2	1	6	2	1	5	7,5%				
	4	4	7	8	8	0	0	11	2	6	0	6	0	0	1	2	1	2	7	9	5	11	3	7,5%				
	5	4	2	9	2	0	0	1	2	5	2	2	1	0	0	0	0	0	1	2	2	1	4	22,5%				
	6	1	4	8	4	0	1	8	2	5	2	3	0	0	1	0	1	0	18	8	11	7	6	4,4%				
	7	4	0	12	2	0	0	4	4	14	0	4	0	0	2	4	0	0	2	14	8	10	4	0,0%				
	8	5	0	12	2	0	0	4	4	14	0	4	0	0	2	4	0	0	3	14	9	10	4	0,0%				
	9	2	1	5	0	0	0	7	3	4	1	1	0	0	0	0	1	0	5	5	4	4	5	14,6%				
	10	8	1	11	2	0	0	10	0	5	0	2	0	0	1	0	2	2	18	12	8	4	7	0,0%				
	11	6	3	13	0	0	4	13	1	25	3	3	0	0	2	0	1	0	18	14	13	9	11	18,0%				
	12	2	2	9	2	0	0	6	1	10	4	8	0	0	1	0	2	2	7	9	6	24	11	3,8%				
	14	0	1	6	1	0	0	2	0	2	2	5	0	0	0	0	2	0	0	5	5	12	3	10,9%				
	16	6	2	10	2	0	0	11	3	16	1	8	0	0	3	1	3	1	16	17	9	8	9	0,0%				
	17	3	2	12	5	0	0	5	3	13	0	5	1	0	0	1	1	0	11	11	6	11	7	0,0%				
	18	6	7	18	4	0	0	6	3	25	2	5	1	1	5	0	2	1	18	13	7	15	10	3,4%				
	19	4	2	9	7	0	0	13	4	4	0	5	0	0	1	0	4	3	16	7	7	4	7	0,0%				
	20	1	0	4	3	0	1	6	4	3	0	3	0	0	0	0	1	1	10	9	6	3	0	1,8%				
	22	3	4	10	5	0	1	6	0	10	2	4	0	0	1	0	0	1	20	14	11	4	12	0,9%				
	23	1	0	3	2	0	0	2	0	6	0	1	0	0	0	0	0	0	21	8	2	3	1	42,0%				
	24	2	2	4	1	0	0	4	0	1	0	2	0	0	0	0	0	0	9	11	1	4	5	23,9%				
	25	2	3	6	2	0	2	10	2	5	2	2	0	0	0	2	1	0	6	5	15	4	3	20,8%				
	26	0	1	4	0	0	2	2	0	3	3	5	0	1	0	0	0	0	0	2	2	45	7	58,4%				
	27	2	3	7	2	0	7	6	2	6	0	6	0	0	2	0	1	0	5	6	5	14	20	21,3%				

Table 6 – Jackknifed correct classification matrix of the linear function discriminant analysis for red mullet (N=1845) between subunits of the GSAs areas based on the selected Principal Components (PC) matrix. The percentages in each row represent the classification into the sampling area given in columns (correct classification in grey square). Overall classification success: 6.39%.

		Predicted area																											Correct classification												
		1a	4b	4c	5a	6a	6c	7a	8ab	9a	10A	10B	11b	11c	11e	12	14	16b	16c	16d	17a	17b	18a	18b	18c	19b	19c	20A		22A	22B	22c	23a	24a	25c	26b	26c	27a	27b		
Actual area	1a	0	4	1	1	1	3	0	1	2	1	0	0	0	1	1	0	1	0	1	0	0	1	0	2	1	1	0	2	5	0	1	4	1	0	1	0	3	0,0%		
	4b	0	3	1	2	2	6	0	0	1	0	2	6	0	0	1	0	0	0	0	0	0	0	2	3	3	2	0	1	1	2	2	3	1	0	2	1	0	0,0%		
	4c	1	2	2	3	2	3	0	0	1	0	0	2	0	0	0	0	3	1	0	1	0	0	2	1	1	1	1	0	2	0	3	3	1	5	2	0	2	17,0%		
	5a	3	0	2	6	0	4	0	0	0	0	2	3	0	3	1	0	0	0	1	0	0	1	1	1	0	0	0	1	0	1	2	1	1	2	0	0	4	2,2%		
	6a	0	0	0	0	3	2	0	0	1	0	2	1	0	1	2	1	0	0	1	0	0	0	2	2	0	0	0	0	0	1	8	4	3	2	2	3	1	13,0%		
	6c	0	3	1	5	2	7	0	1	1	0	0	4	0	2	0	1	0	0	1	0	0	1	0	0	0	3	0	1	0	0	2	0	6	3	1	0	1	0	3	20,5%
	7a	0	4	0	6	4	2	0	0	0	0	4	4	0	6	0	4	2	0	2	0	0	6	0	2	4	0	0	4	6	2	2	6	4	4	6	0	4	2,2%		
	8ab	0	4	0	6	5	2	0	0	0	0	4	4	0	6	0	4	2	0	2	0	0	6	0	2	4	0	0	4	7	2	2	6	5	4	6	0	4	10,5%		
	9a	0	3	2	0	2	1	1	0	0	0	3	5	0	1	0	0	0	2	0	1	0	0	1	1	0	0	2	0	4	0	0	5	2	4	0	3	0	0,0%		
	10A	3	2	1	3	0	4	0	0	4	0	0	0	0	3	1	1	0	0	0	0	0	0	1	0	0	0	1	1	2	3	0	2	4	3	0	2	2	1	0,0%	
	10B	0	1	0	0	5	4	0	0	5	0	0	1	0	4	2	0	0	0	0	0	0	0	1	1	3	0	0	2	0	8	5	3	0	0	0	0	4	0,0%		
	11b	1	3	0	2	0	2	0	1	1	0	0	8	0	2	2	0	1	0	1	0	0	3	1	0	0	0	0	2	2	0	1	6	2	1	2	0	3	0,0%		
	11c	1	3	0	1	1	0	0	2	0	0	1	3	1	1	1	0	3	0	3	0	0	3	0	1	0	1	0	0	2	0	6	3	4	2	2	0	1	4,0%		
	11e	0	1	1	4	1	2	0	1	2	0	1	6	0	6	0	0	1	0	1	0	0	3	0	2	0	0	0	0	1	1	1	3	3	1	2	2	0	0,0%		
	12	1	4	0	5	5	4	0	0	1	0	1	5	0	4	9	4	4	0	0	0	3	1	2	0	1	1	3	2	1	6	7	4	7	13	0	8	0,0%			
	14	0	1	0	4	1	2	0	0	0	1	0	0	0	2	3	1	3	0	0	0	0	2	0	2	0	0	1	1	0	0	5	2	5	5	0	4	0,0%			
	16b	0	0	1	2	0	2	0	0	0	0	0	0	0	5	6	1	4	0	1	0	0	2	0	0	1	0	0	1	0	0	1	0	0	1	3	2	1	1	0,0%	
	16c	3	1	1	4	1	4	0	0	0	0	0	1	0	0	0	0	1	0	3	0	0	2	0	1	1	0	0	2	1	0	4	7	1	1	1	0	3	0,0%		
	16d	1	2	0	0	3	1	0	0	5	0	2	4	0	5	1	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	4	6	1	0	1	0	1	0,0%		
	17a	0	3	1	2	4	2	0	0	2	1	2	4	0	2	1	0	2	0	0	0	2	1	3	0	0	1	0	2	4	3	1	1	1	1	0	4	3,2%			
	17b	0	1	0	4	2	2	0	0	2	0	1	2	0	2	2	3	0	2	0	0	2	1	1	0	0	0	0	2	2	3	3	2	4	0	3	0,0%				
	18a	1	1	0	2	0	2	0	0	0	0	1	5	0	3	3	1	6	0	1	0	0	2	0	0	1	0	0	1	4	0	6	4	1	4	1	0	0	2,2%		
	18b	0	3	0	2	0	5	0	0	0	1	2	1	0	3	4	0	1	0	2	1	1	0	0	0	1	0	0	1	0	1	3	5	3	1	3	0	5	22,0%		
	18c	1	1	0	3	1	10	0	0	1	0	0	3	0	0	1	0	1	0	2	0	0	2	1	0	2	0	2	0	2	1	5	3	1	1	1	0	5	19,6%		
	19b	1	2	0	2	5	5	0	0	2	0	2	2	0	2	1	2	0	0	1	0	0	0	0	1	0	0	0	1	1	2	3	7	2	2	1	0	2	13,9%		
	19c	0	2	1	3	3	0	0	2	0	1	3	0	1	0	1	0	0	1	0	0	0	3	1	0	0	0	0	3	2	1	4	4	3	2	2	0	3	7,3%		
	20A	0	0	0	4	1	1	0	1	3	0	2	1	0	0	1	0	2	0	5	0	0	1	0	8	0	0	0	2	1	2	6	6	5	1	2	0	0	47,2%		
	22A	0	0	0	1	1	2	0	1	1	0	0	3	0	1	3	0	1	0	1	0	0	0	0	2	1	0	0	0	1	0	0	1	3	3	1	0	2	2	1,9%	
	22B	1	0	0	0	3	5	0	0	1	1	0	4	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0	0	0	2	1	3	2	0	0	3	21,4%		
	22c	0	1	0	2	1	7	0	0	1	0	0	1	0	5	1	0	0	0	0	0	0	0	1	6	0	0	0	0	0	1	9	4	2	1	1	1	0	6,4%		
23a	0	2	0	1	3	1	0	0	0	0	4	2	0	3	1	0	1	0	2	0	1	0	1	0	1	0	1	0	1	1	4	11	7	1	0	2	0	0	4,3%		
24a	1	2	0	2	3	2	0	0	1	1	1	3	0	0	1	0	0	0	0	0	0	1	1	2	1	1	0	1	0	1	5	9	0	4	1	0	2	15,0%			
25c	2	3	0	3	6	4	0	2	6	1	3	2	0	2	0	1	3	0	2	0	0	0	0	3	2	0	0	2	2	1	3	3	10	0	4	0	2	7,1%			
26b	0	1	1	2	0	1	0	2	1	0	0	1	0	1	4	2	1	0	0	1	2	0	0	0	0	0	0	0	1	0	0	1	1	3	14	0	1	14,6%			
26c	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	3	17	0	5	0,0%				
27a	0	1	0	2	2	1	0	7	1	0	1	5	0	0	4	0	2	0	0	0	0	1	4	2	0	0	0	1	1	0	3	3	4	0	2	1	4	0,0%			
27b	0	3	0	1	1	1	0	0	1	0	0	3	0	0	6	3	1	0	0	0	0	1	0	0	0	0	0	0	3	1	0	1	1	0	3	3	0	9	10,4%		

The misclassification percentage for all sampling areas gives indications of major geographical groups but does not clearly identify the boundaries of these groups. Consequently, hypotheses based on the knowledge of the species biology, of the hydrodynamics and the results of preliminary LDA and clustering analyses were identified. The mainly hypothesis were compared to finalise the structure of the groups (Tab. 7).

Table 7 –Results of Linear Discriminant Analysis according to tested different hypotheses presenting the composition of subunits and the correct classification rate for the red mullet dataset.

Hypotheses	Subregions	GSA Composition	Number	Correct classification rate
1	western med	gsa 1-16	1000	37.56%
	central med	gsa 17-20	398	
	eastern med	gsa 22-27	447	
2	western med	gsa 1-14	828	35.61%
	central med	gsa 16-20	570	
	eastern med	gsa 22-27	447	
3	western med	gsa 1-12	828	33.22%
	adriatic sea	gsa 17-18	246	
	central & eastern med	gsa 14, 16 & 19-27	771	
4	western med	gsa 1-16	1000	35.66%
	adriatic sea	gsa 17-18	246	
	central & eastern med	gsa 19-27	599	
5	western med	gsa 1-16	1000	35.82%
	central med	gsa 17-19	343	
	eastern med	gsa 20-27	502	
6	western med	gsa 1>16	1000	35.77%
	central med	gsa 17>20 without 19b	349	
	central & eastern med	gsa 19b & 22>27	496	
7	western med	gsa 1-16 without 14	954	35.88%
	central med	gsa 17-20	398	
	central & eastern med	gsa 14 & 22-27	493	

The optimised classification (correct classification rate = 37.567%) presented 3 groups which are distributed as follows (Fig. 9):

- Western med (from GSA1 to GSA16)
- Central Mediterranean Sea (from GSA17 to GSA20)
- Eastern med (from GSA22 to GSA27)

The overlapped otolith shape of these 3 groups showed 3 difference areas particularly between the rostrum and antirostrum (Fig. 10; Tab. 8).

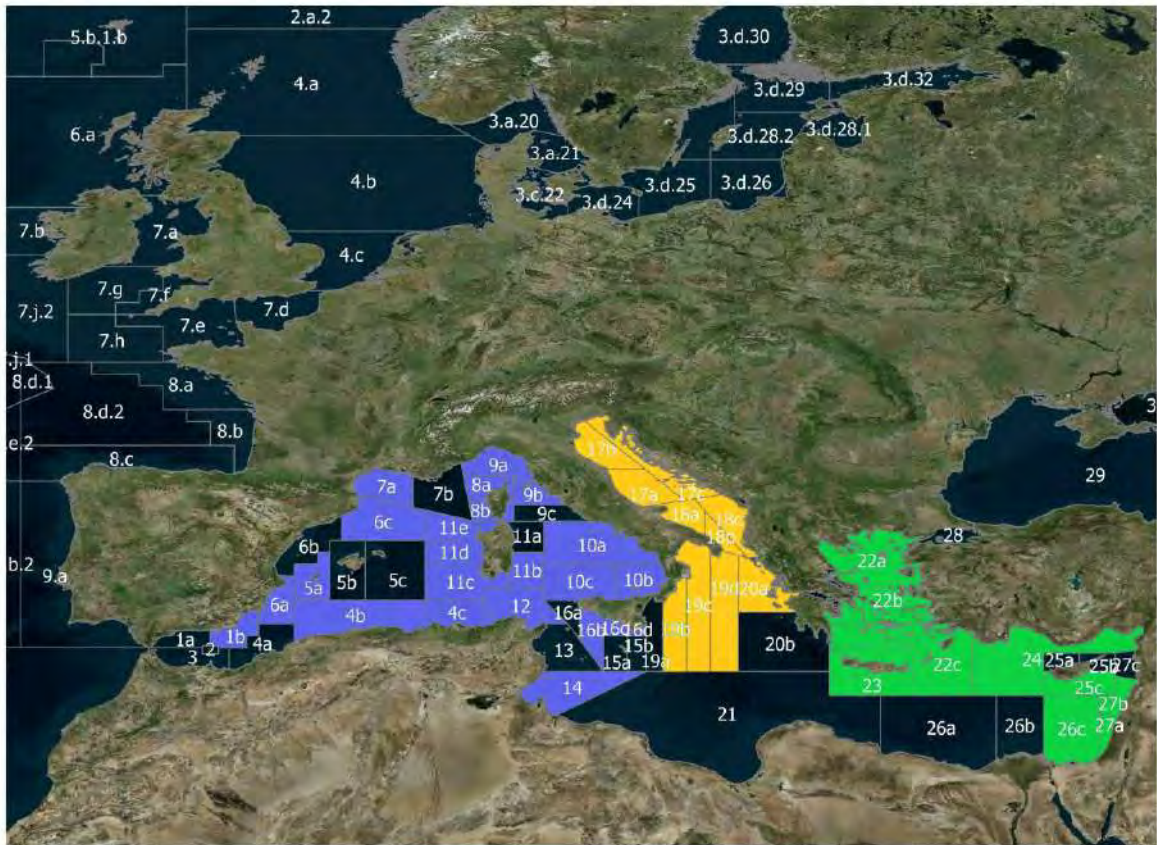
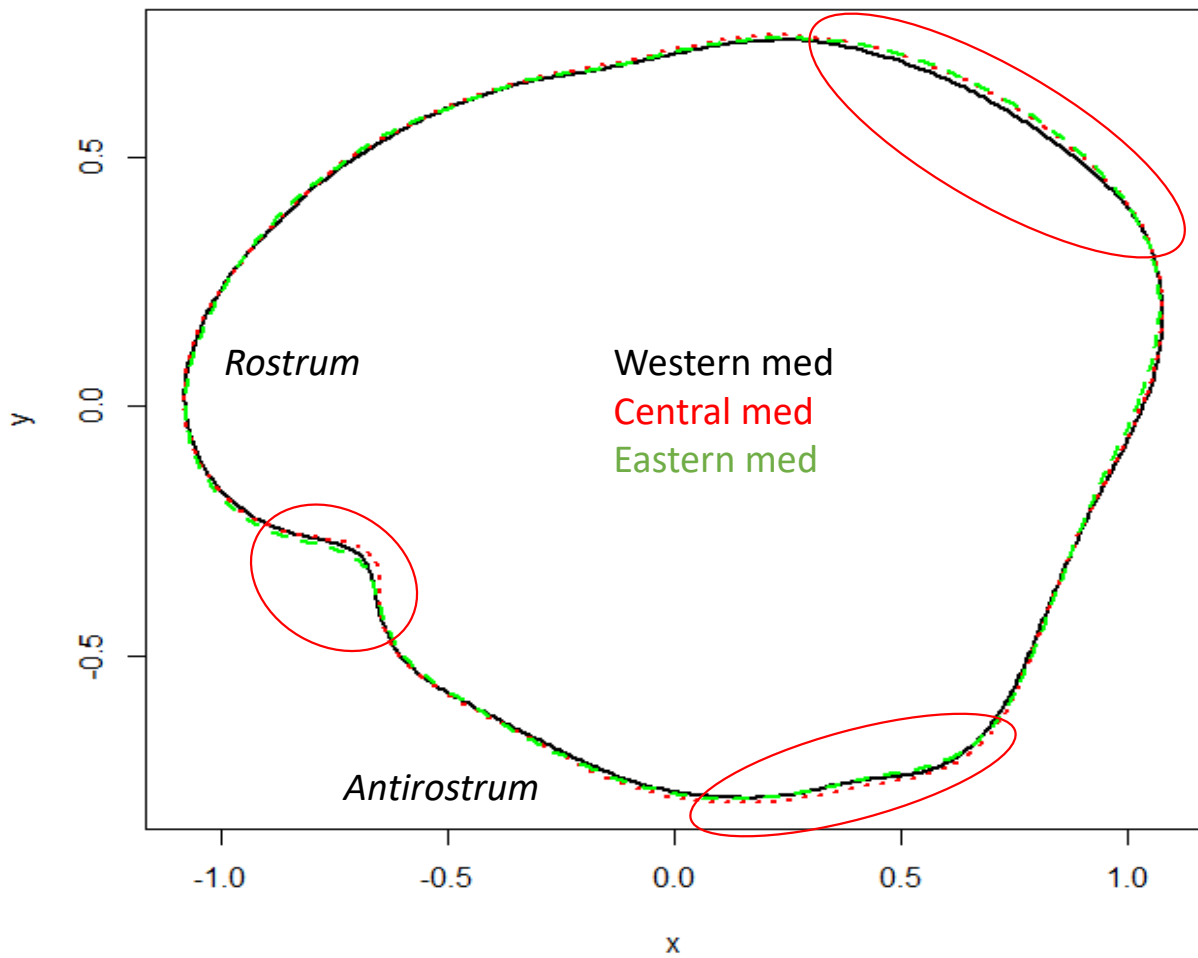


Figure 9 - Identification of boundaries of stocks for red mullet from otolith shape analysis.



	West-Med	Central-Med	East-Med
West-Med			
Central-Med	1,837		
East-Med	1,654	1,346	

Figure 10 - Overlapped otolith shape of the 3 identified groups for red mullet in the Mediterranean sea (red area showed the highest difference between groups) (A.) and the average percentages of non-overlapping surface between the two reconstructed otolith shapes between 3 identified groups (B.)

8.2 Identification of boundaries of stocks for hake

The results of the hierarchical clustering (Unsupervised Machine Learning) showed an optimized distribution of individuals in 4 clusters (Fig. 11). The distribution of individuals in the 4 clusters does not seem to indicate clearly geographical patterns (Tab. 8).

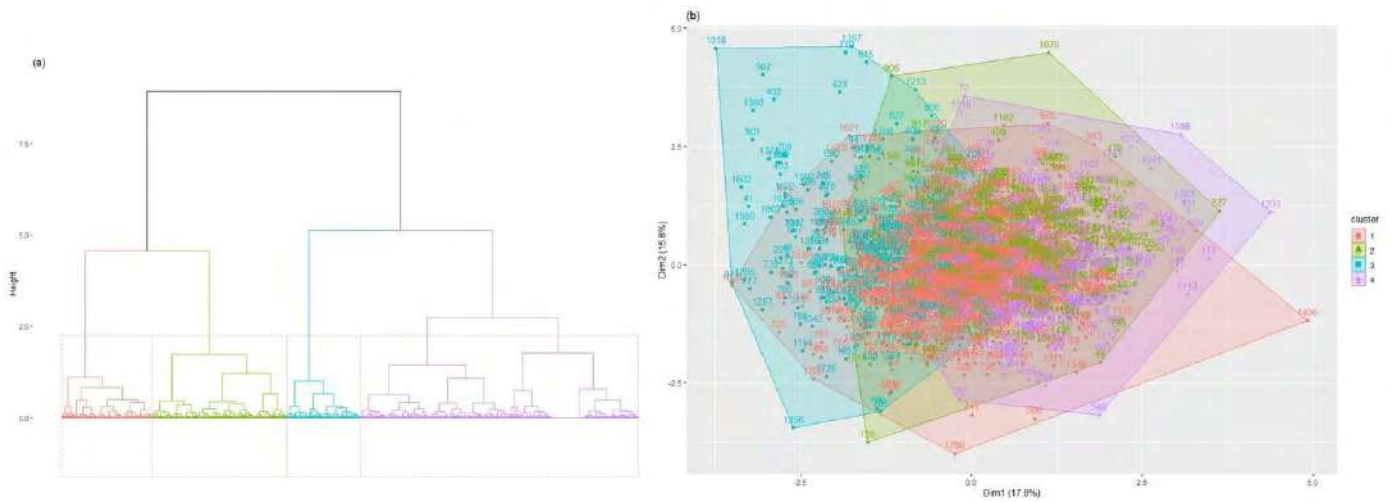


Figure 11 – Hierarchical clustering analysis applied on the otolith shape of the hake dataset.

Table 8 – Classification matrix of hake from each sampling location (by GSA : A. ; by Subunits of GSA : B) among the different clusters (grey case showing the cluster with the highest number of individuals).

A)

Area	Cluster 1	Cluster 2	Cluster 3	Cluster 4
IV	32	16	9	10
VIII	75	26	11	17
1	13	5	2	5
4	32	23	7	9
5	33	6	6	14
6	61	27	17	14
7	17	7	1	3
8	26	16	7	5
9	44	26	11	13
10	47	18	20	13
11	79	51	38	27
12	37	34	5	22
13	13	6	5	2
16	34	5	5	21
17	45	41	23	17
18	46	24	13	14
19	35	23	17	19
20	37	19	8	12
22	62	33	10	32
23	41	8	3	13
25	35	7	11	4
26	24	9	4	3
27	34	6	4	4

B)

Area	Cluster 1	Cluster 2	Cluster 3	Cluster 4
IVa	32	16	9	10
VIIIa	19	7	6	7
VIIIb	25	11	3	2
VIIIC	31	8	2	8
1b	13	5	2	5
4b	12	6	6	2
4c	20	17	1	7
5a	4	2	1	2
5b	29	4	5	12
6a	6	13	1	4
6b	22	8	8	7
6c	33	6	8	3
7a	17	7	1	3
8a	14	15	7	3
8b	12	1	0	2
9a	26	8	8	5
9b	18	18	3	8
10A	22	13	9	6
10BC	25	5	11	7
11b	22	12	12	4
11c	23	12	9	4
11d	15	12	12	11
11e	19	15	5	8
12a	37	34	5	22
13	13	6	5	2
16b	27	2	5	11
16d	7	3	0	10
17a	15	22	5	8
17b	13	6	6	7
17C	17	13	12	2
18a	13	14	12	8
18c	33	10	1	6
19b	14	18	5	7
19c	3	2	6	6
19d	18	3	6	6
20A	37	19	8	12
22A	20	20	5	12
22B	19	8	3	10
22c	23	5	2	10
23a	41	8	3	13
25c	35	7	11	4
26a	24	9	4	3
27b	34	6	4	4

The second analysis (LDA) is supervised to predict the position of individuals according to their origin geographical area and thus evaluate the percentage of correct classification (predicted area=actual area). Sampling area was used as an explanatory variable in the subsequent LDA. The overall jackknifed classification success was 9.26% and 7.01% at the geographical scale of the GSA (Tab. 9) and subunits of GSA (Tab. 10) respectively. The analysis showed significant differences

among groups of hake sampled in different areas of Mediterranean Sea (at GSA scale: Wilks'λ = 0.695; χ²=673.74; p=0.001 ; at Subunits of GSA scale : Wilks'λ = 0.5609; χ²=1065.0 ; p=0.001).

Table 9 – Jackknifed correct classification matrix of the linear function discriminant analysis for hake (N=1862) between GSA areas based on the selected Principal Components (PC) matrix. The percentages in each row represent the classification into the sampling area given in columns (correct classification in grey square). Overall classification success: 9.26%.

		Predicted area																				Correct classification			
		IV	VIII	1	4	5	6	7	8	9	10	11	12	13	16	17	18	19	20	22	23		25	26	27
Actual area	IV	6	0	2	7	9	2	2	1	0	3	3	3	1	7	1	2	2	0	0	4	5	1	6	9,0%
	VIII	36	0	2	8	10	3	4	4	0	2	1	4	5	11	1	1	5	2	2	3	10	11	4	0,0%
	1	0	0	1	0	1	2	3	0	0	2	1	1	2	1	0	2	4	0	0	1	1	0	3	4,0%
	4	0	1	1	11	1	1	1	9	0	6	1	3	3	3	0	4	2	0	1	3	6	10	4	15,5%
	5	1	0	2	0	15	2	6	2	0	2	0	0	3	5	0	1	0	0	1	1	9	4	5	25,4%
	6	0	1	7	3	15	8	5	10	1	0	4	5	10	15	0	2	12	0	2	1	11	4	3	6,7%
	7	12	0	0	0	0	0	3	0	0	0	2	0	0	1	0	2	4	0	0	3	1	0	0	10,7%
	8	10	0	2	2	1	1	1	8	0	3	0	4	3	5	0	0	4	1	0	2	4	3	0	14,8%
	9	3	0	4	9	5	3	8	2	0	3	5	3	10	10	1	6	5	0	0	1	4	7	5	0,0%
	10	0	1	4	4	8	2	1	5	0	9	7	1	5	3	1	3	3	1	0	4	10	18	8	9,2%
	11	2	0	11	10	11	7	9	5	0	10	14	7	14	13	1	6	25	1	6	6	6	21	10	7,2%
	12	7	0	8	5	3	4	11	5	0	1	3	3	0	5	0	4	13	0	2	1	4	12	7	3,1%
	13	0	0	1	0	2	2	1	5	0	1	0	1	4	0	0	0	1	0	0	0	0	4	4	15,4%
	16	0	0	2	4	6	1	1	2	0	4	1	1	0	16	0	4	6	2	3	1	6	3	2	24,6%
	17	1	0	10	11	9	2	9	11	0	8	5	3	4	10	1	8	3	0	2	0	5	15	9	0,8%
	18	1	1	2	8	4	5	8	2	0	3	3	5	6	10	0	3	8	0	2	3	10	9	4	3,1%
	19	0	0	6	1	2	3	8	6	0	4	3	3	3	6	0	3	23	0	0	1	5	2	15	24,5%
	20	2	0	7	2	2	0	3	4	0	2	2	2	2	10	0	4	11	1	1	2	6	3	10	1,3%
	22	5	0	5	4	8	1	14	6	0	6	6	2	4	16	0	4	7	1	3	11	10	14	10	2,2%
	23	7	2	5	1	9	1	2	0	0	2	1	2	3	2	0	1	3	0	2	7	0	6	9	10,8%
25	2	0	0	4	1	1	0	5	0	2	2	0	3	2	0	1	1	0	3	3	11	7	9	19,3%	
26	3	0	1	3	4	0	1	2	0	2	1	0	1	4	0	1	3	0	1	1	3	8	1	20,0%	
27	10	0	2	0	0	0	3	0	0	0	1	1	0	0	0	0	6	0	0	0	4	3	18	37,5%	

Table 10 – Jackknifed correct classification matrix of the linear function discriminant analysis for hake (N=1862) between subunits of GSAs based on the selected Principal Components (PC) matrix. The percentages in each row represent the classification into the sampling area given in columns (correct classification in grey square). Overall classification success: 7.01%.

		Predicted area																											Correct classification rate																		
		IVa	VIIIa	VIIIb	VIIIc	1b	4b	4c	5a	5b	6a	6b	6c	7a	8a	8ab	9a	9b	10A	10B	11b	11c	11d	11e	12	13	16b	16d		17a	17b	17C	18a	18c	19b	19c	20A	22A	22B	22c	23a	25c	26a	27b			
Actual area	IVa	6	0	0	0	0	1	1	3	4	7	0	1	0	1	0	1	0	2	0	2	0	1	5	0	1	1	1	1	1	4	1	2	0	0	5	0	0	1	0	4	4	0	6	9,0%		
	VIIIa	13	0	0	0	0	1	0	0	1	1	0	1	2	1	0	0	0	1	0	0	0	0	1	0	0	0	2	4	0	1	4	1	0	1	0	0	0	1	1	1	2	0	0	0,0%		
	VIIIb	9	0	1	0	0	1	0	2	2	1	4	0	0	2	0	0	1	1	0	0	0	2	0	0	1	1	1	2	2	0	0	1	0	1	1	0	0	1	0	0	2	1	2,4%			
	VIIIc	14	0	0	0	0	2	1	0	6	4	0	1	1	2	0	0	1	0	1	0	0	0	0	0	0	2	2	2	3	2	0	0	0	0	0	0	0	1	1	2	0	1	0,0%			
	1b	0	0	0	0	0	0	0	2	3	0	1	2	0	0	0	0	2	1	0	0	0	0	0	0	1	0	1	0	1	2	0	1	2	0	0	0	0	1	1	0	1	0,0%				
	4b	0	0	3	0	0	8	0	1	0	2	0	0	0	1	0	0	0	2	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	2	3	0	30,8%			
	4c	0	0	3	0	1	0	0	0	3	7	0	0	0	0	4	0	0	0	2	1	1	0	3	0	0	0	0	1	2	2	1	2	1	0	3	0	0	0	1	1	1	2	3	0,0%		
	5a	1	0	0	0	0	1	0	0	1	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0,0%		
	5b	0	0	0	0	0	3	0	1	10	3	2	1	2	1	1	0	0	4	1	0	0	0	0	0	0	0	0	2	0	4	0	0	1	0	0	0	0	1	2	1	4	1	5	20,0%		
	6a	0	0	0	0	1	0	1	1	6	0	0	2	0	0	0	1	1	0	0	1	1	0	0	0	1	0	0	0	1	1	1	1	0	1	1	2	0	1	0	0	0	0	0	25,0%		
	6b	0	0	0	0	0	0	1	1	1	1	4	4	0	1	0	1	1	0	5	0	0	1	4	0	2	2	5	0	2	2	0	1	1	1	2	0	0	2	1	1	1	1	1	2,2%		
	6c	0	0	2	0	0	0	1	2	8	3	0	0	0	0	0	0	1	2	0	0	0	1	2	0	0	1	0	3	3	1	2	2	2	0	0	0	7	0	0	1	0	0	4	0	1	0,0%
	7a	12	0	0	0	0	0	0	1	0	1	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	2	2	0	2	0	0	0	2	1	0	0	7,1%	
	8a	2	0	0	0	0	7	0	2	0	7	1	1	0	3	0	0	0	1	2	1	0	1	1	0	1	1	0	2	0	1	1	0	0	3	0	0	0	0	0	0	1	1	0	0	7,7%	
	8ab	8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0,0%	
	9a	3	0	0	0	2	1	1	2	2	1	0	2	0	0	0	0	1	1	1	0	0	0	3	0	3	1	3	2	2	1	2	1	2	2	0	0	0	0	3	0	1	1	3	0,0%		
	9b	0	0	0	0	5	0	5	3	0	0	0	2	1	0	0	0	5	1	1	0	0	5	0	0	0	7	2	1	0	1	1	1	0	2	0	1	0	2	0	0	0	0	1	1	1	0,0%
	10A	0	0	1	0	1	2	0	3	2	2	1	1	1	0	0	0	2	6	1	2	0	1	4	0	2	1	4	0	1	0	1	0	0	0	0	0	0	1	1	0	3	3	3	12,0%		
	10Bc	0	0	2	0	0	9	0	0	5	2	2	0	0	3	0	0	2	2	1	1	0	0	1	0	0	1	0	2	4	0	2	2	0	1	0	0	1	0	0	0	1	0	2	0	3	2,1%
	11b	0	0	0	1	0	1	0	2	1	4	1	0	1	0	0	0	2	0	1	3	1	3	0	0	0	2	1	4	4	1	0	1	0	1	0	1	0	1	2	1	0	1	7	3	2,0%	
	11c	1	0	0	0	3	2	5	1	3	2	1	0	0	0	0	4	1	4	0	2	0	3	1	2	1	1	2	1	1	0	1	0	1	1	0	1	0	1	0	3	0	0	2	1	0,0%	
	11d	0	0	3	1	2	0	0	2	1	3	2	1	0	0	0	0	1	1	0	0	2	0	0	0	0	0	3	0	1	2	2	0	10	1	7	0	0	0	2	1	0	1	1	0,0%		
	11e	0	0	1	0	0	0	0	0	1	2	2	0	1	1	0	0	3	0	2	0	1	7	0	1	1	3	0	1	2	0	0	5	0	2	0	0	0	1	2	2	2	4	14,9%			
	12	7	0	1	0	1	4	1	4	1	11	1	0	5	5	0	0	1	3	1	0	0	5	0	0	0	6	1	8	2	2	2	11	1	4	0	0	0	2	0	3	1	4	0,0%			
	13	0	0	0	0	0	3	0	3	1	1	2	0	1	2	0	0	1	1	0	0	1	0	0	1	0	0	1	0	2	1	0	0	1	0	2	0	0	0	0	0	0	0	0	3	0,0%	
	16b	0	0	0	0	0	0	3	3	0	0	1	2	0	0	0	0	1	0	0	1	0	0	1	0	0	0	3	6	2	0	0	1	2	1	5	1	0	1	2	0	5	1	1	6,7%		
	16d	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0	6	2	0	0	0	1	0	1	0	0	1	0	0	1	0	2	1	0	30,0%
	17a	1	0	0	0	3	1	0	3	6	2	0	1	3	0	0	1	0	2	0	1	2	0	0	0	3	6	2	0	4	3	1	0	0	1	0	0	0	0	0	0	0	3	1	4,0%		
	17b	0	0	1	0	2	2	0	0	5	1	0	2	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	2	2	0	2	0	1	0	1	0	0	0	0	0	2	0	4	6,3%		
	17C	0	0	0	0	1	4	0	1	0	3	3	2	2	2	1	0	0	6	1	0	0	1	2	0	1	0	2	1	0	1	1	1	0	0	1	0	0	0	1	0	1	0	1	0	4	2,3%
	18a	0	0	0	0	1	6	1	3	0	7	0	0	3	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	2	0	0	5	1	3	0	4	0	0	0	1	1	2	0	0	10,6%	
	18c	0	0	1	1	0	3	0	4	1	4	2	0	0	1	0	0	0	4	1	0	0	3	0	0	0	3	0	0	5	2	0	0	1	0	3	0	4	0	0	0	5	0	2	0	3	0,0%
	19b	0	0	0	0	0	0	0	2	1	3	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	0	0	0	0	0	19	0	3	0	0	0	0	1	0	7	43,2%	
	19c	0	0	0	0	0	1	0	0	0	3	1	0	1	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	2	0	0	0	0	0	1	0	2	0,0%	
	19d	0	0	0	0	2	3	0	0	0	1	1	0	1	0	1	0	0	2	0	0	0	1	0	0	0	0	2	0	3	0	0	4	0	8	0	0	0	0	2	1	1	24,2%				
20A	2	0	2	0	3	2	0	3	2	3	1	2	2	2	0	0	0	1	0	0	5	0	0	0	2	6	1	1	3	3	0	10	2	4	0	0	1	3	0	1	1	8	0,0%				
22A	0	0	1	1	0	3	0	3	2	3	0	1	6	3	0	0	0	1	0	0	4	0	0	4	0	0	2	9	1	3	2	1	0	2	0	0	0	0	0	1	3	1	4	0,0%			
22B	0	0	0	0	1	0	0	2	3	0	1	0	1	0	0	0	1	0	0	0	2	0	0	0	0	4	1	0	4	1	0	0	0	5	1	1	4	0	3	1	0	4	10,0%				
22c	5	0	1	2	1	3	0	1	1	2	0	0	1	1	0	0	0	3	0	0	0	0	1	0	0	0	6	1	0	1	0	0	0	2	0	0	1	1	2	2	1	1	2,5%				
23a	7	0	0	2	2	1	1	5	5	1	0	0	1	0	0	0	1	2	0	1	0	1	0	1	0	1	6	1	3	0	0	4	0	2	0	1	0	6	3	0	2	6	4,6%				
25c	2	0	0	0	0	7	0	0	1	1	1	0	1	1	0	0	0	3	1	1	0	0	1	0	0	0	3	0	2	1	0	1	0	0	3	0	0	0	4	3	8	3	9	14,0%			
26a	3	0	1	0	0	9</																																									

Table 11 –Results of Linear Discriminant Analysis according to tested different hypotheses presenting the composition of subunits and the correct classification rate for the hake dataset.

Hypothesis	Subunit	GSA Composition	Number	Correct classification rate
1	Atlantic	ICES IV-VIII	196	38.54%
	western med	GSA1-16	932	
	central med	GSA17-20	393	
	eastern med	GSA22-27	347	
2	Atlantic	ICES IV-VIII	196	39.61%
	western med	GSA1-13	867	
	central med	GSA16-20	458	
	eastern med	GSA22-27	347	
3	Atlantic	ICES IV-VIII	196	39.07%
	western med	GSA1-13	951	
	central med	GSA16-22	595	
	eastern med	GSA23-27	210	
4	Atlantic	ICES IV-VIII	196	38.00%
	western med	GSA1-16	932	
	central med	GSA17-22	530	
	eastern med	GSA23-27	210	
5	Atlantic	ICES IV-VIII	196	35.12%
	western med	GSA1-12	841	
	central med	GSA13-20	484	
	eastern med	GSA22-27	347	
6	Atlantic	ICES IV-VIII	196	33.46%
	western med	GSA1-12	841	
	central med	GSA13-22	621	
	eastern med	GSA23-27	210	

The optimised classification (correct classification rate = 39.61%) presented 4 groups which are distributed as follows (Fig. 12):

- Atlantic ocean (from ICES IV to ICES VIII)
- Western med (from GSA1 to GSA13)
- Central med (from GSA16 to GSA20)
- Eastern med (from GSA22 to GSA27)

The overlapped otolith shape of these 4 groups showed that the highest difference was between Atlantic ocean and the Mediterranean sea. Moreover, the average otolith shape among 4 groups

showed the evaluation gradient from the west with Atlantic ocean to east with Eastern Mediterranean sea (Fig. 13).

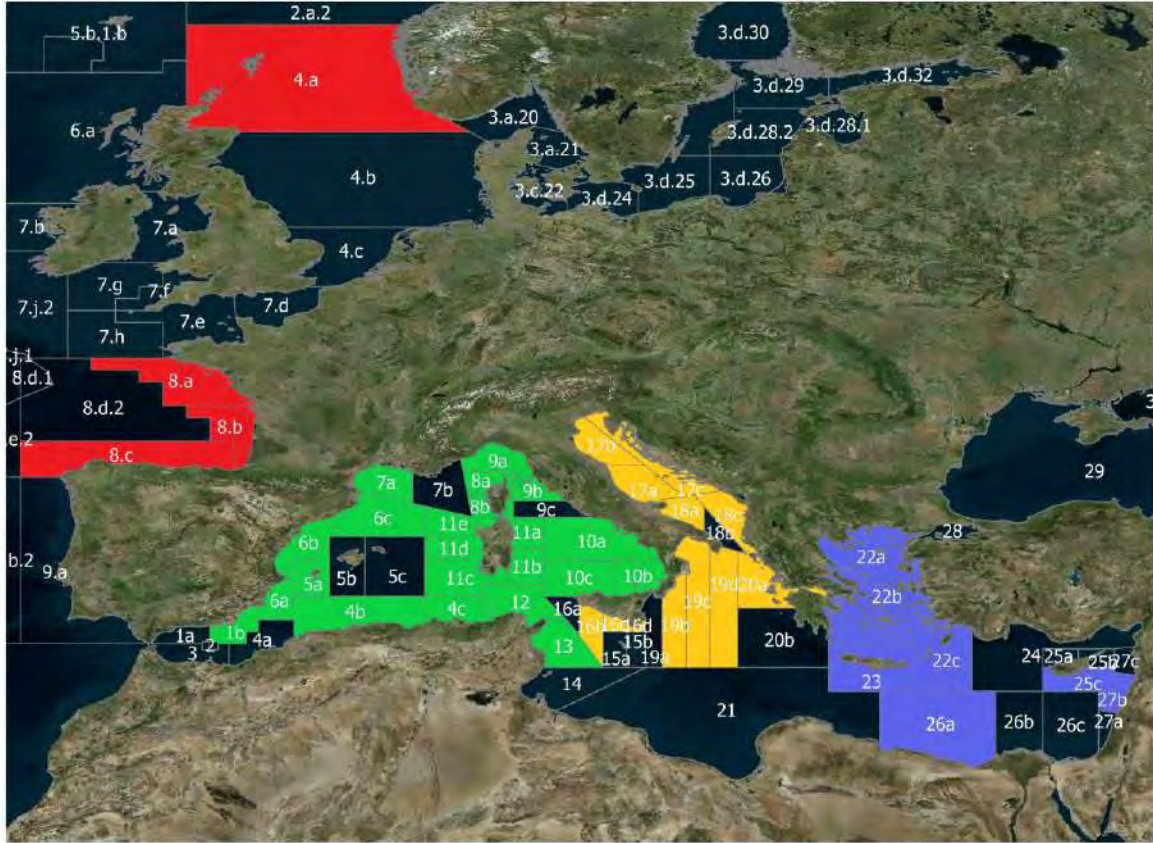
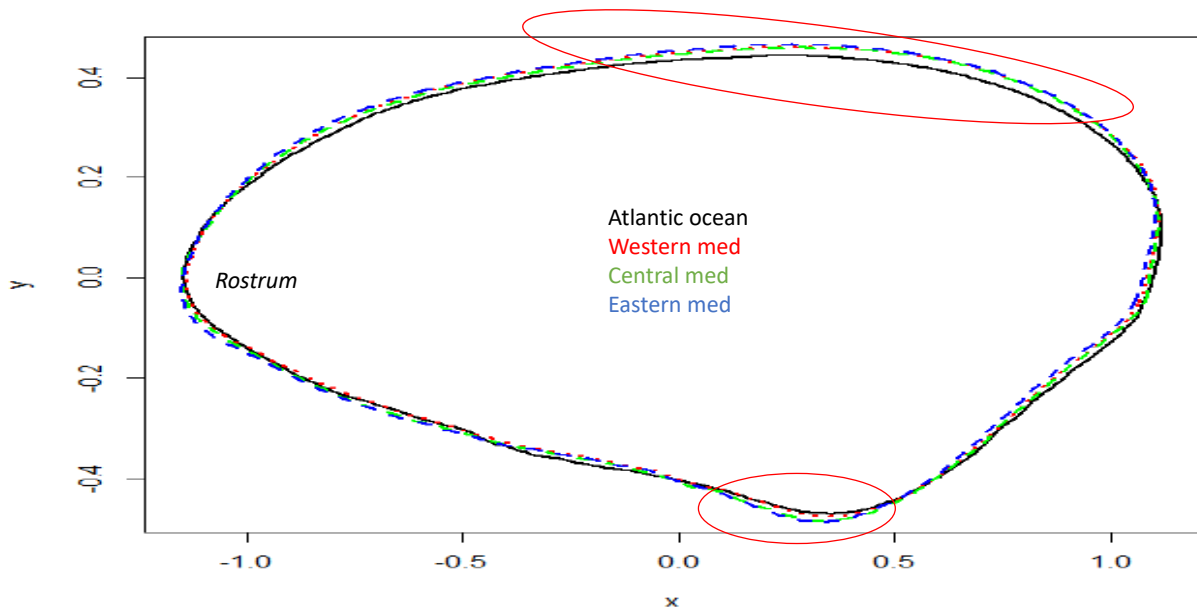


Figure 12- Identification of boundaries of stocks for hake from otolith shape analysis.



	Atlantic	GSA1-13	GSA16-20	GSA22-27
Atlantic				
GSA1-13	3, 47			
GSA16-20	3, 79	1, 29		
GSA22-27	5, 14	2, 00	2, 11	

Figure 13 - Overlapped otolith shape of the 4 identified groups for hake in the Mediterranean sea and Atlantic ocean (A. red area showed the highest difference between groups) and the average percentages of non-overlapping surface between the two reconstructed otolith shapes between 3 identified groups (B.).

9. New approach: 3D analysis of otolith shape

During the Med-Units project, we have started to develop new techniques for the analysis of otolith shape variation. These methods are coming from 3D techniques applied to neuro-imaging analysis and adapted to other biological objects. They are based on Spherical Harmonics (3D Fourier). Similarly to the well-known 2D outlines analysis techniques, they consist on the characterisation on the surface of the biological objects. Variables of size, on one side, and shape, on the other side, are extracted and for the analysis of variation between specimens, or populations.

Initially developed for Matlab, the Spherical Harmonics methods were translated and adapted in order to be employed under R. They are still under development but we conducted a preliminary study to test their reliability. We identified some points to solve and some solutions that could be useful for a detailed understanding of otoliths shape and size variation.

We conducted a preliminary study on specimens of some populations of red mullet. The otolith morphology is characterised with the help of a μ CT scanner and reconstructed using the Spherical Harmonics functions. 3D reconstructions are accurate models of the otolith shape (Fig. 14). Thus, the coefficients of the reconstructed models are used as variables for multivariate statistical analysis, such as Principle Component Analyses. This allows us to quantify shape variation between populations, but also visualise shape variation along PCs (Fig. 15). Other statistical methods, such as clustering methods, can be used on shape variables. Moreover, variation in size can be analysed independently from shape, but also in association to shape to identify patterns of allometries for each population.

This new method for the analysis of otolith size and shape variation in 3D constitutes a high methodological advance, but also new insights in the characterisation and analysis of otoliths.

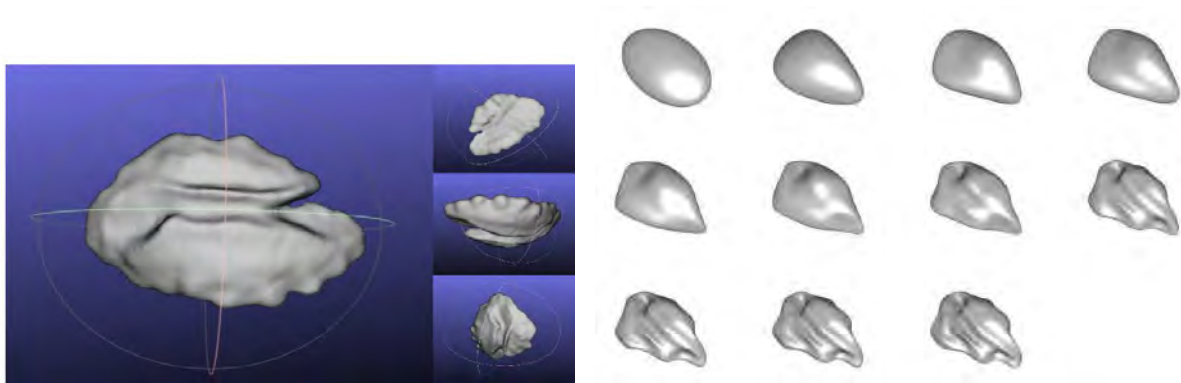


Figure 14: Left: 3D characterisation of otolith shape. Right: This shape is reconstructed by increasing the degree of Spherical Harmonics analysis from a simple ellipsoid (first degree) to an accurate model (last degree).

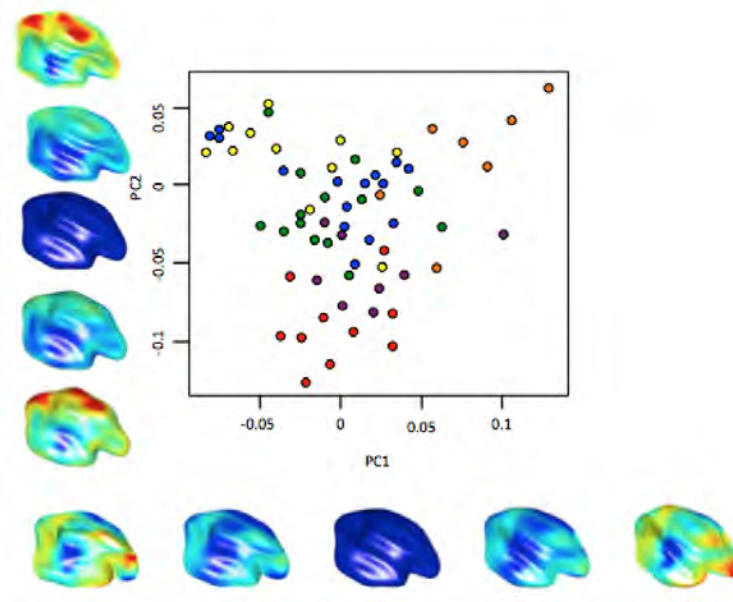


Figure 15: Principle Component Analysis of otolith shape coefficient from Spherical Harmonics method. This preliminary study is conducted on specimens of some population of red mullet, identified by coloured dots. Shape variation along PCs axes are visualised in 3D.

The first 3D analyses of stock discrimination are ongoing for red mullet. However, the first trend from the 3D data correlated the results from the 2D data.

10. References

Campana, S. E., and Casselman, J. M. (1993). Stock Discrimination Using Otolith Shape Analysis.

Canadian Journal of Fisheries and Aquatic Sciences, 50(5), 1062–1083.

Campana, S.E., 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and

applications. *Mar. Ecol. Prog. Ser.* 188, 263–297.

Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W. 2009. Genetic population structure of

marine fish: mismatch between biological and fisheries management units. *Fish and*

Fisheries, 10: 361–395



Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-
Project MED_UNITS

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP2- OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

D.2.3. Protocol for analysis and data treatment of trace elements in otoliths of European hake and red mullet

Compiled by: Sílvia Pérez Mayol (IMEDEA), Beatriz Morales-Nin (IMEDEA) and Marco Stagioni (CoNISMa)

January 2020

INDEX

EXECUTIVE SUMMARY.....	3
LIST OF ACRONYMS.....	4
1. INTRODUCTION.....	6
2. EUROPEAN HAKE (<i>Merluccius merluccius</i>)	6
2.1. Samples analysed	6
2.2. Sample pre-processing.....	6
Initial cleaning.....	6
Sectioning and mounting otoliths	7
Mapping and second cleaning.....	7
2.3. LA-ICPMS analysis	8
LA-ICPMS running conditions.....	8
Sample analytical procedure.....	8
Elements analysed.....	9
2.4. Treatment of data: from cps to metal concentrations.....	9
2.5. Matrix of results and repositories	10
3. RED MULLET (<i>Mullus barbatus</i>)	11
3.1. Samples analysed	11
3.2. Sample pre-processing.....	11
Initial cleaning.....	11
Sectioning and mounting otoliths	11
Mapping and second cleaning.....	12
3.3. LA-ICPMS analysis	12
LA-ICPMS running conditions.....	12
Sample analytical procedure.....	13
Elements analysed.....	13
3.4. Treatment of data: from cps to metal concentrations.....	14
3.5. Matrix of results and repositories	14
4. REFERENCES	14

EXECUTIVE SUMMARY

The Deliverable D2.3. Defines the protocols to be followed for otolith microchemistry analysis of both *Merluccius merluccius* and *Mullus barbatus*.

The protocols are based on previous studies done in IMEDEA (UIB/CSIC) facilities with European hake otoliths and in pilot studies specifically developed for red mullet otoliths in CoNISMa.

For both species, the general protocol includes the procedures for the following major steps: i) otolith sample pre-processing, ii) running LA-IPCMS analysis, iii) calibration and treatment of data, and iv) structure of the matrix of results and repositories.

The sample pre-processing includes an initial cleaning and mounting of 25 left otolith per site following clean methods, sagittal grinding until core exposure and polishing to mirror appearance. Finally, otolith sections are transferred and glued to petrographic glass slides in random order to avoid analytical bias. Otoliths are then photographed and mapped to decide laser shot location, and then subjected to a final cleaning.

The LA-ICPMS analyses are defined according to the species and equipment availability. In the case of European hake, analyses will be performed using a CETAC Laser Ablation System LSX-213 G2+ coupled to a Thermo-Finnigan ICPMS Element XR; linear scans of 30 µm diameter and ~200 µm length will be performed on the core and edge of each otolith. For red mullet a GEOLAS 102 Laser Ablation System coupled to a triple quadrupole 8900 Agilent will be used; ablation spots of 50-80 µm diameter will be shot on the core and edge of the otoliths. In both cases, several Certified Reference Materials (CRM) will be used following the “bracketing protocol”. The list of isotopes analyzed differs depending on the species and according to previous studies, in both cases, Ca will be used as internal standard (IS).

To obtain final concentrations of the isotopes analyzed, data treatment will entail several steps such as LA-ICPMS signal review, data transformation and reduction, machine drift correction, IS normalization, and concentration quantification using the different CRMs. These common steps will be conducted following specific methods depending on the laboratory.

For both species, matrices of results, include isotope concentration related to Ca by distance to core, fish age, sampling location and sampling date, and values for accuracies and precisions of the analyses, will be produced. Raw data will be available in institutional repositories.

LIST OF ACRONYMS

Ar	Argon
B	Boron
Ba	Barium
bp	bais pairs
Ca	Calcium
CoNISMa	Consorzio Nazionale Interuniversitario per le Science del Mare
CSIC	Spanish Research Council
cm	centimetre
cps	counts-per-second
CRM	Certified Reference Material
Co	Cobalt
Cu	Copper
Fe	Iron
G	gram
GSA	geographical subarea
He	Helium
Hz	hertz
IGG	Istituto di Geoscienze e Georisorse
IMEDEA	Mediterranean Institute for Advanced Studies
IS	internal standard
J	joule
K	Potassium
LA-ICPMS	laser ablation inductively coupled plasma mass spectrometry
Li	Lithium
LPM	litres per minute
LSTM	long short-term memory
MFC	mass flow controller
Mg	Magnesium
Mn	Manganese
Mpix	megapixel
µg	microgram
µm	micrometre
Na	Sodium
Ni	Nickel
nm	nanometre
ns	nanosecond
P	Phosphorus
Pb	Lead
pg	picograms
ppm	parts per million
Rb	Rubidium
s	second
Sc	Scandium

Si	Silicon
Sn	Tin
Sr	Strontium
U	Uranium
UIB	University of the Balearic Islands
Zn	Zinc

1. INTRODUCTION

Deliverable 2.3 was due at month 13, albeit the analysis of the samples had not yet started (due at month 16). The protocol developed as D.2.3 is based on previous work on European hake carried out at IMEDEA (UIB/CSIC) and already published (see references in D.2.1) and on pilot trials on red mullet carried out at CoNISMa. Therefore, some minor modifications might be necessary during the development of the analyses.

The otoliths sample collections were optimized in coordination with WP0 and WP1. According to Milestone 2.1, the otolith microchemistry study would use the otoliths of *Merluccius merluccius* and *Mullus barbatus* collected for the otolith shape analysis. For hake, they would be juvenile fish (females <28 cm TL and males <16 cm TL); in the case of red mullet, adult fish of both sexes (TL >12 cm). For microchemistry, only females of both species will be used. A subsample of otoliths (25 otoliths per target species) from 8 geographical areas will be analysed.

2. EUROPEAN HAKE (*Merluccius merluccius*)

2.1. Samples analysed

According to the genetic study and the otolith availability, in the plenary meeting on November 2019, it was agreed to include 8 GSA for the otolith analysis. To the date, the selected GSA's are: 6b, 9c, 16b, and 18a. Four more have to be determined to try to cover the whole Mediterranean area. Moreover, two more sites would be included, if otolith samples are available, from the Bay of Biscay and the Black Sea that will be used to compare the results with those produced in the Mediterranean. The number of otoliths for each area has been increased to 30 in order to have replacements in case some otoliths are broken during the preparatory process but a total of 25 will be effectively analysed.

2.2. Sample pre-processing

Initial cleaning

According to protocol explained on Milestone 2.1, after their extraction from fish's head using TPFE tweezers, the sagittal otoliths were cleaned with MilliQ water. The organic tissue from the otic capsule was removed with H₂O₂ 5% suprapur for 2-3 min and afterward rinsed with MilliQ water. Once dried the otoliths were stored in clean plastic vials.

Sectioning and mounting otoliths

Hake otoliths are flat structures that generally are analysed using sagittal sections across the otolith core. For microchemistry purposes, left *sagitta* otoliths will be used, but right ones can be also used if any left is broken. To obtain a sagittal section the otolith will be glued sulcus up with Crystalbond to a cover glass and then glued to a glass slide using a small amount of Crystalbond. When dry, the otolith will be manually ground with silicon carbide lapping papers of decreasing grain thickness (FEPA P1200, P2500 and P4000) using MilliQ water as lubricant till the section reaches the otolith core. Finally, the sections will be manually polished with decreasing grain size diamond suspensions (3 µm and 1µm) till the surface will have the mirror appearance needed to avoid laser interferences.

Once cleaned with MilliQ water and dried, the sections will be released and the cover glass with the section on it will be transferred to a petrographic glass slide and glued it in place with Loctite in random order (Figure 1).

Mapping and second cleaning

Images of each otolith and glass slide will be taken with a Stereomicroscope Leica M165C coupled to an AVT Marlin 10 Mpix digital camera. Using these images, the positions of each laser scan will be determined according to otolith growth increments.

Every glass slide will be sonicated in a MilliQ bath for 1 min, submerged in HNO₃ 2% suprapur for 15sec, triple rinsed in MilliQ water and finally sonicated again in a clean MilliQ water bath for 1min. Once dried in a laminar flow hood, each glass slide will be stored in double zip plastic bags.

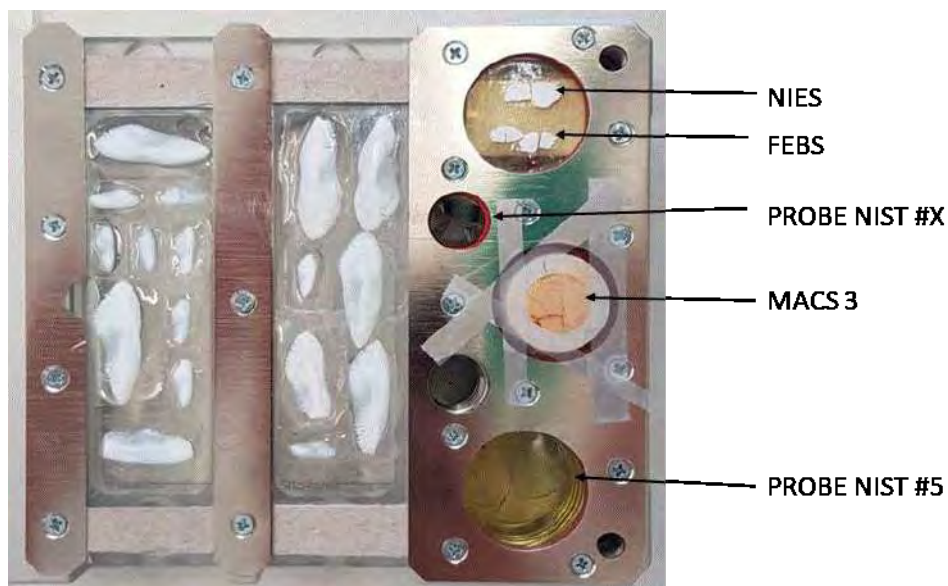


Figure 1. Two petrographic slides with hake otolith sections and reference materials placed on the laser chamber.

2.3. LA-ICPMS analysis

LA-ICPMS running conditions

The laser-ICPMS analysis of hake otoliths will be carried out at the Plasma-Mass Unit of the *Universidade de A Coruña*. A CETAC Laser Ablation System LSX-213 G2+ (Table 1) coupled to a Thermo-Finnigan ICPMS Element XR will be used. The laser will be operated using a laser beam of 30 μm in diameter with a laser output energy at about 30% ($\sim 3.30 \text{ J}\cdot\text{cm}^{-2}$). Ablation occurs in an atmosphere of pure He to minimise condensation of ablated material and the ablated material is transported to the ICPMS mixed with Ar carrier gas. The ICPMS will measure in Medium Resolution mode to avoid mass interferences. NIST 610 (National Institute of Standards and Technology) standard will be fired to tune the ICPMS trying to reduce oxides and increase the intensities as much as possible.

Table 1. CETAC Laser Ablation System Characteristics

- Q-Switched laser Nd:YAG 213 nm
- > 4mJ pulse energy and < 5 ns pulse amplitude
- variable energy exit and variable repetition frequency of the pulse (1 – 20 Hz)
- diameter patterns between 4 and 200 μm
- Chamber “HelEx II Active 2-volume”
- He and Ar parameters are automatically set to: MFC1 = 0.670 LPM and MFC2 = 0.400 LPM

Sample analytical procedure

Ablation line scans of $\sim 200 \mu\text{m}$ length and 30 μm width will be performed across the core (natal signature) and parallel to the otolith edge (fish capture location signature) on otolith ventral axis (Figure 2). The lines will be ablated with a scan speed of 3-10 $\mu\text{m}\cdot\text{s}^{-1}$ (to be defined according to the ICPMS method of acquisition).

A set of Certified Reference Materials (CRM) of known composition (Table 2) will be used in the analysis using the same laser conditions as in otoliths: NIST612, NIST614, NIST616 (National Institute of Standards and Technology), MACS-3 (International Association of Geoanalysts), FEBS-1 (National Research Council Canada, (Sturgeon et al. 2005)) and NIES-22 (National Institute for Environmental Studies, (Yoshinaga et al. 2000)). The CRMs will be analysed twice using the “bracketing protocol” for which all the CRMs, except MACS-3, are analysed at the beginning and the end of a working session (day) and every 20 line scans on otoliths. MACS-3 will be analysed only at the beginning and end of the working session.

Table 2. Certified Reference Materials.

- NIST 612: 5-100 ppm crystal
- NIST 614: 0.5-50 ppm crystal
- NIST 616: 0.008-30 ppm crystal
- MACS-3: pressed powder of calcium carbonate
- FEBS-1 and NIES-22: pressed powder of otoliths (homemade)

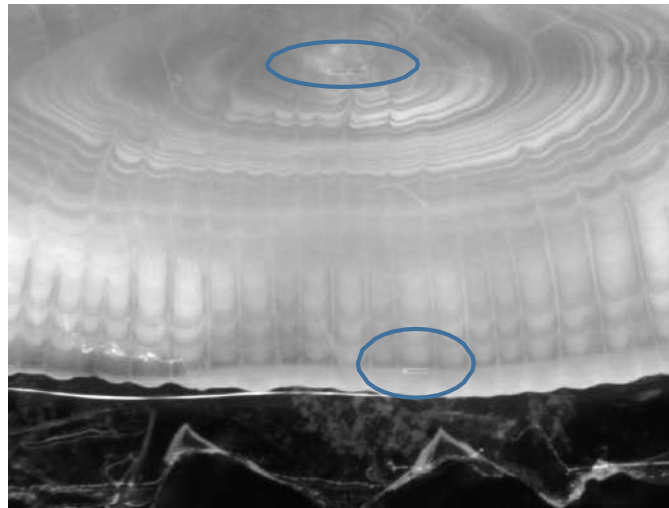


Figure 2. Sagittal section of an European hake otolith showing the ablated lines in the core and the edge.

Elements analysed

A modification of (Tiepolo et al. 2003) will be followed to reduce the acquisition time, by increasing the number of ICPMS runs and decreasing the mass window. This method allows to measure a wide variety of elements: lithium (^6Li and ^7Li), sodium (^{23}Na), magnesium (^{24}Mg and ^{25}Mg), silicon (^{29}Si), phosphorus (^{31}P), potassium (^{39}K), calcium (^{42}Ca , ^{43}Ca –used as internal standard- and ^{44}Ca), manganese (^{55}Mn), iron (^{56}Fe), cobalt (^{59}Co), nickel (^{60}Ni), copper (^{63}Cu), zinc (^{66}Zn), strontium (^{86}Sr and ^{88}Sr), rubidium (^{85}Rb), barium (^{137}Ba and ^{138}Ba), and lead (^{206}Pb , ^{207}Pb and ^{208}Pb). The isotopes to be included in the final statistical analysis will be determined based on the quality of the data, their accuracy and precision.

2.4. Treatment of data: from cps to metal concentrations

The raw intensities (counts-per-second, cps) obtained from the LA-ICPMS system have to be transformed into metal concentrations (ppm). Following IMEDEA protocols, this procedure requires several steps: *i)* the selection of “blank” and “plateau” intervals of every LA-shot, *ii)* data transformation and reduction, *iii)* the machine drift correction, *iv)* the normalization by ^{43}Ca as Internal Standard, and *v)* the quantification of estimated element concentrations on CRMs and otoliths using Bayesian analyses.

This entails two different stages, the first one is conducted using a Long Short-Term Memory (LSTM) Network built in MatLab, and developed by the *Data Processing Group of the University of Vic*, and the second by running an automatic R script developed by the *Fish Ecology Lab in IMEDEA*. See Martí-Puig et al. (in press) and Palmer et al. (in prep) for details.

The network was trained using previous supervised shots so that it is able to automatically select the “blank” and “plateau” intervals of the raw data for one specific LA-shot. In the second step, a hierarchical model developed in R is automatically applied in a supervised way to estimate Element concentrations (ppm, $\mu\text{g}_E/\text{g}_{\text{sample}}$) and their uncertainties from the LSTM network output. Several external standards are sampled in a bracketing way along with a given session to be able to estimate the concentrations. The concentrations of the standards used correspond to the certified concentrations (given for the whole element) and corrected by their naturally occurring isotope abundances. Besides, measured cps along a given session are often affected by a (usually linear) temporal instrumental drift. The internal standard used corresponds to a major element of the sample and it is assumed to be homogeneously distributed in it; the internal standard normalization is used to correct for matrix effects. To normalise, the difference plateau-blank for any given element is normalized by the difference plateau-blank for ^{43}Ca .

For a given session and a given element, the parameters of the model are estimated using a Bayesian approach. The parameterized model is then used for estimating all the concentrations of each LA-shot firstly on CRMs and then on otoliths. The uncertainties of the estimates for any concentration properly propagates all the uncertainty levels considered. The Bayesian approach also includes several quality controls with the aim to supervise how the estimations of the concentrations have been produced, being rejected if their quality is poor.

2.5. Matrix of results and repositories

A matrix with the obtained concentrations for each LA-shot on otoliths and MACS-3 will be produced. MACS-3 will be used as quality control at two different levels. For any given isotope, at the shot level MACS-3 will be used to calculate shot recovery (as an estimation of shot accuracy); at the session level, MACS-3 results will be used to calculate session accuracy and precision. Otolith LA-scans (in core and edge) will then be matched to otolith age, location and sampling date. Therefore, for each sampled location we will have a matrix with 50 rows (25 for the core and 25 for the edge) and columns depending on the isotopes with good response.

As requested in the contract, the raw data obtained from the LA-CIPMS system will be deposited in the CSIC institutional repository **digital csic**.

3. RED MULLET (*Mullus barbatus*)

3.1. Samples analysed

According to the genetic study and the otolith availability, in the plenary meeting on November 2019, it was agreed to include 8 GSA for the otolith analysis. To the date, the selected are: GSA5, GSA9, GSA17 (two sites), GSA18 (two sites) (as an alternative one site can be replaced by GSA19), GSA22, GSA25. However, these sites need to be verified according to the genetic studies. Moreover, two more sites may be included, if available, from outside Mediterranean sea to compare like outgroup. The number of otoliths for each area has been increased to 30 in order to have replacements in case some otoliths are broken during the preparatory process but a total of 25 will be effectively analysed. Left otoliths will be preferred but if not available will be replaced with right one.

3.2. Sample pre-processing

Initial cleaning

According to protocol explained on Milestone 2.1, after their extraction from fish's head using plastic tweezers, the sagittal otoliths were cleaned with MilliQ water. The organic tissue from the otic capsule was removed with H₂O₂ 5% suprapur for 2-3 min and afterward rinsed with MilliQ water. Once dried the otoliths were stored in clean plastic vials. If necessary cleaning procedure will be repeated to ensure the best fix of otoliths for the next steps.

Sectioning and mounting otoliths

Due to the small size of the otoliths these will be mounted and processed on the sagittal plane to exploit the maximum available surface. To obtain a sagittal section the otolith will be glued to a cover glass strip and then glued to a glass slide using a small amount solvent free high viscosity cyanoacrylate glue (superglue). When dry (about after 24 hours), the otolith will be manually ground with silicon carbide dry lapping papers of decreasing grain thickness (P2500 and P4000) till the section reaches the otolith core. Finally, the sections will be manually polished with decreasing grain size alumine lapping plastic sheets (3 µm, 1 µm and 0.1 µm) till the surface will have the mirror appearance needed to avoid laser interferences. Between each steps sample and paper will be air blown to minimize cross-contamination and lengthen the efficiency of the abrasives. Finally the sections will be released and the cover glass with the section on it will be transferred to an half glass slide and glued it in place with Loctite the same previous superglue in random order (Figure 3).

Mapping and second cleaning

Images of each otolith and glass slide will be taken with a Stereomicroscope Zeiss Stemi 2000-C coupled to an Moticam 1080 10 Mpix digital camera. Using these images, the positions of each laser spot will be determined according to otolith features.



Figure 3. Glass slide with multiple samples to be placed on the laser chamber.

3.3. LA-ICPMS analysis

LA-ICPMS running conditions

The laser-ICPMS analysis of red mullet otoliths will be carried out at the IGG (Istituto di Geoscienze e Georisorse) CNR Pavia. A triple quadrupole 8900 Agilent (Table 3) will be used. The laser will be operated using a laser beam of 50-80 μm in diameter (will be determined by preliminary test on definitive samples) with a laser output energy at about 4-8 $\text{J}\cdot\text{cm}^{-2}$. Ablation occurs in an atmosphere of pure He to minimise condensation of ablated material and the ablated material is transported to the ICPMS mixed with Ar carrier gas.

Table 3. Laser Ablation System Characteristics

- GEOLAS 102 laser 193 nm
- 3 mJ pulse energy
- frequency of the pulse (10 Hz)
- diameter patterns between 5 and 120 μm
- He and Ar parameters are automatically set near to 0.45 LPM and 0.84 LPM

Sample analytical procedure

Ablation spots of ~50-80 µm will be performed across the core (natal signature – pelagic phase) and near to the otolith edge (fish capture location signature) (Figure 4). A set of Certified Reference Materials (CRM) of known composition will be used in the analysis using the same laser conditions as in otoliths: NIST612 (National Institute of Standards and Technology), BCR-2 (USGS Geochemical Reference Materials) (Table 4). The CRMs will be analysed twice using the “bracketing protocol” for which NIST612 is analysed at the beginning, at the end of a working session and every 20 spots on otoliths. BCR-2 will be analysed only at the beginning and end of the working session.

Table 4. Certified Reference Materials

- NIST612
- BCR-2

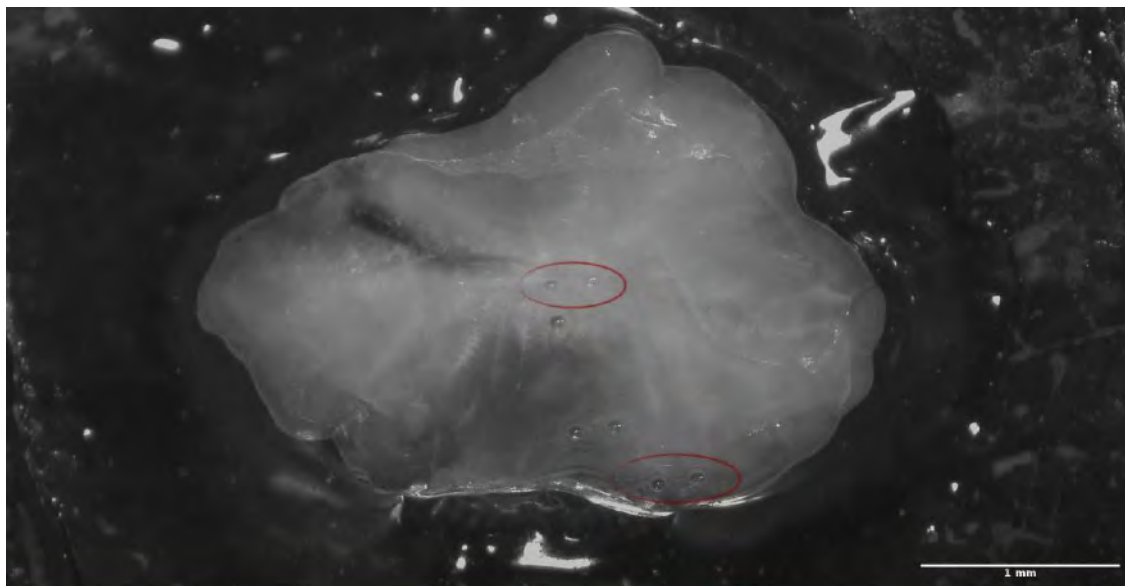


Figure 4. Sagittal section of a red mullet otolith showing the ablated spots in the core and the edge.

Elements analysed

From testing samples previously done, a preliminary list of isotopes is provided: ^7Li , ^{11}B , ^{23}Na , ^{25}Mg , ^{29}Si , ^{31}P , ^{39}K , ^{43}Ca , ^{44}Ca , ^{45}Sc , ^{55}Mn , ^{57}Fe , ^{59}Co , ^{60}Ni , ^{63}Cu , ^{66}Zn , ^{85}Rb , ^{88}Sr , ^{118}Sn , ^{138}Ba , ^{208}Pb , ^{238}U . The number and type of isotopes to be included in the final statistical analysis will be determined based on the quality of the data, their accuracy and precision.

3.4. Treatment of data: from cps to metal concentrations

The raw intensities (counts-per-second, cps) obtained from the LA-ICPMS system have to be transformed into metal concentrations (ppm). Following IGG protocols, this procedure requires several steps: *i)* the selection of “blank” and “plateau” signals of every LA-shot, *ii)* data transformation and reduction, *iii)* the machine drift correction, *iv)* the normalization by ^{44}Ca (or ^{43}Ca) as Internal Standard, *v)* and the quantification of estimated element concentrations on CRMs and otoliths using supervised spectra analysis on Glitter software environment.

3.5. Matrix of results and repositories

A matrix with the obtained concentrations for each LA-shot on otoliths (values mediated on the two spots of the same region) and CRMs will be produced. Otolith LA-spot (in core and edge) will then be matched to otolith age, location and sampling date. Therefore, for each sampled location we will have a matrix with 50 rows (25 for the core and 25 for the edge) and columns depending on the isotopes with good response.

As requested in the contract, the raw data obtained from the LA-ICPMS system will be deposited in an institutional repository.

4. REFERENCES

- Martí-Puig, P.; Pérez-Mayol, S.; Serra-Serra, M.; Palmer, M. (In press). Automatic interval selection on LA-ICPMS isotopic profiles using an Extreme Learning Machine (ELM) network. *Frontiers in Artificial Intelligence and Applications*, Ed. IOS Press.
- Sturgeon, R.E., Willie, S.N., Yang, L., Greenberg, R., Spatz, R.O., Chen, Z., Scriver, C., Clancy, V., Lam, J.W., and Thorrold, S. 2005. Certification of a fish otolith reference material in support of quality assurance for trace element analysis. *J. Anal. At. Spectrom.* **20**(10): 1067–1071.
- Tiepolo, M., Bottazzi, P., Palenzona, M., and Vannucci, R. 2003. A laser probe coupled with ICP - double-focussing sector-field mass spectrometer for in situ analysis of geological samples and U-Pb dating of zircon. *Can. Mineral.* **41**(2): 259–272.
- Yoshinaga, J., Nakama, A., Morita, M., and Edmonds, J.S. 2000. Fish otolith reference material for quality assurance of chemical analyses. *Mar. Chem.* **69**(1–2): 91–97.

The logo for MED_UNITS, featuring the text "MED_UNITS" in white on a blue rectangular background. Below the logo is a decorative horizontal bar with a repeating pattern of DNA double helix structures.

Specific Contract No. 03 implementing Framework Contract No.
EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-

Project MED_UNITS

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP2- OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

Deliverable 2.5. Report on European hake and red mullet microchemistry analysis

Compiled by: Sílvia Pérez-Mayol, Beatriz Morales-Nin, Ignacio A. Catalán and Miquel Palmer (IMEDEA), and Marco Stagioni (CoNISMa-UniBIO).

Responsible: Beatriz Morales-Nin (IMEDEA) and Marco Stagioni (CoNISMa-UniBIO)

Date: 28th June 2021

INDEX

1. SUMMARY	3
2. INTRODUCTION	4
3. EUROPEAN HAKE (<i>Merluccius merluccius</i>) MICROCHEMISTRY ANALYSES.....	4
4. RED MULLET (<i>Mullus barbatus</i>) MICROCHEMISTRY ANALYSES	25
5. REFERENCES.....	36

1. SUMMARY

Due to covid-19 related problems that impede the access to laboratory and analytical facilities in Italy, red mullet could not be analysed on time for the date due for this Deliverable. Therefore, on January 2021 a Draft Deliverable was presented only reporting the results of the microchemistry analysis of *Merluccius merluccius* otoliths.

This new final Deliverable 2.5 includes now the microchemistry analyses for both *M. merluccius* and *Mullus barbatus* otoliths. In the case of European hake 279 otoliths were analysed, all of them belonging to 10 Mediterranean management units (hereafter, GSAs), plus two additional areas in the NE Atlantic added for comparison. Red mullet analysis was restricted to a subsample of 250 otoliths from 10 different GSAs. All otoliths were analysed using Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry (LA-ICP-MS), following the protocols previously defined within the project (D2.3).

For European hake, a data matrix of concentration (ppm) for 25 isotopes corresponding to the otolith core and otolith edge for each individual was obtained using a methodology developed at IMEDEA and using 6 Certified Reference Materials (CRMs) for isotopic quantification. The data matrix was analysed using classification and linear multivariate models. In spite that GSAs differed in otolith microchemistry, GSA membership has been correctly predicted from only 30% of the fish when using otolith edge data. Such a pattern suggests that between-GSA differences in the water mass where a given fish is actually living is not fully reflected in the otolith edge's chemical composition, at least at the spatial scale considered. The lack of links between a GSA and the microchemical composition of the otolith edge precludes to use the core data for safely estimating the natal origin of a given fish or linking specific fishing grounds to nursery areas, at least at the spatial scale considered. The percentage of correct classification increased to 63% when using only Western, Central and Eastern pooled areas but this increase should be interpreted with caution.

Red mullet elemental quantification was obtained for 19 isotopes, with samplings performed at both the core and edge of the otolith. Four (4) CRMs were analysed to obtain isotope concentrations (ppm) using commercial software *iolite* for data reduction. Results were analysed using Discriminant Analysis of Principal Components (DAPC) treating core and edge sampling sites independently in order to assign microchemical information to GSA natal and fishing origin, respectively. Correct allocation was 63% for edges and 66% for cores, albeit a high individual variability that decreased the classification power.

The results reported here can be explained by at least three non-incompatible hypotheses: (1) otolith microchemistry may only reflect water mass features at another spatial scale, (2) the limits of biological populations may include several management units, and (3) alternative processes related with growth

rate may also be affecting the microchemical composition and mask the link between water mass features and otolith composition.

2. INTRODUCTION

The present *Deliverable 2.5* should have been presented at month 18 (June 2020), however a project extension due to COVID-19 pandemic allowed to present it at month 24 (January 2021). Despite of this extension, several problems encountered for Italian restrictions due to the pandemic impeded the compliance of the red mullet microchemistry analysis on time, so on January 2021 a Draft Deliverable was presented only containing the information for hake analysis. The Final Deliverable presented here includes the results for both committed species.

The information reported here follows the protocol developed in *Deliverable 2.3* regarding the methodology applied to prepare and analyse European hake and red mullet otolith samples; however, some changes have been done and will be explained and justified below.

3. EUROPEAN HAKE (*Merluccius merluccius*) MICROCHEMISTRY ANALYSES

3.1. Samples analysed

According to Milestone 2.1, the otolith microchemistry study was restricted to a subsample of 230 otoliths (N=205 from the Mediterranean, 25 from Atlantic ICES VIIIa) selected from the otoliths used for shape analysis. We also included additional data from 49 otoliths analysed using the same protocols and coming from ICES VIIIc. The selection of the same otoliths from the shape analysis was due to the need to conduct joint analyses to test if both, otolith shape and microchemistry, could better describe the European hake stock identity (*Deliverable 2.6*).

The subsample comprised 25 juvenile females per GSA (maturity stages 1 or 2), with total lengths (TL) up to 280 mm, although some fish with higher TL were included in the case of Eastern Mediterranean areas to achieve a sufficient number of otoliths by GSA. The characteristics of the analysed otoliths are summarized in Table 1.

<i>Geographic subunits</i>	<i>Fish analysed</i>	<i>TL (mm)</i>	<i>Maturity stage</i>
1b	25 females	≤ 280	1 or 2a
9b	25 females	< 280	1
11c	25 females	< 280	1
16b	25 females	< 280	1 or 2a
18a	25 females	< 280	1
20a	24 females	< 280	1 or 2
22b	25 females	< 280	1 or 2
25 c	20 females	< 286	1 or 2a
26a	19 females	< 327	1 or 2
27b	17 females	< 399	1 or 2
ICES VIIIa	25 females	≤ 280	1, 2 and 2a
ICES VIIIc	49 females	< 280	juvenile
TOTAL	279 otoliths		

Table 1. List of GSAs analysed for hake otolith microchemistry and fish descriptors. Maturity stage corresponds to the scale according to MEDITS (2017).

The initial contract proposal contemplated a total of 8 GSAs to be evaluated and selected according to the genetic pilot study results and the otolith availability. However, those GSAs did not cover the complete longitudinal Mediterranean range and no non-European countries were included. For that reason, IMEDEA, in accordance with the Project Coordinator, proposed to increase the number of GSAs to be analysed for a better coverage of the species geographical distribution in the whole Mediterranean. The selection of the additional GSAs was conducted based on the Mediterranean Sea circulation patterns and basins (Fig. 1) to cover possible connectivity patterns. Moreover, otolith samples from Atlantic Ocean (ICES VIIIa) provided by IFREMER (France) were also prepared and analysed; in addition, the results obtained from the Spanish project DREAMER (CTM2015-66676-C2-1-R), corresponding to individuals from the Galician coast (ICES VIIIc), were also included in the data analysis (Table 1).

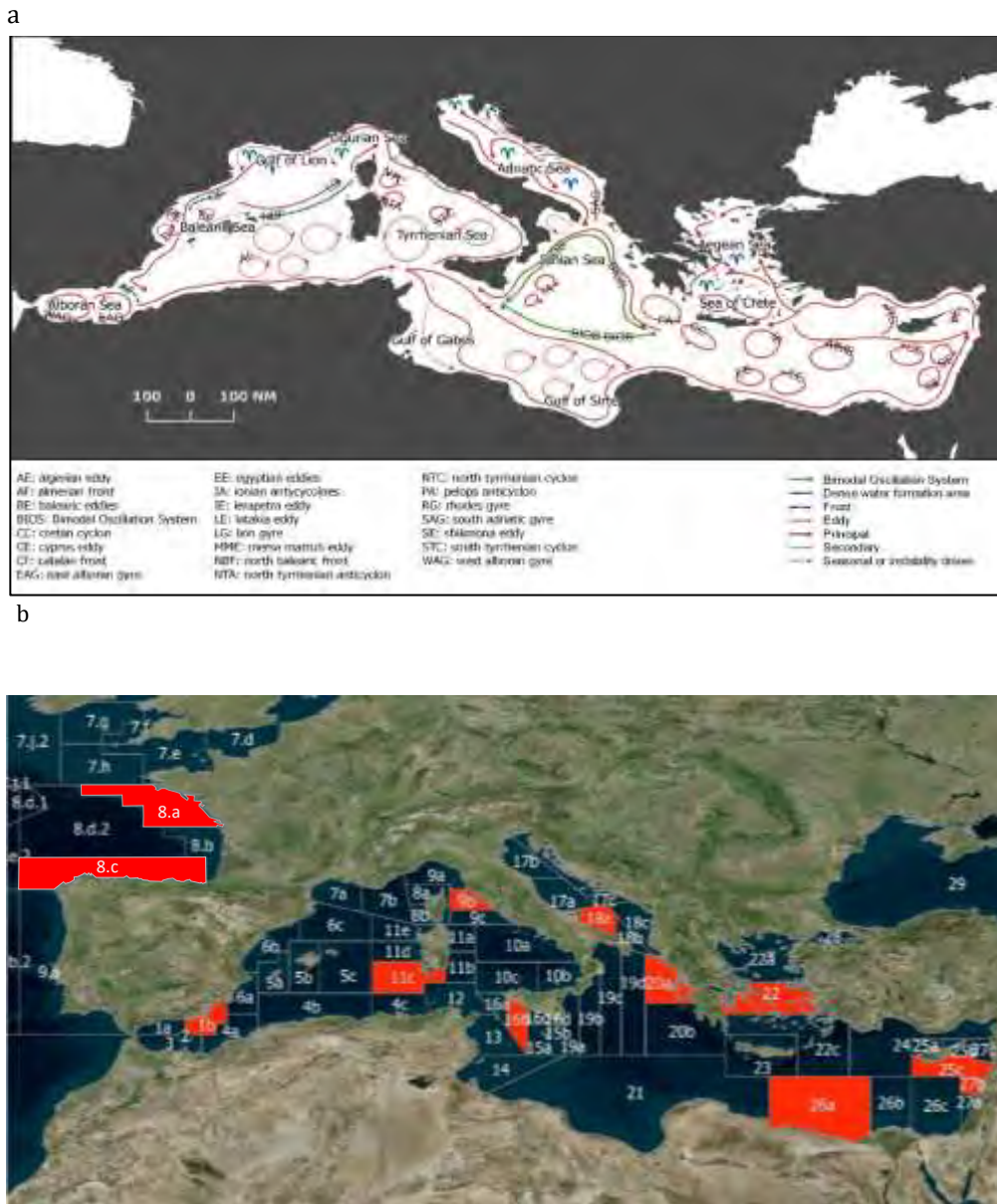


Figure 1. a) Conceptualized Mediterranean Sea circulation; b) Position of the GSAs and Atlantic areas analysed for hake otolith microchemistry (in red).

3.2. Sample preparation procedures

The detailed sample preparation protocols are developed in *Deliverable 2.3- Protocol for analysis and data treatment of trace elements in otoliths of European hake and red mullet*. Briefly, after extracting, otoliths were cleaned from organic tissue with 5% hydrogen peroxide and stored in plastic vials. At IMEDEA laboratory and with the objective to obtain sagittal otolith thin sections with exposed core and ventral edge areas appropriate for microchemistry analysis, left *sagitta* otoliths were included in epoxy resin blocks, ground with silicon carbide lapping papers (from FEPA P1200 to P4000) and polished with 3 μ m and 1 μ m diamond suspensions. Thereupon, otoliths were randomly transferred and glued to 23 different petrographic glass slides using double-sided tape; in addition, for better identification of the ablation

areas, images of each otolith were obtained using a microview system. Finally, glass slides were decontaminated and cleaned with nitric acid 2% and sonicated in MilliQ water; once dried, glass slides were stored in double zip plastic bags (Fig. 2).



Figure 2. Example of two petrographic glass slides with some identified otolith sections ready to analyse.

3.3. LA-ICP-MS analysis

Otolith composition was carried out at the Plasma-Mass Unit of the Universidade de A Coruña (Spain) in July 2020. The equipment used consisted of a CETAC Laser Ablation System LSX-213 G2+ coupled to a Thermo-Finnigan Element XR Inductively Coupled Plasma-Mass Spectrometer. This system uses He gas in the ablation cell and Ar as carrier gas of the ablated material.

The analytical method was defined by the use of Medium Resolution (MR) as acquisition procedure, the length of the ablation lines and the number of isotopes to be analysed. A total of 28 different isotopes was included: ^6Li , ^7Li , ^{11}B , ^{23}Na , ^{24}Mg , ^{25}Mg , ^{27}Al , ^{28}Si , ^{31}P , ^{42}Ca , ^{43}Ca , ^{44}Ca , ^{45}Sc , ^{55}Mn , ^{56}Fe , ^{59}Co , ^{60}Ni , ^{63}Cu , ^{66}Zn , ^{85}Rb , ^{88}Sr , ^{118}Sn , ^{137}Ba , ^{138}Ba , ^{206}Pb , ^{207}Pb , ^{208}Pb , and ^{238}U . In *Deliverable 2.3* it was indicated to analyse ^{39}K , but it was not possible because, when acquiring at MR, ^{39}K presents an interference with ^{40}Ca , so that isotope is only measurable at High Resolution (a procedure that gives very low signal intensities not applicable on surface-based ICP-MS techniques); in addition, ^{29}Si was changed by ^{28}Si since it is the most abundant isotope and can be easily quantified at MR.

Equipment tuning (i.e., signal intensity, oxides %, mass offset and fractionation events) was conducted firing on NIST610 (National Institute of Standards and Technology) Certified Reference Material (CRM). An initial proof to determine the best laser energies and frequencies was done, concluding that 60% of laser energy and 20 Hz of repetition rates (producing a fluency of $\sim 7.5 \text{ J} \cdot \text{cm}^{-2}$) gave the best results in

terms of signal intensity and laser performance on the ablated material (i.e., clean craters, inexistence of big and badly ablated fragments).

Otolith scans were defined as ablation straight lines 200 μm length and 30 μm width (circular section), across the core (natal signature) and parallel to the otolith edge (fish capture location signature) at the ventral axis (Fig. 3). Scan speed was fixed at 5 $\mu\text{m} \cdot \text{sec}^{-1}$, so the laser was firing each scan during 40 sec. With these settings, the method total duration was 66 sec, taking into account the initial and final background (blank) signal measurements.

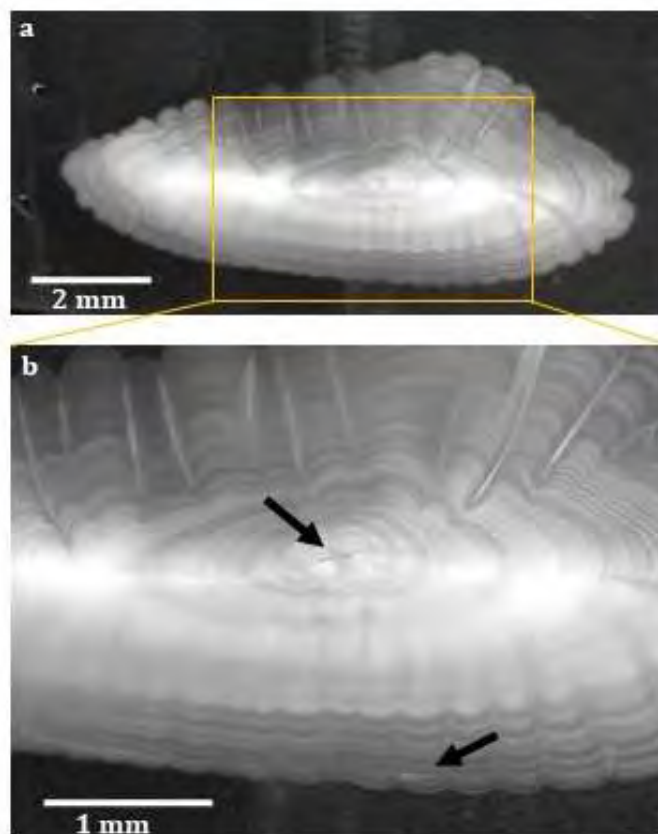


Figure 3. Sagittal section of a European hake otolith showing the ablated lines (black arrows) on the core and ventral edge.

A set of 6 Certified Reference Materials (CRM) of known composition (Table 2) was fired using the same laser conditions as in otoliths: NIST612, NIST614, NIST616 (National Institute of Standards and Technology), FEBS1 (National Research Council Canada, (Sturgeon et al. 2005)), NIES22 (National Institute for Environmental Studies, (Yoshinaga et al. 2000)), and MACS-3 (International Association of Geoanalysts). The CRMs were analysed twice using the “bracketing protocol” for which all the CRMs, except MACS3, were analysed at the beginning and at the end of a working session (day) and every 20 line scans on otoliths. MACS-3 was analysed twice only at the beginning and end of the working session. A total of 902 LA-shots on both otoliths and CRMs was performed.

After analysis, digital images using a microview system were obtained with the objective to verify the laser-scan appropriate position (Fig. 3).

CRM	Concentration range (ppm)	Matrix
NIST612	10 – 80	silicon crystal
NIST614	0.5 – 50	silicon crystal
NIST616	0.008 – 30	silicon crystal
FEBS1	0.002 – 2600	otolith pressed powder (homemade)
NIES22	0.002 – 2400	otolith pressed powder (homemade)
MACS3	1 – 7000	calcium carbonate pressed powder

Table 2. Certified Reference Materials attributes.

3.4. Treatment of data I: from cps to elemental concentrations

Two protocols were applied to calculate elemental concentration from raw data. The IMEDEA protocol and the *iolite* protocol. The former is more restrictive and thus judged more reliable, whereas the second may yield a higher number of data points (see further in the text). The results obtained with the IMEDEA and the *iolite* protocols were qualitatively equivalent, thus in sake of simplicity here we report the results corresponding to the IMEDEA method only.

3.4.1. IMEDEA protocol

For each one of the laser-shots and for every measured isotope, the spectrometer provides a temporal profile (intensities over time of counts per second, cps) that is composed by a background interval at which no sample is ablated (*blank*) and a *plateau* interval at which intensity values reach a noisy but steady level after the signal (cps) returns to the blank level again. These raw intensities have to be transformed into isotope concentrations ($\mu\text{g}_{\text{isotope}} \cdot \text{g}_{\text{sample}}^{-1}$ or ppm). This procedure requires several steps: *i*) the selection of *blank* and *plateau* intervals of every LA-shot (or LA-scan), *ii*) data transformation and reduction, *iii*) the machine drift correction, *iv*) the normalization by ^{42}Ca as Internal Standard, and *v*) the quantification of estimated element concentrations on CRMs and otoliths using Bayesian analyses.

The first step (*i*) was conducted using a Long Short-Term Memory (LSTM) Network built in MatLab and developed by the *Data Processing Group of the University of Vic* and IMEDEA (Martí-Puig et al., 2019); the network was trained using previous supervised shots so that it is able to automatically identify and select the *blank* and *plateau* intervals of the raw data for one specific LA-shot.

The second part of the data processing (steps *ii* to *v*) was implemented into a hierarchical model developed in R by the *Fish Ecology Lab* at IMEDEA. The procedure works as follows: The distribution of

count per second (cps) for a given scan and isotope was determined by subtracting the distribution of cps at the *plateau* from the distribution of cps at the *blank*. When the *plateau-blank* difference was not different from zero (prob > 0.05), the scan concentration of the corresponding isotope was considered not detectable (missing data). Next, the difference *plateau-blank* for any given isotope was normalized (ratio) by the difference *plateau-blank* for ^{42}Ca . This normalization (known as *internal standard normalization*) is a widely applied procedure for correcting intensity disparities related only with structural differences between (synthetic and homogeneous) CRMs and biogenic samples (otoliths) (known as matrix effect); the internal standard used corresponds to a major component of the sample and it is assumed to be homogeneously distributed in it. When a ratio was not different from zero (prob > 0.05), the scan concentration of the corresponding isotope was considered also as not detectable (missing value). Finally, the element: ^{42}Ca ratio from the CRMs (as dependent variable) is assumed to be a lineal combination of the certified concentration for each element (ppm) corrected by its naturally occurring isotope abundances, the session (categorical variable accounting for session-specific random effects common to all the LA-shots from a given working session) and the time elapsed since the session's start and end (to account for any lineal temporal drift of signal intensities throughout a given session, which is a common technical issue in LA-ICP-MS).

For a given session and a given element, the parameters of the statistical model including all the above variables, were estimated using a Bayesian approach. The parameterized model was then used to estimate the concentration (ppm) of each LA-scan firstly on CRMs and then on otoliths, after accounting for the effects of session and temporal drift. In case of bad convergence of the Bayesian analyses, the corresponding isotope scan concentration was also considered missing data. The uncertainties of the estimates for any concentration properly propagates all the uncertainty levels considered.

Moreover, one of the CRMs (MACS3) was treated as a sample of unknown concentration. In case of discrepancy between the certified and the estimated concentration, the LA-shot concentration of the corresponding isotope was also considered as missing data.

Once the concentration values of each isotope for all otolith core and edge scans were obtained, a joined matrix was built. It included the biological parameters for each fish, the sampling information (i.e., GSA, date, depth) and the concentration (ppm) for each isotope at the specific otolith location. The matrix also included the results obtained from the DREAMER project in relation to ICES VIIIc area. Note that missing values (obtained due to the above mentioned quality controls) were also included, reported as NA; any value was obtained for ^6Li and ^{238}U , so they were not included in the final matrix and a total of 25 elements were reported.

3.4.2. *Iolite* protocol

A second set of concentration calculations were performed with the aim to compare IMEDEA protocol for obtaining isotope concentrations and those obtained with a data reduction commercial software. The raw intensities for all shots and isotopes were analysed using the commercial software *iolite* (v4) (Paton et al., 2011). After an interactive identification and selection of the *blank* and *plateau* of each isotope profile, the *blank* subtraction, the internal standard normalization and correcting for machine (linear fit) drift, the interface calculates elemental concentrations following Longerich et al. (1996) method. Unlike the IMEDEA method, *iolite* uses one single CRM to obtain elemental concentrations. The certified concentration value for a given element is used to directly extrapolate the concentration value of that element on the otolith according to the obtained intensities, so no calibration line is obtained to estimate concentrations, conversely to the IMEDEA method.

Provided that the results obtained using the two methods are qualitatively the same, hereafter, we reported the results of the IMEDEA method only because this method is more conservative in the sense described above (i.e., the criteria for considering that a given reading is unreliable are stricter). The results obtained from both methods can be considered comparable based in a number of facts. For example, between-otolith (di)similarity (as depicted by Euclidean distance calculated from the scaled chemical composition) obtained using the two methods are correlated. The Mantel test calculated from a sample of 200 otoliths rendered an r value of 0.43. The probability of obtaining this value by chance is less than 0.001, thus the chemical composition estimated by the two methods should be considered comparable at the fish level. Moreover, all the multivariate analyses detailed below rendered very similar patterns at the GSA level. For example, the IMEDEA methods reached a 30% of correct classification vs 36% using *iolite*.

3.5. Treatment of data II: statistical analyses and results

3.5.1. General overview of the statistical analyses

The general goals of the statistical analyses were: (1) to describe the multivariate microchemical composition of each GSA, (2) to evaluate the classification capability of microchemical composition (i.e., to predict subarea membership of a new fish from the multivariate microchemical composition of its otolith), and (3) to test for the existence of between-GSA differences in multivariate microchemical composition.

When doing multivariate analyses, it is not possible to include otoliths with missing values for some element. After a preliminary data exploration, it was evidenced a large number of missing values in the multivariate data. Conventional approaches for dealing with this problem are deleting either variables

(elements) or samples (otoliths) with missing values from the analyses. However, this was not reliable in that case because it would reduce the data to a very few otoliths and elements and would compromise the objectives of the project. Therefore, we adopted a mixed strategy: in a first step, variables with less than 30% of valid data and otoliths with four or more missing data were deleted. Note that even using these undemanding thresholds, the number of remaining missing values were still very high. Therefore, in a second step, a method for missing data imputation was used for filling the gaps (see below).

The resulting matrix was then analysed using two complementary statistical strategies. First, classification methods were used to evaluate the capability of predicting the GSA membership from the microchemical composition of its otolith. Second, multivariate linear models were used for testing between-GSA differences and displaying the patterns of between-GSA similarity (ordination analysis).

According to the results obtained (fully detailed below), edge data cannot be used for predicting GSA membership. No links were found between the features of the water mass (at the GSA scale) where a fish was actually fished with the microchemical composition of the otolith edge, which is assumed to be formed just before fishing. The lack of identifiable links between a water mass and the microchemical composition of the otolith precluded, at the spatial scale considered, to use the core data for safely estimating population natal origin of a given fish or linking specific fishing grounds to nursery areas. Similarly, between-GSA connectivity could not be deduced from the otolith microchemical composition. Accordingly, two independent analyses (i.e., edge and core) were completed.

Finally, an additional classification analysis was completed after pooling the Mediterranean subareas into three larger areas (Eastern, Middle and Western Mediterranean).

3.5.2. Missing data and missing data imputation

The pattern of missing data was similar for the core and the edge data. Concentration (ppm) for most of the elements analysed did not pass the quality controls (Fig. 4).

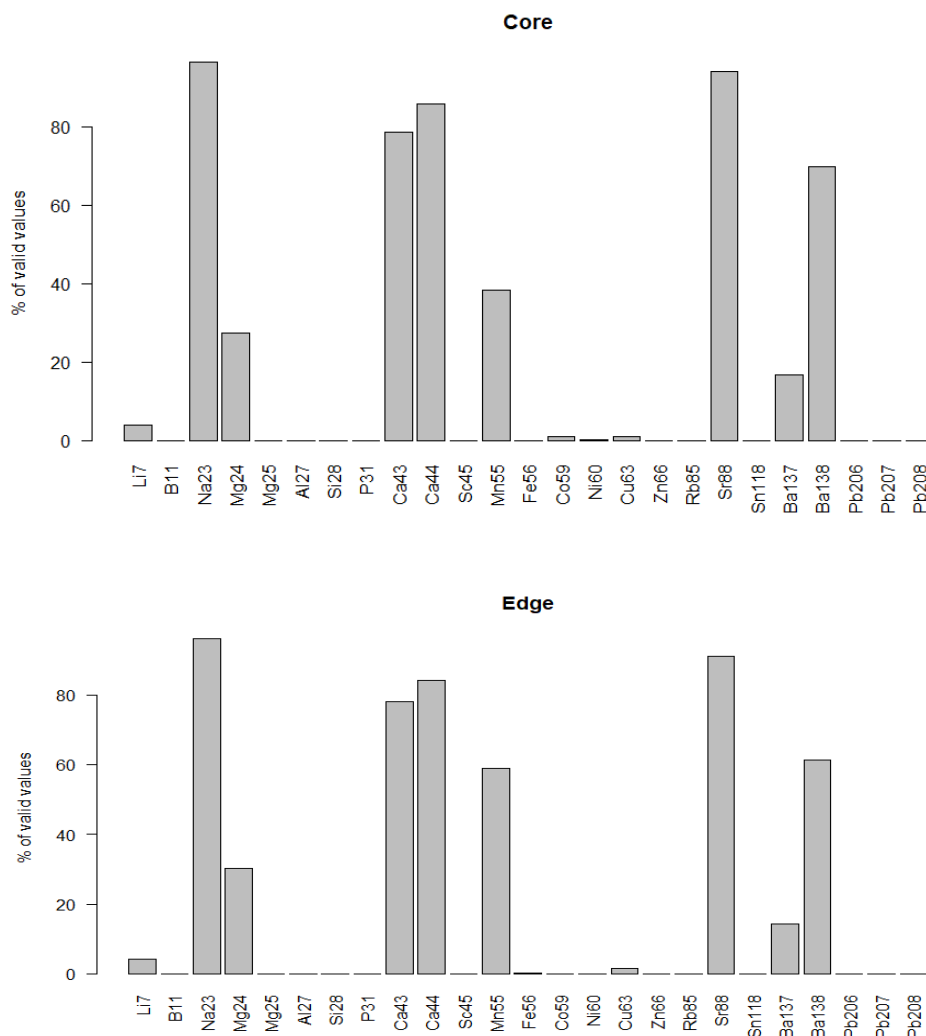


Figure 4. Ratio of valid estimations for the 25 elements initially measured at both core (up) and edge (down) otolith areas.

According to these results, the isotopes retained for further analyses (those with more than 30% of valid results) were: ^{23}Na , ^{24}Mg , ^{43}Ca , ^{44}Ca , ^{55}Mn , ^{88}Sr and ^{138}Ba .

These retained elements showed relevant between-GSA variability in both edge and core otolith regions. ^{23}Na , ^{24}Mg and ^{55}Mn showed high variations in both core and edge samplings; ^{138}Ba variability was also high on cores (Fig. 5). ^{23}Na concentration, which has been proposed as an indicator of metabolism, was higher in the otolith cores of Atlantic samples and low in the Eastern Mediterranean (i.e.; GSAs 25c, 26a, 27b), whilst in the edges were more similar in all GSAs. It is noteworthy the high levels of ^{55}Mn and ^{24}Mg in the otolith edges and ^{55}Mn on the cores of European hake from GSA 20a.

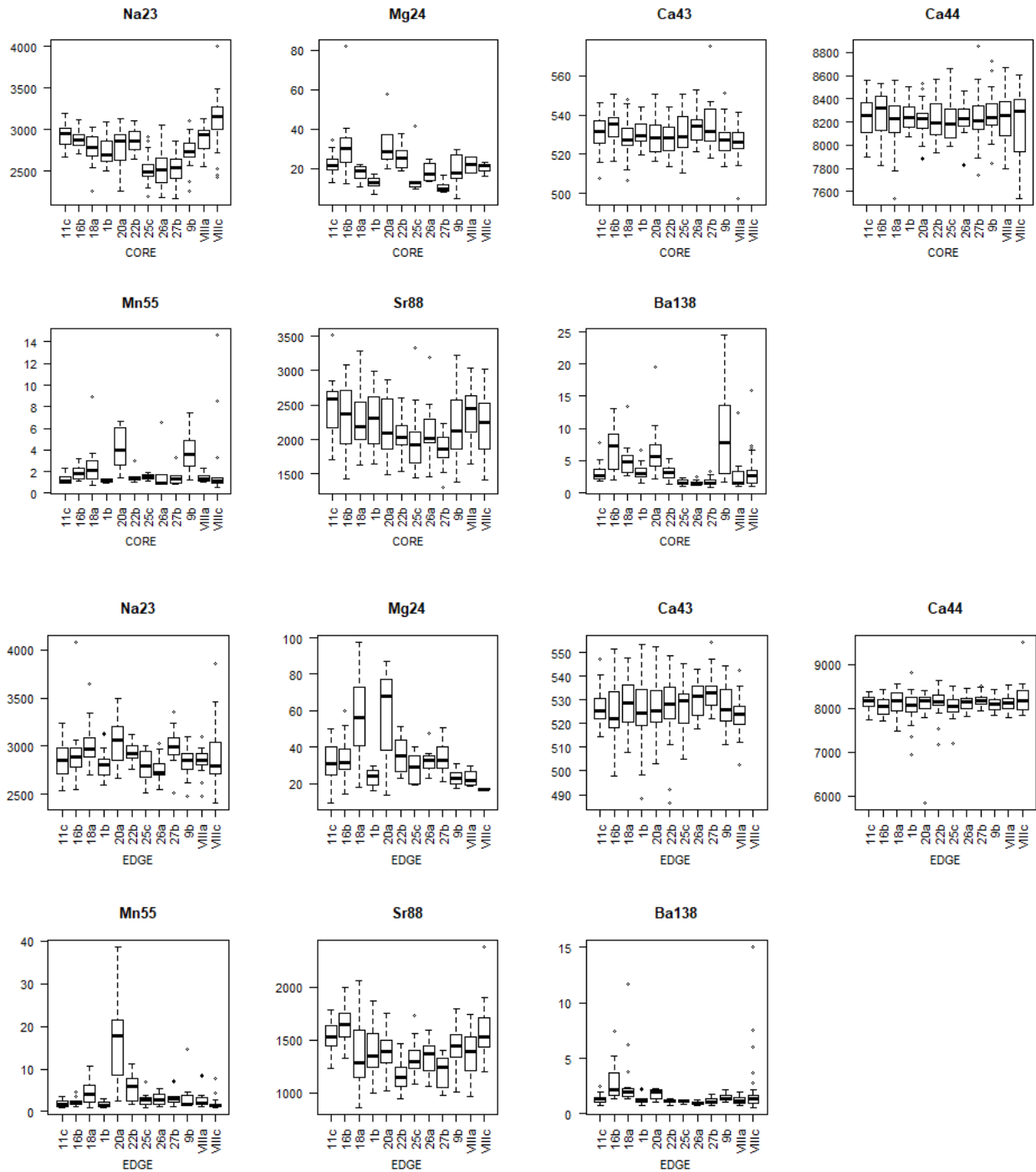


Figure 5. Between-GSA variability for the seven elements analysed in both core (up) and edge (down) otolith areas. Concentration units (y-axis) are ppm.

At this stage, the number of missing values per scan was still very large, so 83 scans (14% of the total) with four or more missing values were deleted from further analyses. Therefore, the number of remaining scans was reduced to 229 for both core and edge data sets. Table 3 summarises the number of scans retained per subarea.

	1b	9b	11c	16b	18a	20a	22b	25c	26a	27b	VIIIa	VIIIc
Core	24	25	23	25	25	24	25	19	18	16	21	17
Edge	24	24	24	24	24	22	25	17	17	17	25	16

Table 3. Number of remaining otoliths after deleting four or more missing values.

After the application of deletions from those previous steps, the number of missing values was 23% for core and 22% for edge. Therefore, a missing data imputation method was applied to fill those gaps (Chiloui et al., 2019; Ćwiklińska-Jurkowska et al., 2005).

Imputation methods for multivariate data can be based (1) on replacing a given missing data for the corresponding value from the most similar case in the matrix, or (2) on estimating the missing values from the variance/covariance matrix. For the first alternative, we used the *impute.knn* function from the *impute* library of the R package. In that case, a *k-nearest neighbors* algorithm with $k = 3$ was used to find the most similar otolith in the matrix and the missing data were replaced with the mean values of those three otoliths (Troyanskaya et al., 2001). For the second alternative, the function *imputeData* from the *mclust* and *mix* packages was used (Schafer et al., 1997). Provided that similar results were achieved using both methods, hereafter only the results obtained with the first method (*knn*) are reported and used for further analyses.

Preliminary trials were also completed to compare missing data imputation at the within-GSA scale and at the overall scale (i.e., including all otoliths and ignoring GSA membership). Since the results in terms of classification capability were almost the same, hereafter, we ignored GSA membership for missing data imputation in order to avoid any circularity.

3.5.3. Predicting GSA membership: Edge

Preliminary multivariate analyses including the Atlantic populations rendered counter-intuitive results, in the sense that the Atlantic populations seems to display otolith chemical composition closer to some of Mediterranean GSAs than the similarity displayed for the later Mediterranean GSAs to other Mediterranean GSAs. For example, the GSA 25c (Cyprus) and the Atlantic VIIIa were closer each other than the GSA 25c to other Mediterranean GSAs. Thus, adding the Atlantic populations may introduce additional noise to the analyses and may mask the true relationships among the Mediterranean GSAs, which is the main objective of the project. Accordingly, hereafter, the multivariate analyses (either, ordination or classification) were restricted to the Mediterranean GSAs and the Atlantic samples were excluded. Note that this decision is also coherent with the geographical extent of other analyses in the project (e.g. genetics), which focussed exclusively in the Mediterranean GSAs.

Classification algorithms have recently experimented an unprecedented inflation. Specifically, computer-intensive methods have become popular (Mercier et al., 2011). However, it has been demonstrated that conventional methods as Linear Discriminant Analysis (LDA) or Quadratic Discriminant Analysis (QDA)

have better or at least similar performance than those fashion, computer-intensive methods when the number of samples is relatively small and when the data fits the underlying assumptions (Jones et al., 2017). Here we compared the predictive capability (measured by the rate of correct predictions after a leave-one-out cross-validation) of 12 computing intensive methods as they are implemented in the *RWeka* library (Witten et al., 2005). LDA and QDA were completed using the functions implemented in the *MASS* library (Venables et al., 2002).

Before those classification analyses, the raw data was tested for multinormality using the function *mvn* from the *MVN* package (Korkmaz et al., 2014). According to the results of those multinormality tests, ^{24}Mg , ^{25}Mn and ^{138}Ba were submitted to the Box-Cox transformation using the function *boxcox* of the *MASS* package (Venables et al., 2002). This function estimates the likelihood profile for a range of λ values. The value of λ showing the maximum likelihood was used for transforming the data using: transformed = $\lambda^{-1}(\text{raw}^\lambda - 1)$.

Provided that the comparison between classification methods rendered similar results for the edge and the core, here we only report the edge results. The results of comparing the performance of different classification methods are depicted in Table 4, where the correct classification rate was simply the ratio between the number of correctly classified cases and the total number of cases (229).

<i>Method</i>	<i>R library</i>	<i>Correct classification rate</i>
LDA	MASS	0.30
QDA	MASS	0.26
J48	RWeka	0.21
LMT	RWeka	0.30
DecisionStump	RWeka	0.08
Logistic	RWeka	0.30
SMO	RWeka	0.27
IBk	RWeka	0.22
AdaBoostM1	RWeka	0.08
Bagging	RWeka	0.26
LogitBoost	RWeka	0.26
JRip1	RWeka	0.13
OneR	RWeka	0.18
PART	RWeka	0.23

Table 4. Correct classification rate after leave-one-out cross-validation.

Linear Discriminant Analysis (LDA), Logistic Model Trees (LMT) and Logistic regression seemed to be the best methods in terms of cross-validated accuracy. Note, however, that even in those cases, the classification performance was small (30%). The 95% interval for correct classification rate after randomly shuffling the otoliths among GSAs (100 bootstrap replications) were found to be between 5%

and 14%. The correct classification rate (30%) was larger than the upper value (14%), thus, suggesting that chemical composition performs better than random but, again, the classification success was discrete.

The cross validated (leave-one-out) confusion matrix for Linear Discriminant Analysis at the edge samplings is shown in Table 5.

GSA	11c	16b	18a	1b	20a	22b	25c	26a	27b	9b
11c	3	2	1	2	1	0	3	0	0	6
16b	3	12	4	4	1	0	0	1	0	5
18a	3	2	6	4	1	2	0	1	1	2
1b	4	1	2	4	1	2	4	2	0	4
20a	3	0	4	0	13	0	1	0	1	1
22b	1	0	5	2	3	12	5	3	7	1
25c	1	0	0	2	1	1	0	2	0	0
26a	1	0	0	3	0	2	3	5	3	1
27b	1	1	1	1	0	5	0	2	4	1
9b	4	6	1	2	1	1	1	1	1	3
	3/21	12/12	6/18	4/20	13/9	12/13	0/17	5/12	4/17	3/21

Table 5. Confusion matrix for edges. The actual GSA membership are at columns and the predicted GSA membership are at rows. At the last row the number of correct predictions are compared with the number of wrong predictions.

3.5.4. Between-GSA differences: Edge

The existence of between-GSA differences in microchemical composition at the edge was tested using Redundancy Analysis (RDA) (Borcard et al., 2018) using the matrix of box-cox transformed data. This matrix was build up by the seven elements retained (^{23}Na , ^{24}Mg , ^{43}Ca , ^{44}Ca , ^{55}Mn , ^{88}Sr and ^{138}Ba) and the actual GSA membership as explanatory variable. RDA was completed as implemented in the *rda* function of the *vegan* library (Oksanen et al., 2019). Variables (element concentration) were standardized (i.e., divided) by its standard deviation, thus the same weight is given to all the element. Significance of between-GSA differences was assessed after 1,000 random permutations. The results (Table 6) show that between-GSA differences were clearly significant.

	Df	F	Prob
GSA	9	1.46	>0.001
Residual	208	5.543	

Table 6. Redundancy analyses results for edges.

The case (i.e., otolith) scores from the RDA were used to estimate each GSA centroid for depicting between-GSA similarity (Fig. 6a). For completeness, an ellipse including 95% of the otolith of each GSA

was added for depicting both between- and within-GSA variability in microchemical composition (Fig. 6b).

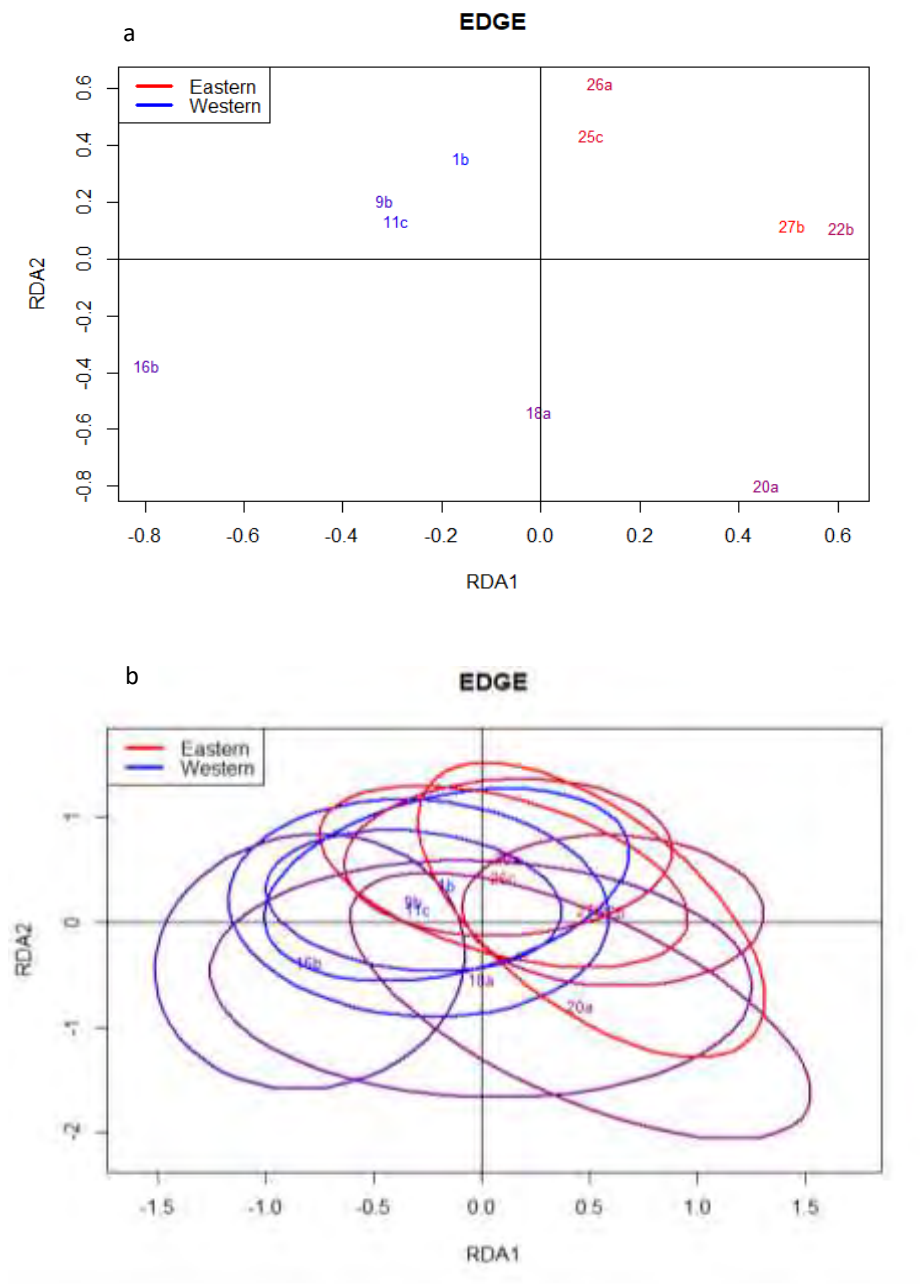


Figure 6. a) Biplot (two first axes of RDA space) showing the position of the centroid of the 10 GSAs analysed. b) Ellipses including 95% of the fish from each GSA are added to the same biplot shown at plot a. The colour in a red to blue gradient of the labels denote the longitude order of each GSA (western GSAs in blue, eastern GSAs in red and central GSAs in intermediate colour).

The pattern depicted by Figure 6a could suggest the existence of some GSA clusters (Western, Central and Eastern Mediterranean; this topic is explored below), although the between-GSA similarity pattern seemed not related with geographical distance. Note, for example, that 1b (Gibraltar) had microchemical composition very close to 25b (Cyprus). Such a pattern also suggested that between-GSA differences in

the water mass where a fish was actually living were not reflected in the chemical composition of the otolith edge, at least at the spatial scale considered. Water circulation at the depth of catch might be more linked to these patterns than GSA distance. Also, as suggested elsewhere (Catalán et al., 2018; Morales-Nin et al., 2014), alternative processes related to growth rate may be also affecting microchemical composition, masking the link between water mass features and otolith composition.

Moreover, within-GSA variability in microchemical composition was large, as depicted in Figure 6b. Thus, despite that GSAs located at the opposite places of the gradients displayed different microchemical composition, between-GSA overlap was large because within-GSA variability was also large. Such a large within-GSA variability was the cause of the poor power of microchemical composition when predicting GSA membership.

Finally, between-element correlation patterns were also plotted (Fig. 7). Note that Ca isotopes are near the center, suggesting that concentration of those isotopes are nearly constant at the between-GSA level (see also Figure 5). For the remaining elements, three clusters could be identified: (1) ^{23}Na , ^{24}Mg and ^{55}Mn , (2) ^{88}Sr and (3) ^{138}Ba .

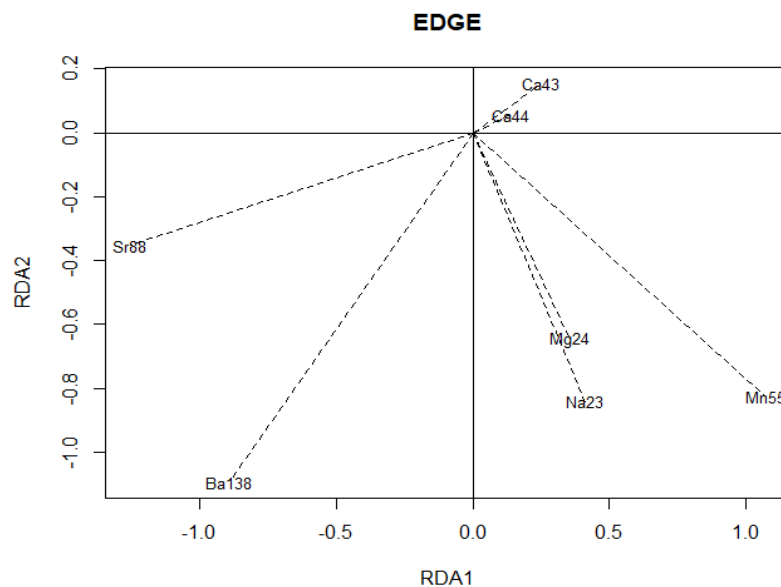


Figure 7: Between-element correlation patterns.

3.5.5. Core analyses

After applying the same procedure detailed above for the edge data (see section 3.5.3), the overall, cross-validated correct classification rate of GSA membership from the core microchemical composition was 21.4%. The corresponding confusion matrix is detailed in Table 7.

GSA	11c	16b	18a	1b	20a	22b	25c	26a	27b	9b
11c	8	3	3	4	2	3	1	1	0	2
16b	1	4	3	1	3	4	0	1	0	3
18a	2	4	2	4	3	0	1	0	0	1
1b	4	1	4	5	2	4	2	0	2	0
20a	1	5	4	2	3	5	0	1	0	8
22b	6	5	4	2	4	6	0	0	1	0
25c	0	0	0	1	2	1	3	4	6	2
26a	0	0	0	1	0	0	5	6	4	1
27b	0	0	0	3	0	0	7	4	3	0
9b	1	3	5	1	5	2	0	1	0	8
	8/15	4/21	2/23	5/19	3/21	6/19	3/16	6/11	3/13	8/17

Table 7. Confusion matrix for core. The actual GSA membership are at columns and the predicted GSA membership are at rows. At the last row the number of correct predictions are compared with the number of wrong predictions.

As in the case of the edge data, such a classification rate should be considered unsuccessful. Note, however, that the smaller performance of the core data may be related to the moment at which it was assumed that the core material was deposited (i.e., larval or juvenile stage). Nevertheless, as it has been explained above, connectivity patterns cannot be extracted from the otolith microchemical composition.

The case (i.e., otolith) scores from RDA were used to estimate each GSA centroid for depicting between-GSA similarity (Fig. 8a). For completeness, an ellipse including 95% of the otolith of each GSA was added for depicting both between- and within-GSA variability in microchemical composition (Fig. 8b).

The pattern depicted by Figure 8a suggested the existence of two GSAs clusters (eastern, and central/western Mediterranean), but the between-GSA similarity pattern seemed not related with geographical distance. Note, for example that, as in the case of Edge data, 1b (Gibraltar) is the central-western GSA displaying the closest microchemical composition with the GSAs from the Eastern Mediterranean.

As in the case of the Edge data, within-GSA variability in microchemical composition was large, as shown in Figure 8b. Such a within-GSA variability was the cause of the poor power of microchemical composition for predicting GSA membership.

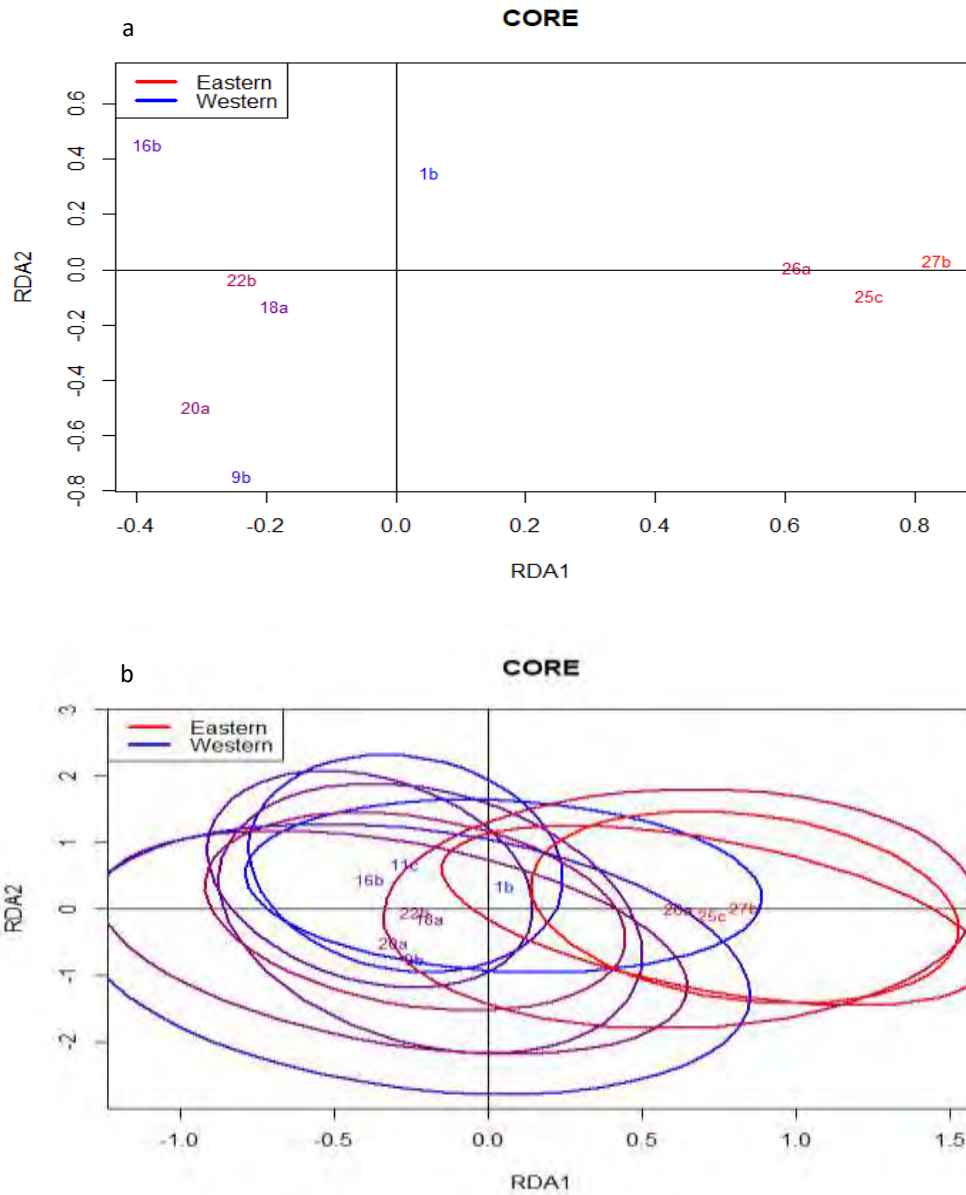


Figure 8. a) Biplot (two first axes of RDA space) showing the position of the centroid of the 10 GSAs analysed. b) Ellipses including 95% of the fish from each GSA are added to the same biplot shown at plot a. The colour in a red to blue gradient of the labels denote the longitude order of each GSA (western GSAs in blue, eastern GSAs in red and central GSAs in intermediate colour).

Finally, between-element correlation patterns were also plotted (Fig. 9). As in the case of edges data, Ca isotopes are near the center.

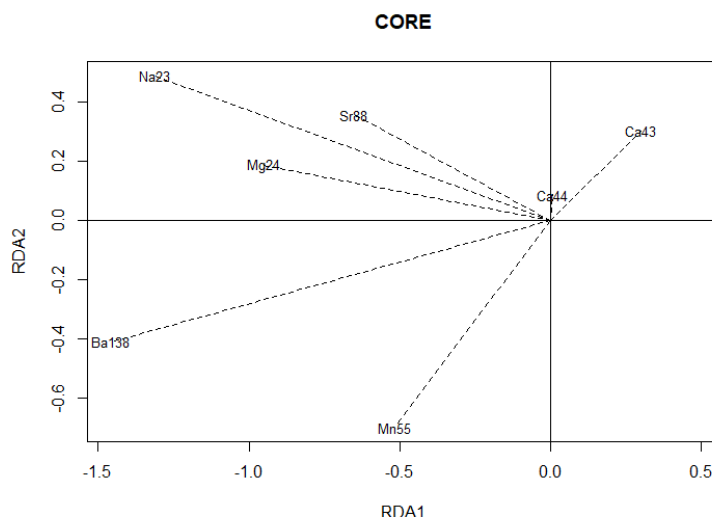


Figure 9. Between-element correlation patterns.

3.5.6. Merging GSAs into larger areas

The results reported above could suggest that after pooling some GSA, membership predictive power of chemical composition may increase, which in turn may suggest that the current GSA boundaries are not supported attending to the microchemical otolith classification.

After *a priori* assuming three areas (Western ["1b"+"11c"+"9b"+"16b"], Middle ["26a"+"25c"+"27b"] and Eastern ["18a"+"20a"+"22b"]), the correct classification rate certainly raised up to 63% when using the edge data set. The corresponding confusion matrix is detailed at Table 8.

	Western	Middle	Eastern
Western	18	8	10
Middle	15	45	11
Eastern	18	18	75

Table 8. Confusion matrix. The actual area membership is at columns and the predicted area membership are at rows.

Note, however, that the correct classification rates with different number of areas were not easily comparable. The correct classification rate achieved with the raw GSA (n=10) and the new areas (n=3) area detailed at Table 9, where the correct classification rates expected when fish are assigned at random to groups allows a better comparison. In both cases, it should be concluded that the achieved correct classification rate was only slightly larger than the values expected at random. In conclusion, there were not clear evidences supporting that classification rate was improving after pooling GSAs into larger areas.

	Actual value	Random permutations		
		<i>0.025 lower quantile</i>	<i>median</i>	<i>0.025 upper quantile</i>
3 areas	63%	33%	37%	43%
GSA	30%	10%	11%	14%

Table 9. Comparing actual values of correct classification rate with those expected at random (95% interval for correct classification rate after randomly shuffling the fish among GSAs; 100 bootstrap replications). Note that the value when considering 3 areas alone was larger, but the expected values when fish are distributed at random into areas were larger too.

3.6. Discussion and conclusions

Otolith chemical composition has been successfully used for predicting migratory movements whenever tagging and tracking are not feasible (Elsdon and Gillanders, 2003; Campana, 2005; Sturrock et al., 2012), comparing population structure or define stock delimitation (Ashford et al., 2006; Higgins et al., 2013; Wright et al., 2018), investigating connectivity during ontogenetic development (Campana et al., 1994; Longmore et al., 2014; Morales-Nin et al., 2014; Régnier et al., 2017; Rogers et al., 2019), tracing natal origin and nursery grounds (Thorisson et al., 2011; Guidetti et al., 2013), determining age and growth patterns (Hüssy et al., 2016; Siskey et al., 2016), and reconstructing environmental exposure histories (Thorrold et al., 1997; Stanley et al., 2015).

Previous studies to define the European hake population structure employed a wide variety of techniques showing a clear differentiation between Atlantic and Mediterranean populations, with a front between the Atlantic and the Alboran Sea (Cimmaruta et al., 2005; Milano et al., 2014). However, inside the Mediterranean there is no agreement on the stock structure. Two main hypotheses have been proposed: *i)* a single stock (genetic data: Cimmaruta et al., 2005; Milano et al., 2014) or *ii)* three stocks corresponding to Western, Central and Eastern Mediterranean basins. This last hypothesis has mounting evidence based on genetic and meristic data (Maurin, 1965; Pla et al., 1991; Roldán et al., 1998), non-neutral SNPs markers (Milano et al., 2014), parasite load studies (Mattiucci et al., 2004) and otolith shape (Torres et al., 2000). Swan et al. (2006) used the otolith composition to address stock separation in hake, identifying the capture locations in the Northeast Atlantic Ocean and Mediterranean Sea with an accuracy of 79% and 65.5%, respectively. Tanner et al. (2012) used the otolith geochemistry, both of the otolith core and edge, showing significant differences among the collection locations in the Atlantic Ocean and the Mediterranean Sea. Individuals were assigned to three different groups in both areas, and the results suggested movement of individuals among local populations within each basin, but little or no movement of hake through the Straits of Gibraltar. In our case, the Atlantic samples showed signatures that mixed Atlantic individuals from local samples from, mainly, the central Mediterranean. Differential growth processes between Atlantic and Mediterranean samples (e.g. Catalán et al., 2018) or the influence of

Atlantic water into central Mediterranean areas through the SE Atlantic current (Fig. 1a) might partly explain these results.

Classification accuracy using otolith microchemistry is moderate and often ranges between 60% and 80% (Arslan and Secor, 2005; Forrester and Swearer, 2002; Patterson et al., 2004; Rooker et al., 2003; Secor et al., 2002; Swan et al., 2006). This accuracy depends to a large extent on the a-priori grouping based on the hypothesis to be tested. For example, if groupings are based upon strong differences in the characteristics of water masses, known to affect some elements (e.g. areas of widely differing salinity or productivity), the classification success is optimal. Our classification success was low even when Atlantic and Mediterranean populations were considered together. However, the hypothesis of an East-West gradient cannot be rejected in the Mediterranean, albeit the individual variability was very high decreasing the classification power of the methodology. A potential confounding process is the existence of fish endogenous effects. All European hake otolith geochemical studies have reported ontogenetic variation in otolith composition (Morales-Nin et al. 2005; 2014). Certainly, it could be argued that exogenous (environmental) influence could also be associated with ontogeny, as fish change their habitats as part of their development. Hake reach deeper waters (hundreds of meters) as they grow older (Recasens, 1993; Recasens et al., 1998). The similarities in the variations of otolith composition may also be caused by the changes in diet that are associated with ontogeny (Bozzano et al., 1997; Guichet, 1995). Nevertheless, there must be a considerable endogenous influence on the elemental variation, which would explain why the hake from areas with *a priori* different water masses had similar patterns. Further analyses within this project should aim to classify water mass characteristics at relevant depths/month per GSA. This would enable to build a priori groupings that were related to functional drivers of otolith microchemical properties. In this work, we have limited the ontogenetic and physiological effects by selecting immature females with similar, when possible, length ranges. However, the poor discrimination power of the otolith edge composition that should represent few weeks of life, compromised the use of the microchemical tags for European hake geographical differentiation.

A similar result was found by Morales-Nin et al. (2014) analysing European hake from the Mediterranean showing uncertainty in the assignment of fish to its capture site. This edge heterogeneity and lack of geographical differences in the otolith edge composition might be further related to limitations imposed by the low number of elements yielding reliable information to be analysed, to low residence time in the location (implying fish movements) or, as commented before, to similarities in the water mass characteristics. European hake is a mobile species undertaking ontogenetic and seasonal migrations to different depths and areas being adults less restricted in depth and location (Recasens et al., 1998). Recruits are located in spatially differentiated high productivity zones that act as nurseries (Maynou et al., 2003). Further, daily migrations have been reported related to their feeding rhythms (Bozzano et al., 1997; Cartes et al., 2009; Orsi-Relini et al., 1986). To this end, existing data from tagging studies evidenced

important individual vertical migrations in the water column and along the slope (de Pontual et al., 2003; 2012; 2013).

4. RED MULLET (*Mullus barbatus*) MICROCHEMISTRY ANALYSES

4.1. Samples analysed

The otolith microchemistry study was restricted to a subsample of 250 otoliths selected from the otoliths used for shape analysis. The selection of the same otoliths from the shape analysis was due to the need to conduct joint analyses to test if both, otolith shape and microchemistry, could better describe the Red mullet stock identity (*Deliverable 2.6*).

The subsample comprised 25 adults per GSA (01a, 06a, 09a, 11c, 12m, 17b, 18b, 19b, 22c, 27a), with total lengths (TL) over to 120 mm.

The initial contract proposal contemplated a total of 8 GSAs to be evaluated and selected according to the genetic pilot study results and the otolith availability. However, those GSAs did not cover the complete longitudinal Mediterranean range and no non-European countries were included. For that reason, CoNISMa-UniBO, in accordance with the Project Coordinator, proposed to increase the number of GSAs to be analysed for a better coverage of the species geographical distribution in the whole Mediterranean. The selection of the additional GSAs was conducted based on the Mediterranean Sea circulation patterns and basins to cover possible connectivity patterns.

4.2. Sample preparation procedure

The detailed sample preparation protocols are developed in *Deliverable 2.3- Protocol for analysis and data treatment of trace elements in otoliths of European hake and red mullet*. Briefly, after extracting, otoliths were cleaned from organic tissue with 5% hydrogen peroxide and stored in plastic vials. At CoNISMa-UniBO laboratory and with the objective to obtain sagittal otolith exposed core and ventral edge surface areas appropriate for microchemistry analysis, *sagitta* otoliths were glued with solvent-free cyanoacrylate resin on precutted microscope coverslip slides (Fig. 10), ground with silicon carbide papers (from 30 to 5 grain micron size) and polished with 5-3-1-0.1 µm alumina lapping sheet.



Figure 10. Single otolith mount for grinding/polishing to expose core level.

Thereupon, 9 otoliths were randomly transferred and glued to 28 different half microscope glass slides on a 3x3 grid using cyanoacrylate resin; in addition, for better identification of the ablation areas, images of each otolith were obtained using a high resolution scanner (Fig. 11).

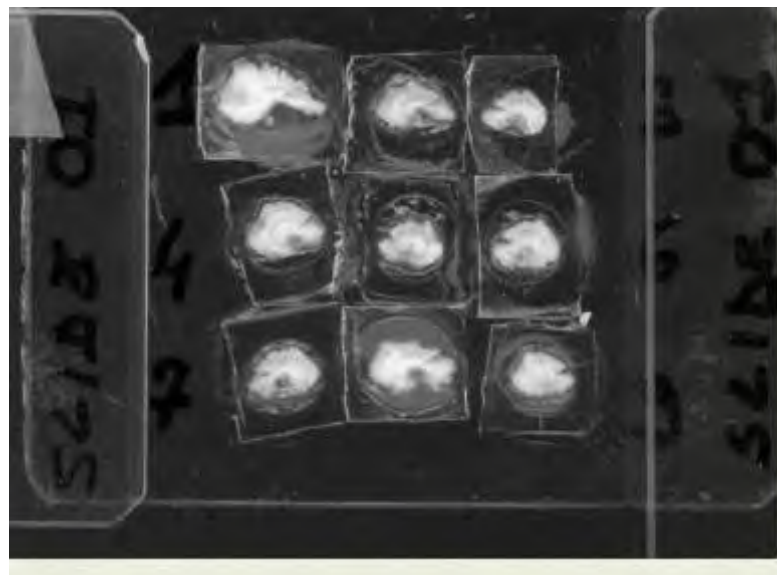


Figure 11. Otoliths multimount (3X3 grid) ready for LA-ICP-MS analyses.

4.3. LA-ICP-MS analysis

Otolith composition was carried out at the LSMAI of the University of Salento (Italy) in April-June 2021. The equipment used consisted of a NewWave Research UP213 with aperture imaging Laser Ablation System coupled to a Thermo-Finnigan ICAP-Q Inductively Coupled Plasma-Mass Spectrometer. This system uses He gas in the ablation cell and Ar as carrier gas of the ablated material.

Equipment tuning (i.e., signal intensity, oxides %, mass offset and fractionation events) was conducted firing on NIST610 (National Institute of Standards and Technology) Certified Reference Material (CRM). The best results in terms of signal intensity and laser performance on the ablated material (i.e., clean craters, absence of badly ablated fragments) were obtained with 65% of laser energy and 4 Hz of repetition rates (producing a fluency of $\sim 13 \text{ J} \cdot \text{cm}^{-2}$).

Otolith scans were defined as ablation spots having nominal 80 μm of diameter (effective 72 μm), centred at the core (natal signature) and near the otolith edge (fish capture location signature). The laser was firing in each scan for 12 sec. (with a preablation at 2 Hz). With these settings, the method total duration was 66 sec, taking into account the initial and final background (blank) signal measurements. Preliminary measurements showed that firing an otolith core for 12 seconds using the laser ablation condition above described produced a hole with a flat profile about 10 μm deep. As a result, the second natal signature measurement was taken after the first shot in the same position but moving the motorised plane up by 10 microns to focus the new surface of the sample core.

In the explorative runs a set of 47 isotopes were acquired: Li7, Be9, B11, Na23, Mg25, Al27, Si29, K39, Ca43, Ca44, Sc45, Ti47, V51, Cr53, Mn55, Fe57, Co59, Ni60, Cu63, Zn66, Rb85, Sr88, Y89, Zr90, Nb93, Mo95, Cd111, Sn118, Sb121, Cs133, Ba138, La139, Ce140, Pr141, Nd146, Sm149, Eu151, Gd157, Tb159, Dy163, Ho165, Er167, Tm169, Yb173, Lu175, Hf177, Ta181, Pb208, Th232, U238 using the standard mode. After these experiments all the isotopes that gave no signal greater than LOD in a consistent number of runs were excluded from the list. A total of 19 different isotopes was included: Li7, B11, Na23, Mg25, Si29, K39, Ca44, Sc45, Mn55, Fe57, Co59, Ni60, Cu63, Zn66, Rb85, Sr88, Sn118, Ba138, Pb208.

A set of 4 Certified Reference Materials (CRM) of known composition (Table 10) was fired using the same laser conditions as in otoliths: NIST612, NIST614, NIST616 (National Institute of Standards and Technology), and FEBS1 (National Research Council Canada) (Sturgeon, 2005). The CRMs were analysed twice using the "bracketing protocol" for which all the CRMs were analysed at the beginning and at the end of a working session (day) and every 20 line scans on otoliths.

After analysis, digital images using a Nikon optical microscope equipped with a 20X objective in dark field mode were obtained with the objective to verify the laser-scan appropriate position (Fig. 12).



Figure 12. Otoliths fired by LA-ICP-MS analyses: core and edge replica.

CRM	Concentration range (ppm)	Matrix
NIST612	10 – 80	silicon crystal
NIST614	0.5 – 50	silicon crystal
NIST616	0.008 – 30	silicon crystal
FEBS1	0.002 – 2600	otolith pressed powder (homemade)

Table 10. CRMs utilised.

4.4. Data treatment I: from cps to elemental concentrations

The *iolite* protocol was applied to calculate elemental concentration from raw data as follow. The raw intensities for all shots and isotopes were analysed using the commercial software *iolite* (v4) (Paton et al., 2011). After an interactive identification and selection of the *blank* and *plateau* of each isotope profile, the *blank* subtraction, the internal standard normalization and correcting for machine (linear fit) drift, the interface calculates elemental concentrations method. This method uses one single CRM to obtain elemental concentrations. The certified concentration value for a given element is used to directly extrapolate the concentration value of that element on the otolith according to the obtained intensities, so no calibration line is obtained to estimate concentrations.

NIST612 and Ca (40 wt.%) were used as external calibration and internal standards, respectively, to convert elemental intensities (counts per seconds) into concentrations. Internal standard also permits to correct variations in ablation and aerosol efficiency. NIST 614 was used as secondary standard.

The CRMs were analysed twice using the “bracketing protocol” for which all the CRMs were analysed at the beginning and at the end of a working session and every 20 spots, corresponding to four otoliths analysis, to correct for instrumental (linear fit) drift. The certified concentration value for a given element is used to directly calculate the concentration value of that element on the otolith according to the obtained intensities. The recoveries of the method calculated on NIST 614 was in the window 85%-115%. Detection limits (LOD) were calculated from the concentration of analyte, yielding a signal equivalent to 3x the standard deviation of the blank signal for each of the elements.

4.5. Data treatment II: statistical analyses and results

4.5.1. General overview of the statistical analyses

The general goals of the statistical analyses were: (1) to describe the multivariate microchemical composition of each GSA, (2) to evaluate the classification capability of microchemical composition (i.e., to predict subarea membership of a new fish from the multivariate microchemical composition of its otolith).

When doing multivariate analyses, it is not possible to include otoliths with missing values for some element. After a preliminary data exploration, it was evidenced only a few number of missing values in the multivariate data: when more than 4 missing elements were found per spot or when the spots were not correctly positioned, they were re-analysed. So then it was sufficient to average the values of the replicas (2 per core and 3 per edge) and fill in the few missing values with the average of the geographical sub-area to keep all 25 samples per GSA.

According to these results, the 16 isotopes retained for further analyses were: Li7, B11, Na23, Mg25, Si29, K39, Mn55, Co59, Ni60, Cu63, Zn66, Rb85, Sr88, Sn118, Ba138, Pb208.

Subsequently the data were tested for normality (by geographic group) with the MVN package and were transformed with the BoxCox function of the Caret package.

The data analyses were kept separate for core and edge.

4.5.2. Predicting GSA membership: Edge

The multivariate analyses (either, ordination and classification) were conducted by Discriminant Analysis of Principal Components (DAPC using the *ade4* package for the R software). DAPC in itself requires prior groups to be defined. However, groups are often unknown or uncertain, and there is a need for identifying clusters before describing them. This can be achieved using *k*-means, a clustering algorithm which finds a given number (say, *k*) of groups maximizing the variation between groups. To identify the optimal number of clusters, *k*-means is run sequentially with increasing values of *k*, and different

clustering solutions are compared using Bayesian Information Criterion (BIC). Ideally, the optimal clustering solution should correspond to the lowest BIC. In practice, the 'best' BIC is often indicated by an elbow in the curve of BIC values as a function of k . For edge data the optimal k values are greater than 10 indicating a strong mix of samples. Various simulations with k from 3 to 10 do not provide clear signs of correlation with the individual GSAs or groups of GSAs, therefore the same GSAs have been used as priors groups (Fig. 13).

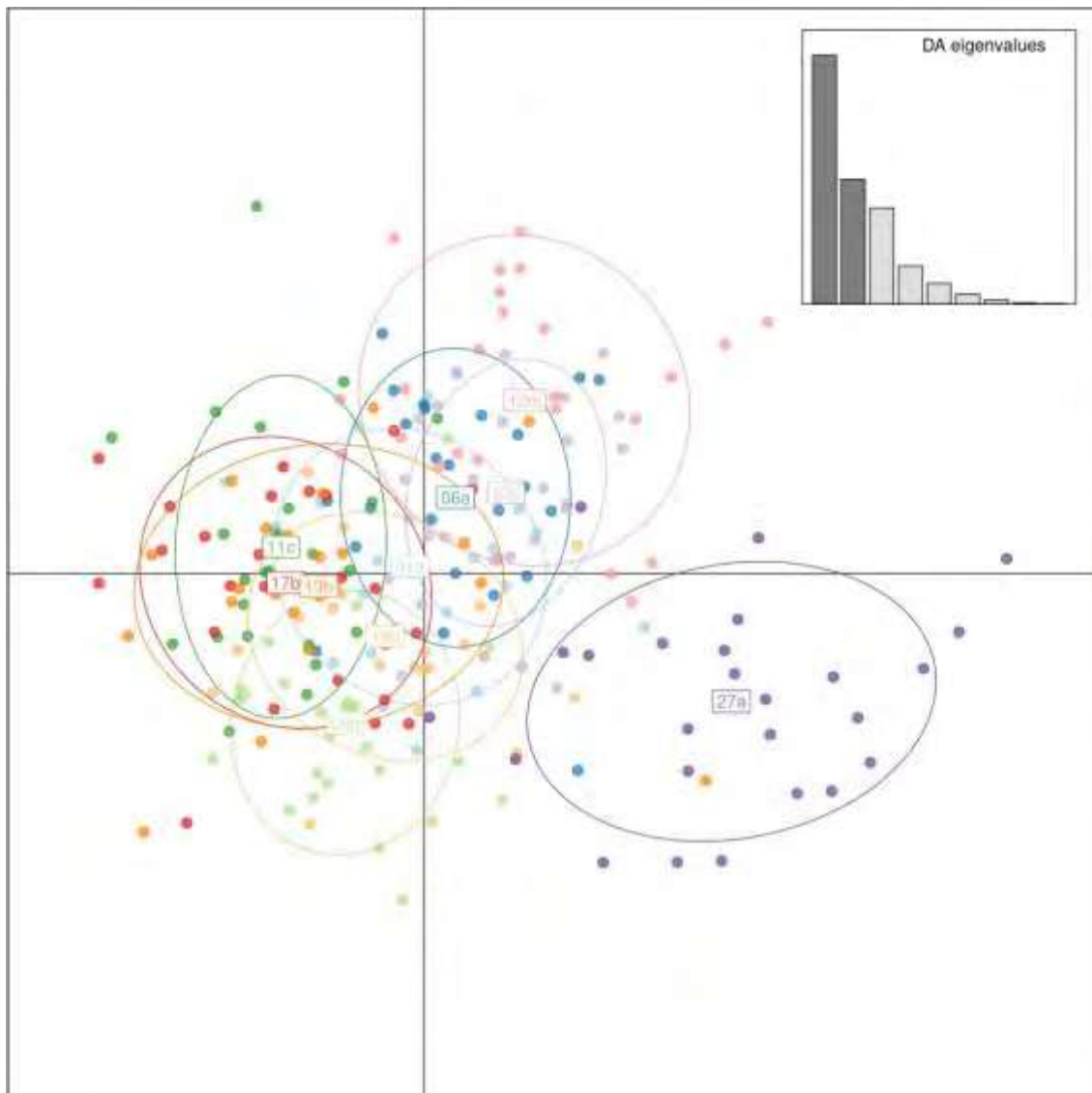


Figure 13. DAPC ordination plot: GSAs priors groups – edge data.

The elements that help discriminate groups are highlighted in Figure 14.

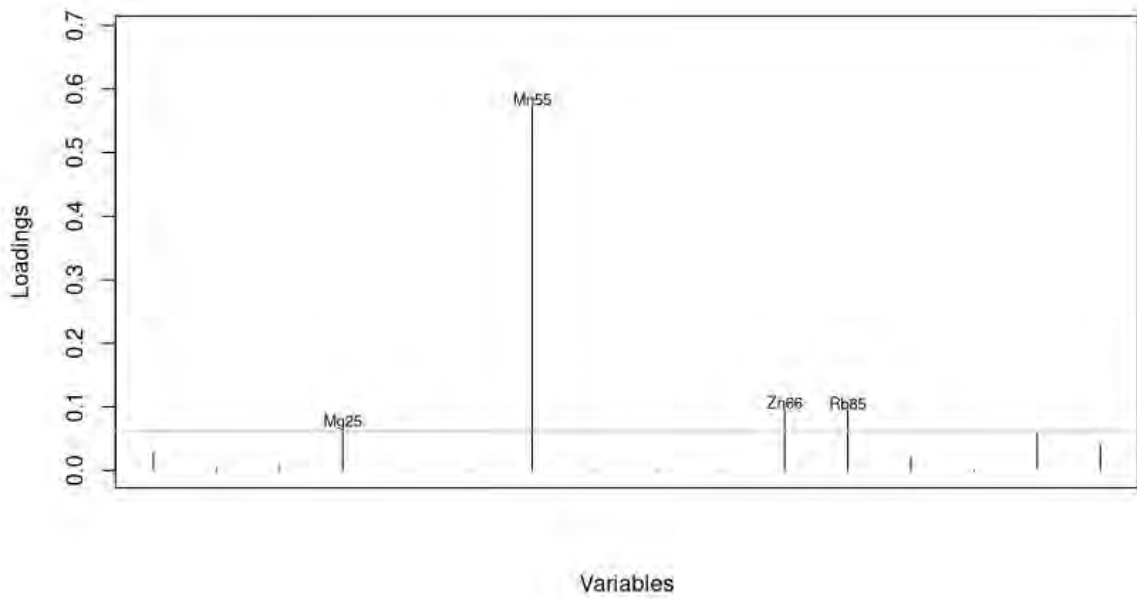


Figure 14. Loadings plot of first DAPC component - edge data.

The average percentage of correct allocation of samples in groups (GSA) is 0.63 with subareas values showed in Table 11 and Figures 15-16. Note that this value is given when no cross validation is done.

01a	06a	09a	11c	12m	17b	18b	19b	22c	27a
0.68	0.48	0.64	0.60	0.80	0.52	0.48	0.48	0.72	0.88

Table 11. Membership probability of correct allocation - edge data.

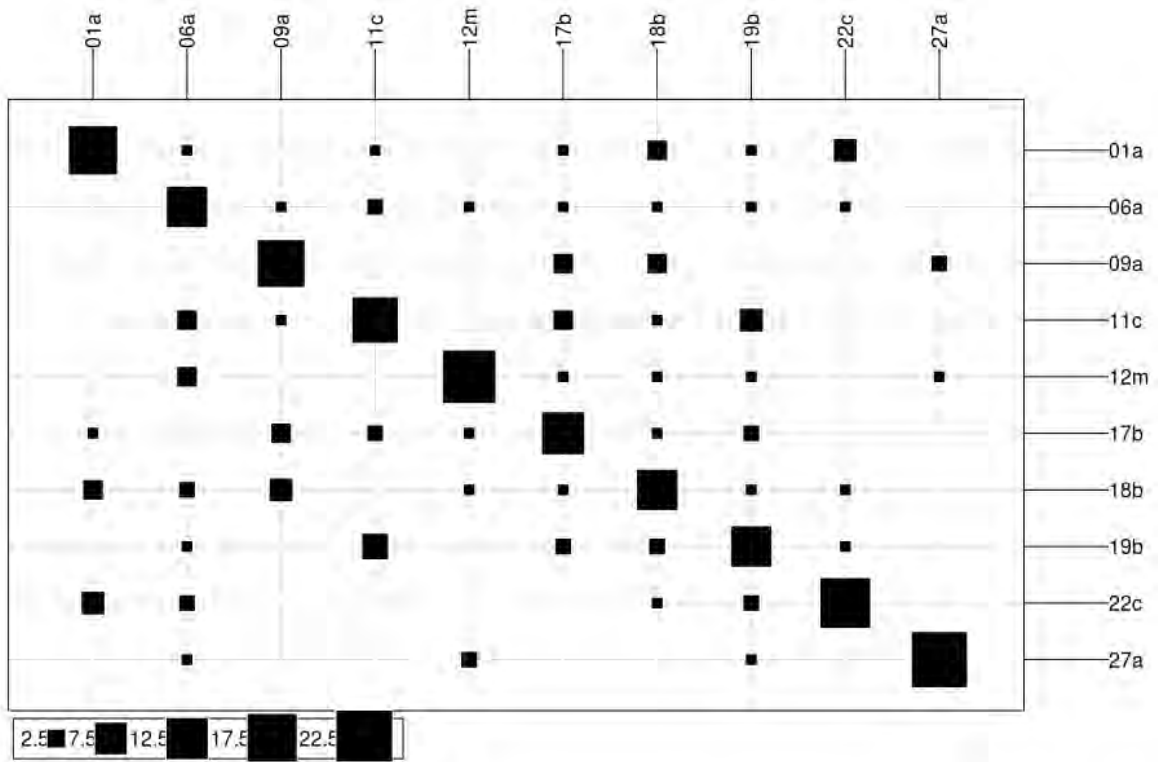


Figure 15. Membership counts plot of correct allocation:columns correspond to actual clusters and rows inferred clusters - edge data.

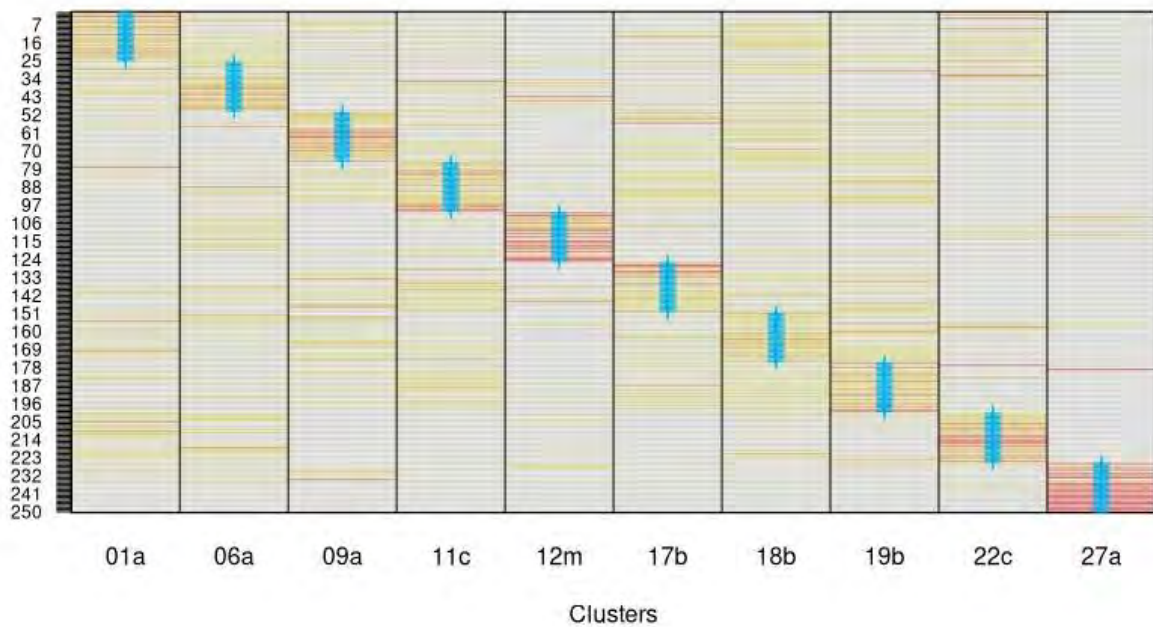


Figure 16. Individuals assignment plot: blue crosses indicate the true GSA, the red bars the highest probability of membership, the white ones the lowest - edge data.

4.5.3. Predicting GSA membership: Core

For the data relating to the core, the same procedure was followed as for the edge.

The average percentage of correct allocation of samples in groups (GSA) is 0.66 with subareas values showed in Table 12 and Figures 17-20. Note that this value is given when no cross validation is done.

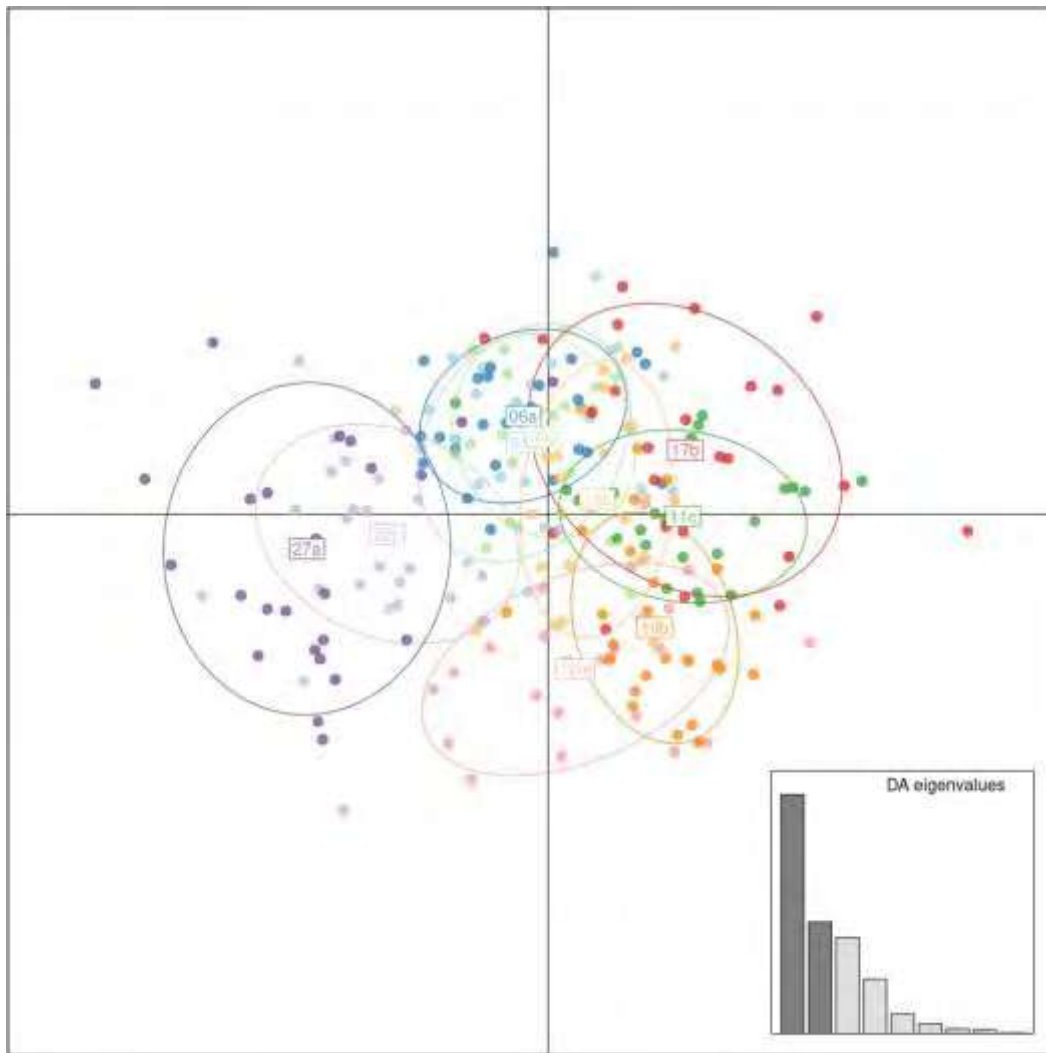


Figure 17. DAPC ordination plot: GSAs prior groups – core data.

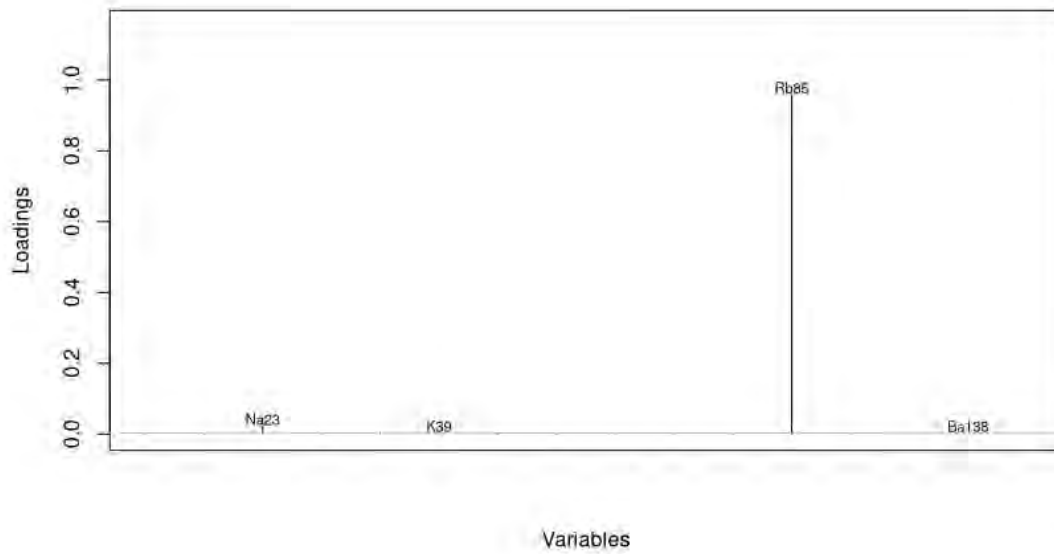


Figure 18. Loadings plot of first DAPC component - core data.

<i>01a</i>	<i>06a</i>	<i>09a</i>	<i>11c</i>	<i>12m</i>	<i>17b</i>	<i>18b</i>	<i>19b</i>	<i>22c</i>	<i>27a</i>
0.44	0.52	0.76	0.80	0.68	0.56	0.44	0.72	0.76	0.88

Table 12. Membership probability of correct allocation - core data.

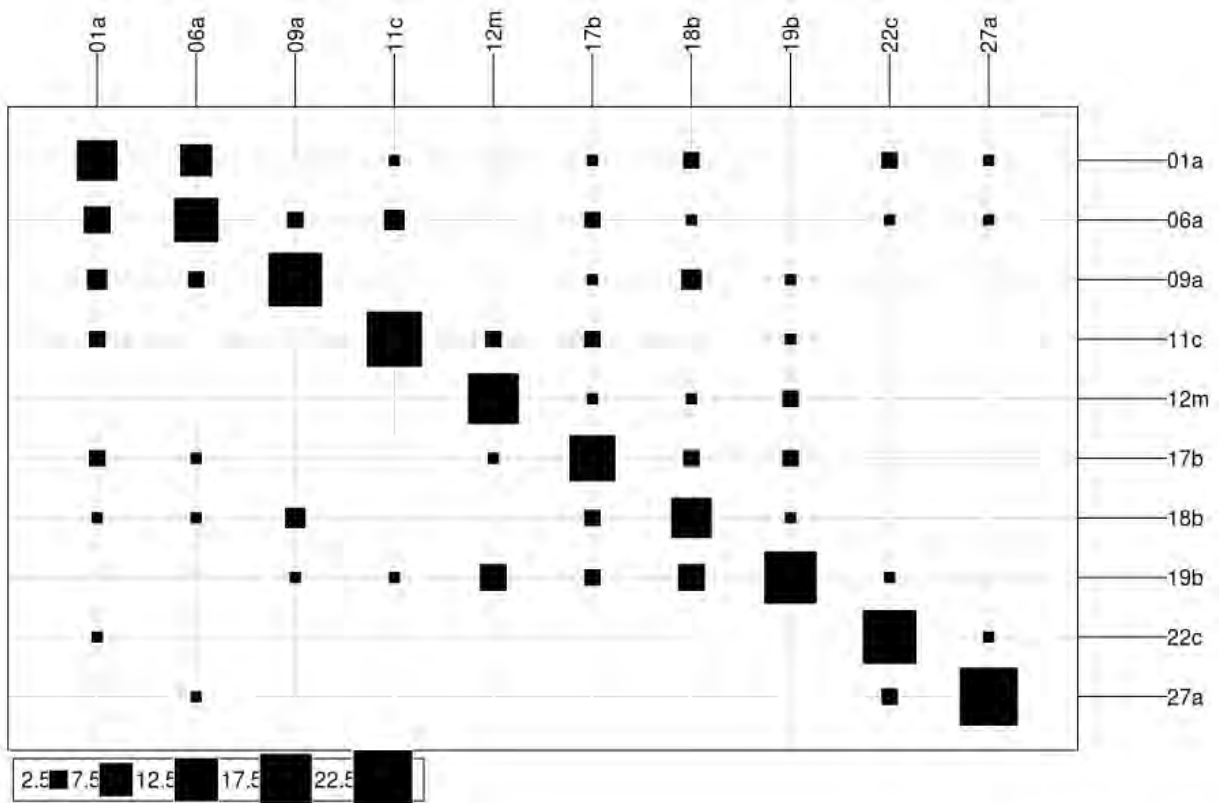


Figure 19. Membership counts plot of correct allocation: columns correspond to actual clusters and rows inferred clusters- core data.

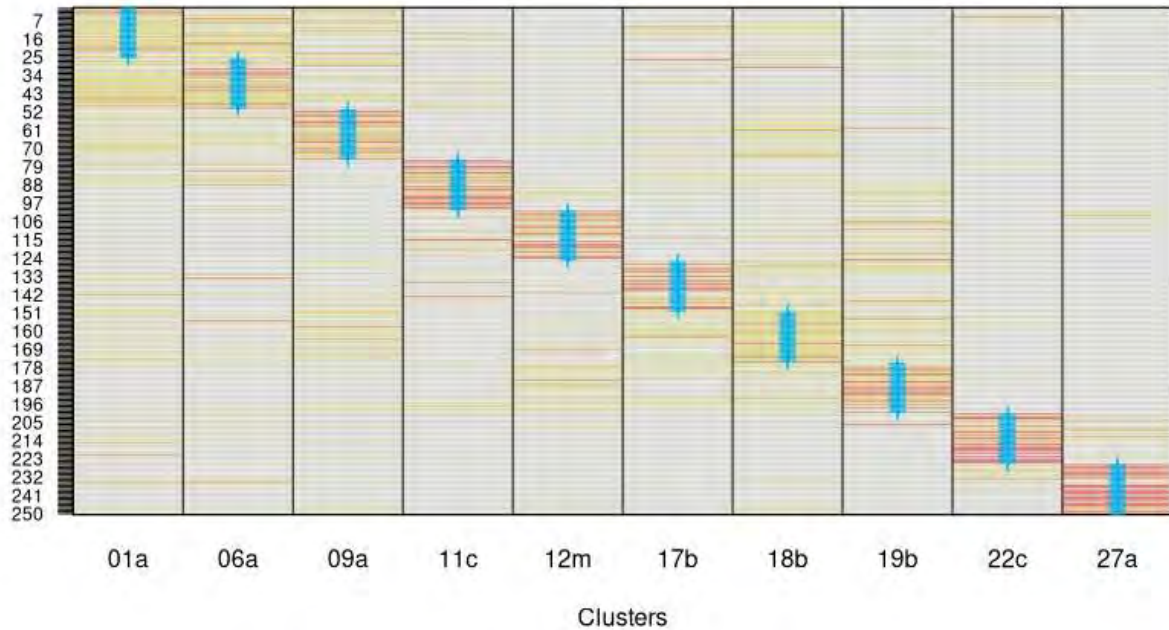


Figure 20. Individuals assignment plot: blue crosses indicate the true GSA, the red bars the highest probability of membership, the white ones the lowest - core data.

4.6. Discussion and conclusions

No previous studies to define the Red mullet population structure in Mediterranean Sea were found so this is the first attempt to define any structures of this stock by microchemistry analyses.

For other species classification accuracy using otolith microchemistry is moderate and often ranges between 60% and 80% (Arslan and Secor, 2005; Forrester and Swearer, 2002; Patterson et al., 2004; Rooker et al., 2003; Secor et al., 2002; Swan et al., 2006). This accuracy depends to a large extent on the a-priori grouping based on the hypothesis to be tested. For example, if groupings are based upon strong differences in the characteristics of water masses, known to affect some elements (e.g., areas of widely differing salinity or productivity), the classification success is optimal.

Classification success was relatively low in general but however, the hypothesis of an East-West gradient cannot be rejected in the Mediterranean, albeit the individual variability was very high decreasing the classification power of the methodology. A potential confounding process is the existence of fish endogenous effects. Certainly, it could be argued that exogenous (environmental) influence could also be associated with ontogeny, as fish change their habitats as part of their development. Red mullet undergoes vertical migrations and with respect to the coast during its development: in the juvenile stages it leads pelagic life for a few weeks and then reaches the bottom and remains in contact with it for the rest of its life but carrying out trophic and reproductive migrations. The similarities in the variations of otolith composition may also be caused by the changes in diet that are associated with ontogeny. Nevertheless, there must be a considerable endogenous influence on the elemental variation, which would explain why

the red mullet from areas with a priori different water masses had similar patterns. Further analyses within this project should aim to classify water mass characteristics at relevant depths/month per GSA. This would enable to build a priori groupings that were related to functional drivers of otolith microchemical properties.

The mean probability of assignment is 0.63 and 0.66 from edge and core respectively but with highest values on some GSAs (see Tables 11-12).

From the edge data, GSA 12m - 22c – 27a are the best differentiated from the others and this pattern is confirmed by core results too.

In respect to edge results, core data better differentiate the other GSAs highlighting a minor admixture of singles individuals (see Figures 15-16 and 19-20).

Instead GSAs 17b, 18b and 19b are those with more admixtures based on both edge and core data.

5. REFERENCES

- Arslan, Z.; Secor, D.H. (2005). Analysis of trace transition elements and heavy metals in fish otoliths as tracers of habitat use by American eels in the Hudson River estuary. *Estuaries*, 28(3): 382–393.
- Ashford, J.R.; Arkhipkin, A.I.; Jones, C.M. (2006). Can the chemistry of otolith nuclei determine population structure of Patagonian toothfish *Dissostichus eleginoides*?. *Fish Biology*, 69: 708-721, doi: 10.1111/j.1095-8649.2006.01144.x.
- Borcard, D.; Gillet, F.; Legendre, P. (2018). Numerical ecology with R. Springer.
- Bozzano, A.; Recasens, L.; Sartor, P. (1997). Diet of the European hake *Merluccius merluccius* (Pisces: Merlucciidae) in the Western Mediterranean (Gulf of Lions). *Scientia Marina*, 61: 1–8.
- Campana, S.E. (2005). Otolith elemental composition as a natural marker of fish stocks. In: S.X. Cadrin; K.D. Friedland; J.R. Waldman (Eds.), *Stock Identification Methods: Applications in Fisheries Science*. Elsevier, Academic Press, New York, pp. 227-245.
- Campana, S.E.; Fowler, A.J.; Jones, C.M. (1994). Otolith elemental fingerprinting for stock identification of Atlantic cod (*Gadus morhua*) using laser ablation ICPMS. *Canadian Journal of Fisheries and Aquatic Sciences*, 51: 1942-1950, doi: 10.1139/f94-196.
- Cartes, J.; Hidalgo, M.; Papiol, V.; Massutí, E.; Moranta, J. (2009). Changes in the diet and feeding of the hake *Merluccius merluccius* at the shelf-break of the Balearic Islands: influence of the mesopelagic-boundary community. *Deep-Sea Research Part I*, 56: 344–365, doi: 10.1016/j.dsr.2008.09.009.

- Forrester, G.E.; Swearer, S.E. (2002). Trace elements in otoliths indicate the use of open-coast versus bay nursery habitats by juvenile California halibut. *Marine Ecology Progress Series*, 241: 201-213.
- Guichet, R. (1995). The diet of European hake (*Merluccius merluccius*) in the northern part of the Bay of Biscay. *ICES Journal of Marine Science*, 52: 21–31, doi: 10.1016/1054-3139(95)80012-3.
- Higgins, R.; Isidro, E., Menezes, G.; Correia, A. (2013). Otolith elemental signatures indicate population separation in deep-sea rockfish, *Helicolenus dactylopterus* and *Pontinus kuhlii*, from the Azores. *Journal of Sea Research*, 83: 202-208, doi: 10.1016/j.seares.2013.05.014.
- Hüssy, K.; Gröger, J.; Heidemann, F.; Hinrichsen, H.H.; Marohn, L. (2016). Slave to the rhythm: seasonal signals in otolith microchemistry reveal age of eastern Baltic cod (*Gadus morhua*). *ICES Journal of Marine Science*, 73: 1019-1032, doi: 10.1093/icesjms/fsv247.
- Jochum, K.P.; Weis, U.; Stoll, B.; Kuzmin, D.; Yang, Q.; Raczek, I.; Jacob, D.E.; Stracke, A.; Birbaum, K.; Frick, D.A., Günther, D.; Enzweiler, J. (2011). Determination of reference values for NIST SRM 610–617 glasses following ISO guidelines. *Geostandards and Geoanalytical Research*, 35 (4): 397-429. doi: 10.1111/j.1751-908X.2011.00120.x.
- Jones, C.M.; Palmer, M.; Schaffler, J.J. (2017). Beyond Zar: the use and abuse of classification statistics for otolith chemistry. *Journal of Fish Biology*, 90(2), 492-504, doi: 10.1111/jfb.13051.
- Korkmaz, S.; Goksuluk, D.; Zararsiz, G. (2014). MVN: An R Package for assessing multivariate normality. *The R Journal*, 6(2):151-162, doi: 10.32614/RJ-2014-031.
- Longerich, H.P.; Simone, E.J.; Günther, D. (1996). Laser ablation inductively coupled plasma mass spectrometric transient signal data acquisition and analyte concentration calculation. *Journal of Analytical Atomic Spectrometry*, 11(9): 899-904, doi: 10.1039/JA9961100899.
- Longmore, C.; Trueman, C.N.; Neat, F.; Jorde, P.E.; Knutsen, H.; Stefanni, S.; Catarino, D.; Milton, J.A.; Mariani, S. (2014). Ocean-scale connectivity and life cycle reconstruction in a deep-sea fish. *Canadian Journal of Fisheries and Aquatic Sciences*, 71: 1312-1323, doi: 10.1139/cjfas-2013-0343.
- Martí-Puig, P.; Pérez-Mayol, S.; Serra-Serra, M.; Palmer, M. (2019). Automatic interval selection on LA-ICPMS isotopic profiles using an Extreme Learning Machine (ELM) network. In: J. Sabater-Mir, V. Torra, I. Aguiló & M. González-Hidalgo (Eds.), *Frontiers in Artificial Intelligence and Applications*; IOS Press., p. 74-83, doi: 10.3233/FAIA190110.
- Mattiucci, S.; Abaunza, P.; Ramadori, L.; Nascetti, G. (2004). Genetic identification of *Anisakis* larvae in European hake from Atlantic and Mediterranean waters for stock recognition. *Journal of Fish Biology*, 65: 495–510, doi: 10.1111/j.0022-1112.2004.00465.x.
- Maurin, C. (1965). Les merlus des mers européennes et nord-ouest africaines. *Rapports et procès-verbaux des réunions CIESM*, 18: 215-220.

- Maynou, F.; Lleonart, J.; Cartes, J.E. (2003). Seasonal and spatial variability of hake (*Merluccius merluccius* L.) recruitment in the NW Mediterranean. *Fisheries Research*, 60, 65–78, doi: 10.1016/S0165-7836(02)00062-0.
- Mercier, L.; Darnaude, A.M.; Bruguier, O.; Vasconcelos, R.P.; Cabral, H.N.; Costa, M.; Lara, M.; Jones, D.L.; Mouillot, D. (2011). Selecting statistical models and variable combinations for optimal classification using otolith chemistry. *Ecological Applications* 21, 1352–1364, doi: 10.1890/09-1887.1.
- Milano, I.; Babbucci, M.; Cariani, A.; Atanassova, M.; Bekkevold, D.; Carvalho, G.R.; Espiñeira, M.; Fiorentino, F.; Garofalo, G.; Geffen, A.J.; Helyar, S.J.; Nielsen, E.E.; Ogden, R.; Patarnello, T.; Stagioni, M.; FishPopTrace Consortium; Tinti, F.; Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23(1): 118–135, doi: 10.1111/mec.12568.
- Morales-Nin, B.; Swan, S.C.; Gordon, J.D.M.; Palmer, M.; Geffen, A.J.; Shimmield, T.; Sawyer, T. (2005). Age-related trends in otolith chemistry of *Merluccius merluccius* from the north-eastern Atlantic Ocean and the Western Mediterranean Sea. *Marine and Freshwater Research*, 56: 599–607, doi: 10.1071/MF04151.
- Morales-Nin, B.; Pérez-Mayol S.; Palmer, M.; Geffen, A.J. (2014). Coping with connectivity between populations of *Merluccius merluccius*: An elusive topic. *Journal of Marine Systems*, 138: 211-219, doi: 10.1016/j.jmarsys.2014.04.009.
- Oksanen, J.; Guillaume Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Henry, M.; Stevens, E.S.; Wagner, H. (2019). *vegan*: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- Orsi-Relini, L.; Fiorentino, F.; Cappanera, M. (1986). The timing of recruitment of *Merluccius merluccius* in the Ligurian Sea. *Rapports et procès-verbaux des réunions CIESM*, 30 (2), 224.
- Paton, C.; Hellstrom, J.; Paul, B.L.; Woodhead, J.; Hergt, J. (2011) Iolite: Freeware for the visualisation and processing of mass spectrometric data. *Journal of Analytical Atomic Spectrometry*, 26: 2508-2518, doi:10.1039/c1ja10172b.
- Patterson, H.M.; McBride, R.S.; Julien, N. (2004). Population structure of red drum (*Sciaenops ocellatus*) as determined by otolith chemistry. *Marine Biology*, 144(5): 855–862.
- Pla, C.; Vila, A.; García-Marín, J.L. (1991). Différentiation de stocks de merlu (*Merluccius merluccius*) par l'analyse génétique: comparaison de plusieurs populations méditerranéennes et atlantiques du littoral espagnol. *FAO, Rapport sur les Pêches*, 447: 87–93.
- Recasens, L. (1993). La merluza del Mediterráneo. La comunidad de que forma parte. Aspectos biológicos de la especie: alimentación, ciclo vital y reproducción. Explotación: pesquería. In: S.A. González-

Garcés & F.J. Pereiro Muñoz (Eds.), *Working group on the present state of knowledge of hake populations which live on the continental shelf on the Atlantic and Mediterranean parts of the European Union, with special attention to the Iberian Peninsula*. Centro Oceanográfico de Vigo (IEO) and The European Community Programme Air, Vigo, Spain, pp. 237–254.

- Recasens, L.; Lombarte, A.; Morales-Nin, B.; Torres, G.J. (1998). Spatiotemporal variation in the population structure of the European hake in the NW Mediterranean. *Journal of Fish Biology*, 53(2), 387–401, doi: 10.1111/j.1095-8649.1998.tb00988.x.
- Régnier, T.; Augley, T.; Devalla, S.; Robinson, C.D.; Wright, P.J.; Neat, F.C. (2017). Otolith chemistry reveals seamount fidelity in a deepwater fish. *Deep Sea Research Part I*, 121: 183-189, doi: 10.1016/j.dsr.2017.01.010.
- Rogers, T.A.; Fowler, A.J., Steer, M.A.; Gillanders, B.M. (2019). Spatial connectivity during the early life history of a temperate marine fish inferred from otolith microstructure and geochemistry. *Estuarine, Coastal and Shelf Science*, 227: 106342, doi: 10.1016/j.ecss.2019.106342.
- Roldán, M.I., García-Marín, J.L.; Utter, F.M.; Pla, C. (1998). Population genetic structure of European hake, *Merluccius merluccius*. *Heredity*, 81: 327–334.
- Rooker, J.R.; Secor, D.H.; Zdanowicz, V.S.; De Metro, G.; Relini, L.O. (2003). Identification of Atlantic bluefin tuna (*Thunnus thynnus*) stocks from putative nurseries using otolith chemistry. *Fisheries Oceanography*, 12(2): 75–84, doi: 10.1046/j.1365-2419.2003.00223.x.
- Schafer, J.L. (1997). Analysis of incomplete multivariate data. *Monographs on Statistics and Applied Probability*, 72. Chapman and Hall.
- Secor, D.H.; Campana, S.E.; Zdanowicz, V.S.; Lam, J.W.H.; Yang, L.; Rooker, J.R. (2002). Inter-laboratory comparison of Atlantic and Mediterranean bluefin tuna otolith microconstituents. *ICES Journal of Marine Science*, 59(6): 1294–1304, doi: 10.1006/jmsc.2002.1311.
- Siskey, M.R.; Lyubchich, V.; Liang, D.; Piccoli, P.M.; Secor, D.H. (2016). Periodicity of strontium: calcium across annuli further validates otolith-ageing for Atlantic bluefin tuna (*Thunnus thynnus*). *Fisheries Research*, 177: 13-17, doi 10.1016/j.fishres.2016.01.004.
- Stanley, R.R.E.; Bradbury, I.R.; Bacco, C.D.; Snelgrove, P.V.R.; Thorrold, S.R.; Killen, S.S. (2015). Environmentally mediated trends in otolith composition of juvenile Atlantic cod (*Gadus morhua*). *ICES Journal of Marine Science*, 72: 2350-2363, doi: 10.1093/icesjms/fsv070.
- Sturgeon, R.E.; Willie, S.N.; Yang, L.; Greenberg, R.; Spatz, R.O.; Chen, Z.; Scriver, C.; Clancy, V.; Lam, J.W.; Thorrold, S. (2005). Certification of a fish otolith reference material in support of quality assurance for trace element analysis. *Journal of Analytical Atomic Spectrometry*, 20(10): 1067–1071, doi: 10.1039/B503655K.

- Sturrock, A.M.; Trueman, C.N.; Daurade, A.M.; Hunter, E. (2012). Can otolith elemental chemistry retrospectively track migrations in fully marine fishes?. *Journal of Fish Biology*, 81(2): 766-795, doi: 10.1111/j.1095-8649.2012.03372.x.
- Swan, S.C.; Geffen, A.J.; Morales-Nin, B.; Gordon, J.D.M.; Shimmield, T.; Sawyer, T.; Massutí, E. (2006). Otolith chemistry: an aid to stock separation of *Helicolenus dactylopterus* (bluemouth) and *Merluccius merluccius* (European hake) in the Northeast Atlantic and Mediterranean. *ICES Journal of Marine Science*, 63(3): 504–513, doi: 10.1016/j.icesjms.2005.08.012.
- Tanner, S.E.; Vasconcelos, R.P.; Cabral, H.N.; Thorrold, S.R. (2012). Testing an otolith geochemistry approach to determine population structure and movements of European hake in the northeast Atlantic Ocean and Mediterranean Sea. *Fisheries Research*, 125–126: 198–205, doi: 10.1016/j.fishres.2012.02.013.
- Thorisson, K.; Jónsdóttir, I.G.; Marteinsdottir, G.; Campana, S.E. (2011). The use of otolith chemistry to determine the juvenile source of spawning cod in Icelandic waters. *ICES Journal of Marine Science*, 68: 98-106, doi: 10.1093/icesjms/fsq133.
- Thorrold, S.R.; Jones, C.M.; Campana, S.E. (1997). Response of otolith microchemistry to environmental variations experienced by larval and juvenile Atlantic croaker (*Micropogonias undulatus*). *Limnology and Oceanography*, 42: 102-111, doi: 10.4319/lo.1997.42.1.0102.
- Torres, G.J.; Lombarte, A.; Morales-Nin, B. (2000). Sagittal otolith size and shape variability to identify geographical intraspecific differences in three species of the genus *Merluccius*. *Journal of the Marine Biological Association of the United Kingdom*, 80: 333–342, doi: 10.1017/S0025315499001915.
- Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R.B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6): 520-525, doi: 10.1093/bioinformatics/17.6.520.
- MEDITS (2017). MEDITS-Handbook, International bottom trawl survey in the Mediterranean, Instruction Manual; G. Tserpes (Coord.), Version 9, 106 pp.
- Venables, W.N.; Ripley, B.D. (2002). Modern applied statistics with S. 4th Edition. Springer, New York. ISBN 0-387-95457-
- Witten, I.H.; Frank, E. (2005). Data mining: Practical machine learning tools and techniques, 2nd edition. Morgan Kaufmann, San Francisco.
- Wright, P.J.; Regnier, T.; Gibb, F.M.; Augley, J.; Devalla, S. (2018). Assessing the role of ontogenetic movement in maintaining population structure in fish using otolith microchemistry. *Ecology and Evolution*, 8: 7907-7920, doi: 10.1002/ece3.4186.
- Yoshinaga, J.; Nakama, A.; Morita, M.; Edmonds, J.S. (2000). Fish otolith reference material for quality

assurance of chemical analyses. *Marine Chemistry*, 69(1-2): 91-97, doi: 10.1016/S0304-4203(99)00098-5.

The logo for MED_UNITS, featuring the text "MED_UNITS" in white on a blue rectangular background. Below the logo is a decorative horizontal bar with a repeating pattern of small, colorful icons.

Specific Contract No. 03 implementing Framework Contract No.
EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-

Project MED_UNITS

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP2- OTOLITH SHAPE AND MICROCHEMISTRY ANALYSES

Deliverable 2.6. Matrix generation for joint analysis of stock delineation

Compiled by: Miquel Palmer, Sílvia Pérez-Mayol and Beatriz Morales-Nin (IMEDEA), Kelig Mahé (IFREMER) and Marco Stagioni (CoNISMa)

Responsible: Miquel Palmer (IMEDEA) and Kelig Mahé (IFREMER)

Date: 14 July 2021

Index

1. Executive summary	3
2. Introduction	3
3. General overview of the statistical analyses	4
4. Predicting population membership	5
5. Between-population differences.....	9
6. Discussion	12
7. References	13

1. Executive summary

Otolith markers (i.e., shape and chemical composition) can be used as stock identification methods, including the current delimitation of the Mediterranean management units (named GSA, Geographical Sub-Areas according to GFCM classification). The hypothesis that otolith chemical composition data and otolith shape data can be combined for improving the capability of predicting GSA membership of a given fish from its otolith features has been tested. Concerning hake, data for both shape and chemistry otoliths are available for 159 fish from 10 GSA subunits (1b, 9b, 11c, 16b, 18a, 20a, 22b, 25c, 26a, 27b). Cross-validated correct predictions of population membership inferred from Linear Discriminant Analysis (LDA) when using only shape data reached 34.6%. The same analysis using only the chemical composition of the otolith edge showed 29.9% of correct prediction. Finally, after merging shape and chemistry data, the percentage of correct prediction of population membership was 42.2%. Concerning red mullet, data for both shape and chemistry otoliths are available for 237 fish from 10 GSA subunits (1a, 6a, 9a, 11c, 12m, 17b, 18b, 19b, 22c and 27a). Cross-validated correct predictions of population membership inferred from Linear Discriminant Analysis (LDA) when using only shape data reached 23.2%. The same analysis using only the chemical composition of the otolith edge showed 47.2% of correct prediction. Finally, after merging shape and chemistry data, the percentage of correct prediction of population membership was 44.7%. Therefore, with this level of sampling, combining the two sources of data implied a slight improvement of accuracy for hake but a slight decline for red mullet. In any case, all the trials completed showed moderate success and it is necessary to increase the spatial coverage and the total number of individuals to improve the stock identification using otolith shape and chemical composition. Analysis of otolith shape alone was performed on larger samples for hake (n= 1868 from 39 subunits of GSAs in the Mediterranean Sea and from 4 ICES areas in the Atlantic Ocean) and red mullet (n= 1845 from 37 subunits of geographical subareas). The time and cost of otolith microchemistry analysis precluded to have the same coverage.

2. Introduction

The approved contract proposal indicated that the present *Deliverable 2.6* should be presented at month 20 (August 2020), but an extension of the contract, due to COVID-19 pandemic, postponed its deliver to January 2021. As indicated in *Deliverable 2.5-Report on European hake and red mullet microchemistry analysis*, Italian laboratory (CoNISMa) in charge of red mullet microchemistry, had not finished the analyses because of pandemic restrictions. For this reason, the Deliverable 2.6 was submitted as a Draft only including European hake. Therefore, this is the final Deliverable reporting the results for both target fish species.

This report has been produced thanks to the collaboration of the institutions in charge of otolith shape analysis (IFREMER) and otolith microchemistry analysis (IMEDEA and CoNISMa) for

European hake and red mullet. Results reported here obtain the data from the previously presented Deliverables D.2.4 and D.2.5 related to shape and microchemistry, respectively. Therefore, it is recommended to refer to the mentioned Deliverables to obtain detailed information about the samples/sites analyzed, the methods applied to obtain the results and the discussion of those results.

3. General overview of the statistical analyses

The general goals of the statistical analyses developed here are:

- a. to evaluate the classification capability of a given fish after merging chemical and shape data (i.e., to predict GSA membership of a fish from the shape/chemical descriptors of its otolith)
- b. to test for the existence of between-GSA differences in the multivariate shape/chemical descriptors
- c. to explore the between-GSA similarity patterns

Otolith combined information (shape and chemical data) were available for 159 European hake and 237 red mullet. Hereafter, all the analyses described are related with this subset of fishes. The number of fish per GSA subunit and subarea location are detailed at Table 1.

A) European hake:

GSA subunit	11c	16b	18a	1b	20a	22b	25c	26a	27b	9b
N	15	22	16	20	16	15	14	7	13	16

B) Red mullet:

GSA subunit	01a	06a	09a	11c	12m	17b	18b	19b	22c	27a
N	22	23	25	22	24	24	25	25	24	23

Table 1. Number of fish for each GSA subunit with both otoliths shape and microchemical data by species.

The analytical procedures for extracting the microchemical and the shape descriptors are described in detail at the *Deliverables 2.4* and *2.5*.

Briefly, external shape was described by 99 elliptical Fourier harmonics corresponding to 396 parameters. The raw data was provided from the work developed for Milestone 2.6. It is assumed that the first harmonic was forced to be invariant in order to remove any size effect. This raw data (393 columns and 159 rows for hake and 393 columns and 237 rows for red mullet) was analyzed following two strategies: First, the data was submitted to a Principal Components

Analysis (PCA) using the *rda* function from the *vegan* library (Oksanen *et al.*, 2019). The number of dimensions retained for further analyses, according with the broken stick model (as implemented by the function *bsDimension* of the *PCDimension* package; Coombes & Wang, 2019) was 6 and 9 PCs from PCA, for respectively hake and red mullet. The second analytical strategy consisted in an one-by-one increasing of the number of harmonics included in the discriminant analysis in order to find the optimal number of harmonics, according with the classification success of a discriminant analysis (details of the classification methods are provided below). The species-specific strategy showing the best classification success was used for analyzing otolith shape.

The chemical composition of a given otolith was assumed to be described by the (Box-Cox transformed) concentration ($g_{\text{element}} \cdot g_{\text{otolith}}^{-1}$ or ppm) of the following isotopes measured at the otolith edge: ^{23}Na , ^{24}Mg , ^{43}Ca , ^{44}Ca , ^{55}Mn , ^{88}Sr and ^{138}Ba for hake and ^7Li , ^{11}B , ^{23}Na , ^{25}Mg , ^{29}Si , ^{39}K , ^{55}Mn , ^{59}Co , ^{60}Ni , ^{63}Cu , ^{66}Zn , ^{85}Rb , ^{88}Sr , ^{118}Sn , ^{138}Ba and ^{208}Pb for red mullet. The different isotopes considered depended on the laser-ICPMS procedures in the different laboratories. The methodological procedures for transforming the raw chemical data (a time-series of isotope intensities obtained from the Laser Ablation Inductively Coupled Plasma-Mass Spectrometer) to the data actually used for the statistical analyses are fully detailed in the corresponding Deliverable 2.5.

4. Predicting population membership

Classification algorithms have recently experimented an unprecedented inflation. Specifically, computer intensive approaches have become popular (Mercier *et al.*, 2011). However, it has been demonstrated that conventional methods as Linear Discriminant Analysis (LDA) or Quadratic Discriminant Analysis (QDA) have better or at least similar performance than those fashion, computer-intensive methods when the number of samples is relatively small and when the data fits the underlying assumptions (Jones *et al.*, 2017). Here, in line with the results reported for chemical composition, only LDA has been completed using the functions implemented in the MASS library of the R package (Venables & Ripley, 2002).

Concerning the chemical data, before the classification analysis, the raw data was tested for normality at the within-group level using the function *mvn* from the *MVN* package (Korkmaz *et al.*, 2014). The variables (i.e., elements) showing non-normal distribution were transformed. In the case of hake, ^{24}Mg , ^{55}Mn and ^{138}Ba were submitted to the Box-Cox transformation using the function *boxcox* of the *MASS* package. In the case of red mullet, all the variables were submitted to the Box-Cox transformation. This function estimates the likelihood profile for a range of λ values. The value of λ showing the maximum likelihood was used for transforming the data using: transformed values = $\lambda^{-1}(\text{raw values}^{\lambda}-1)$. In spite that LDA is supposed to be robust against including in the analyses non-informative or correlated variables, in addition to the LDA analyses of the Box-Cox transformed data, a second series of LDA analyses were completed. In that case, the Box-Cox transformed data were first submitted to a PCA and the resulting PCA scores were

submitted to a series of LDAs. The number of PCA dimensions included in this series were one-by-one increased. This procedure ensures that noise (i.e., non-informative elements) and redundant information (i.e., pairs of highly correlated elements) are not affecting classification capability of LDA. Nevertheless, the results achieved using the PCA-LDA approach were qualitatively the same than those obtained with LDA. Therefore, here we are reporting only the latter results (i.e., LDA on element composition) because the between-population differences can be more easily related with changes in the concentration of specific elements.

4.1 Results for European hake

In the case of hake, shape data (harmonics coefficients) was first submitted to a PCA and the first 6 axes were retained according with the broken-stick criterion. Cross validated correct predictions of population membership inferred from Linear Discriminant Analysis only on shape data reached 34.6%. Note that this percentage of correct classification rate is simply the ratio between number of correctly classified cases and total number of cases (n=159). The same figure for only the chemical composition of the otolith edge reached 29.9%. Finally, after merging shape and chemistry data, the percentage of correct prediction of population membership was 42.2%. The confusion matrices of the three analysis (i.e., only shape, only chemistry and both) are presented in Table 2.

A) Chemistry data only; correct classification rate of 29.9%.

	11c	16b	18a	1b	20a	22b	25c	26a	27b	9b
11c	2	2	1	2	2	0	2	0	0	4
16b	5	11	5	3	1	0	0	0	0	3
18a	0	1	1	2	0	0	0	0	1	1
1b	3	5	3	3	1	2	6	1	1	4
20a	1	0	1	0	11	0	0	0	0	1
22b	0	0	3	4	1	9	3	1	5	0
25c	1	1	0	3	0	0	2	3	0	1
26a	1	0	0	1	0	1	0	1	1	0
27b	1	1	1	2	0	3	1	1	5	2
9b	1	1	1	0	0	0	0	0	0	0
	2/15	11/22	1/16	3/20	11/16	9/15	2/14	1/7	5/13	0/16

B) Shape data only; correct classification rate of 34.6%.

	11c	16b	18a	1b	20a	22b	25c	26a	27b	9b
11c	2	2	0	0	0	1	1	0	0	0
16b	5	12	0	0	5	3	2	2	0	9
18a	1	1	6	1	1	0	2	0	0	4
1b	0	1	4	9	1	1	1	3	2	0
20a	3	4	1	2	7	2	1	0	0	1
22b	1	3	0	3	1	5	1	0	0	1
25c	0	1	2	2	1	3	5	3	2	0
26a	0	0	1	1	1	0	0	0	0	0
27b	0	0	0	0	0	0	1	0	8	0
9b	3	0	3	1	1	0	1	0	0	1
	2/15	12/22	6/16	9/20	7/16	5/15	5/14	0/7	8/13	1/16

C) Chemistry and Shape data; correct classification rate of 42.2%.

	11c	16b	18a	1b	20a	22b	25c	26a	27b	9b
11c	3	2	1	1	1	1	0	0	0	4
16b	5	9	2	4	0	0	1	0	0	2
18a	1	2	4	1	0	0	0	0	0	3
1b	0	1	4	9	2	1	1	2	2	1
20a	2	1	2	0	11	1	2	0	0	1
22b	0	1	1	2	0	10	1	1	1	0
25c	0	0	0	1	2	2	4	2	1	1
26a	1	0	0	2	0	0	1	2	0	0
27b	0	0	0	0	0	0	2	0	9	0
9b	3	6	2	0	0	0	2	0	0	4
	3/15	9/22	4/16	9/20	11/16	10/15	4/14	2/7	9/13	4/16

Table 2. European hake confusion matrices for chemistry (A), shape (B) and combined data (C). The actual population membership is at columns and the predicted population membership are at rows. At the last row the number of correct predictions are compared with the number of fish.

4.2 Results for red mullet

In the case of red mullet, shape data (harmonics coefficients after Fourier decomposition of the otolith outline) was first submitted to a PCA and the first 9 axes were retained according with the broken-stick criterion. However, this strategy rendered a very small classification success after LDA (9%), which strongly suggests that between-GSA shape is not retained by PCA. Therefore, the alternative analytical strategy mentioned above (i.e., progressive increasing of the number of harmonics) was preferred for red mullet. The optimal number of harmonics after Fourier decomposition of the otolith outline was 8 (i.e., 32 variables). Using this alternative strategy, cross validated correct predictions of population membership inferred from Linear Discriminant Analysis only on shape data reached 23.2%. Note that this percentage of correct

classification rate is simply the ratio between number of correctly classified cases and total number of cases (n=237). The same figure for only the chemical composition of the otolith edge reached 47.2%. Finally, after merging shape and chemistry data, the percentage of correct prediction of population membership was 44.7%. The confusion matrices of the three analysis (i.e., only shape, only chemistry and both) are presented in Table 3.

A) Chemistry data only; correct classification rate of 47.2%.

	<i>01a</i>	<i>06a</i>	<i>09a</i>	<i>11c</i>	<i>12m</i>	<i>17b</i>	<i>18b</i>	<i>19b</i>	<i>22c</i>	<i>27a</i>
<i>01a</i>	9	3	0	1	0	0	3	1	4	0
<i>06a</i>	3	6	1	2	2	0	1	1	2	1
<i>09a</i>	1	0	14	2	0	5	6	1	0	2
<i>11c</i>	0	5	0	8	0	5	0	4	0	0
<i>12m</i>	0	3	1	1	16	2	1	1	1	1
<i>17b</i>	1	0	3	2	2	6	1	2	0	0
<i>18b</i>	4	2	5	1	1	1	10	2	0	0
<i>19b</i>	1	1	1	5	1	4	2	10	1	1
<i>22c</i>	3	3	0	0	0	1	1	2	16	1
<i>27a</i>	0	0	0	0	2	0	0	1	0	17
	9/22	6/23	14/25	8/22	16/24	6/24	10/25	10/25	16/24	17/23

B) Shape data only; correct classification rate of 23.2%.

	<i>01a</i>	<i>06a</i>	<i>09a</i>	<i>11c</i>	<i>12m</i>	<i>17b</i>	<i>18b</i>	<i>19b</i>	<i>22c</i>	<i>27a</i>
<i>01a</i>	5	2	1	0	4	1	1	1	3	0
<i>06a</i>	1	4	1	0	3	0	4	4	1	2
<i>09a</i>	2	3	12	2	0	3	4	2	0	2
<i>11c</i>	1	1	3	6	1	5	5	1	0	4
<i>12m</i>	3	4	0	1	6	8	1	1	3	4
<i>17b</i>	1	2	3	5	3	1	3	4	2	2
<i>18b</i>	3	1	2	1	1	4	5	4	2	1
<i>19b</i>	3	2	2	2	3	1	1	5	5	1
<i>22c</i>	3	2	1	1	0	1	1	1	7	3
<i>27a</i>	0	2	0	4	3	0	0	2	1	4
	5/22	4/23	12/25	6/22	6/24	1/24	5/25	5/25	7/24	4/23

C) Chemistry and Shape data; correct classification rate of 44.7%.

	01a	06a	09a	11c	12m	17b	18b	19b	22c	27a
01a	8	3	1	1	0	2	4	0	5	2
06a	3	8	1	1	5	1	2	1	3	3
09a	0	0	15	1	0	4	5	2	0	0
11c	0	2	0	6	0	3	1	2	0	1
12m	0	4	0	2	17	1	1	0	0	2
17b	2	0	2	5	1	7	1	2	0	0
18b	6	3	5	1	0	2	5	2	0	0
19b	0	2	1	5	0	3	5	12	1	1
22c	3	1	0	0	0	1	1	3	15	1
27a	0	0	0	0	1	0	0	1	0	13
	8/22	8/23	15/25	6/22	17/24	7/24	5/25	12/25	15/24	13/23

Table 3. Red mullet confusion matrices for chemistry (A), shape (B) and combined data (C). The actual population membership is at columns and the predicted population membership are at rows. At the last row the number of correct predictions are compared with the number of fish.

5. Between-population differences

The existence of between-population differences has been tested using Redundancy Analysis (RDA, Borcard *et al.*, 2018) using the merged data (i.e., chemistry and shape) as response matrix and the actual population membership as explanatory variable. Note that RDA in that case (factor as explanatory variable) is equivalent to a multivariate analysis of variance (MANOVA). RDA was completed as implemented in the *rda* function of the *vegan* library (Oksanen *et al.*, 2019). Variables were standardized (i.e., divided) by standard deviation, thus the same weight is given to any variable. Significance of between-population differences was assessed after 1,000 permutations. The results are shown at Table 4. Between-population differences are clearly significant.

hake	Df	Variance	P
Population	9	2.78	>0.001
Residual	144	10.21	

red mullet	Df	Variance	P
Population	9	3.83	>0.001
Residual	227	44.16	

Table 4. MANOVA (redundancy analysis) results for merged data.

Case (i.e., otolith) scores from RDA were used to estimate each Mediterranean subarea centroid for depicting between-population similarity. For completeness, an ellipse including 95% of the

otolith of each population has been added at Figure 1 for depicting both, between- and within-population variability.

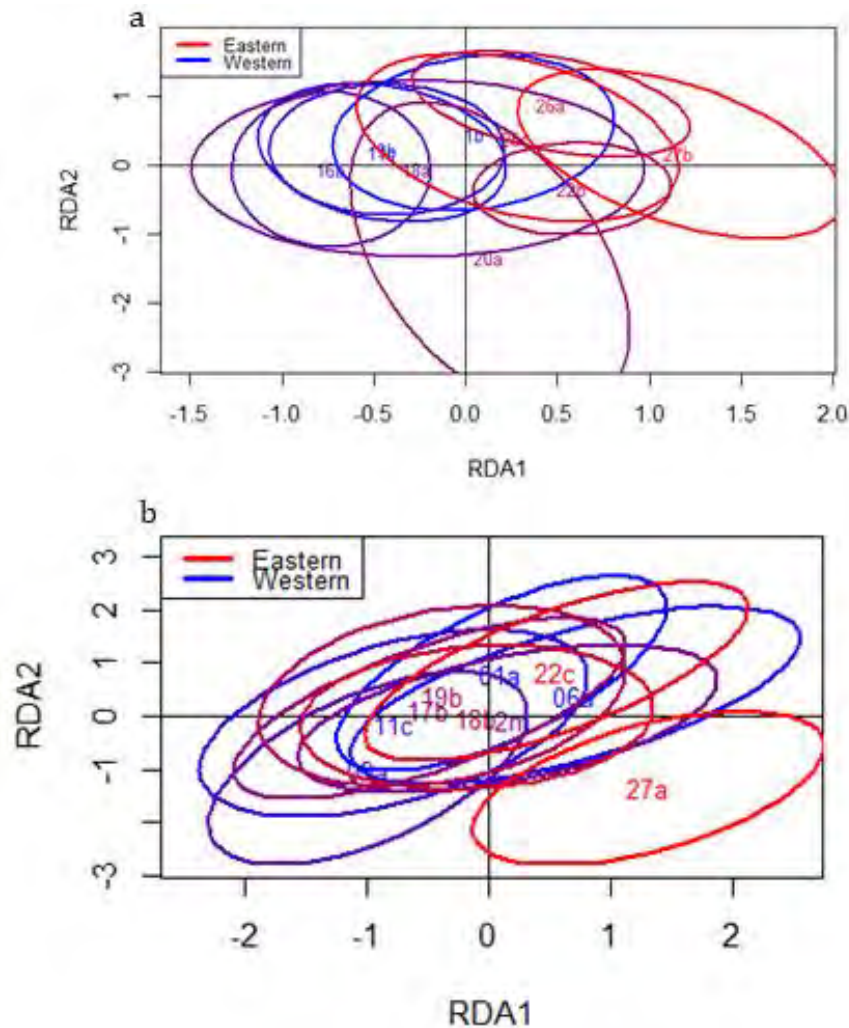


Figure 1. European hake (a) and red mullet (b) biplots (two first axes of RDA space) showing the position of the centroid of the 10 GSAs analyzed. Ellipses including 95% of the fish from each GSA subunits are added. The color in a red to blue gradient of the labels denote the latitude order of each GSA (Western GSA subunits in blue, Eastern GSA subunits in red and Central GSA subunits in purple).

In spite that the pattern depicted by Figure 1 could suggest the existence of some clusters (Western, Central and Eastern GSA subunits), between-GSA overlap was large and the between-GSA similarity pattern seems not related with geographical distance. For instance, in the case of hake, note that population 1b (Gibraltar) has microchemical composition very close to 25b population (Cyprus). At the report on microchemical composition (Deliverable 2.5), it is suggested that between-population differences in the water mass where a fish is actually living are not reflected to the chemical composition of the otolith edge, at least at the spatial scale considered. As it has been suggested (Catalán *et al.*, 2018), alternative processes related with growth rate may be also affecting microchemical composition, masking the link between water mass features and otolith composition. The large between-GSA variability is probably the main

cause of the poor power for predicting GSA membership, even after merging chemical and shape data.

Finally, the patterns of between-GSA similarity were explored by means of cluster analyses (Fig. 2 and 3). For hake the results confirm that some of the most distant populations pairs displayed close similarity in otolith shape and chemical composition (e.g., Sardinia [11c] and Cyprus [25b]). Moreover, it should be noted that the between-GSA similarity patterns suggested by shape alone and chemical composition alone differed each other and with the patterns suggested by the merged data. Therefore, the two data sets seem to reflect at least partially independent features, as it is also confirmed by the non-significant correlation between shape and chemical composition: The Mantel statistic (Manly, 2006) based on Pearson's product-moment correlation between the two matrices of similarities was $r = -0.188$ (prob = 0.795), which is clearly below the 90% upper value obtained by 1,000 random permutations (0.287). Mantel r and its significance were estimated using the *mantel* function from the *vegan* library of the R package. The results for red mullet are qualitatively similar: no clear groups or GSAs are evidenced and some pairs of distant populations display close similarity in otolith shape and chemical composition.

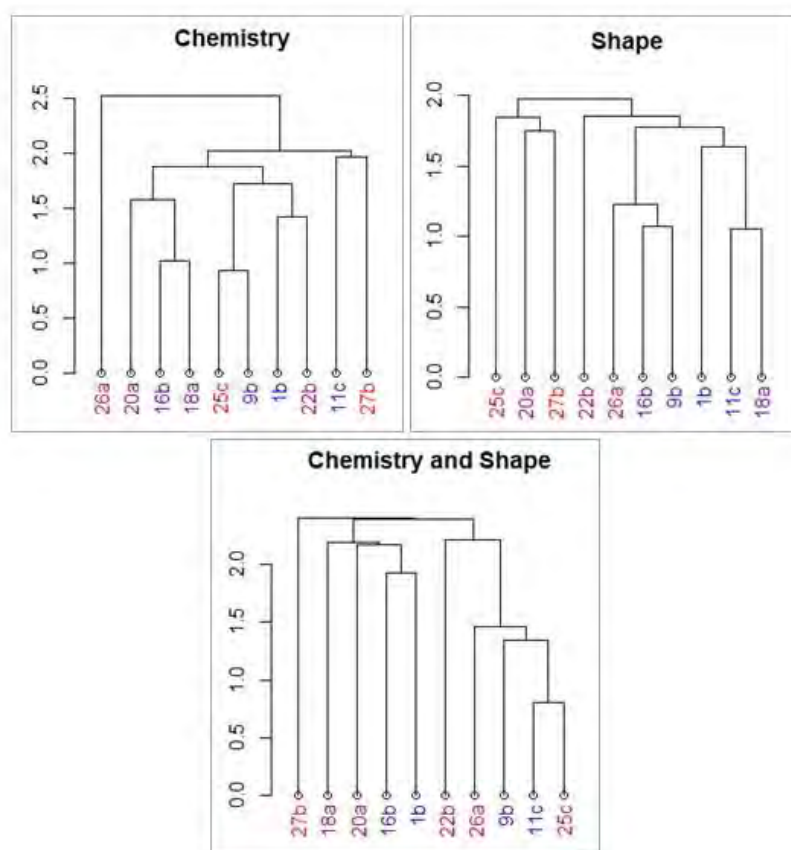


Figure 2. European hake cluster analysis of between-GSA similarities, as depicted by shape alone, chemical composition alone and by the merged data. The color in a red to blue gradient of the labels denote the latitude order of each GSA (Western GSA subunits in blue, Eastern GSA subunits in red and Central GSA subunits in purple). Note that geographical distance is better reflected by Shape alone, but that the other two data sets seem to reflect patterns that are independent from location.

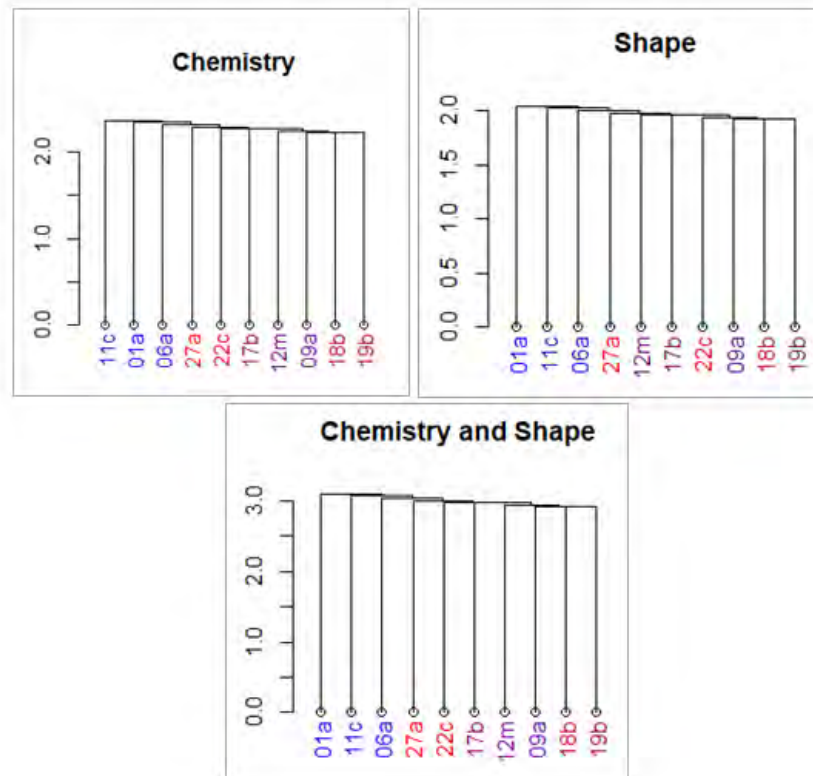


Figure 3. Red mullet cluster analysis of between-GSA similarities, as depicted by shape alone, chemical composition alone and by the merged data. The color in a red to blue gradient of the labels denote the latitude order of each GSA (Western GSA subunits in blue, Eastern GSA subunits in red and Central GSA subunits in purple). Note that geographical distance is better reflected by Shape alone, but that the other two data sets seem to reflect patterns that are independent from location.

6. Discussion

The otolith shape reflects the genetic information of a given fish (species specific component) and the phenology of each individual (i.e. environmental and biological influences). The biological influences have been reduced by the selection of specific gender and sexual development stages in each species. However, the life stage for each species was not the same, with juveniles for hake and mature adults for red mullet.

In both species the classification success has been discrete. In the case of the chemical composition of the otolith edge, that might reflect the characteristics of the water mass on which the fish was captured registering the last weeks of the fish life, the classification success was lower in hake than in red mullet. When considering both indicators, shape and chemistry, the results of the classification improved for hake but slightly decreased in red mullet regarding the results for the chemical analysis alone.

The discrete success of the classification rate could be caused by complex combinations of both GSA delimitations -based on administrative reasons and not based on oceanographic and geographic reasons- and to lack of clear boundaries in water temperature and chemical properties. A West-East gradient is suggested but with some specific GSAs breaking this general pattern. Moreover, migrations along the slope in the case of hake and other ontogenetic factors might be stronger than regional characteristics.

The effect of genetic and environmental conditions combined with the otolith shape and microchemistry factors upon the stock structure will be presented in Deliverable 4.2.

Regarding the differences in the statistical treatment of red mullet data in relation to the treatment presented in *Deliverable 2.5*, they were caused by the need of harmonizing the treatment of shape and chemical data. Specifically, we have found that the maximum classification capability of shape is reached after including the coefficients of the first 8 harmonics after Fourier decomposition of the otolith outline. Moreover, here we have used the element concentrations (and not the otolith scores on a PCA space) because the results achieved in terms of classification capability using the PCA-LDA approach were qualitatively the same than those obtained with LDA and, thus, between-population differences can be more easily related with changes in the concentration of specific elements. In any case, irrespective of the specificities of the analytical strategies, the general conclusion is the same for the two species: neither chemistry nor shape are good predictors of GSA membership for this level of sampling (159 European hake and 237 red mullet from 10 subunits of geographical subareas). In the *Deliverable 2.4*, the shape otolith is a good descriptor for stock identity for both species (1845 red mullet from 37 subunits of geographical subareas; 1868 European hake from 39 subunits of GSAs in the Mediterranean Sea and from 4 ICES areas in the Atlantic Ocean). The number of samples for otolith microchemistry was lower than for shape due to the time and cost involved in the analysis, thus the reduced number of samples decreased the capability to detect population boundaries.

7. References

- Borcard, D.; Gillet, F.; Legendre, P. (2018). Numerical ecology with R. Springer.
- Catalán, I.A.; Alós, J.; Díaz-Gil, C.; Pérez-Mayol, S.; Basterretxea, G.; Morales-Nin, B.; Palmer, M. (2018). Potential fishing-related effects on fish life history revealed by otolith microchemistry. *Fisheries Research*, 199: 186-195, doi: 10.1016/j.fishres.2017.11.008.
- Coombes, K. R. & Wang, M. (2019). PCDimension: Finding the Number of Significant Principal Components. R package version 1.1.11. <https://CRAN.R-project.org/package=PCDimension?>
- Jones, C.M.; Palmer, M.; Schaffler, J.J. (2017). Beyond Zar: the use and abuse of classification statistics for otolith chemistry. *Journal of Fish Biology*, 90(2), 492-504, doi: 10.1111/jfb.13051.
- Korkmaz, S.; Goksuluk, D.; Zararsiz, G. (2014). MVN: An R Package for assessing multivariate normality. *The R Journal*, 6(2):151-162, doi: 10.32614/RJ-2014-031.
- Manly, B. F. J. (2006). Randomization, bootstrap and Monte Carlo methods in biology. Vol. 70. CRC press.
- Mercier, L.; Darnaude, A.M.; Bruguier, O.; Vasconcelos, R.P.; Cabral, H.N.; Costa, M.; Lara, M.; Jones, D.L.; Mouillot, D. (2011). Selecting statistical models and variable combinations for optimal classification using otolith chemistry. *Ecological Applications* 21, 1352–1364, doi: 10.1890/09-1887.1.

Oksanen, J.; Guillaume Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Henry, M.; Stevens, E.S.; Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.

Venables, W.N.; Ripley, B.D. (2002). *Modern applied statistics with S*. 4th Edition. Springer, New York. ISBN 0-387-95457-

**Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/
SI2.793201 – SC03- Project MED_UNITS**

**“Study on Advancing fisheries assessment and management advice
in the Mediterranean by aligning biological and management units
of priority species”**

WP3-DELINEATE FISHING GROUNDS AND STOCK ASSESSMENT

Task 3.1 - Fishing grounds delineation

**Deliverable 3.1. – Maps on: (a) the fishing grounds
in temporal and spatial scale, (b) hot and cold
spots**

Responsible: T. Russo, CoNISMa – UNITORV; A. N. Tassetti CNR

Compiled by: Tommaso Russo, Anna Nora Tassetti, Lorenzo D’Andrea, Carmen Ferrà,
Simone Franceschini, Enrico Nicola Armelloni, Stefanos Kavadas, Irida Maina

February 2020

Version: DRAFT #1.0

Executive summary

Executive summary	2
Introduction	4
Objectives	5
Workplan	6
EU and non-EU fleets equipped with tracking devices.	6
Materials and methods	7
Analysis of AIS data: approach #1	7
Analysis of AIS data: approach #2	14
Comparison with the EMODNET footprint	18
Validation of output #1: Comparison with fishing footprint by Global Fishing Watch	18
Validation of output #2: Comparison with fishing footprint by MINOUW project (AIS and VMS data combined)	19
Multi-Criteria Decision Analysis for both large- and small-scale components of the fleets without AIS/VMS	21
Estimation of the suitability index (Sc)	22
Estimation of the activity index (Ac)	25
Data . General Fisheries Commission for the Mediterranean - GFCM fleet register	25
Quality assessment of Activity index (Ac)	31
Cascaded Multilayer perceptron network applied on environmental and fleet data for some component of the fleet without AIS/VMS	33
Reconstruction of fluxes from fishing grounds to harbors	33
Results	34
Analysis of AIS data: approach #1	34
Analysis of AIS data: approach #2	38
Comparison with the EMODNET footprint	41
Validation of output #1: Comparison with fishing footprint by Global Fishing Watch	42
Validation of output #2: Comparison with fishing footprint by MINOUW project (AIS and VMS data combined)	45
General comparison	47
Multi-Criteria Decision Analysis for both large- and small-scale components of the fleets without AIS/VMS	49

Cascaded Multilayer perceptron network applied on environmental and fleet data for some component of the fleet without AIS/VMS _____	52
Hot spots of fishing activity _____	53
Reconstruction of fluxes from fishing grounds to harbors _____	58
<i>Discussion (or Conclusion)</i> _____	62
<i>Deviation from the Workplan</i> _____	64
<i>Remedial actions</i> _____	64
<i>References</i> _____	65

Introduction

In the previous decades, one of the limitations for understanding the fishers' behavior and estimating fishing effort was the lack of sufficient data. The only available information was provided by the onboard observations, as part of established national or international programs, but this information was not enough for monitoring the dynamics of fishing fleets, since, these observations could not cover all vessels and spatiotemporal scales. In recent years, the application of a Vessel Monitoring System (VMS) in Europe allowed an analysis of the spatio-temporal distribution of fishing effort in high resolution (Lee *et al.* 2010), for vessels with total length >12 m (EC 2003). Despite the promising methodological developments and an increasing number of applications and tools, the extensive use of VMS data for scientific purposes is hindered by the difficulty of accessing control data for scientific purposes (Natale *et al.* 2015). An alternative to the VMS is represented by the Automatic Identification System (AIS), which in the EU became compulsory in May 2014 for all fishing vessels of length above 15 meters. AIS data is not related to control purposes and is exchanged also in public domains that expands its availability in respect of the VMS data and offers a very useful opportunity to analyse fishing behaviour at very detailed scale (James *et al.* 2018) and to extend the analysis at supra- national level.

Despite the high significance of small-scale fisheries in the Mediterranean Sea, the vast majority of the small-scale vessels do not meet the criteria of carrying the VMS or AIS. Therefore, there is no actual estimate of their fishing footprint on a spatial scale. Moreover, the coverage of VMS and AIS data for the southern/eastern Mediterranean (non-EU) countries is also not sufficient to provide an estimation of fishing effort for bottom trawlers. Given that primary data on fishing vessels locations (VMS, AIS) were not available for small-scale vessels and bottom trawlers on all spatial scales, a methodological approach based on a Multi-Criteria Decision Analysis (MCDA - Kavadas *et al.* 2015) has been employed to estimate a fishing pressure index for the small-scale fishery (SSF-FPc) and bottom trawlers (OTB-FPc) in the Mediterranean Sea (OTB-FPc was estimated for non-EU countries and SSF-FPc for both EU and non-EU countries as well).

Objectives

The objective of this task is to identify and characterize fishing grounds over the Mediterranean waters, including where possible non-EU fleets, which is an aspect of great importance nowadays. Also the outputs of the present task will be used in the framework of task 3.3 to better define the input data and modelling approaches of stock assessments to be carried out for the target species. Although no explicit and unambiguous definition of “fishing ground” exists, it is largely acknowledged that a fishing ground represents an area where fishing activity is routinely carried out (on a yearly or seasonal scale) as a result of a fisherman strategy aimed at maximizing yields of some target species. To identify fishing grounds different methods are available. The data to be used for the implementation of this task mainly concerns Automatic Identification System (AIS) and Vessel Monitoring System (VMS), experimental data from surveys, fleet observer data, fisheries statistics data (i.e. fishing fleet capacity, distribution by port, etc) and expert judgment approaches. The analysis of the fishing footprint, for the fleets equipped by remote tracking devices (VMS, AIS or both) will be mainly based on AIS data (Natale *et al.* 2015), which will be processed using standard routines developed in R (Russo *et al.* 2014; Ferrà *et al.* 2018). Considering that AIS data are characterized by different coverage issues (i.e. incomplete representativeness of fishing operations, poor coverage for some fleets - see Russo *et al.* 2016; Shepperson *et al.* 2018), also VMS data will be used where needed to improve, at the best possible level, the source dataset.

The results of the effort spatial analysis will be validated by comparing them with fishing footprint of the following data set information: EMODNET (<https://www.emodnet.eu/>), Global Fishing Watch (<https://globalfishingwatch.org/>) and MINOUW project (<http://minouw-project.eu/>).

To estimate a fishing pressure index for non-EU countries for both large- and small-scale components of the fleets, Multi-Criteria Decision Analysis (MCDA) will be applied in the case that AIS and VMS data are not available according to a stepwise proposed by Kavadas *et al.* 2015. As regards estimation of the suitability and activity indexes (Sc and Ac acronyms respectively) the following criteria will be taken into account: bathymetry, distance from coast, environmental information and legislation. Moreover, since the coverage of the VMS for the length class [12–15 m] of the major part of the European fleet operating in the Mediterranean Sea is far from being complete; a Cascaded Multilayer Perceptron Network (CMPN) by (Russo *et al.* 2019b) will be applied to predict the spatial distribution of fishing effort for this class.

In order to evaluate potential clusters in the distribution of the fishing activity of each gear and identify statistically significant hot spots, a Hot Spot Analysis will be performed using the Getis-Ord G_i^* statistic. In addition, a modelling approach will be applied to reconstruct the fluxes of catches from fishing grounds to harbours using the additional information provided by logbook data or other information about landings (Gerritsen and Lordan 2011; Russo *et al.* 2018).

Workplan

According to the original program, the assessment of fishing intensity (i.e. as spatial pattern of effort on a yearly base) should be performed for EU and non-EU fleets and for fishing vessels with and without tracking devices (i.e. VMS and AIS).

The methods to pursue this aim are summarized in [Table 1](#).

Table 1 -Summary of the selected methods

Tracking devices	With AIS/VMS	<ul style="list-style-type: none"> • VMSbase platform (Russo et al., 2014; Russo et al., 2016, Russo et al., 2019a); • The procedure applied in Galdelli et al., 2019, Tassetti et al., 2019, Ferrà et al., 2018
	Without AIS/VMS	<ul style="list-style-type: none"> • Multi-Criteria Decision Analysis (MCDA) applied on environmental and fleet data (Kavadas et al., 2015) for both large- and small-scale components of the fleets; • Cascaded Multilayer perceptron network applied on environmental and fleet data (Russo et al., 2019b) for the large-scale component of the fleets.

EU and non-EU fleets equipped with tracking devices.

The analysis of the fishing footprint, for the fleets equipped by remote tracking devices (VMS, AIS or both) will be mainly based on AIS data. A preliminary analysis of all the available AIS data for the year 2017 allowed to identify fishing vessels belonging to almost all the Mediterranean countries, but the coverage of AIS data for non-EU countries is scant. However, the AIS dataset will be processed as a whole (including data for both EU and non-EU countries). The analysis of AIS data will be firstly based on the VMSbase platform (Russo et al. 2011a, b, 2014, 2016, 2019a). This will allow validating the output, in some areas, by comparing the pattern obtained using only AIS data with the one obtained integrating AIS and VMS data. Secondly, the procedure applied in Ferrà et al. 2018; Tassetti et al. 2019; Galdelli et al. 2019 will be applied on AIS data. The final output of these analyses will be represented by a series of grids, coherent with the standard defined by the GFCM (<http://www.fao.org/gfcm/data/maps/grid/en/>), and populated with different metrics of fishing effort (i.e. fishing time in hours, fishing days, and nominal effort) and disaggregated by country, gear type (*métier* of level 4¹) and fleet segment.

According to the workflow described in the project documents and in the Inception Report, the output obtained from the analysis of AIS data will be compared, at least in some areas, with those obtained in other projects through the analyses of integrated VMS/AIS dataset (i.e. the H2020 Project MINOUW) and with previous studies based on AIS data (i.e. FP7/H2020/RTD projects, EMODnet MedSea Checkpoint²).

¹ <https://datacollection.jrc.ec.europa.eu/wordef/fishing-activity-metier>

² http://www.emodnet-mediterranean.eu/mainstream_and_grey_literature_publications

Materials and methods

Analysis of AIS data: approach #1

The VMSbase platform (Russo et al. 2014) was used to process the AIS dataset provided by the CNR and containing the AIS pings for the whole Mediterranean fishing fleet. First of all, the AIS dataset was inspected, and some summary statistics were prepared (Table 2).

Table 2 - Main statistics for the native AIS dataset

Statistic	Value
Num of MMSI	5082
Num of Callsigns	7281
Num of Names	5585
Num of pings	70815465

According to the "M.585: Assignment and use of identities in the maritime mobile service"³ the initial digits of an MMSI categorize the identity. In particular, the meaning of the first digit is:

First digit	Meaning
0	Ship group, coast station, or group of coast stations
1	For use by SAR aircraft
2	Europe (e.g., Italy has MID 247; Denmark has MIDs 219 and 220)
3	North and Central America and Caribbean (e.g., Canada, 316; Panama, 351 through 357, plus 370 through 373; United States, 303(Alaska), 338(domestic), plus 366 through 369)
4	Asia (not the southeast) (e.g., PRC, 412, 413, and 414; Maldives, 455)
5	Oceania, & (Australia, 503; New Zealand, 512)
6	Africa (Eritrea, 625)
7	South America (Peru, 760)
8	Handheld VHF transceiver with DSC and GNSS
9	Devices using a free-form number identity: <ul style="list-style-type: none"> ◦ Search and Rescue Transponders (970yyzzzz) ◦ Man overboard DSC and/or AIS devices (972yyzzzz) ◦ 406 MHz EPIRBs fitted with an AIS transmitter (974yyzzzz) ◦ craft associated with a parent ship (98MIDxxxx) ◦ navigational aids (AtoNs; 99MIDaxxx)

Using this reference, the vessels in the AIS dataset were grouped by area of origin (Fig. 1).

³<http://www.legislation.gov.uk/eur/2019/838/annex/division/1/division/1.2/data.xht?view=snippet&wrap=true>

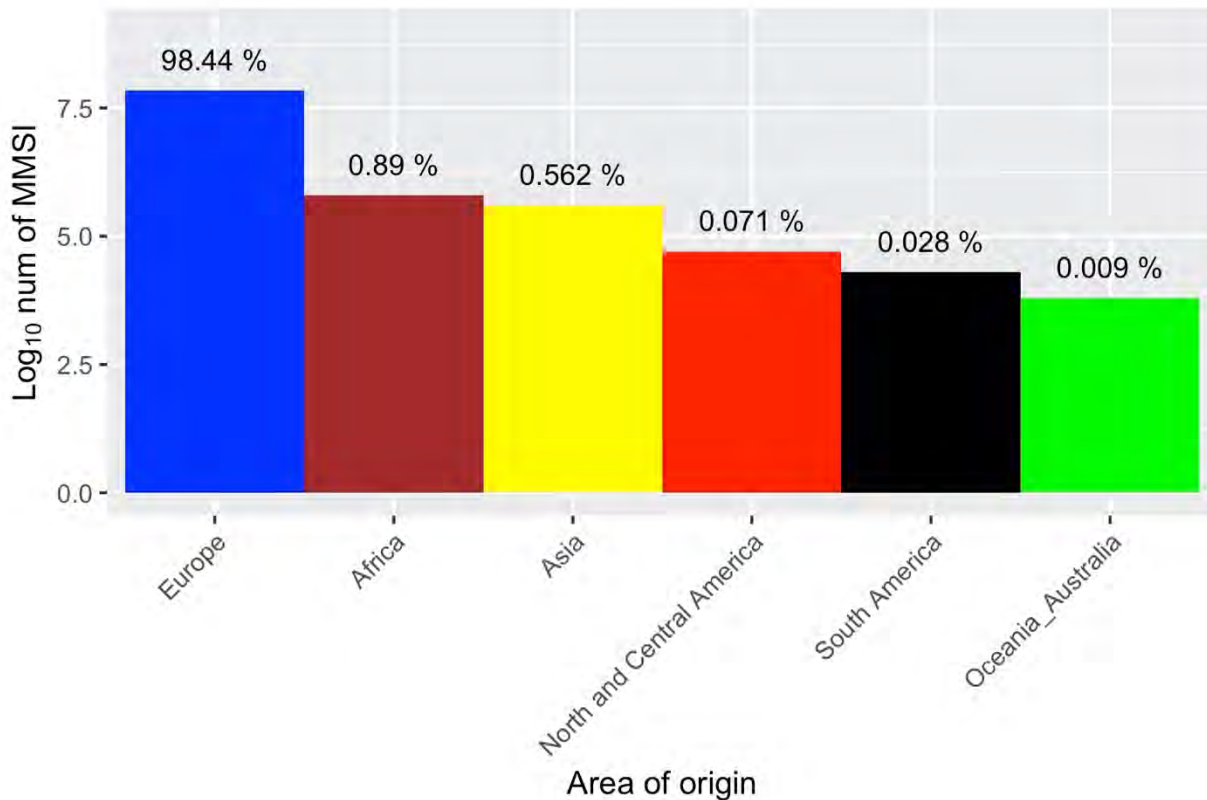


Figure 1 – Barplot of the number of AIS records by Area of origin, as assigned using the first digit of the MMSI code.

However, a preliminary graphical inspection of data by vessels allowed to detect that this kind of reference is not reliable and “stable”, and thus the assignment of vessels to their respective area of origin could not be based on this formal approach.

A Register of AIS vessels in this dataset, as identified by univocal MMSI, was compiled. The analysis of this dataset immediately allowed to detect some major issues in the structure and quality of the native AIS dataset (Table 3).

Table 3 – Some examples of the main issues detected in the AIS dataset

MMSI	NAME	CALLSIGN	ISSUE
237333000	ANEMOS	SV6340	Same vessel with different MMSI, corresponding to different area of origin, and different CALLSIGN
237333700	ANEMOS	6340	
237333700	ANEMOS	SV6340	
637333000	ANEMOS	SV6340	
637333700	ANEMOS	6340	
247139500	M/P AFRODITE PESCA	IVFT	Same vessel with different MMSI corresponding to different area of origin
347139500	M/P AFRODITE PESCA	IVFT	

In particular:

The first digit of the MMSI cannot be used to automatically split the AIS data by area of origin;

Several fishing vessels are associated to different MMSI, CALLSIGN, NAMES or combinations of these fields;

The previous issue determines the occurrence of “duplicated” vessels, that is vessels occurring in the native AIS dataset with different identifiers.

A two-step procedure was applied to fix these issues and prepare a “stable” version of the AIS dataset.

In the first step, an automatic procedure was applied to identify the AIS records to be checked. In particular, a measure of similarity between different MMSI codes, different NAMES and different CALLSIGNS was computed using the function “stringdist” of the R package “stringdist” (van der Loo et al. 2019). This function allows to compute pairwise string distances between strings considering deletion, insertion, substitution and transposition of letters. For each pair of univocal combinations of AIS MMSI, NAME and CALLSIGN codes, the three distances were computed together with their sum (the “global” distance) – Fig. 2.

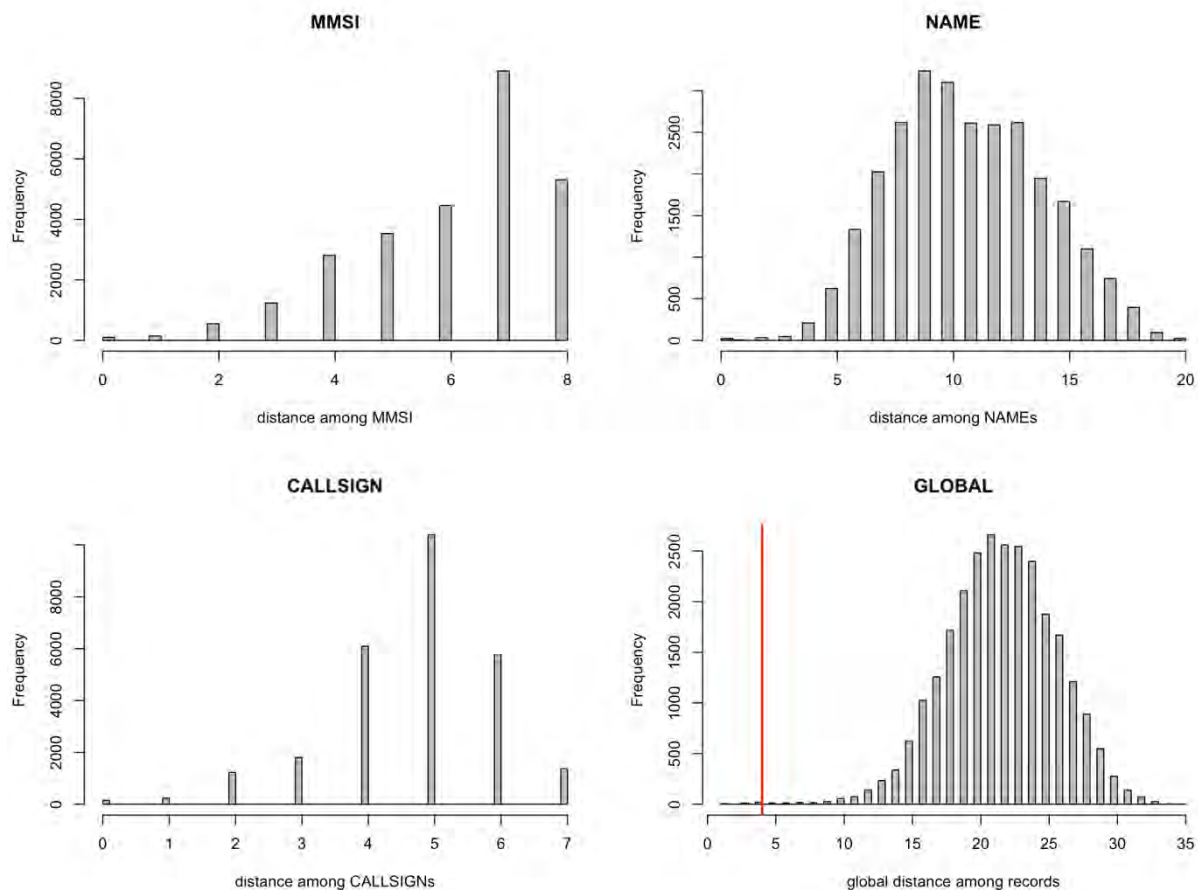


Figure 2 -Histograms of values for the three vectors of distances computed, on the pairs of univocal combinations of AIS MMSI, NAME and CALLSIGN codes, using the R function stringdist. The GLOBAL distance among pairs was computed as the sum of distances for MMSI, NAME and CALLSIGN.

A number of 122 pairs was selected, from the whole set of 27225 pairs of univocal combinations of AIS MMSI, NAME and CALLSIGN codes, using a threshold of four on the global distance among pairs. The codes of AIS record for these 122 pairs were manually inspected and compared, also considering the corresponding spatial and temporal patterns of AIS pings (see next step);

A visual approach was adopted to individually check and compare the patterns of AIS pings and the corresponding activity, in space and time, of each MMSI code and, in particular, for the 122 pairs of vessels identified above. Some examples of the “summary map” used for this visual check are reported in Fig. 3.

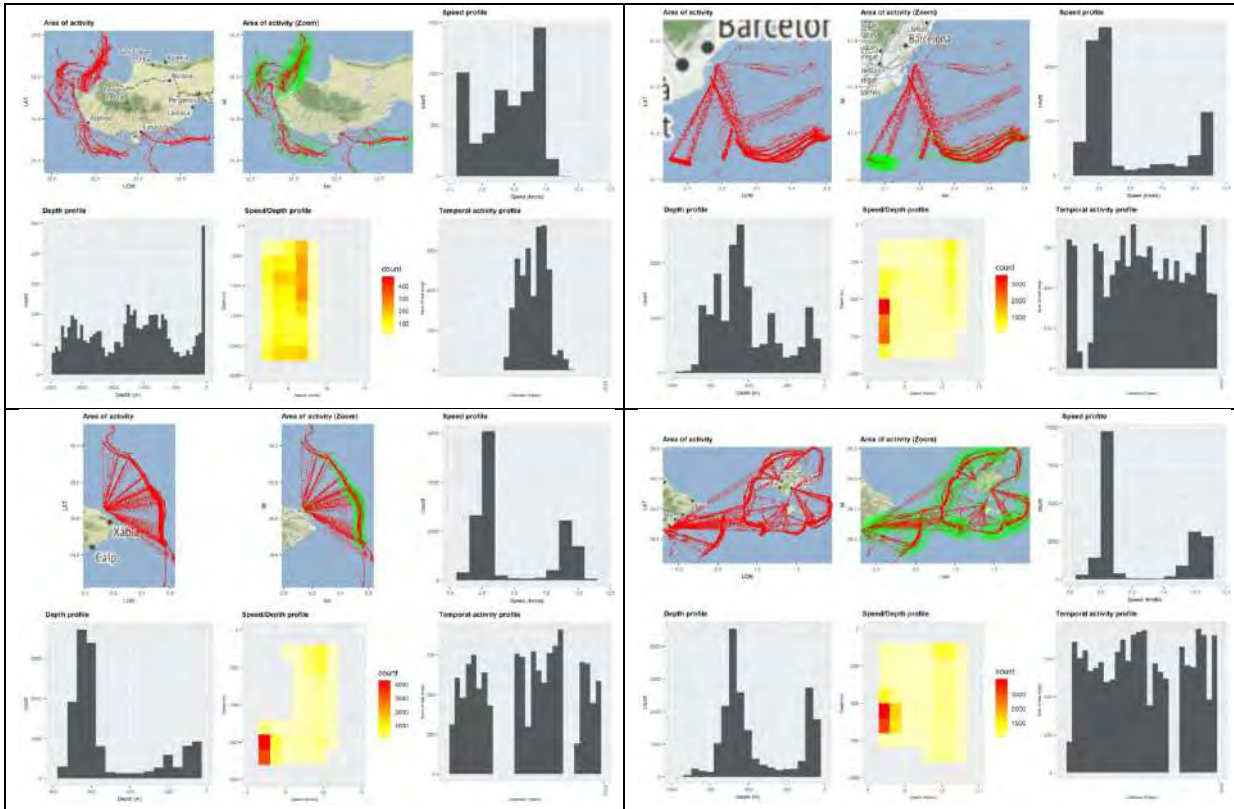


Figure 3 -Samples of the summary maps used to visually check the AIS data by vessels and pairs

Each of these summary maps was composed ad a panel of different subplots. The first subplot (Area of activity) provides a shot of the distribution of pings, while the second subplot (Area of activity (Zoom)) integrates a density plots aimed to help the reader in the identification of the fishing gear and of the main fishing grounds. Speed and depth profiles (see the corresponding subplots) show the distribution of pings with respect to their values in these fields, while the combined Speed/depth profile subplot shows the matrix plot of these combined fields. The last subplot (Temporal activity profile) described the temporal distribution of pings during the year 2017.

The whole AIS dataset was cleaned, checked and validated using this supervised procedure in which the researchers carefully used all these inputs to fix the issues described above and including “duplicated” vessels and “mismatch” between MMSI digits and area of activity.

It is worth noting that, during this cleaning, a number of strange AIS records were detected, mainly belonging to fishing vessels from Asia and Oceania.

Table 4 – Some main statistic for the Europa and Africa AIS datasets, for the year 2017, as obtained at the end of check and cleaning two-step procedure described above

Area	Statistic	Value
Europa	Number of vessels	4,395
	Number of pings	69,716,966
	Mean of the number of pings by vessel	15084/16041.05
	Mean /sd number of days with activity by vessel	136.10/99.53
Africa	Number of vessels	128
	Number of pings	624161
	Mean/sd number of pings by vessel	4876.29/12022.37
	Mean/sd number of days with activity by vessel	28.93/55.84

At the end of this cleaning phase, the two AIS datasets for Europa and Africa were submitted to a standard processing through the platform VMSBASE (Russo et al. 2014, 2016).

The details of this processing are extensively described and applied in different papers (Russo et al. 2011a, b, 2014, 2016).

A summary of the workflow could be itemized as follows:

1. Data Cleaning: identification and flagging of duplicates and erroneous pings;
 2. Track Cutting: temporal sorting of pings for each single vessel, followed by the identification of in-harbour positions and then by the identification of "tracks", that is, series of temporally-ordered pings describing a single vessel trip, starting and ending in a given harbour;
 3. Track Interpolation: standardization of pings frequency (from native/variable pings rate to a fixed value of 5 minutes). This step also allows synchronizing the temporal coordinates of the pings for each area/case study;
 4. Assign bathymetry: Automatic download of the provided data by means of the NOAA (National Oceanic and Atmospheric Administration of the U.S. Department of Commerce) web servers and assignment of an estimated value of sea bottom depth to each ping in the databases. The NOAA bathymetric web servers will be accessed through the functionalities provided by the marmap R package (Pante et al. 2019).
 5. Fishing Points Identification: classification of interpolated pings as "Fishing" or "not fishing" (which corresponds to various activities, including steaming) on the basis of case study-specific filter for speed and depth. This step allows identifying hauls within each track (fishing trip)
5. Assignment of Fishing gear: this critical step was carried out using multiple sources of information. Namely:
- The case study-specific Fleet registry prepared for the MINOUW projects;
 - The global database of AIS vessels provided by Global Fishing Watch⁴;
 - The output of the Artificial Neural Network described in Russo et al. 2011b
 - The vessel-specific summary maps described above;

⁴ <https://globalfishingwatch-org>

The photograph (if any) of the fishing vessel available through the VESSEL DATABASE of the Vessel Finder website⁵. In some cases (see Fig. 4), these photos help because the fishing gear is directly observable.



Figure 4 -Samples of photographs from VESSEL DATABASE of the Vessel Finder website (<https://www.vesselfinder.com/vessels>).

Each vessel in the Africa and Europe datasets was classified with respect to the schema of *métiers* defined within the Data Collection Framework for the Mediterranean and Black Sea⁶. In particular, the classification was limited to the Level 3 (Gear groups) and the following Classes were considered (Table 3):

⁵ <https://www.vesselfinder.com/vessels>

⁶ <https://datacollection.jrc.ec.europa.eu/wordef/fishing-activity-metier>

Table 5 – Extract of the schema of *Métiers* defined within the Data Collection Framework for the Mediterranean and Black Sea (<https://datacollection.jrc.ec.europa.eu/wordef/fishing-activity-metier>). In particular, the classification was limited to the Level 3 (Gear groups) and only the Gear groups were considered. In particular, Purse Seiners (PS) and Mid-water Pair Trawlers (PTM) were not considered since they target small pelagics while Beam Trawl (TBB) was not considered since it does not target the species of interest for this study

Level 1	Level 2	Level 3	Level 4
Activity	Gear classes	Gear groups	Gear type
Fishing activity	Trawls	Bottom trawls	Bottom otter trawl [OTB]
			Multi-rig otter trawl [OTT]
			Bottom pair trawl [PTB]
			Beam trawl [TBB]
	Hooks and Lines	Longlines	Drifting longlines [LLD]
			Set longlines [LLS]
	Nets	Nets	Trammel net [GTR]
			Set gillnet [GNS]
			Driftnet [GND]
	Traps	Traps	Pots and Traps [FPO]
			Fyke nets [FYK]
			Stationary uncovered pound nets [FPN]

After gear identification and classification of fishing vessels with respect to their gear groups, interpolated pings were classified with respect to the following classes: “in harbor”, “steaming”, and “fishing”. The last class comprises the fishing set positions, identified using gear-specific speed and depth filters.

The number of fishing day by cell was computed, as a measure of the total annual fishing effort, using the procedure of the R package fecR (Scott et al. 2017).

The final output of this analysis was represented by a set of four maps representing the distribution of the annual fishing effort (year 2017) with respect to the 1 Km square grid used for the EMODNET project.

Analysis of AIS data: approach #2

The procedure described in Ferrà *et al.* 2018; Tasseti *et al.* 2019; Galdelli *et al.* 2019 was applied on terrestrial AIS (t-AIS) data to yield a single database of fishing vessels, trips and fishing operations for the whole Mediterranean basin.

After a preliminary data inspection, data were prepared and pre-processed as follow:

1. Raw Data Cleaning: identifying and editing outliers, such as pings in land and/or far away from preceding and following transmissions;
2. Fishing trip identification: partitioning strings of consecutive AIS records for each unique MMSI into individual trips, from the time vessels leave a port to the time they return to a port. An ad-hoc algorithm was developed in Matlab to identify individual trips (including calculation of fields, algorithmic rules, and integration of data sources) and it required a number of nuanced steps to fix data issues, such as data gaps, missing AIS transmissions in close proximity to ports and vessels steaming near the coast.

Fishing trips were stored in the spatial database along with inherited geographical attributes (ports, countries and GSAs of departure and landing) and additional trip metrics were computed such as number of pings, distance, duration and percentage of data gaps due either to intentional switch off by fishermen or transmission issues (i.e. in regions with high maritime traffic or missing terrestrial receivers along the coastlines).

Some summary statistics were prepared for the final fleet that was classified:

Table 2 - Main statistics for the cleaned and pre-processed AIS dataset

Statistic	Value
Num of MMSI	3980
Num of fishing trips	342919
Num of departure/landing ports	1058
Num of GSAs	28
Num of departure/landing Countries	19

It is worth noting that, during this cleaning, a number of strange AIS records were detected, mainly belonging to fishing vessels from China and other non-Mediterranean countries (classified as other in Fig. 5) as well as a few very inactive likely fishing vessels broadcasting for only a few hours or a few days in the year.

By fleet, Italy has the most fishing vessels adopting AIS (39%), followed by Spain (16%), Turkey (14%) and Greece (10%).

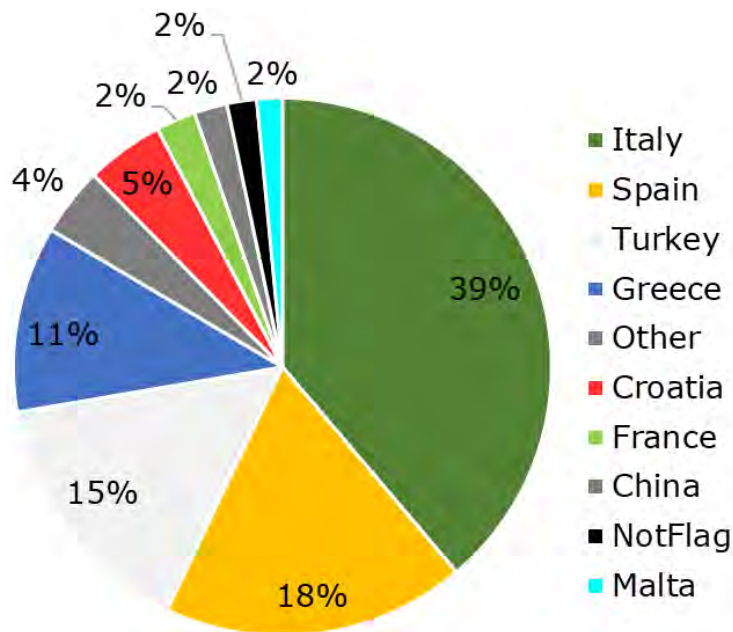


Figure 5 - Analyzed fleet composition by country

At the end of this pre-processed phase, rule-based machine learning was used to characterize single trips and to identify the type of fishing when it occurs, based on GPS tracks and gear-specific movement patterns without the need of matching official registers (Fig. 6).

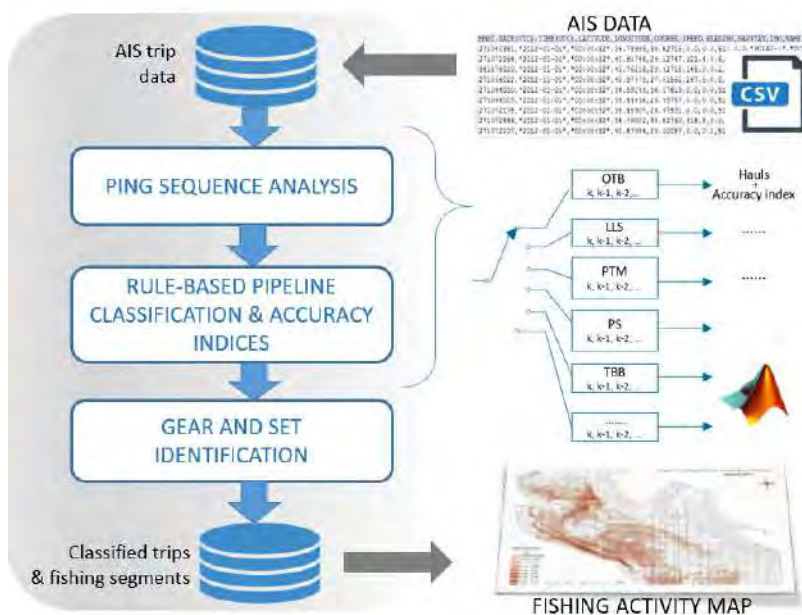


Figure 6 - Simplified workflow used to process AIS data

A set of algorithms were developed using Matlab software environment to incorporate prior domain knowledge for different gear types and executed in sequence to label single trips according to pre-defined gear classes: bottom otter trawl, beam trawl, pelagic trawl, purse seine, longline, and "other" fishing (including nets and traps). At the current stage of development, the algorithms perform well

at classifying the most common gear types among larger vessels, while they do less well at differentiating small coastal vessels operating gear types such as set gillnets, pots and traps. For this reason, these gear types were grouped together into fishing vessels of “other” class.

Trip by trip, the system executed all algorithms designed for each of the pre-defined gear classes and then a rule-based decision was taken to uniquely assign a gear type to each single trip on the basis of answers (given by single algorithms) and trip percentage of lost signals. In particular, the decision was taken on a monthly basis and weighting only “stable” fishing trips that were more robustly classified, such as those with at least 30 transmissions and with a percentage of data gaps non-exceeding 90%.

Given the above, the current rule-based system was able to question trip-level outliers (i.e. wrong gear detections due to a poor data coverage), impose the predominant monthly gear type in use and/or validate real changes in fishing behavior, such as when vessels rotate fishing gears. It could assign to each MMSI more than one gear type, limiting the ability to classify the type of fishing when vessels change gears on a voyage or between voyages during the same month.

Once trips had been labeled according to the gear type they use, segments were extracted when fishing vessels are likely engaging in fishing operations. In particular, in order to prevent duplicating the effort, for passive gears fishing segments are extracted only when fishing vessels are likely hauling back their gear (Fig. 7).

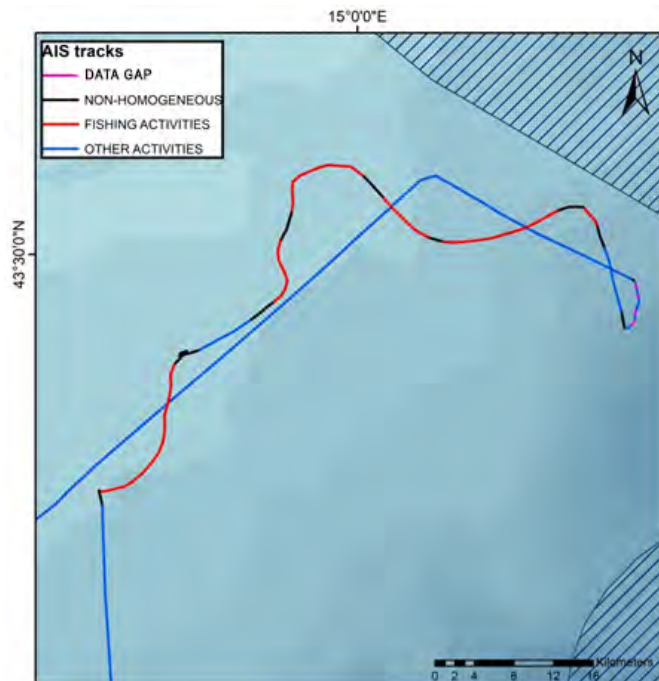


Figure 7 –Classified AIS-based track for a longliner. Extracted fishing segments are red, while longline sets are blue and classified as other activities.

Extracted fishing segments were stored in the single spatial database along with their attributes (number of pings, speed class, distances, duration, vertex coordinates) and related parent tables (trips and raw pings), and they were ready to be aggregated and binned into grid cells.

For the purpose of this task, bottom trawlers' and longlines' fishing operations were directly extracted from the populated database, while vessels classified as "other fishing" were further distinguished in nets and traps linking publicly available vessel registry databases (EU and GFCM Fleet registers) with AIS data. This matching was performed in R and implemented in order to solve problems of misspelling by means of measures of similarity between different MMSI codes, NAMES and CALLSIGNS and according to the method described by Natale *et al.* (2015). When possible, mismatches were manually fixed using additional sources of information such as Vessel Finder⁷ and Marine Traffic vessels' databases⁸. Once all trips were labelled according to the gear in use, some summary statistics by gear type and flag state were prepared for the final fleet engaging in fishing operations in 2017 (Fig. 8).

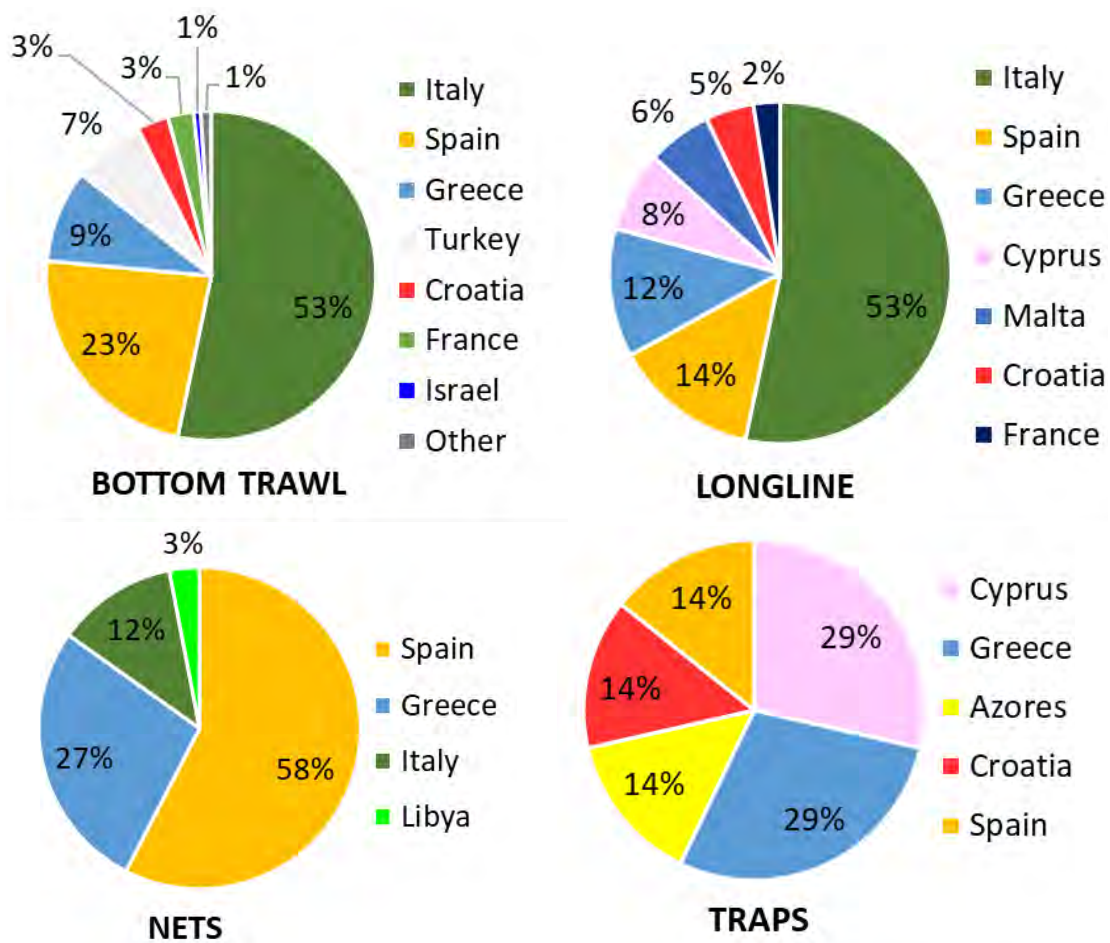


Figure 8 –AIS-identified and classified fleets broadcasting AIS during their fishing operations in the Mediterranean Sea by gear type and flag state

Finally, the intensity of fishing operations was mapped on a yearly base and the final output was:

⁷ <https://www.vesselfinder.com/vessels>

⁸ <https://www.marinetraffic.com/>

- Represented by 2 different grids: 1km x 1km grid based on EMODnet targeted products and 1nm x 1nm grid coherent with the standard defined by the GFCM (GFCM Statistical grid⁹);
- Disaggregated by 4 gear types: Bottom trawls (excluding beam trawls), Longlines, Traps and Nets. It was done according to the schema of Metiers defined within the DCF for the Mediterranean and Black Sea and limiting the classification to the Level 3 (Gear groups) as detailed in the previous [Table 5](#);
- Populated with Fishing days (days fishing vessels spent operating gear in each grid cell). In particular, the procedure of the R package fecR (Scott *et al.* 2017) was applied, allocating 1 fishing day in each grid cell intersected by a fishing segment on the same fishing trip. This AIS-based metric was chosen to align with fishery statistical standards.

Sample resulting maps depicting fishing footprints for trawlers, longliners, nets and traps are provided below in Section Results.

Comparison with the EMODNET footprint

Within the scope of the European Marine Observation and Data Network (EMODnet) MedSea Checkpoint project, an analysis on the spatial extension of trawl fishing grounds was developed using t-AIS data, in order to identify areas which were most disturbed and evaluate the change of bottom trawling activity (including beam trawls), hence impacting the sea bed, over the years¹⁰.

For the period 2012-2014, levels of aggregated AIS-based bottom trawl fishing effort information were produced and made freely available as Targeted Products (Fabi 2017). In particular, thirty-six monthly maps were produced at the 0.01°x0.01° grid resolution to cover the whole Mediterranean basin MEDSEA_CH5_Product_6, (Tassetti *et al.* 2016a; Ferrà *et al.* 2018)), while one map was produced to estimate annual changes in the level of trawling disturbance $\Delta Fe_{2014-2013}$, MEDSEA_CH5_Product_7, (Tassetti *et al.* 2016b; Ferrà *et al.* 2018). The intensity of fishing operations was described in swept kilometres (kilometres fishing vessels swept operating gear in each grid cell).

According to the workflow described in the project documents and in the Inception Report, the fishing footprints of trawlers obtained from the analysis of AIS data 2017 were compared with EMODnet MEDSEA_CH5_Product_6.

Considering that the adoption of the AIS by different fishing fleet segments was implemented starting from 2012, when there were still very few vessels equipped with this system, only year 2014 was selected to make the comparison.

In particular, MEDSEA_CH5_Product_6 2014 was reprocessed by CNR in order to estimate Fishing Days and allow the direct comparisons with the fishing footprints 2017 computed in this project.

Validation of output #1: Comparison with fishing footprint by Global Fishing Watch

⁹ <http://www.fao.org/gfcm/data/maps/grid/en/>

¹⁰ <http://www.emodnet-mediterranean.eu/portfolio/fisheries/>

For the period 2012-2016, daily levels of aggregated fishing effort information were published in Science (Kroodsma *et al.* 2018) and made freely available by the Global Fishing Watch independent and non-profit organization (GFW, <https://globalfishingwatch.org/>). AIS-based aggregated data derived from satellite providers are downloadable in 2 forms: Fishing Effort and Vessel Presence at 100th Degree Resolution by Flag State and Gear Type, and Fishing Effort and at 10th Degree Resolution by MMSI. Fishing vessels are classified by neural net into six categories: drifting longlines, purse seines (both pelagic and demersal), trawlers (including all types), fixed gear (including set longlines, set gillnets, and pots and traps), squid-jiggers (mostly large industrial pelagic operating vessels) and other fishing (including unknown fishing gear and less common gears such as pole and line).

GFW uses two core algorithms to process satellite AIS (s-AIS) data, one to identify fishing vessels and a second to identify fishing activity. They are described in (Kroodsma *et al.* 2018), while the dataset is improved in (Taconet *et al.* 2019).

According to the workflow described in the project documents and in the Inception Report, the fishing footprints of trawlers and longliners obtained from the analysis of AIS data were compared with the corresponding estimates provided by Global Fishing Watch and adapted to the same grid used for the other methods.

Validation of output #2: Comparison with fishing footprint by MINOUW project (AIS and VMS data combined)

Some outputs of the H2020 Project MINOUW¹¹ were used to assess whether the assessment of fishing effort based on an integrated AIS + VMS dataset could lead to more realistic estimations in terms of fishing footprint and coverage of the fishing grounds exploited by the fleets. A detailed description of the procedure is provided in the Deliverable D5.9 – Spatial estimates of Fishing Footprint.

A complete series of AIS data covering these areas (namely the GSA 5, 6, 9, 16, and 22), for the period 2012-2016, and related to all the available fishing vessels (ship type 30) were purchased from ASTRA Paging, a private data provider¹². For each case study, only the AIS data pertaining to the fishing vessels of the respective country were collected (i.e. only Portuguese fishing vessels for the case studies 1.2 and 2.2, only Spanish fishing vessels for the case study 1.4, only Italian fishing vessels for the case studies 1.5, 1.6/1.8, and 2.1). While AIS data are not characterized by confidentiality issues and the identity of the fishing vessels is always associated to each AIS ping as MMSI (Maritime Mobile Service Identity) and Maritime Callsign codes, VMS data are confidential, and the access to the dataset for each area/case study was allowed by the researchers previously identified as the reference person in charge to assist the University of Rome Tor Vergata (UTV), through a one-to-one communication, during the processing of AIS/VMS data and the validation of the outcomes.

¹¹ <http://minouw-project.eu/>

¹² <http://www.astrapaging.com/>

According to the methodology (Deliverable WP5-T5.1), AIS data were transferred to the reference researchers for each case study in order to allow (Fig. 9):

- The inspection of the AIS data in terms of fleet coverage;
- The integration of AIS data with VMS data. In fact, given that VMS data are confidential, they were not shared in their native format: each case study reference researchers encrypted the identifier field in the VMS data (i.e. the Common Fleet Register code - CFR), merged the encrypted VMS data with the AIS data and returned a VMS/AIS dataset for each case study;
- The selection of the fishing vessels to be analysed for each case study. In fact, only some subsets of the whole fleets represented by VMS/AIS data (for each area) were validated, in terms of used gear and fishing activity, to the fisheries of the different case studies. In this way, the case study reference researchers returned a list of selected vessels with their respective characteristics in terms of length-over-all (LOA, in meters) and engine power (PW in Kw).

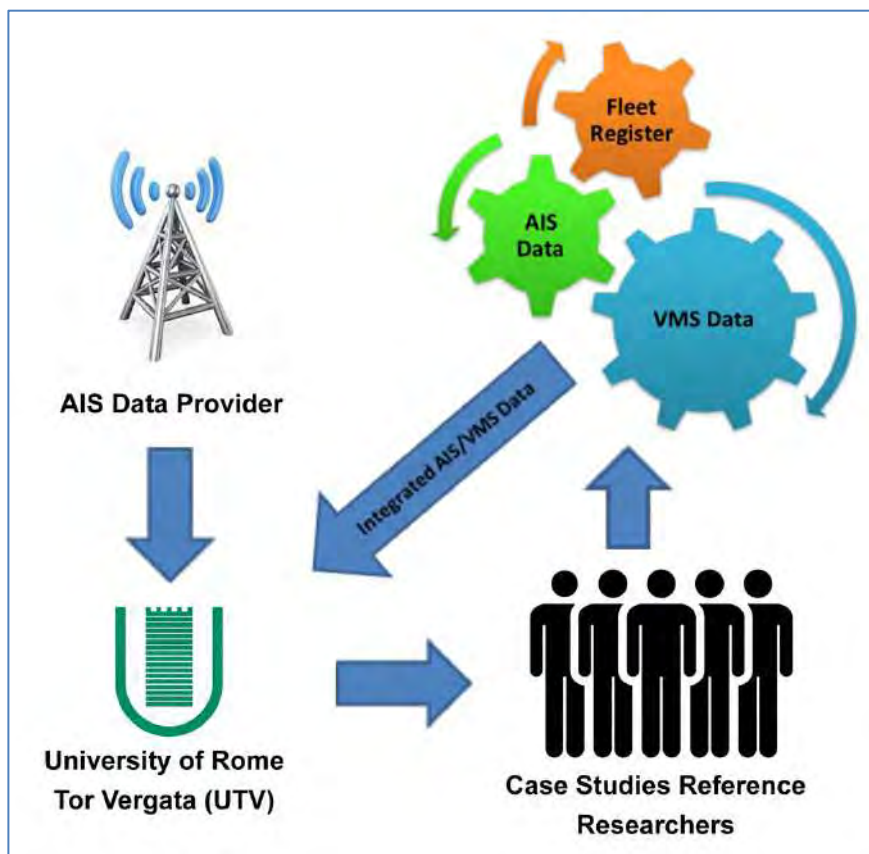


Figure 7 - Representation of the processing for the integration of AIS and encrypted VMS data. The process involved UTV and the Case Studies Reference researchers through a one-to-one data flow.

For the purposes of this project, the following case studies from MINOUW were considered (Table 6):

Table 6 - List of the case studies from the MINOUW project used for comparative purposes in the MED UNITS project. All these case studies were related to bottom trawling.

Case Studies	Code	Area
SEA BOTTOM TRAWL IN CATALONIA	1.4	GSA06
BOTTOM TRAWLING IN SICILY, ITALY	1.5	GSA16
AEGEAN SEA BOTTOM TRAWL	1.7	GSA22
BOTTOM TRAWL FISHERIES IN LIGURIAN AND N. TYRRHENIAN SEA	1.6/1.8	GSA09

The same 1 Km square grid from EMODNET, and applied for the AIS analysis of MED UNITS, was used to re-grid the MINOUW output. This analysis, limited to the year 2016 (the nearest one to the year 2017 used for MED UNITS), returned a map of trawling effort, in fishing days, for the following GSA: 5, 6, 9, 16, and 22.

Multi-Criteria Decision Analysis for both large- and small-scale components of the fleets without AIS/VMS

The MCDA method produces fisheries footprint by taking into consideration several interactions with other anthropogenic and environmental factors. The methodological approach and the application on SSF is further described on Kavadas *et al.* 2015, while this method has been also applied to estimate a fishing pressure index for bottom trawlers (see Hidalgo *et al.* 2020; Mérigot *et al.* 2020).

Several methods and processes such as the Analytic Hierarchy Process (AHP) and Fuzzy logic were applied in an effort to solve the multiple criteria problem. FP_c was perceived as the fuzzy product of two indices: the fishery suitability index (S_c) and the activity index (A_c) based on the spatial distribution of registered fishing vessels in the Mediterranean Sea according to the formula:

$$FP_c = S_c \times A_c$$

A general overview of the criteria and methods of the MCDA tool used is given in [Fig. 8](#).

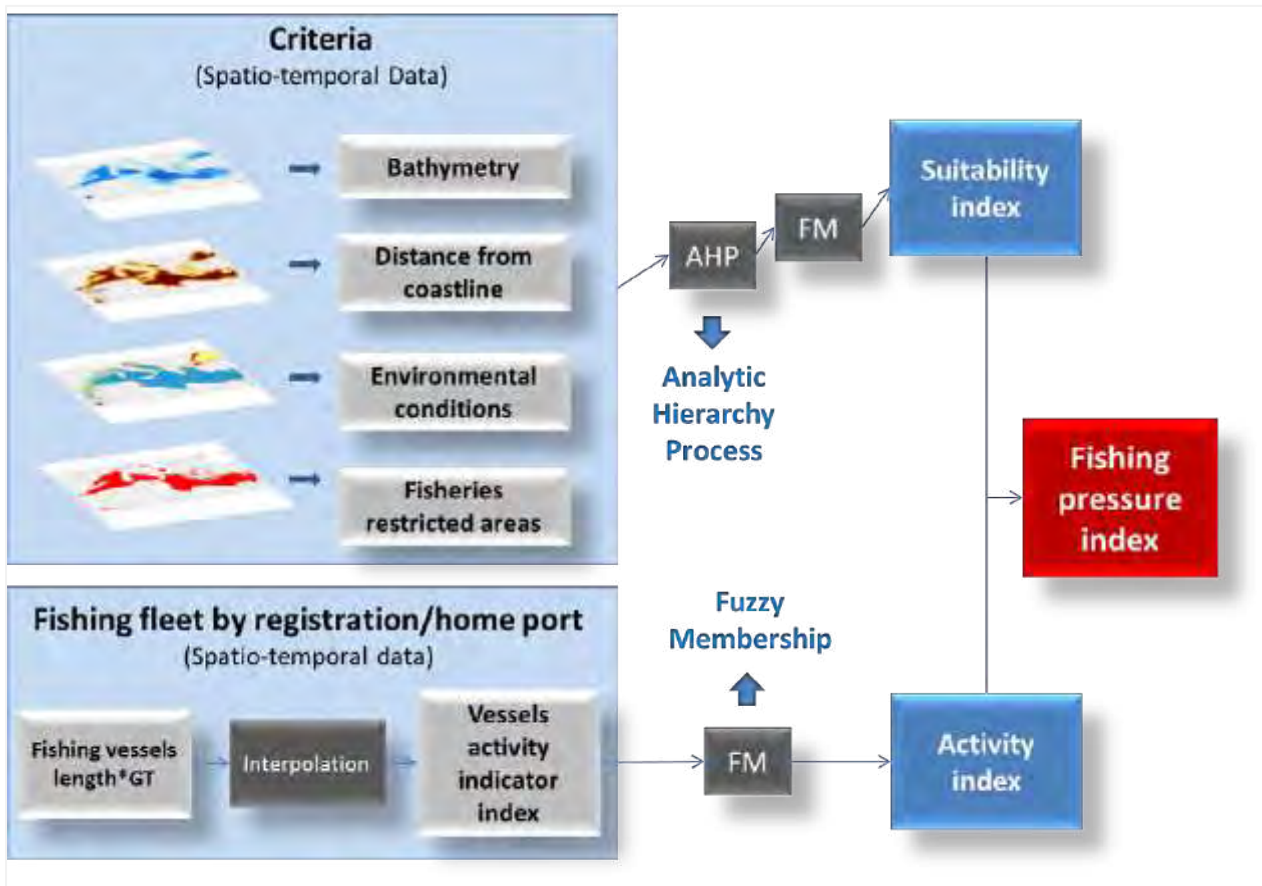


Figure 8 - Flowchart to estimate the Fishing Pressure index for small-scale fisheries and bottom trawlers in the Mediterranean Sea (see details in Kavadas et al., 2015)

Estimation of the suitability index (S_c)

In order to estimate the suitability index (S_c), the most influential components and criteria affecting small-scale fishing and trawling were identified. The criteria used were shown in Table 7. In each criterion a grading value by expert judgment was assigned (i.e. a rank of order of importance). The final rankings used for all criteria under study are shown in Tables 8 and 9. The estimation of the S_c was then calculated based on the following steps:

- creation of spatial information and calibration of each criterion according to a scale of evaluation and formation of the hierarchical structure of the multiple criteria problem (Tables 8 and 9);
- implementation of the AHP to estimate the relative importance of the evaluation criteria (Tables 10 and 11);
- application of the Weighted Linear Combination method (WLC) using the weights (priority vectors) to estimate the suitability index;
- standardization on a scale from 0 to 1 with linear Fuzzy Membership (FM)

Table 7 - Criteria used for the estimation of the suitability index (Sc) Criterion Figure Bathymetry (source: EMODNET) Distance from coastline (ESRI, 2015) Chl-a annual concentration (period: 2015- 2017¹³) Fisheries restricted areas (legislation, source: Mediseh(Petza et al. 2017))

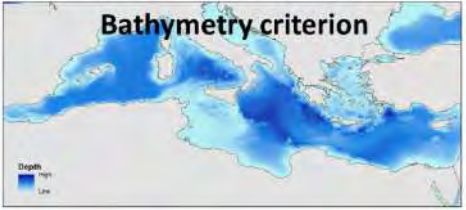
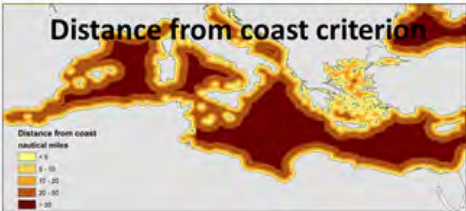


Criterion	Figure
Bathymetry (source: EMODNET)	
Distance from coastline (ESRI, 2015)	
Chl-a annual concentration (period: 2015-2017)	
Fisheries restricted areas (legislation, source: Mediseh; Petza et al., 2017)	

Table 8 - Ranking* of the criteria taken into account in MCDA for bottom trawlers per length (LOA) category (*The higher the grade, the most likely this area is for trawling activities)

Gear	LOA	Bathymetry (m)	Grade	Distance from coast (nm)	Grade	CHL (mg/m3)	Grade	Legislation	Grade	no-take zones
OTB	<1 2 m	0-50	5	1.5	5	>1	4	available areas	5	banned areas (annually) bathymetry = 500 m
		50-100	5	1.5-3	5	0.73-1	5	ban <2	4	
		100-200	3	3-6	4	0.46-0.73	4	ban 2-6	3	
		200-500	1	6-12	3	0.23-0.46	3	ban 6-11	2	
		500-800	0	>12	2	0.1-0.23	2	ban 12	0	
		0-50	5	1.5	5	>1	4	available areas	5	

¹³ <http://oceancolor.gsfc.nasa.gov/cms/>

12-24 m	50-100	5	1.5-3	5	0.73-1	5	ban 2	4	banned areas (annually) bathymetry = 1000 m
	100-200	4	3-6	5	0.46-0.73	4	ban 4	3	
	200-500	3	6-12	4	0.23-0.46	3	ban 6	2	
	500-800	1	>12	3	0.1-0.23	2	ban 12	0	
>24 m	0-50	3	1.5	4	>1	4	available areas	5	banned areas (annually) bathymetry = 1000 m
	50-100	4	1.5-3	4	0.73-1	5	ban 2	4	
	100-200	5	3-6	5	0.46-0.73	4	ban 4	3	
	200-500	5	6-12	5	0.23-0.46	3	ban 6	2	
	500-800	2	>12	4	0.1-0.23	2	ban 12	0	

Table 9 - Ranking* of the criteria taken into account in MCDA for small-scale fisheries per length (LOA) category. *The higher the grade, the most likely this area is for small-scale fishing activities

Gear	LOA	Bathymetry (m)	Grade	Distance from coast (nm)	Grade	CHL (mg/m ³)	Grade	no-take zones
SSF	<12 m	0-50	5	1.5nm	5	>1	5	banned areas (annually) bathymetry = 500 m
		50-100	4	1.5-3	4	0.73-1	4	
		100-200	3	3-6nm	3	0.46-0.73	3	
		200-500	2	6-12nm	2	0.23-0.46	2	
		500-800	0	>12nm	0	0.1-0.23	1	
	12 - 24 m	0-50	5	1.5nm	5	>1	5	Banned areas (annually) bathymetry = 800 m
		50-100	4	1.5-3	4	0.73-1	4	
		100-200	4	3-6nm	3	0.46-0.73	3	
		200-500	3	6-12nm	2	0.23-0.46	2	
		500-800	2	>12nm	1	0.1-0.23	1	

Table 10 - The weights (priority vectors) for each criterion based on AHP and expert judgement per length (LOA) category for bottom trawlers

Gear	LOA	Criterion	Weights based on AHP
OTB	<12 m	Bathymetry	0.4
		Distance from coast	0.2
		CHL	0.15
		Legislation	0.25
	>12 m	Bathymetry	0.43
		Distance from coast	0.23
		CHL	0.16
		Legislation	0.18

Table 11 - The weights (priority vectors) for each criterion based on AHP and expert judgement per length (LOA) category for small-scale fisheries

Gear	LOA	Criterion	Weights based on AHP
SSF	<24 m	Bathymetry	0.43
		Distance from coast	0.37
		CHL	0.2

Estimation of the activity index (A_c)

Data . General Fisheries Commission for the Mediterranean - GFCM fleet register

The commercial fleet data set¹⁴ reports public data as transmitted by CPCs to the GFCM Secretariat in line with the requirements set in Recommendation GFCM/33/2009/5 on the establishment of the GFCM regional fleet register. A total of 70712 fishing vessels are registered in the Mediterranean ports where 27857 of them belong to non-EU countries (Table 12). A total of 17375 fishing vessels are registered in the ports along the Black Sea.

Table 12 - Total number of fishing vessels by gear and non-EU country (* 10195 fishing vessels operating in Black Sea, 5154 are operating in the Mediterranean Sea)

Gear	Country									
	ALB	DZA	EGY	LBN	LBY	MAR	MNE	SYR	TUN	TUR
Beach seines		135					17			
Boat dredges	5						24			31
Bottom beam trawls										658
Bottom otter trawls	2	12	981							418
Bottom shrimp trawls		1							331	
Bottom trawls	166	485						18	98	
Cast nets				3						
Combined gillnets-trammel nets			1681	10					447	
Drifting longlines		8		552			3			
Encircling gillnets				118					2	
Fixed gillnets (on stakes)								1		
Gear not known or not specified				4				4		
Gillnets (not specified)	427	680		153					2140	
Handlines and pole-lines (hand operated)		61		868						
Hooks and lines (not specified)	6	1								
Longlines (not specified)		46	1294	917		111		1		

¹⁴ <http://www.fao.org/gfcm/data/fleet/register>

Midwater otter trawls		1								
Midwater pair trawls										195
Midwater trawls		42								
Miscellaneous gear	11									13615
Not reported	1	1038		11	3422				27	
One boat operated purse seines	1									432
Otter trawls (not specified)						112				
Pots				152						
Seine nets (not specified)	1	13								
Set gillnets (anchored)	1	43					52			
Set longlines		558		779			9			
Trammel nets		987		1118			73			
Traps (not specified)							1	1		
With purse lines (purse seines)	10	1364	244	183		113	17	6	499	
TOTAL	631	5475	4200	4868	3422	336	196	31	3544	15349*

In the data quality control process, a number of 7191 missing GT values were observed (Table 13). The missing values were filled using a simple regression model (variables: vessel total length, vessel GT).

Table 13 - Number of missing GT values in non EU countries

Country										
	ALB	DZA	EGY	LBN	LBY	MAR	MNE	SYR	TUN	TUR
Missing GT values		106	1357	2365	3363				522	

The estimation of the activity index (Ac) for each fishing and LOA category by registration port was based on vessels' length, Gross tonnage (GT) and for the year 2016 and is estimated by the formula:

$$VAI_p = \sum_{v=1}^n (L \times GT)$$

where L= Length, GT= Gross tonnage, v= vessels and n= the total number of vessels at each fishing port. A general overview of the VAI_p index and of the total number of fishing vessels by registration port is given in Figures 9-12.

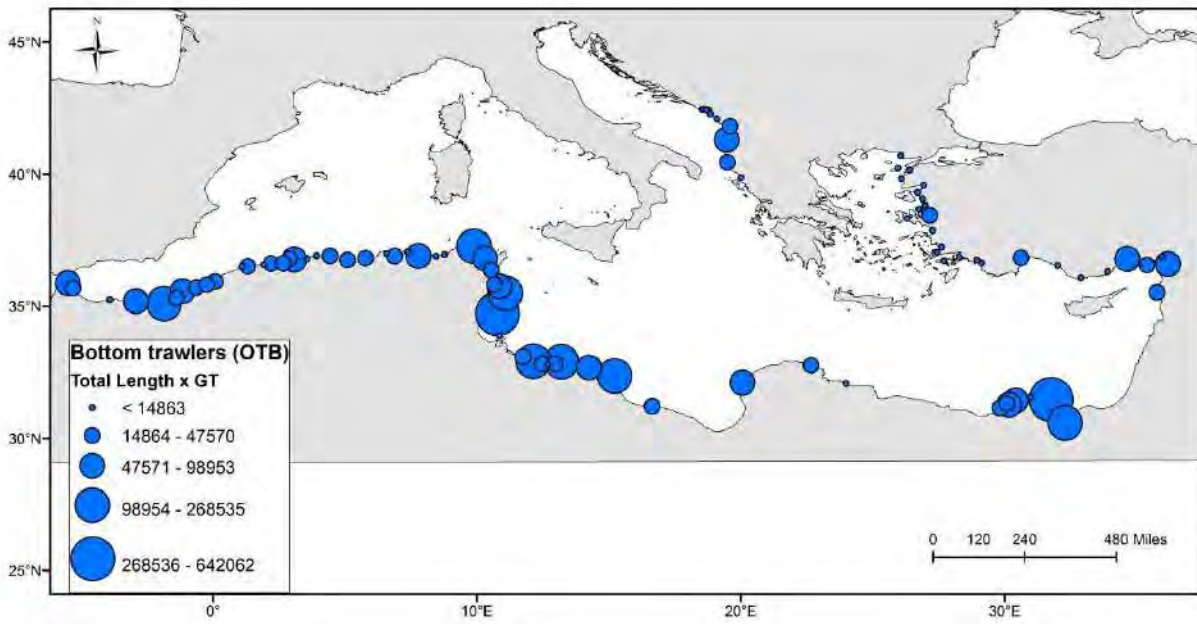


Figure 9 - VAIP index for bottom trawlers indicated for the registration ports for the non-EU Mediterranean countries.

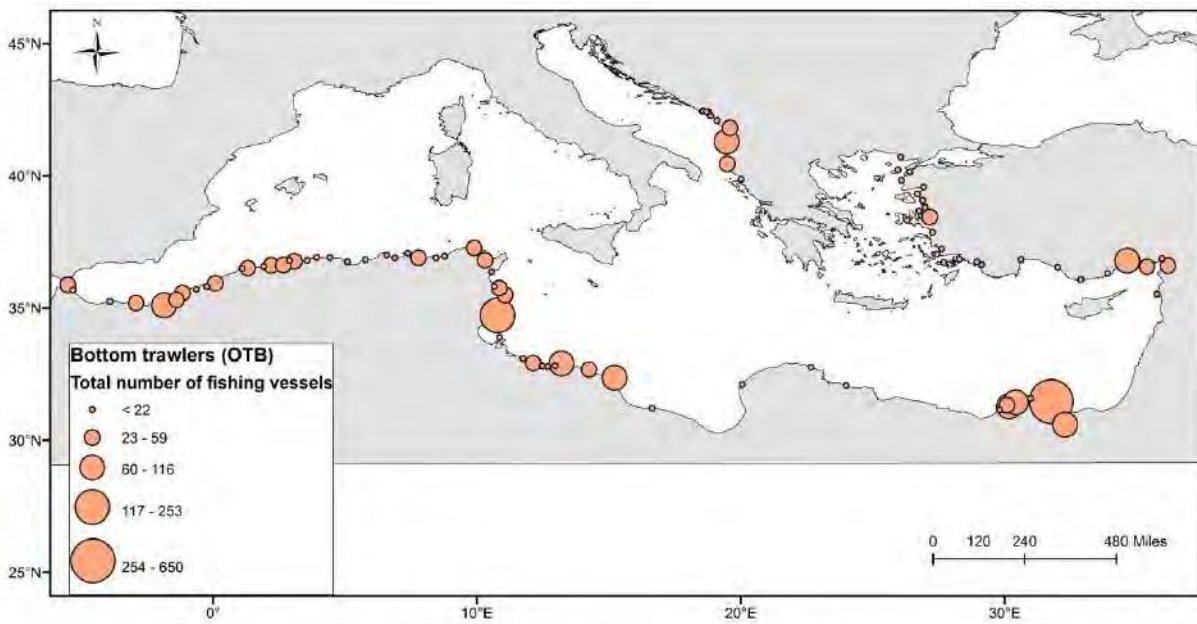


Figure 10 - Total number of bottom trawlers by registration port for the non EU Mediterranean countries.

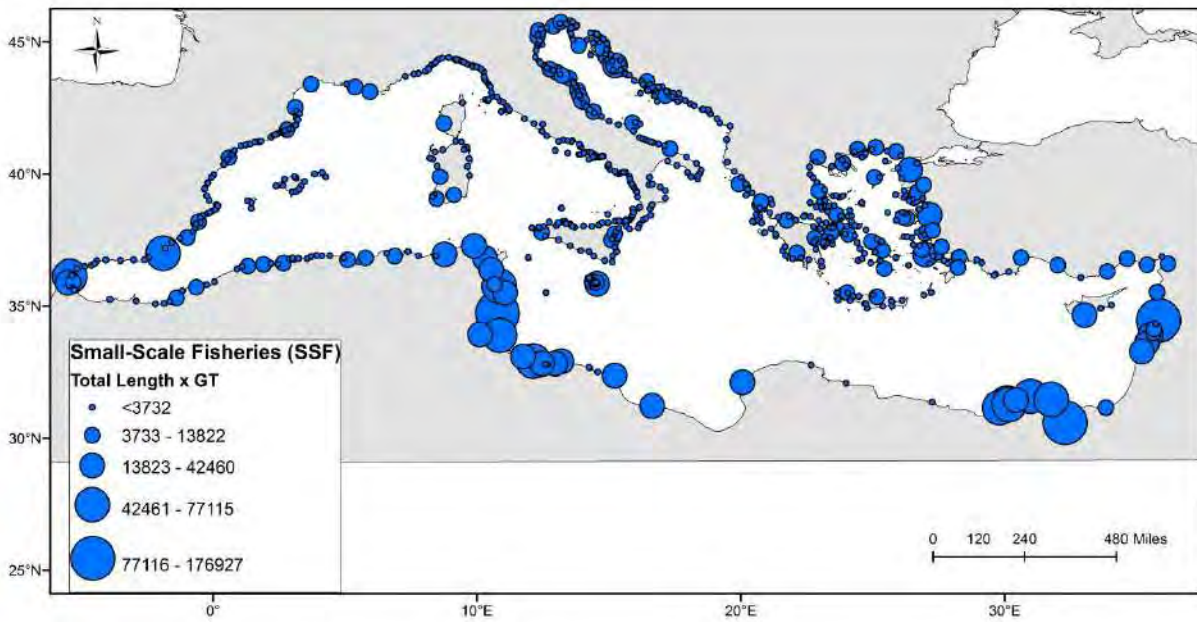


Figure 11 - VAIp index for small-scale vessels by registration port in the Mediterranean Sea.

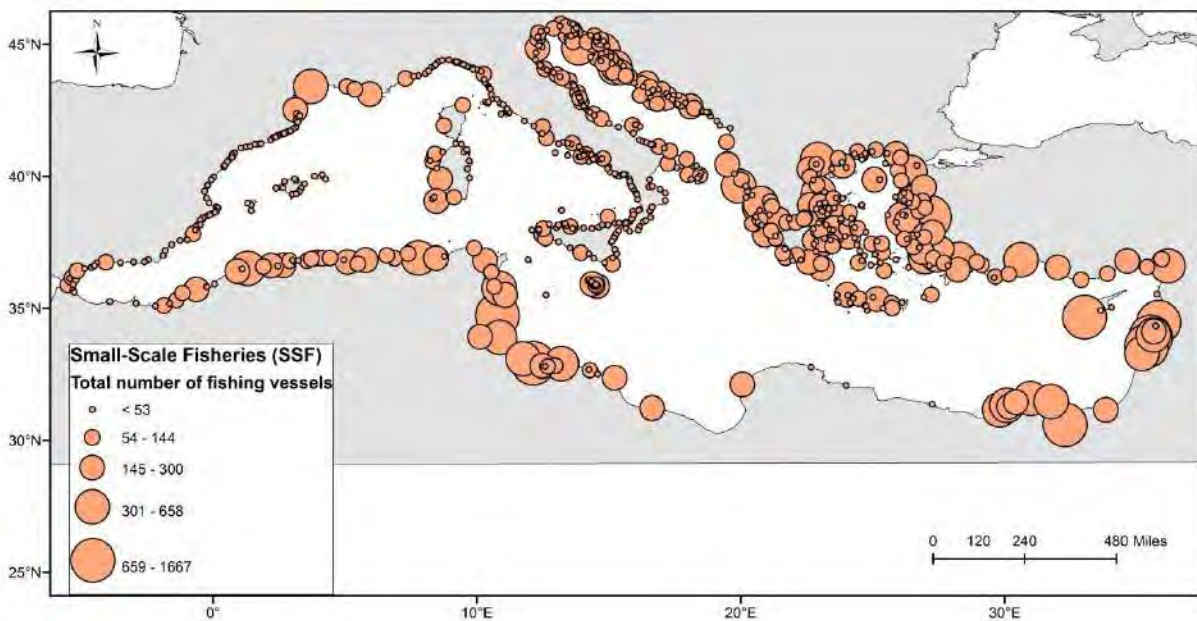


Figure 12 - Total number of small-scale vessels by registration port in the Mediterranean Sea.

The total number of the registration ports is higher per EU Mediterranean country compared to the non-EU Mediterranean countries. Given that the spatial allocation of the registered fishing ports is differentiated while, at the same time, the GT and the LOA of the fishing vessels varied between EU and non-EU Mediterranean countries, a standardization procedure was performed in the VAIp index.

Generalized Additive Modelling (GAMs - Hastie and Tsibarani 1990), which employ non-linear and non-parametric techniques for regression modelling, was used in

order to standardize the VAIp index. Modelling was performed through the *mgcv* library of R. The main advantage of GAMs over traditional regression methods is their ability to deal with highly non-linear and non-monotonic relationships between a response and a set of explanatory variables (Wood 2006). Herein the VAIp index was used as a response variable while the total number of vessels, the mean LOA and the mean GT per port were used as explanatory variables. Given that taking into account the distances between the neighboring fishing ports is of critical importance in order to consider the observed spatial differences between the EU and non-EU registration procedures, a proximity assessment was performed in the basis of Thiessen polygons. Thiessen polygons are generated from a set of sample points (herein sample points are referred to the registered fishing ports) such that each polygon defines an area of influence around its sample point, so that any location inside the polygon is closer to that point than any of the other sample points (Thiessen 1911). The area based on Thiessen polygons estimated for the registered fishing ports in the study area was also included as an explanatory variable in the final GAM.

The detection of collinearity issues between the explanatory variables was based on a Variance Inflation Factor (VIF) analysis, using the *USDm* library in R. Moreover, to avoid over-fitting and to simplify the interpretation of the results, the degree of smoothing of each predictor was chosen based on the REstricted Maximum Likelihood (REML); a method able to protect against bias in variance component estimates (Wood, 2011). The VAIp data were modelled using a Tweedie error distribution (Tweedie, 1984), with a Tweedie index parameter estimated by the *tw()* function supported by the *mgcv* R package, and a logit link function. The final GAM models and their deviance explained were shown in Table 14 and in Fig 13.

Table 14 - Results for factors affecting VAIp index based on the final Generalized Additive Model. *s*: smooth function represented using penalized regression splines; DE%: Deviance Explained. NFVs: Total number of fishing vessels; AVG_LOA: mean LOA; area: total area based on Thiessen polygon applied per port.

Gear	LOA	Formula	DE%
OTB	<12 m	$LxGT \sim s(NFVs, k=5) + s(AVG_LOA, k=5) + s(area, k=5), family= Tweedie(p=1.949), method="REML"$	90.3%
	12-24 m	$LxGT \sim s(NFVs, k=5) + s(AVG_LOA, k=5) + s(area, k=5), family= Tweedie(p=1.58), method="REML"$	92.8%
	>24 m	Non-significant	-
SSF	<12 m	$LxGT \sim s(NFVs, k=5) + s(AVG_LOA, k=5) + s(area, k=5), family= Tweedie(p=1.01), method="REML"$	77.7%
	12-24 m	$LxGT \sim s(NFVs, k=5) + s(AVG_LOA, k=5) + s(area, k=5), family= Tweedie(p=1.99), method="REML"$	72.1%

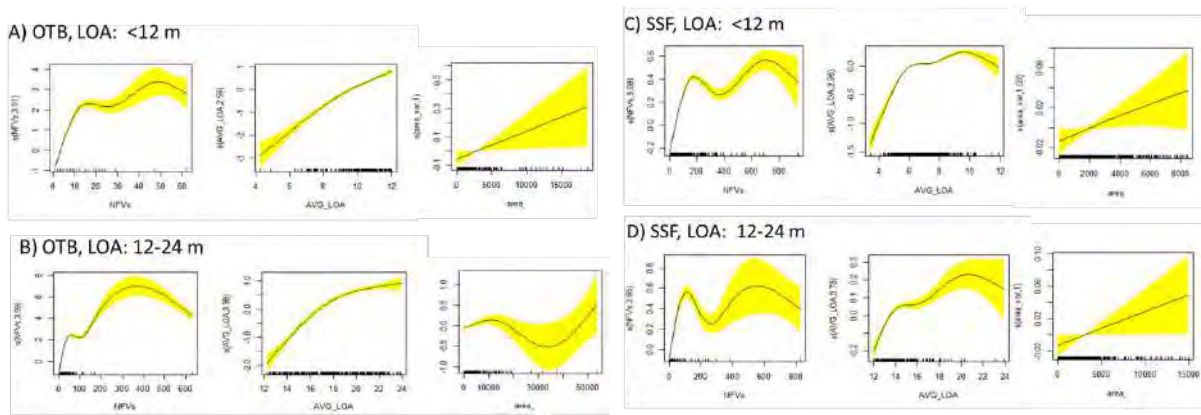


Figure 13 - Estimated smooth terms for the final GAMs for VAIp index. For the main effects per fishing gear and LOA category (a-d) the corresponding 95% confidence intervals are indicated with yellow. The numbers next to the covariate name in the y-axis titles are the estimated degrees of freedom of the smooth term.

In a subsequent step, the final GAMs, were applied in a predictive mode to provide the standardized VAIp estimates per fishing port. To avoid the effect of spatial differences between the EU and non-EU registration ports, the variable area_Thiessen was fixed to its median value in the predictions.

Furthermore, the methodology used to estimate Ac consisted of the following steps: Implementation of the optimal interpolation method on the standardized VAIp to estimate values at a spatial cell level (VAIc);

Implementation of the optimal FM in the standardized VAIc, to represent numerically the degree to which a given measure of criteria within a grid cell belongs to a fuzzy set.

Finally, the Sc and Ac per gear and LOA category were combined using a product fuzzy overlay procedure while, the estimation of the overall fishing effort per gear (i.e. bottom trawlers and small-scale fisheries) was based on a sum fuzzy overlay procedure. Fuzzy logic explores the interaction of the possibility of a phenomenon belonging to multiple sets. In the fuzzy logic technique, combination is defined as a superclass of connectives that is used for fuzzy overlay (Jiang 1996). In fuzzy overlay, there are specific techniques for investigating this relationship and quantifying the interaction. The combination approach used in this work was the Fuzzy Product or Fuzzy Algebraic product (Zimmermann and Zysno 1980). The Fuzzy Product overlay type will, for each cell, multiply each of the fuzzy values for all the input criteria. The Sum fuzzy overlay procedure used to estimate the overall fishing effort per gear type, is an increasing linear combination function that is based on the number of criteria entered into the analysis and is estimated as follows:

Fishing Pressure index for bottom trawl

$FPI_{OTB} = 1 - \text{product}(1 - Fpi_{OTB}(LOA <12m), 1 - Fpi_{OTB}(LOA 12-24m), 1 - Fpi_{OTB}(LOA >24m))$

Fishing Pressure index for small-scale fisheries

$FPI_{SSF} = 1 - \text{product}(1 - Fpi_{SSF}(LOA <12m), 1 - Fpi_{SSF}(LOA 12-24m))$

The study area was gridded with a spatial resolution of 1×1 mile. The corresponding values for each of the MCDA modelled criteria were assigned in each of these grid cells.

Quality assessment of Activity index (Ac)

A qualitative assessment on the data used to estimate the Ac index was based on a pedigree matrix (Weidema and Wesnæs 1996) that describes aspects of data quality influencing the reliability of the overall MCDA result. In our case, the specific pedigree matrix has been adjusted in order to specify the reliability, completeness, temporal correlation, geographical correlation and data collection process quality of the data used for Ac estimation (Table 15). Each characteristic is divided into five quality levels with a score between 1 (high quality) and 5 (low quality). The “reliability” indicator relates to the sources, the acquisition methods and verification procedures used to obtain the data. The “completeness” indicator relates to the statistical properties of the data: how representative is the sample collected by registration port and whether this sample includes a sufficient number of data. The “temporal correlation” indicator represents the time correlation between the year of study and the year the data were obtained. The “geographical correlation” indicator illustrates the geographical correlation between the defined area and the location of origin of the data used for estimating any missing information needed to calculate the Ac. Finally, the “data collection process quality” indicator refers to the methods used during the collection process to characterize the gear type by fishing vessel and port. The pedigree matrix can serve as a data quality management tool providing information about the data quality, finding sources of uncertainty and point out improvements in data quality and in model outcomes. As mentioned above, the scores in the pedigree matrix are ‘semi-quantitative’, but this is actually a qualitative way to assess the quality of the Ac index (and therefore the inherent bias in the overall MCDA procedure) and to indicate, in a transparent way, where there might be a problem.

A modelling framework based on simple regression and Generalized Additive Models has been used to handle data gaps in the model, aiming to improve the Fishing Pressure index based on MCDA. In a quality assessment context, a map has been also created per GSA area. This map indicates the weaknesses (data gaps) but taking into account when these weaknesses considered in the MCDA model. The main weaknesses of Ac i.e. (a) the fishing categories (gear/types) missing, (b) the GT per registration port and fishing vessel missing and (c) the differences in the spatial quality of registration ports (based on the estimated Thiessen polygons) were quantified and normalized in a scale from 0-1. Then, the estimated performance based on the models aiming to improve the FPI outcomes i.e. (d) the R² for simple regression and (e) the average DE% based on GAMs were also normalized in a scale from 0-1. In a context of a ‘semi-quantitative’ process a map showing the weaknesses minus this proportion of weaknesses removed from FPI estimations was constructed based on the formula: $(a+b)*(1-d) + c*(1-e)$.

Table 15: Pedigree matrix used for the quality assessment of the data used to estimate the Ac in the Mediterranean Sea.

Indicator Score	1	2	3	4	5
Reliability	Measured Data	Verified data partly based on assumptions	Non-verified data partly based on qualified estimates	Qualified estimate (e.g. by scientific expert)	Non- estimate
Completeness	Representative data for all registration ports for the study area	Representative data for >50% of the registration ports for the study area	Representative data for a subset of the registration ports (<50%) relevant to the study area	Representative data for only one registration port for the study area (or unknown location) or missing data on important variables (e.g. fishing gear characterization, GT..)	Representativeness unknown
Temporal correlation	Less than 2 years of difference to year of study	Less than 3 years difference	Less than 4 years difference	Less than 5 years difference	Age of data unknown or more than 10 years of difference
Geographical correlation	Data from area under study	Estimated data from larger area including the study area having similar fisheries characteristics	Estimated data from area with slightly similar fisheries characteristics	Estimated data from areas with unknown similarities on fisheries characteristics	Estimations not possible
Data collection process quality	Data from targeted research conducted by similar methods; Fishing activity by gear type and by region - as Level 4 (Commission Decision 2010/93/EU)	Data from targeted research conducted by similar methods; Fishing activity by gear groups and by region - as Level 3 (Commission Decision 2010/93/EU)	Data from targeted research conducted by similar methods; Fishing activity by gear classes and by region - as Level 2 (Commission Decision 2010/93/EU)	Data from common research conducted with different methodologies and incompatible fishing gear characterization	No data available

Cascaded Multilayer perceptron network applied on environmental and fleet data for some component of the fleet without AIS/VMS

A cascaded multilayer perceptron network (CMPN) combining environmental data and fleet structure was applied to predict the spatial distribution of fishing effort for the length class [12–15 m]. This method was specifically developed for the Mediterranean basin and applied on the Adriatic Sea (Russo *et al.* 2019b). The trained CMPN was applied for all the GSAs of the Mediterranean Sea.

Input data required for model development were extracted for each GSA. Environmental characteristics were obtained by EUSeaMap2 Broad-Scale Predictive Habitat Map¹⁵, while the fleet data described in the section “Estimation of the activity index (Ac) - Data General Fisheries Commission for the Mediterranean - GFCM fleet register” of this deliverable were used to defined the size (number of vessels) by country for the following length class: [12–15 m), [15–18), [18–24 m), and [24–40 m).

Reconstruction of fluxes from fishing grounds to harbors

According to the methodology described in Russo *et al.*, 2018, a simple modelling approach was applied to estimate the fluxes of catch by species, from each GSA to EU-member country. The output data from Approach #2 (CNR) described above were used to quantify the total amount of trawling effort, by GSA and country and combined with data about total landings in 2017 by species and country from the 2019 Annual Economic Report (AER) on the EU Fishing Fleet (STECF 19-06) (Carvalho *et al.* 2019).

Under the assumption that the “mean” characteristics (efficiency, catchability of nets) of trawling fleets operating in the Mediterranean Sea were similar among countries, the Non-Negative Least Square regression approach described in Russo *et al.*, 2018 was fit to estimate the mean annual productivities (Landing Per Unit Effort - LPUE) of the different GSAs, for each species.

Then, the total annual landings by species of each country were partitioned and associated to each GSA. A graphical approach (riverplot) was used to compare the estimates returned by this modelling approach and the official origin of landings reported in the AER.

¹⁵ http://data.adriplan.eu/layers/geonode%3Aeunismedscale_4326

Results

Analysis of AIS data: approach #1

The distribution of the total fishing effort, in days of fishing, for the whole Mediterranean fleets operating with the Bottom trawls system, as obtained by VMSbase processing of the AIS dataset, is reported in [Fig. 14](#).

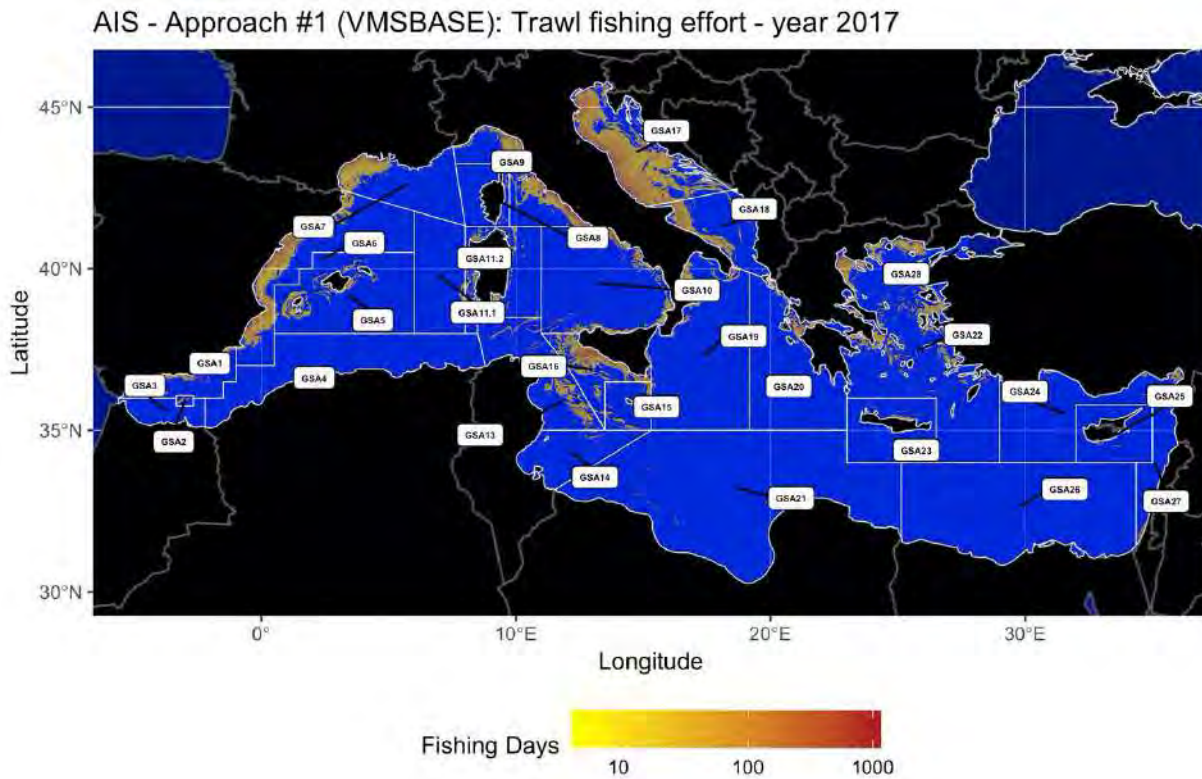


Figure 14 -Fishing footprint for trawlers from AIS data, as obtained using VMSbase and the related procedures.

It is possible to observe that, while fishing effort was detected along all the coasts of EU member states, few fishing grounds could be detected along the coasts of North Africa. In particular, some fished areas are present in GSA21.

From a general point of view, the Adriatic Sea (GSA17) the Strait of Sicily (GSA16), the Tyrrhenian Sea (GSA9 and GSA10), the Gulf of Valencia (GSA6), the Gulf of the Lion (GSA7) and the Aegean Sea (GSA22), are the area in which most of the effort is concentrated. The values of the effort range between 0 and around 1500 fishing days.

The distribution of the total fishing effort, in days of fishing, for the whole Mediterranean fleets operating with longlines, as obtained by VMSbase processing of the AIS dataset, is reported in [Fig. 15](#).

AIS - Approach #1 (VMSBASE): Longlines fishing effort - year 2017

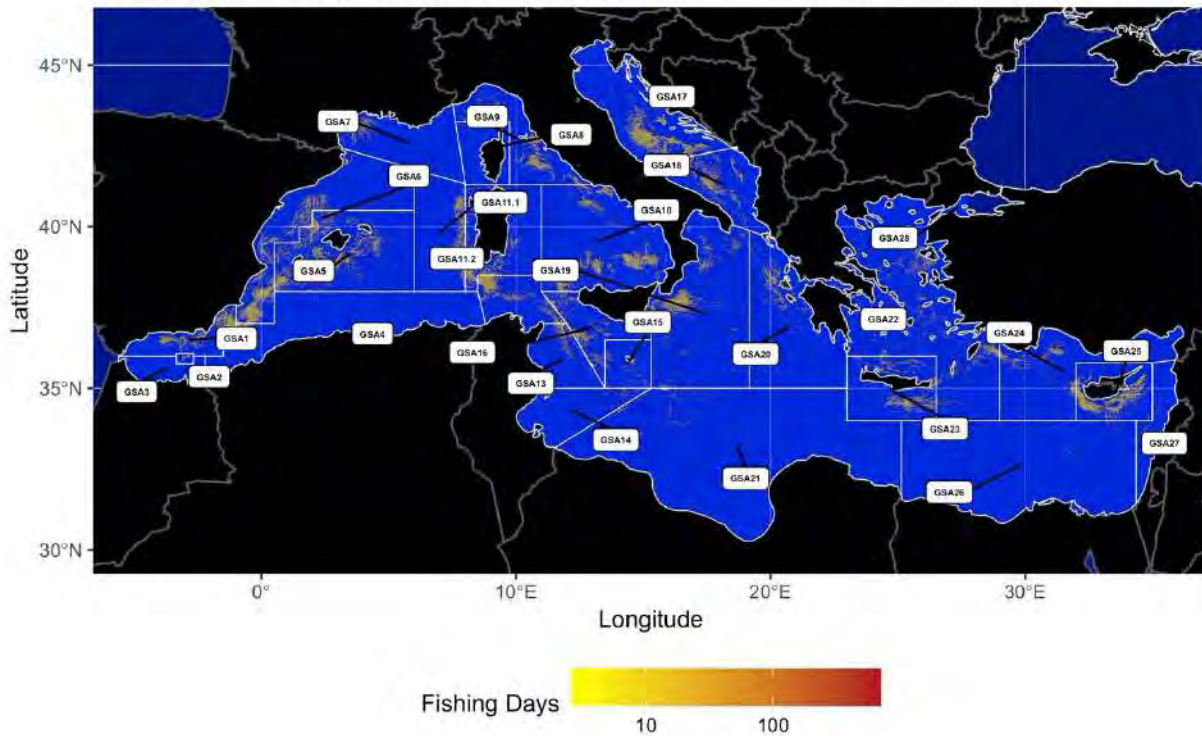


Figure 15 - Fishing footprint for longliners from AIS data, as obtained using VMSbase and the related procedures.

The main fishing areas are located in GSA1, GSA5, GSA9, GSA10, GSA11, GSA13, GSA16, GSA17, GSA18, GSA19, GSA23 and GSA25. The values of the effort range between 0 and around over 1000 fishing days.

The distribution of the total fishing effort, in days of fishing, for the whole Mediterranean fleets operating with traps, as obtained by VMSbase processing of the AIS dataset, is reported in Fig. 16.

In this case, few areas host fishing grounds. Namely, effort could be observed in GSA6 and GSA15 (Fig. 17). The values of the effort range between 0 and around 10 fishing days.

The distribution of the total fishing effort, in days of fishing, for the whole Mediterranean fleets operating with nets, as obtained by VMSbase processing of the AIS dataset, is reported in Fig. 18. Fishing areas could be detected in GSA6, along the southern coast of Spain, around Sicily (GSA10 and GSA16), in the Strait of Sicily (GSA13), around Malta (GSA15), in the Tyrrhenian Sea (GSA9), and in the Adriatic Sea (GSA17). The values of the effort range between 0 and around 100 fishing days.

AIS - Approach #1 (VMSBASE): Traps fishing effort - year 2017

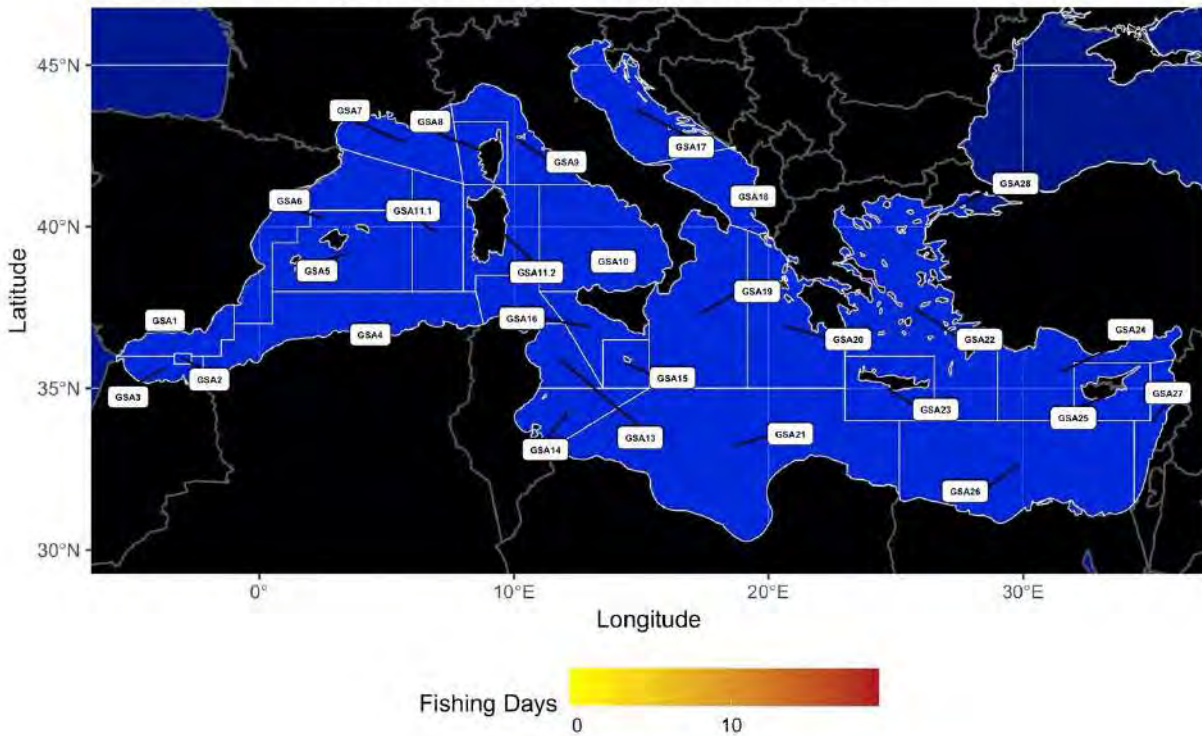


Figure 16 - Fishing footprint for traps from AIS data, as obtained using VMSbase and the related procedures

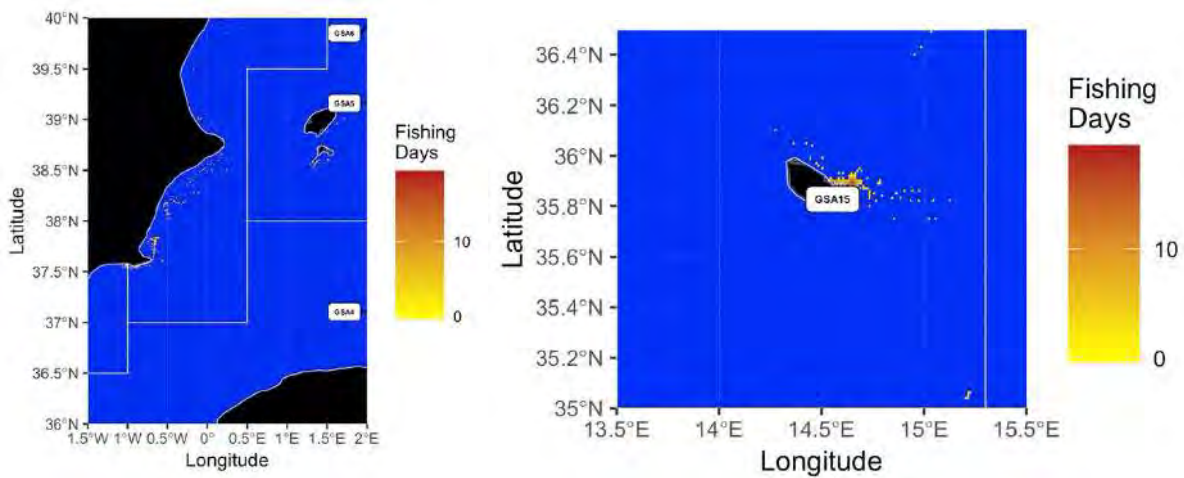


Figure 17 – Zooms on some area in which fishing activity with traps was detected. Fishing footprint for traps from AIS data, as obtained using VMSbase and the related procedures

AIS - Approach #1 (VMSBASE): Nets fishing effort - year 2017

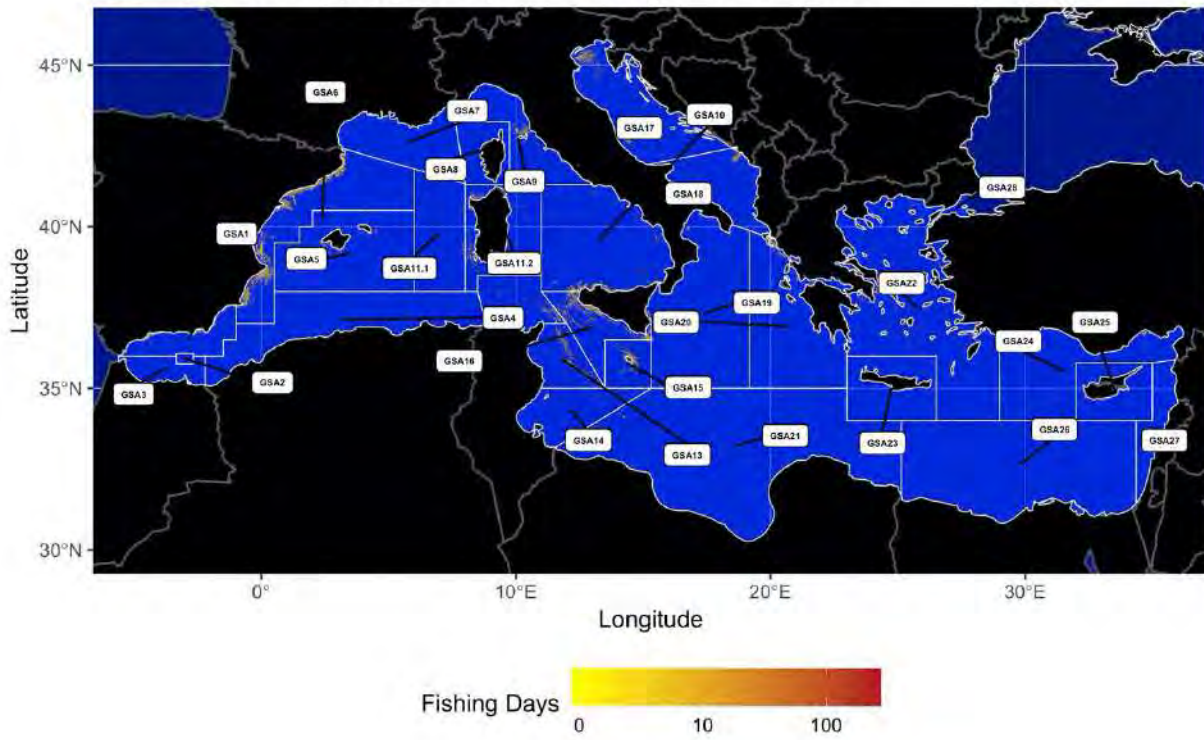


Figure 18 - Fishing footprint for nets from AIS data, as obtained using VMSbase and the related procedures

Analysis of AIS data: approach #2

The AIS-based activity in days of fishing of bottom trawlers, longliners, nets and traps operating in the whole Mediterranean in 2017, as obtained by using the approach developed by CNR the AIS dataset, is reported in Fig. 19-22.

AIS - Approach #2 (CNR): Trawl fishing effort - year 2017

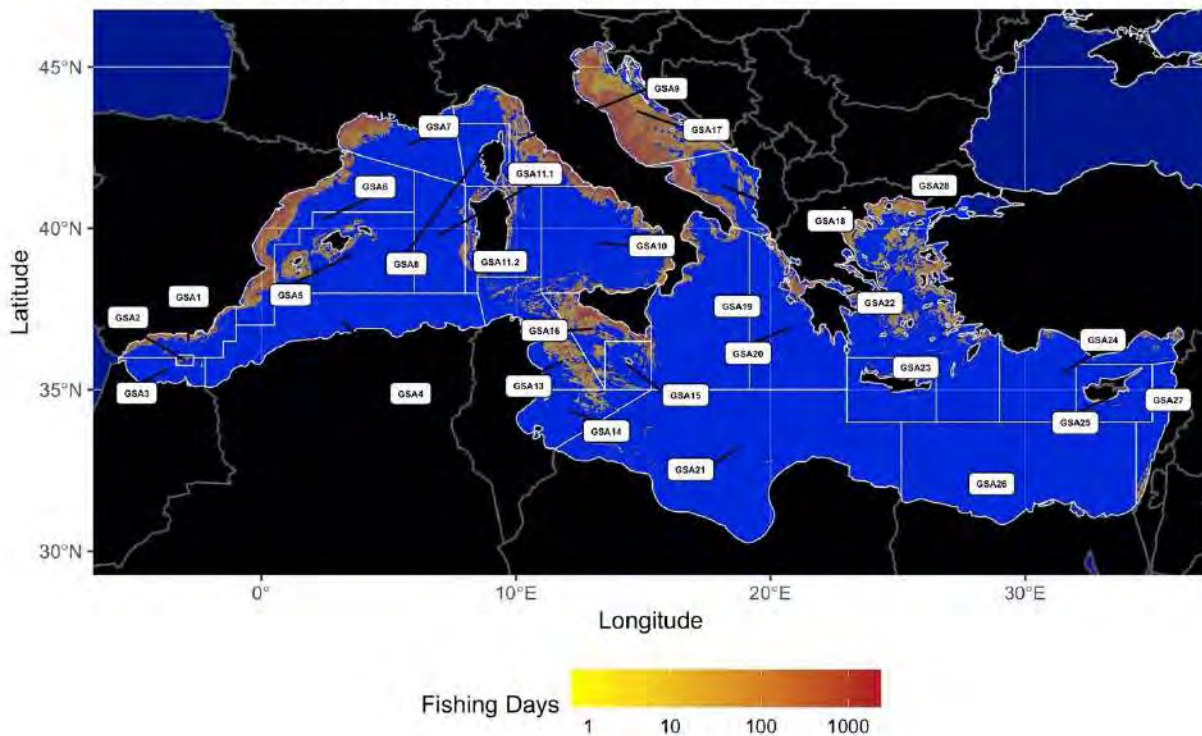


Figure 19 – AIS-based fishing footprints of trawlers during 2017, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

AIS appears to well capture a large fraction of trawlers and longliners (mostly industrial set longlines), while it underestimates fishing activity for nets and traps. It is likely due to the fact that these vessels (operating with these nets and traps) are mostly small-scale vessels not adopting AIS.

With regards to vessels equipped with AIS (above 15 m), European trawlers are dominant in the northern area of the Adriatic Sea (GSA17) and the coast of northeastern Spain (GSA6), the Aegean Sea (GSA22), and southern Sicily (GSA16). By contrast, North African countries have extremely low number of fishing trawlers broadcasting AIS and only a few operations in the southern parts of the Mediterranean Sea is due to non-European vessels (GSA21 and GSA27).

AIS detects longline operations in the western and eastern edges of the Mediterranean basin, especially off the Spanish and Cypriot coasts (respectively GSA1 and GSA25), as well as in the areas south of Sicily and Sardinia (respectively GSA11.1 and GSA11.2). The fishing footprint of longliners is also visible in the Central and Southern Adriatic Sea (GSA17 and GSA18), as well as in the Tyrrhenian and Ionian areas.

Some activity of vessels operating with nets is found in (GSA 17), while fishing footprints of traps are insignificant.

AIS - Approach #2 (CNR): Longlines fishing effort - year 2017

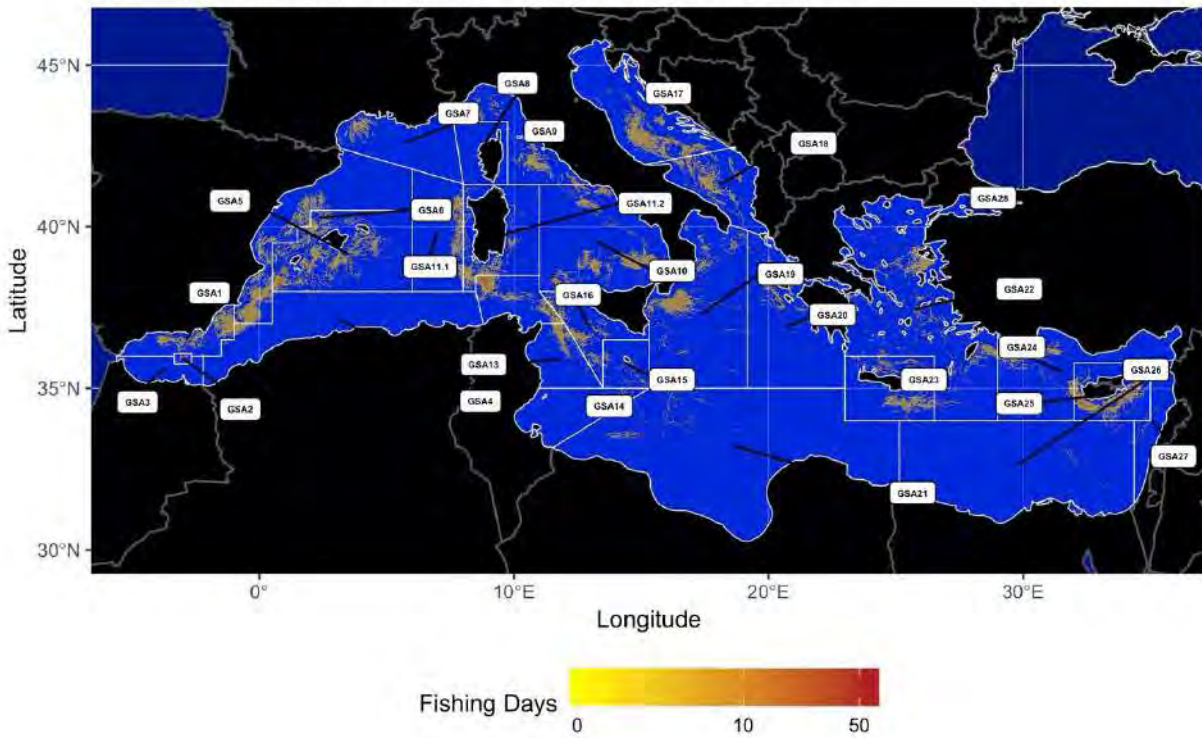


Figure 20 – AIS-based fishing footprints of longliners during 2017, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

AIS - Approach #2 (CNR): Nets fishing effort - year 2017

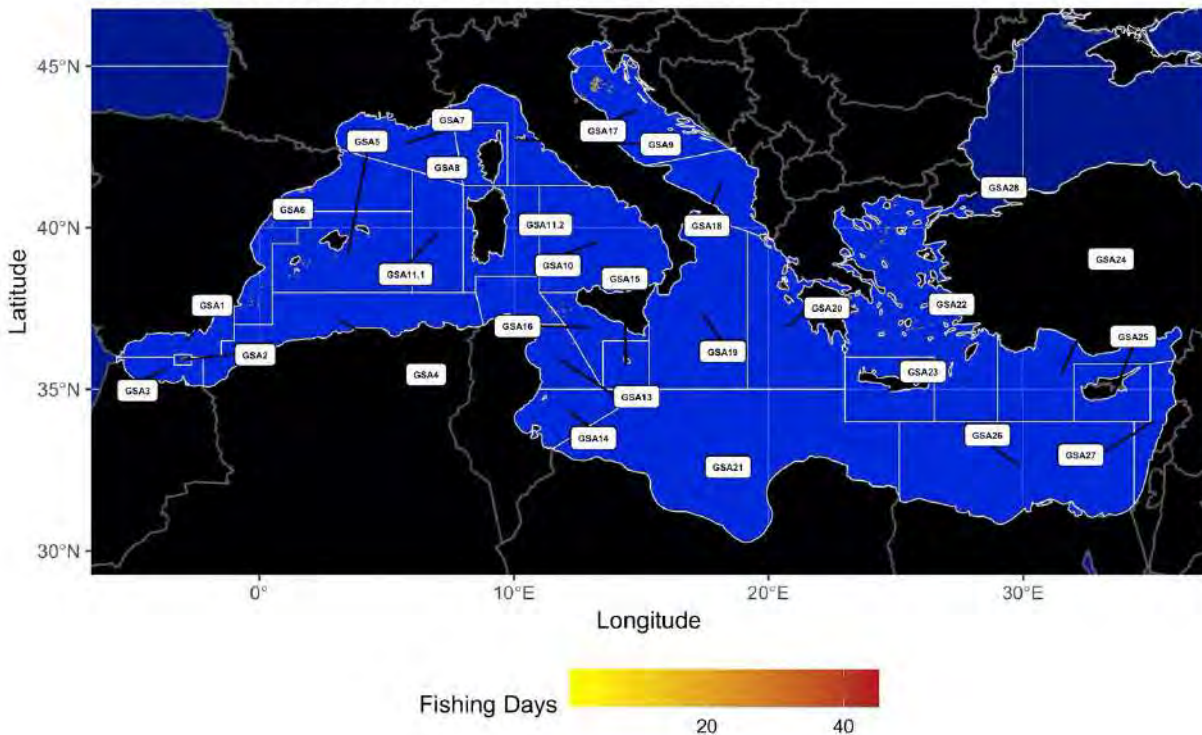


Figure 21 – AIS-based fishing footprints of nets during 2017, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

AIS - Approach #2 (CNR): Traps fishing effort - year 2017

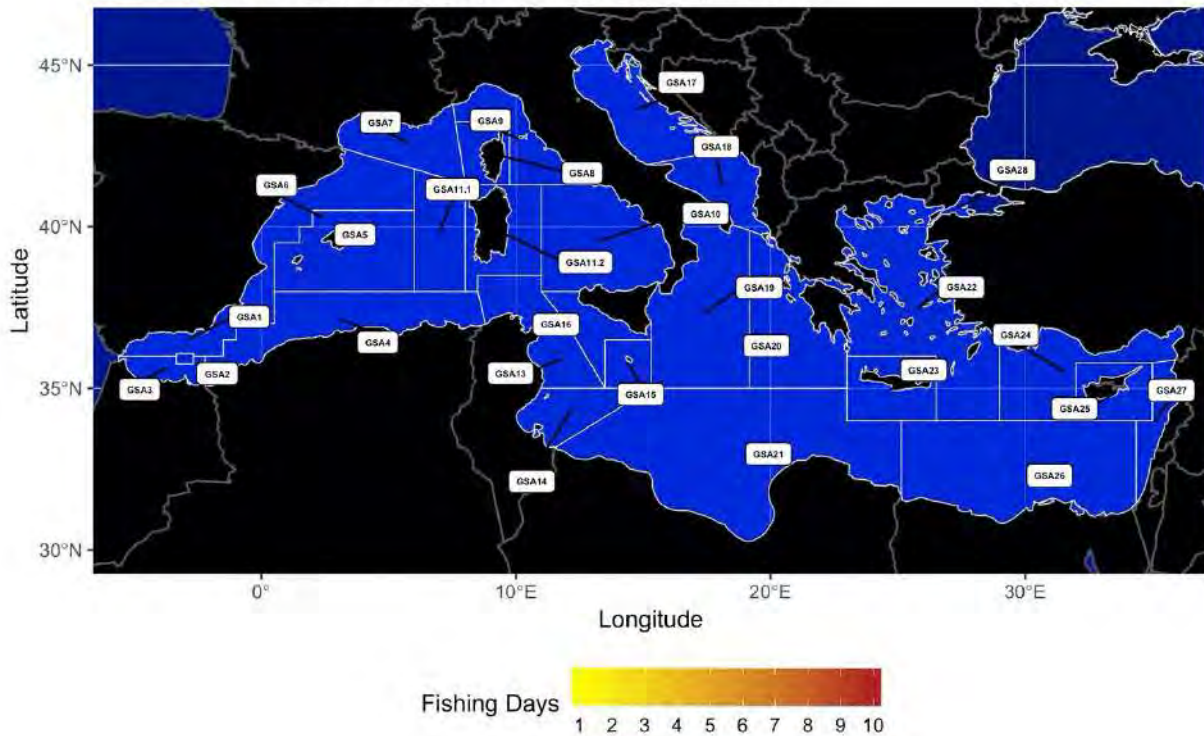


Figure 22 – AIS-based fishing footprints of traps during 2017, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

The value of the effort ranges between 0 and around 1500 fishing days for trawlers, while, at a lower scale, between 0 and around 60 for longliners. Fishing activity considerably decreases for nets and traps, reaching a maximum of 44 fishing days for vessels operating nets and only 4 fishing days for vessels operating traps.

Comparison with the EMODNET footprint

The distribution of the total fishing effort, for the year 2014, for the whole Mediterranean fleets operating with the Bottom trawls system, as obtained by EMODNET dataset, is reported in Fig. 23. Main fishing grounds are generally located in the same areas evidenced by the previous approaches. In fact, the Adriatic Sea (GSA17) the Strait of Sicily (GSA16), the Thyrrenian Sea (GSA9 and GSA10), the Gulf of Valencia (GSA6), the Gulf of the Lion (GSA7) and the Aegean Sea (GSA22) continue to be the areas in which most of the fishing effort is concentrated.

Even in this case, effort information along the coasts of North Africa does not appear to be present, except for some fished areas evidenced near the Tunisian coast (GSA13).

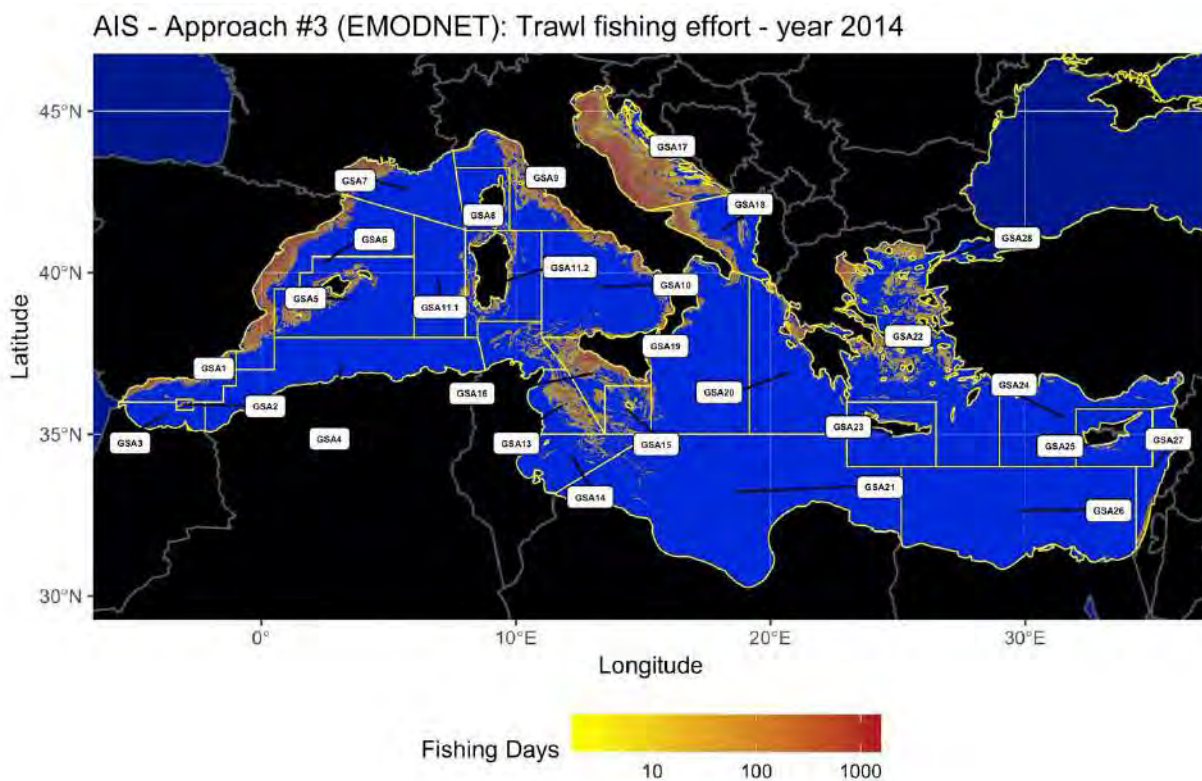


Figure 23 – AIS-based fishing footprints of trawlers during 2014, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

In Fig. 24 the comparisons between the trawl fishing effort evidenced by EMODNET and the VMSBASE and CNR approaches respectively are shown. Scatter plots evidenced differences between VMSBASE processed information and EMODNET data. In particular, a greater fishing effort – in number of days – is estimated by the EMODNET footprint. Similar values occur in the case of CNR approach, even if also in this case EMODNET footprint tends to maintain greater values of fishing effort.

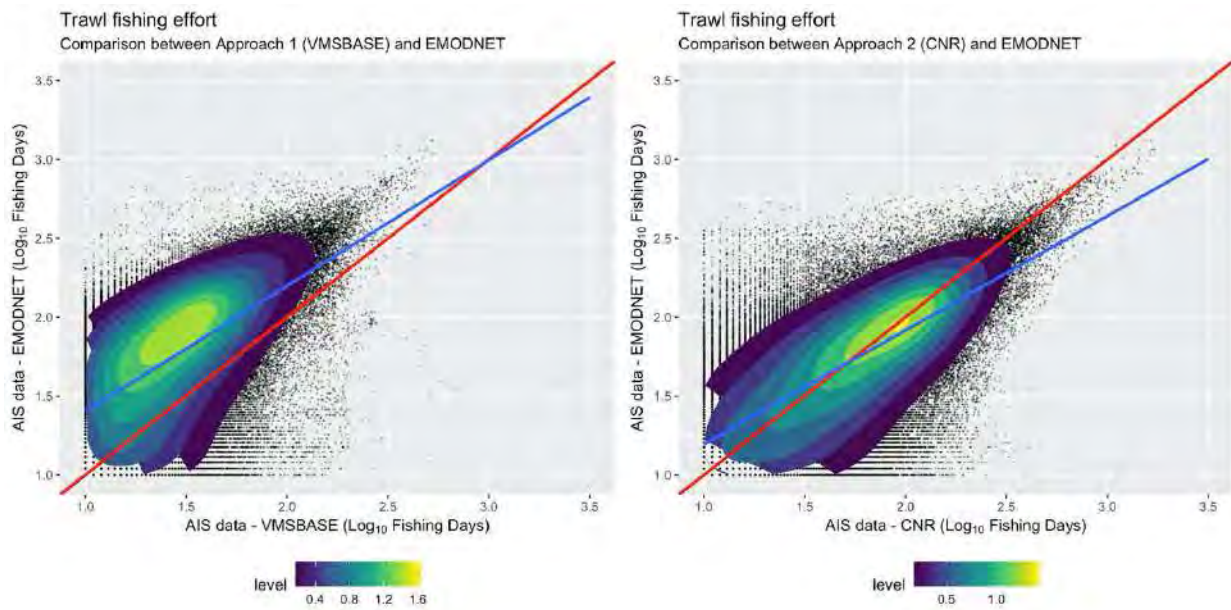


Figure 24 – Scatter plots representing the comparisons between trawl fishing effort evidenced by EMODNET information and VMSBASE and CNR approaches respectively

Validation of output #1: Comparison with fishing footprint by Global Fishing Watch

The distribution of the total fishing effort, for the year 2016, for the whole Mediterranean fleets operating with the Bottom trawls system, as obtained by Global Fishing Watch footprint, is reported in Fig. 25. Fishing effort estimation was obtained by following the procedure described by the Global Fishing Watch (<https://globalfishingwatch.org/>). Even in this case, main fishing grounds find spatial correspondence with the previous approaches. The values of the effort range are between 0 and around 1000 fishing days.

Global Fishing Watch: Trawl fishing effort - year 2017

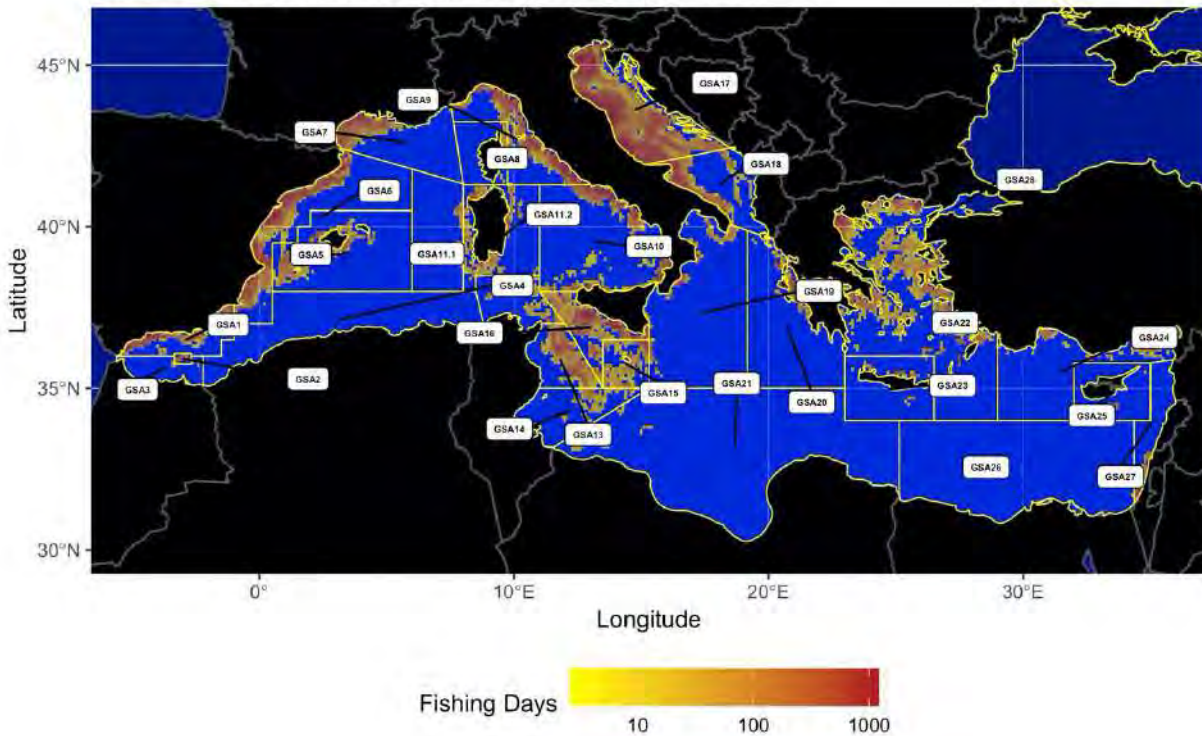


Figure 25 - AIS-based fishing footprints of trawlers during 2016, as obtained using dataset and the procedures of Global Fishing Watch

In Fig. 26 the comparisons between the trawl fishing effort evidenced by the Global Fishing Watch and the VMSBASE and CNR approaches respectively are shown. Scatter plots evidenced that Global Fishing Watch footprint clearly has greater fishing effort values in comparison to both approaches.

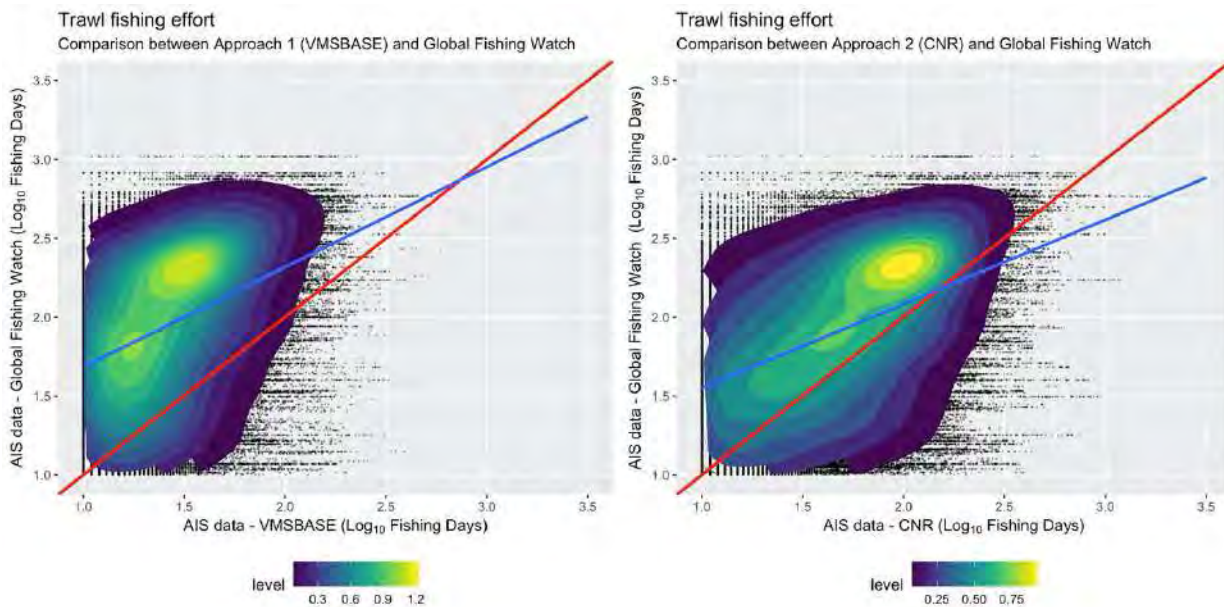


Figure 26 – Scatter plots representing the comparisons between trawl fishing effort evidenced by Global Fishing Watch information and VMSBASE and CNR approaches respectively.

The distribution of the total fishing effort, for the year 2016, for the whole Mediterranean fleets operating with longlines, as obtained by Global Fishing Watch footprint, is reported in Fig. 27. The main fishing areas are located in GSA1, GSA5, GSA11, GSA15, GSA19, GSA21, GSA23 and GSA25. The values of the effort range are between 0 and around 100 fishing days. From a spatial point of view, fished areas by longlines evidenced by the Global Fishing Watch footprint find agreement with those of previous approaches only in the cases of the Northern Alboran Sea (GSA1) and the Cyprus area (GSA25).

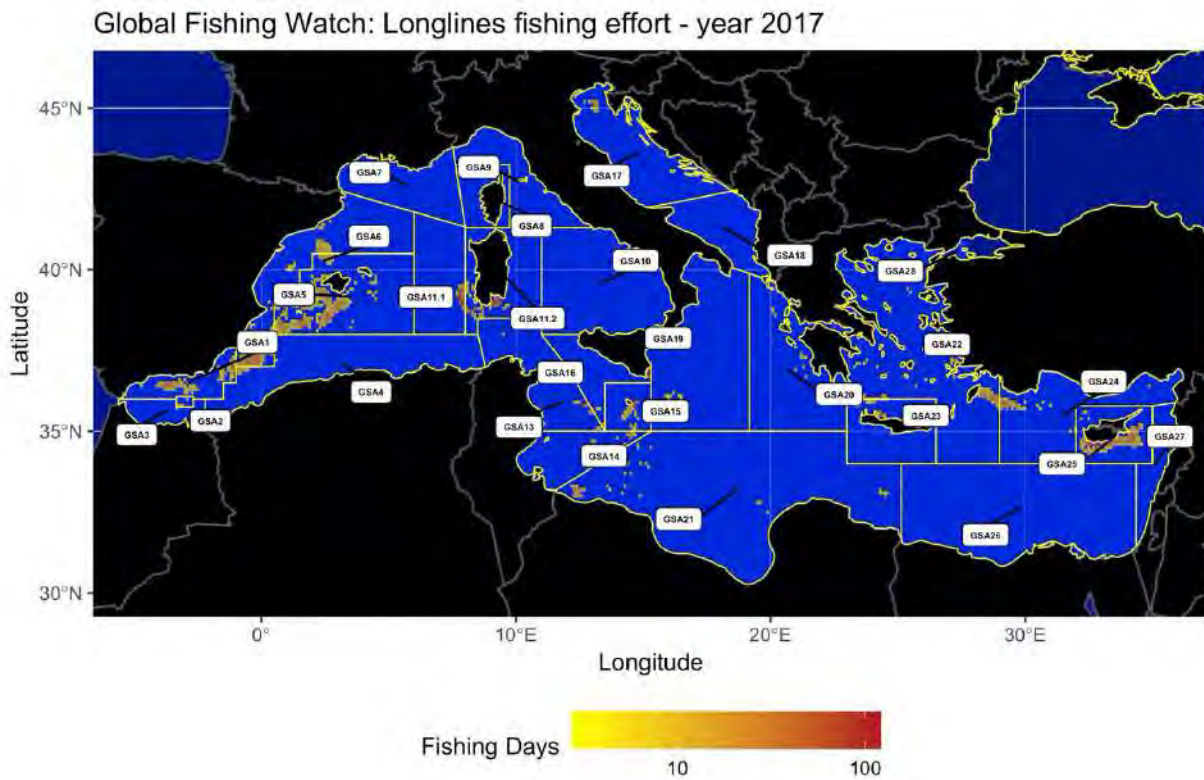


Figure 27 - AIS-based fishing footprints of longlines during 2016, as obtained using dataset and the procedures of Global Fishing Watch

In Fig. 28 the comparisons between the longlines fishing effort evidenced by the Global Fishing Watch and the VMSBASE and CNR approaches respectively are shown. In both cases, Global Fishing Watch showed higher fishing effort values.

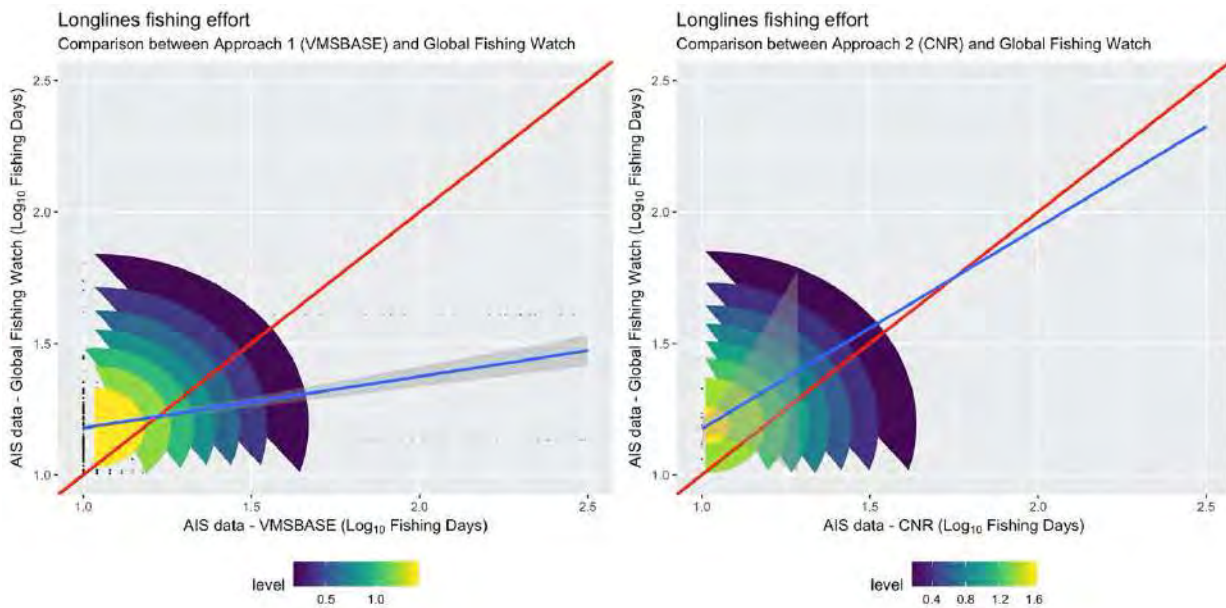


Figure 28 – Scatter plots representing the comparisons between longlines fishing effort evidenced by Global Fishing Watch information and VMSBASE and CNR approaches respectively.

Validation of output #2: Comparison with fishing footprint by MINOUW project (AIS and VMS data combined)

The pattern of fishing footprint obtained, for the year 2016, from the processing with VMSbase of some outputs of the MINOUW projects (namely the trawling effort for the case studies 1.4 - SEA BOTTOM TRAWL IN CATALONIA GSA06, 1.5 - BOTTOM TRAWLING IN SICILY GSA16, 1.7 - AEGEAN SEA BOTTOM TRAWL GSA22, 1.6/1.8 - BOTTOM TRAWL FISHERIES IN LIGURIAN AND N. TYRRHENIAN SEA GSA09) is represented in Fig. 29. It is important to remember that this output was obtained combining, by area of study, the VMS and AIS data of each fleet.

It is worth noting that all these MINOUW case studies correspond to areas in which the level of effort is very high.

However, Fig. 30 indicates that, when compared with both previous approaches (VMSBASE and CNR), MINOUW project information generally showed greater values of fishing effort in terms of fishing days in relation to the VMSBASE processing approach and lower values to the CNR approach.

AIS + VMS - Approach #3 (MINOUW-VMSBASE): Trawl fishing effort - year 2016

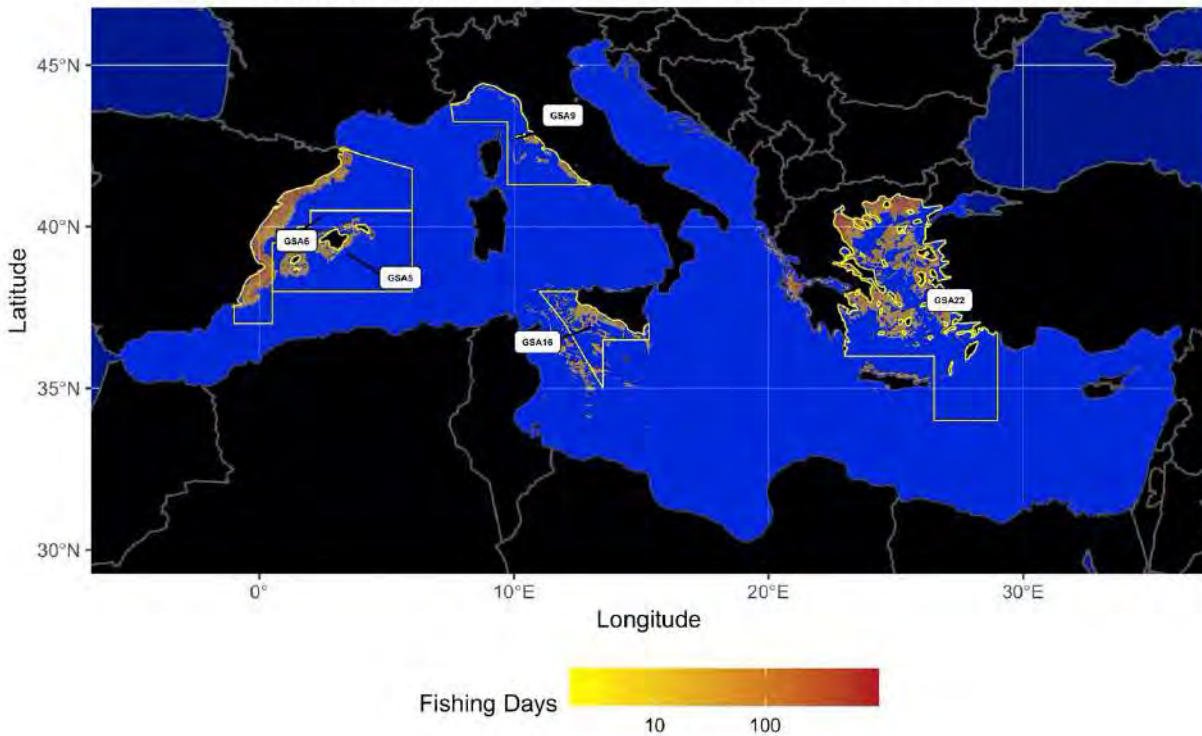


Figure 29 - Fishing footprint from AIS and VMS data, as obtained using VMSbase and the related procedures, for the MINOUW Project

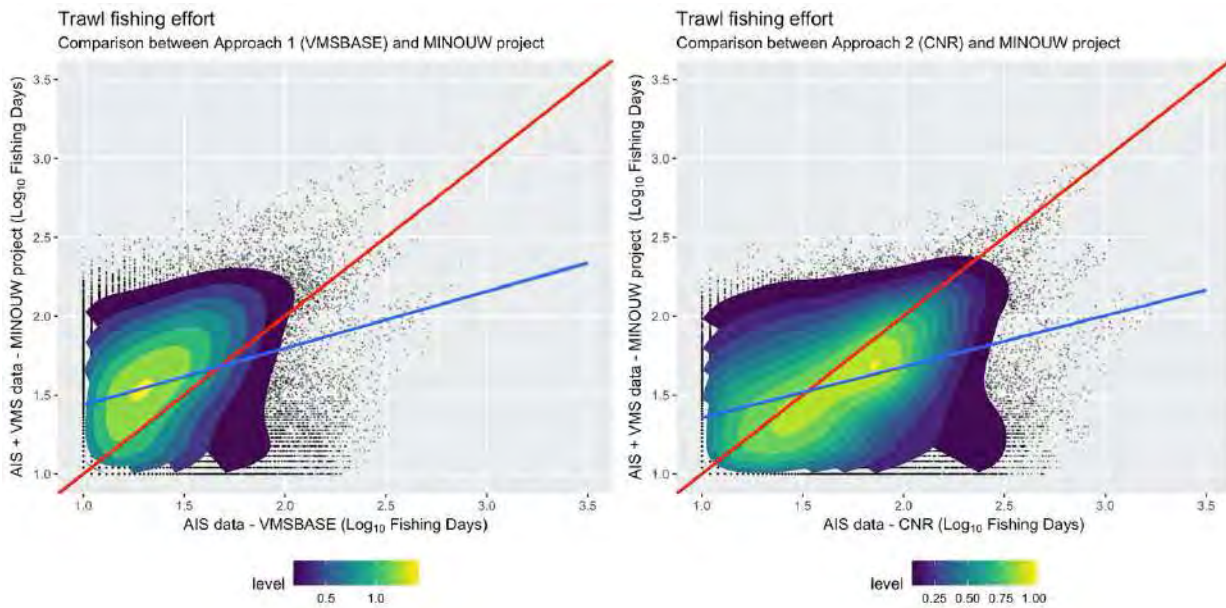


Figure 30 – Scatter plots representing the comparisons between trawl fishing effort evidenced by MINOUW Project information and VMSBASE and CNR approaches respectively.

In Fig. 31 the comparisons between the VMSBASE processing and CNR approach are shown for each gear type (i.e. trawl, longlines, traps and nets). While greater values of trawl fishing effort occur for CNR approach, they instead resulted to be lower for the longlines fishing effort in relation to the VMSBASE processing approach.

Both nets and traps fishing effort resulted in higher values in the case of the CNR-AIS approach.

General comparison

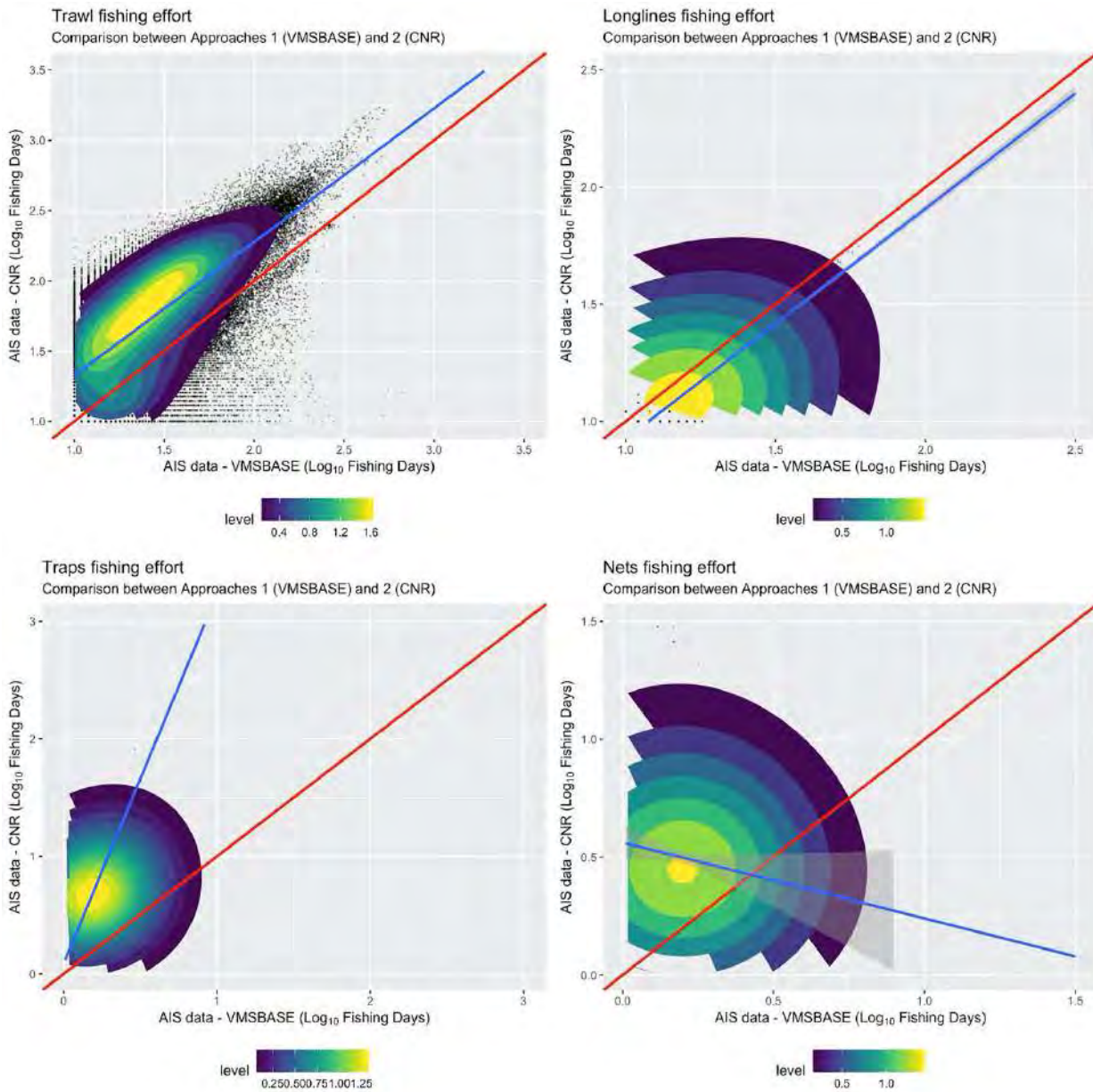


Figure 31 – Scatter plots representing the comparisons between fishing effort evidenced by the VMSBASE and CNR approaches respectively. Results are showed for each gear type (i.e. trawl, longlines, traps and nets).

By looking at the spatial extension of the trawl fishing activities for each GSA (Fig. 31), several considerations can be taken into account. First of all, Global Fishing Watch AIS information generally estimates wider areas of fishing activity. In fact,

for each GSA (except GSA4 and GSA28), the number of cells resulted by the Global Fishing Watch far exceeds that of the other approaches.



Figure 32 – Bar plot of spatial extension of trawl fishing activities for each GSA in relation to the presented approaches. Spatial information is expressed as number of cells where fishing effort occur.

Spatial extension, in terms of number of grid cells, resulted about the same among the GSAs for the other methods (i.e. VMSBASE, CNR and EMODNET). Overall, AIS information provided by the VMSBASE and CNR processing resulted in an underestimation of the spatial extent of trawl fishing effort if compared to the Global Fishing Watch. Nevertheless, for both approaches, fishing effort estimated in areas

closest to the coast resulted higher than that highlighted by the Fishing Global Watch information.

Spatial extension provided by the MINOUW Project information resulted to have agreement with other methods in the GSA5, GSA6 and GSA20 while, on the other hand, wider extension resulted for the Aegean Sea and Crete (GSA22 and GSA23).

Multi-Criteria Decision Analysis for both large- and small-scale components of the fleets without AIS/VMS

FPc for bottom trawlers and small-scale fisheries was derived based on the indices Sc and Ac. Values close to 1 indicate areas with an elevated likelihood of intense fishing pressure. Fig. 33 indicates that bottom trawling fishing pressure is usually distributed in higher depths. Fig. 34 indicates that small scale fishing fleet usually operates mostly along the coastline.

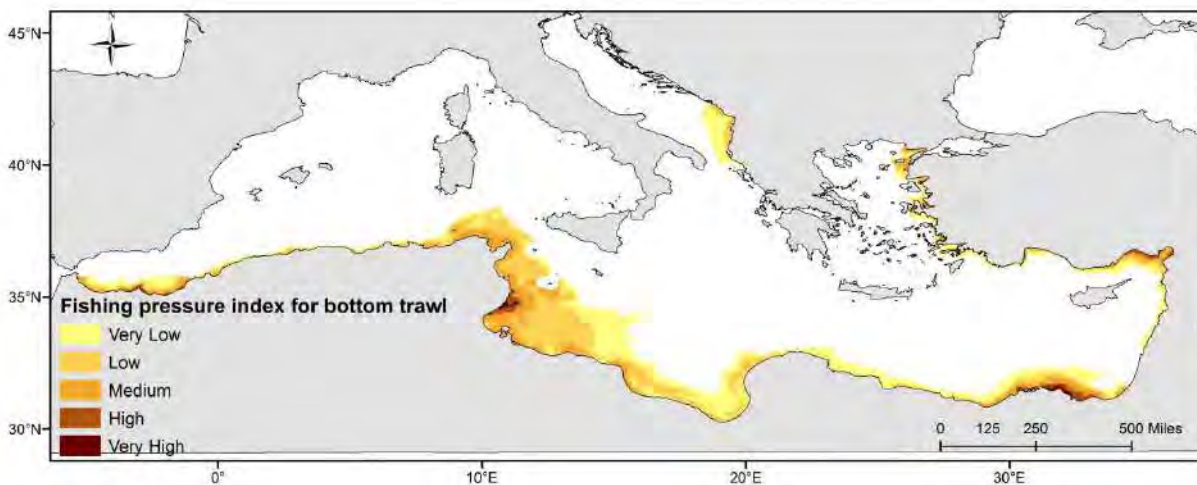


Figure 33 - Fishing pressure index of bottom trawlers estimated by an MCDA approach for the non-EU Mediterranean countries

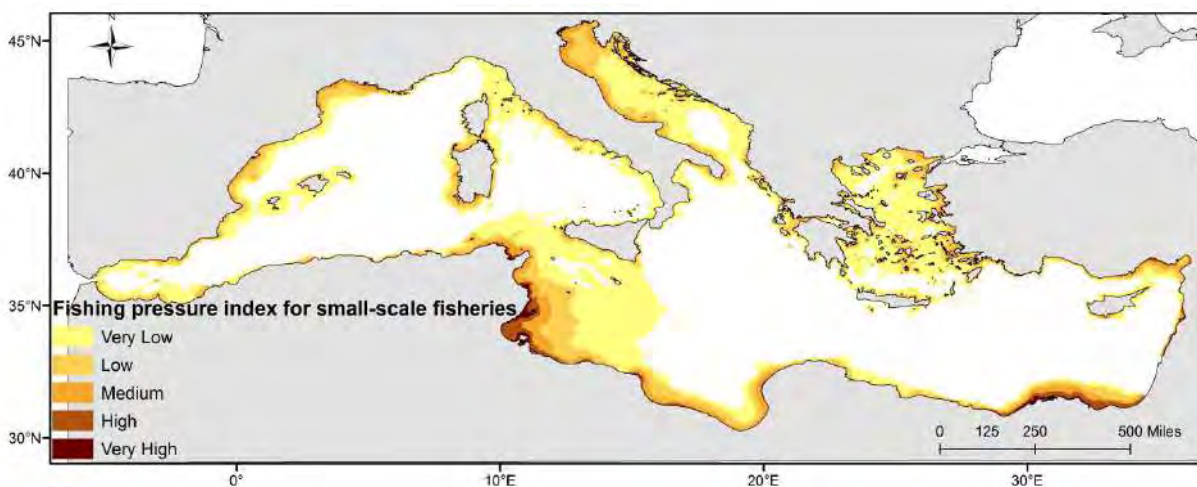


Figure 34 - Fishing pressure index of small-scale fisheries estimated by an MCDA approach for the Mediterranean Sea

An evaluative account of different qualitative aspects of the data describing the Ac has been attempted using the modified pedigree matrix presented above (Table 15). The pedigree matrix reveals that the weakest indicators of the data used for the Ac estimation were: completeness and data collection process quality, since most of the components used for the estimation of Ac were scored with high values (low quality) for these attributes. In an attempt to visualize the qualitative information gained from implementation of the pedigree matrix, radar diagrams were also produced (Fig. 35) presenting the score for each indicator used in the qualitative assessment of the Ac. In addition, a map showing an evaluation of the estimated FPI index accuracy based on weaknesses of Ac minus the weaknesses taken into account to the MCDA process is indicated in Fig. 36.

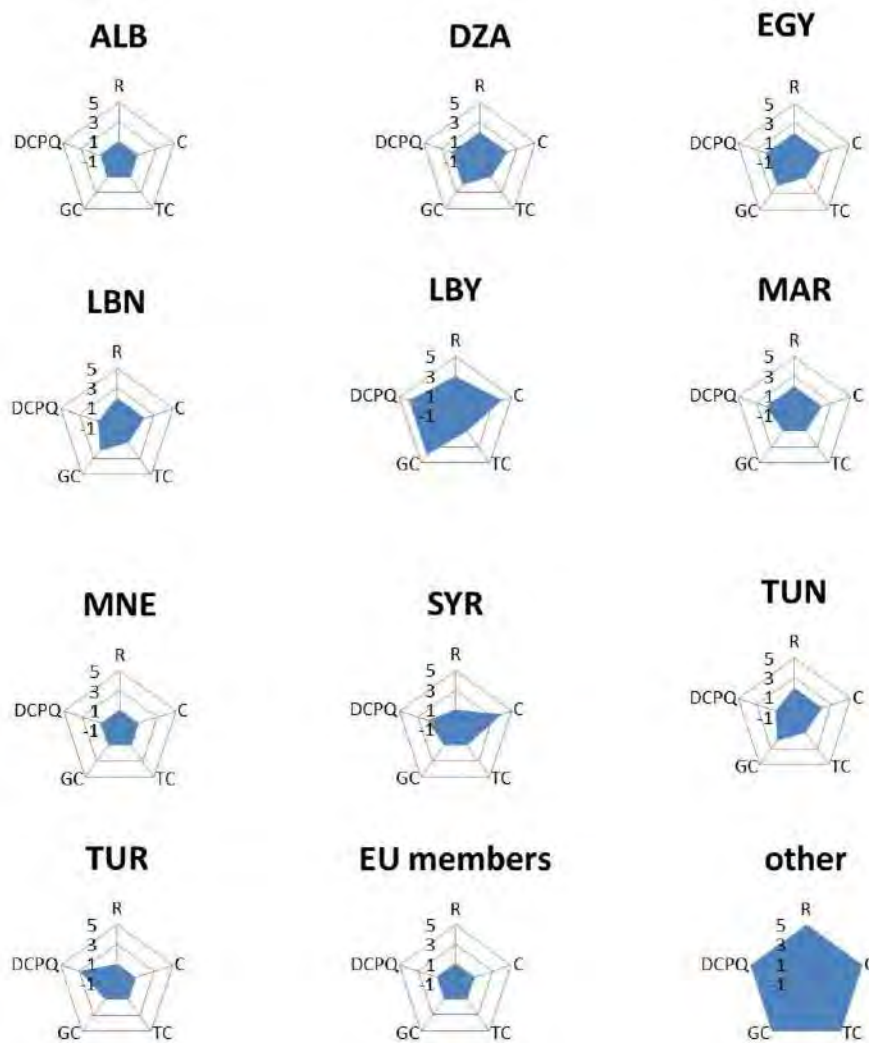


Figure 35 - Radar diagrams based on the pedigree matrix indicating the scores for each indicator (see Table 15). R: Reliability; C: Completeness; TC: Temporal Correlation; GC: Geographical Correlation and DCPQ: Data Collection Process Quality

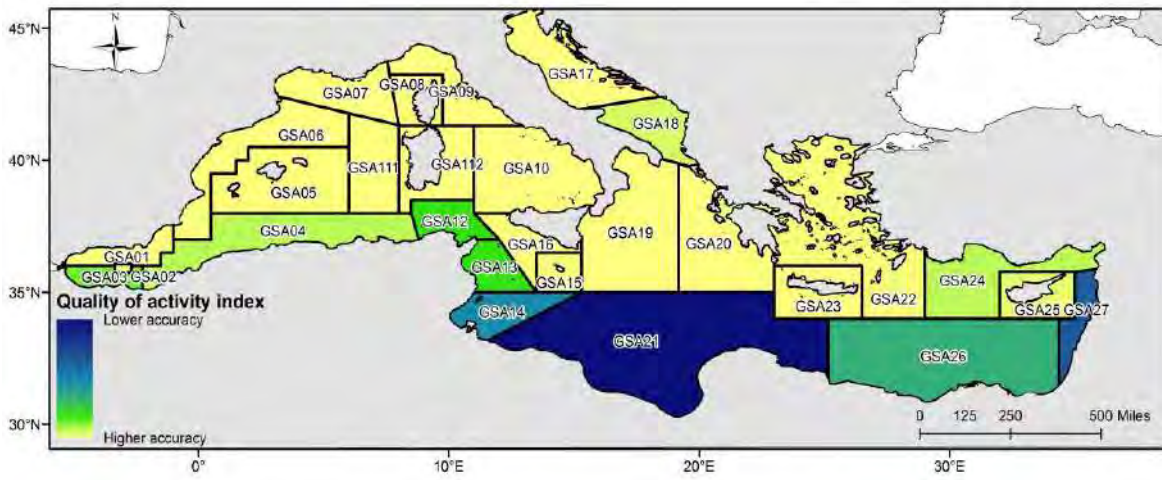


Figure 36 - Estimated quality of Activity index (Ac) per GSA

Cascaded Multilayer perceptron network applied on environmental and fleet data for some component of the fleet without AIS/VMS

The output of the application of the cascaded multilayer perceptron network (CMPN) trained on the Adriatic Sea (Russo *et al.* 2019b), are represented in Fig 37. It is important to stress that this prediction has been restricted to the subset of trawling fleets with LOA between 12 and 15 m. Anyway, this portion of the fleet is very important because the number of vessels belonging to this fleet segment is very high, and they represent, literally, the largest trawlers not equipped with tracking devices (i.e AIS or VMS).

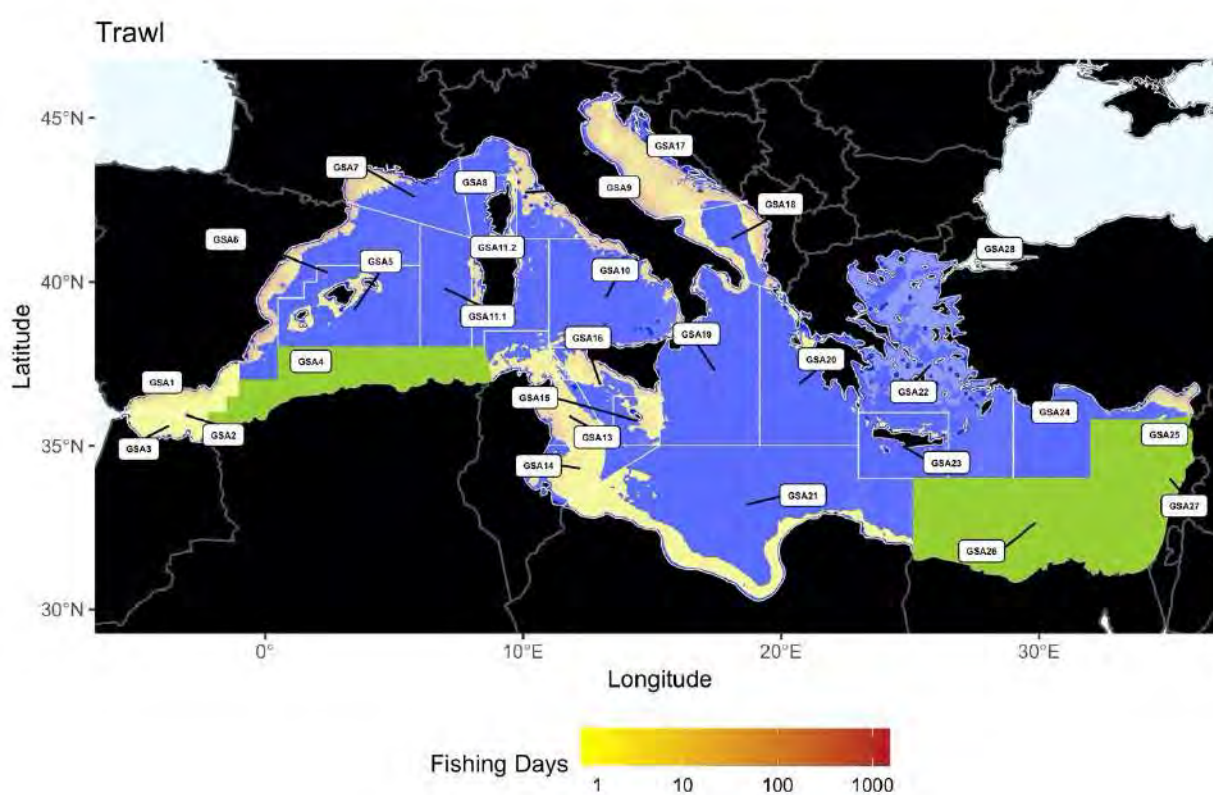


Figure 37 – Predicted fishing footprint for trawlers with LOA between 12 and 15 m, as obtained using the CMPN described in Russo *et al.*, 2019b on the official data about fleet structure in the Mediterranean Sea (GFCM). GSAs in which no data about vessels belonging to this fleet segment were available are colored in green.

The pattern obtained indicates that the activity of this subset is relevant in some areas, including the Strait of Sicily, the Sirte's Gulf, the Adriatic Sea, the Aegean Sea and the southern coasts of Spain and France. According to the topology of the Mediterranean Sea, the operative range of these vessels is shorter in GSAs with small shelf. The magnitude of the predicted effort is comparable with the that of results presented in the previous sections.

Hot spots of fishing activity

Hot Spot Analysis was run to evaluate potential clusters of high or low fishing effort values, and identify statistically significant hot spots using the local Getis-Ord G_i^* statistic (Getis and Ord 2010), by conceptualizing spatial relationships through a fixed bandwidth of 10 km. The statistical significance/confidence interval of the data is given in confidence level bin (G_i _Bin), which ranges from -3 (Cold Spot – 99% Confidence) to 3 (Hot Spot – 99% Confidence), with 0 being non-significant.

Results show statistically significant hot spots in trawl and longline footprints (Fig. 38, 39, 42 and 43). In terms of the interpretation of the results, it is worth noting that whenever a spot reaches a high value it automatically becomes relevant from the point of view of the fishing activity, although it may not be a statistically significant hotspot; i.e., if it is isolated.

The results of the hot spot analysis for traps and nets mostly reflect no spatial clustering and show only some small homogeneous groupings in the Mediterranean Sea (Fig. 40, 41, 44 and 45). It is also worthy of note that the areas that are not significant span a vast zone, due to the heterogeneous nature of the AIS-based variable under analysis and, as a result, the difficulty of making homogeneous groupings. Anyway, although the spatial distribution of fishing activity for nets and traps does not show a clear territorial pattern, it cannot be attributed to a real random effect rather to the inability of AIS data to capture these gears.

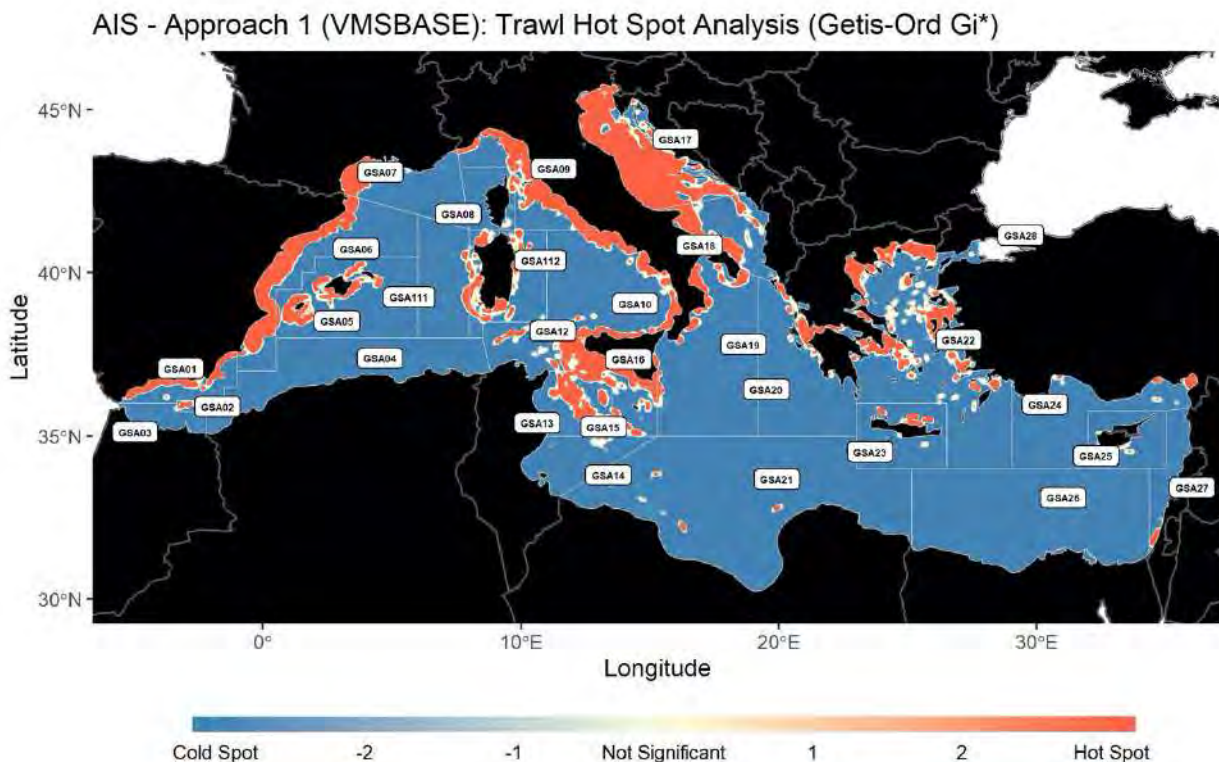


Figure 38 - Hot Spot Analysis for trawl fishing activity as obtained using the VMSbase approach (grid resolution 1kmx1km)

AIS - Approach 1 (VMSBASE): Longline Hot Spot Analysis (Getis-Ord G_i^*)

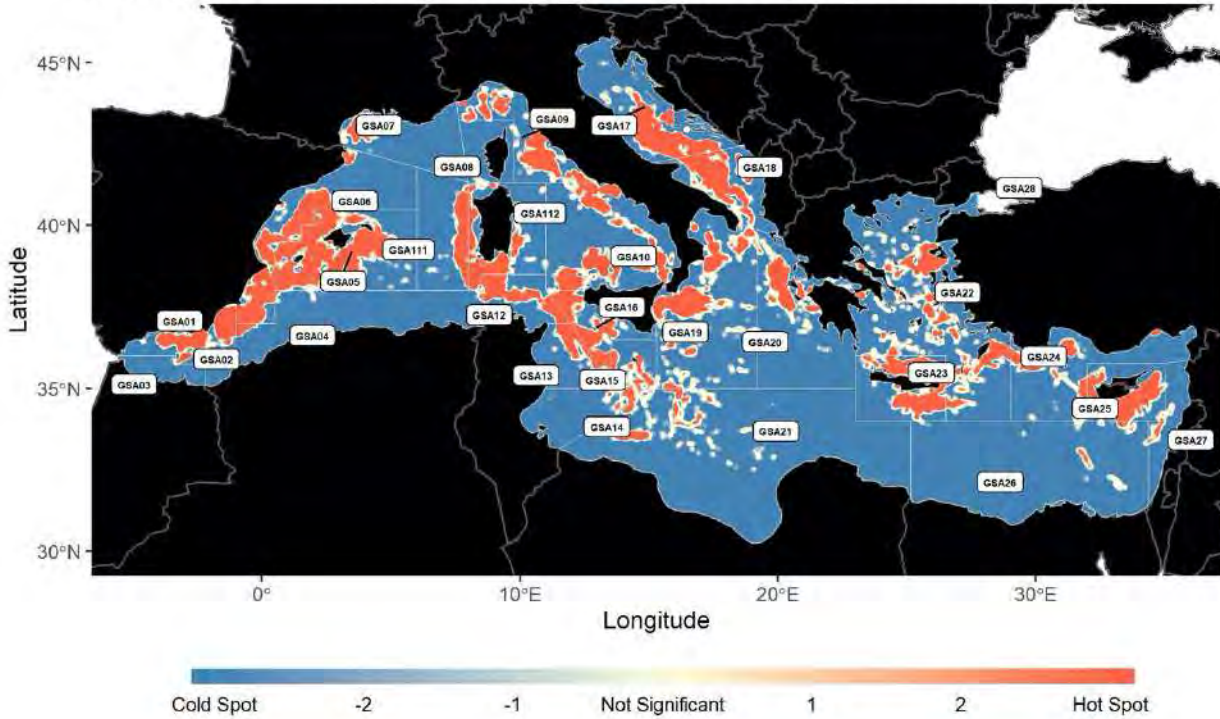


Figure 39 - Hot Spot Analysis for Longline fishing activity, as obtained using the VMSbase approach (grid resolution 1kmx1km)

AIS - Approach 1 (VMSBASE): Nets Hot Spot Analysis (Getis-Ord G_i^*)

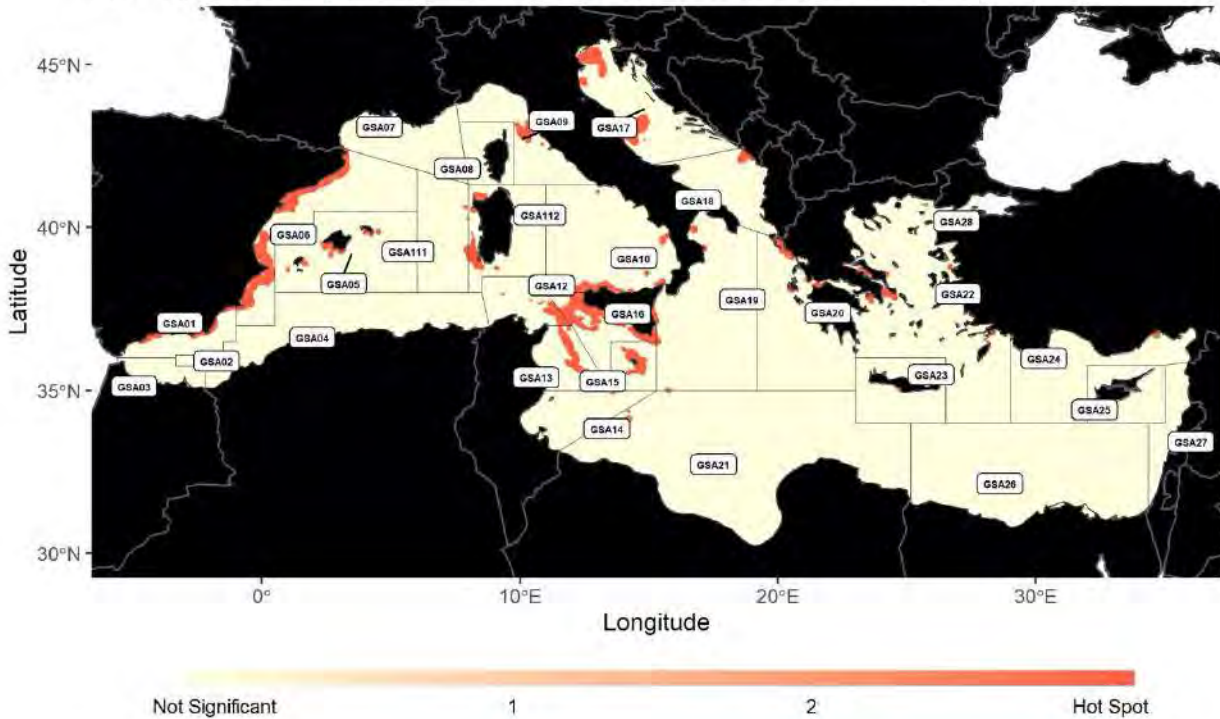


Figure 40 - Hot Spot Analysis for Nets fishing activity as obtained using the VMSbase approach (grid resolution 1kmx1km)

AIS - Approach 1 (VMSBASE): Traps Hot Spot Analysis (Getis-Ord G_i^*)

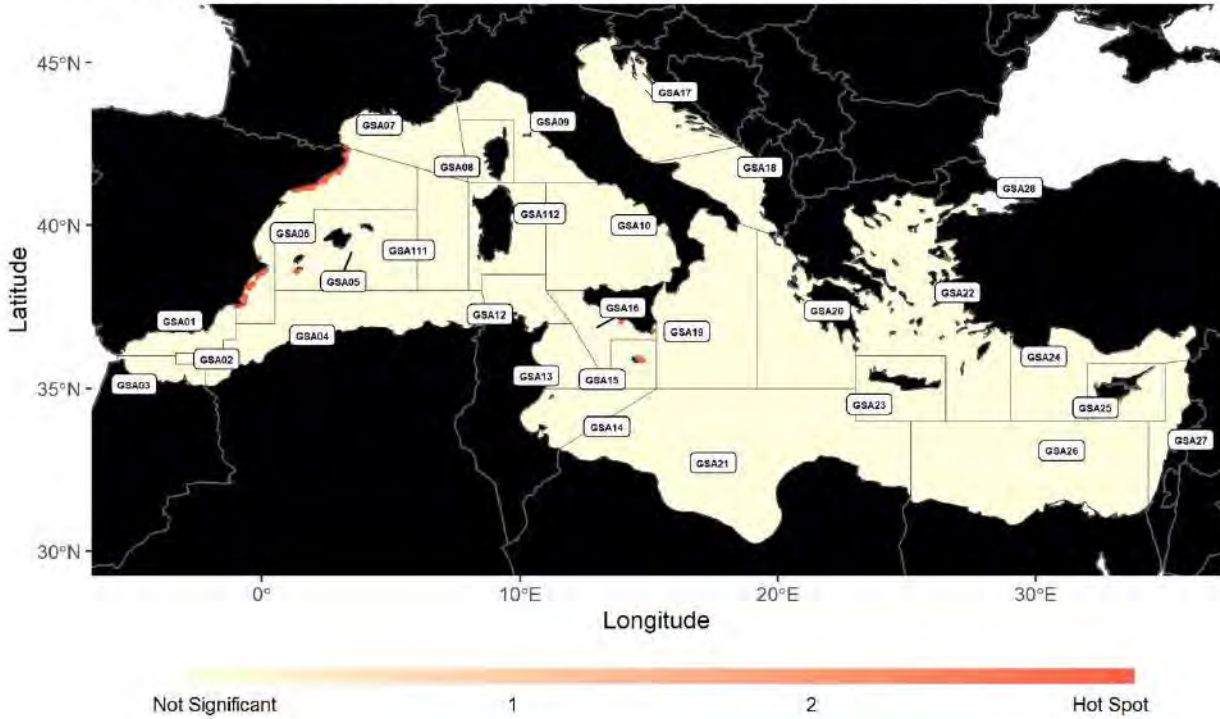


Figure 41 - Hot Spot Analysis results for Traps fishing activity, as obtained using the VMSbase approach (grid resolution 1kmx1km)

AIS - Approach 2 (CNR): Trawl Hot Spot Analysis (Getis-Ord G_i^*)

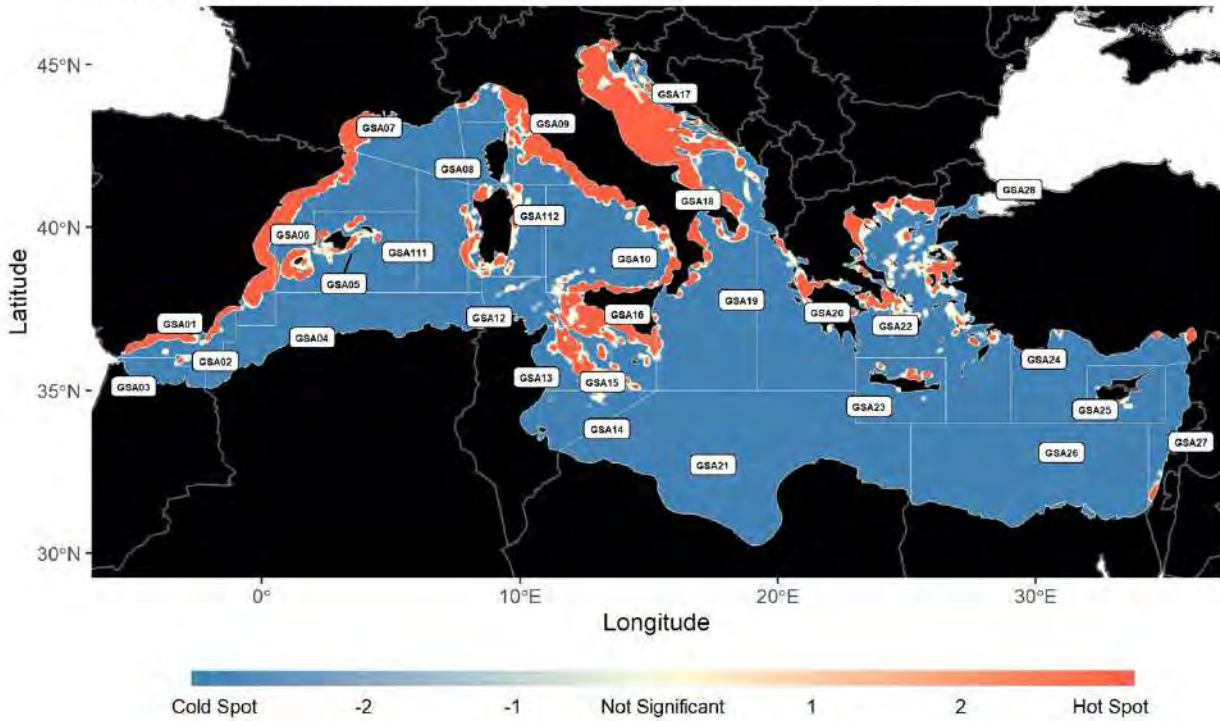


Figure 42 - Hot Spot Analysis for Trawl fishing activity, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

AIS - Approach 2 (CNR): Longline Hot Spot Analysis (Getis-Ord G_i^*)

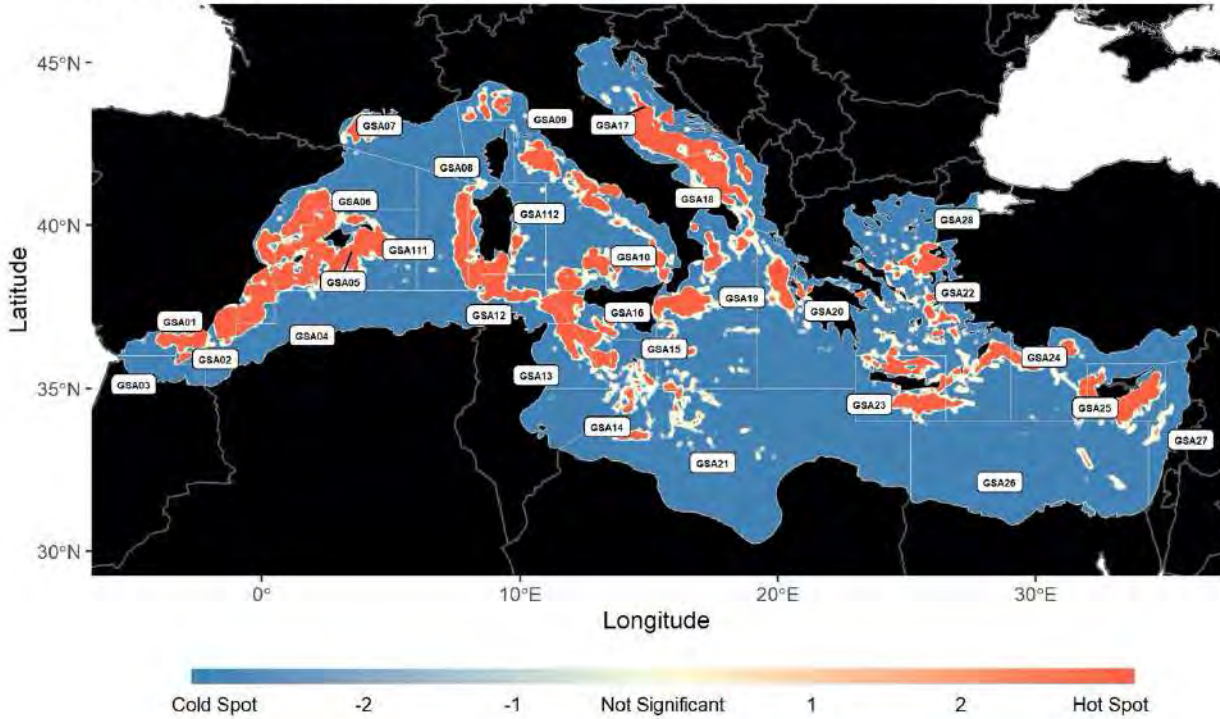


Figure 43 - Hot Spot Analysis for Longline fishing activity, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

AIS - Approach 2 (CNR): Nets Hot Spot Analysis (Getis-Ord G_i^*)

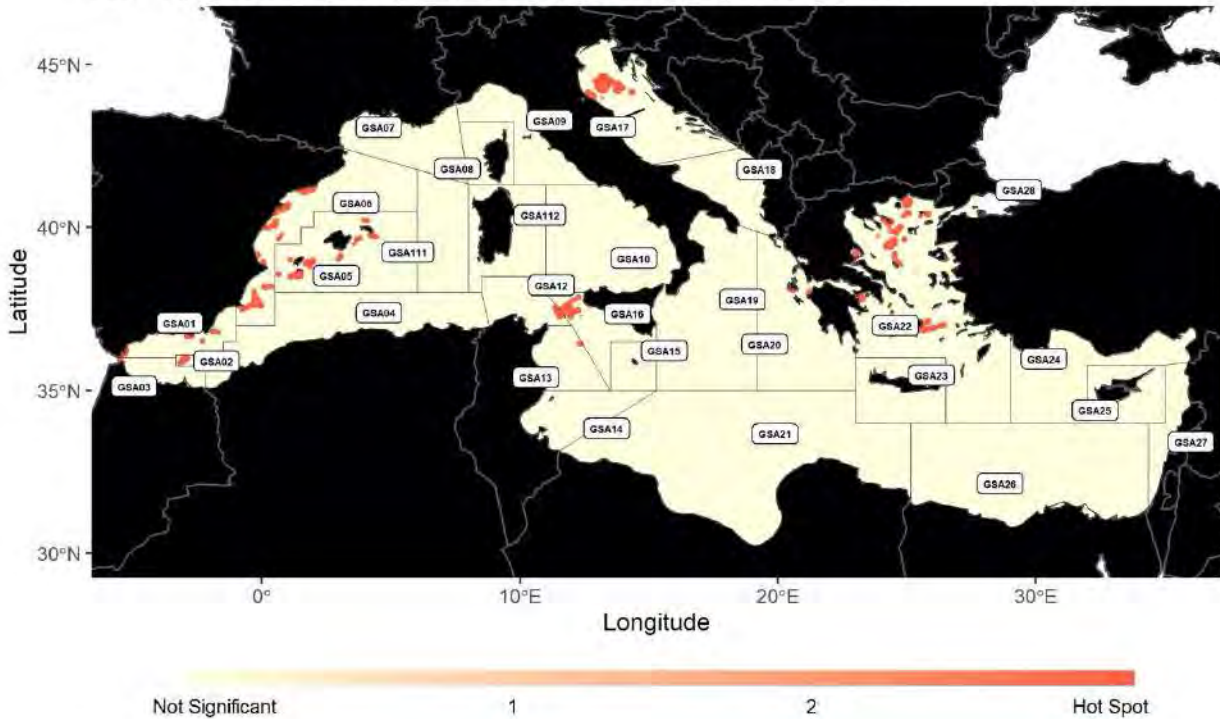


Figure 44 - Hot Spot Analysis for Nets fishing activity, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

AIS - Approach 2 (CNR): Traps Hot Spot Analysis (Getis-Ord G_i^*)

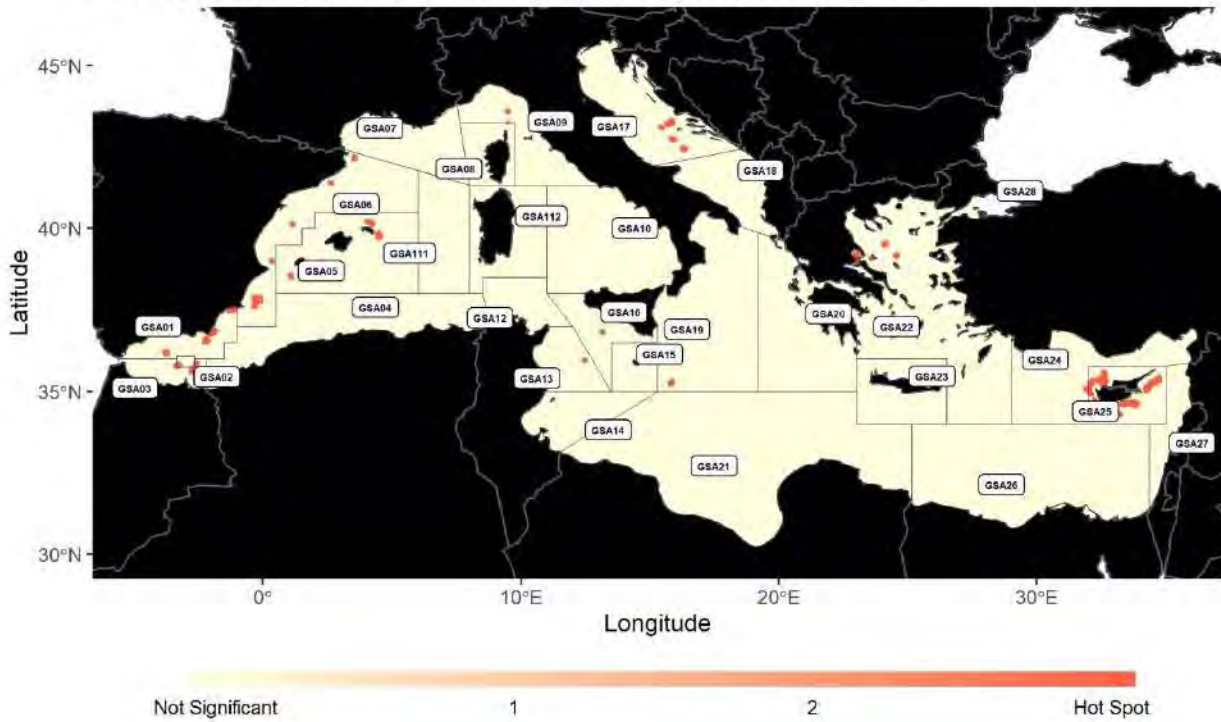


Figure 45 - Hot Spot Analysis results for Traps fishing activity, as obtained using the approach developed by CNR (grid resolution 1kmx1km)

Reconstruction of fluxes from fishing grounds to harbors

The NNLS-based regression approach applied on 1) effort data from Approach #2 (CNR), and 2) total landings in 2017 by species and country from the 2019 Annual Economic Report (AER) on the EU Fishing Fleet (STECF 19-06) (Carvalho et al. 2019), returned both estimates of landings by country and by GSA that were compared with the official ones. The aggregated (without reference to the GSAs) values are showed in Fig. 46.

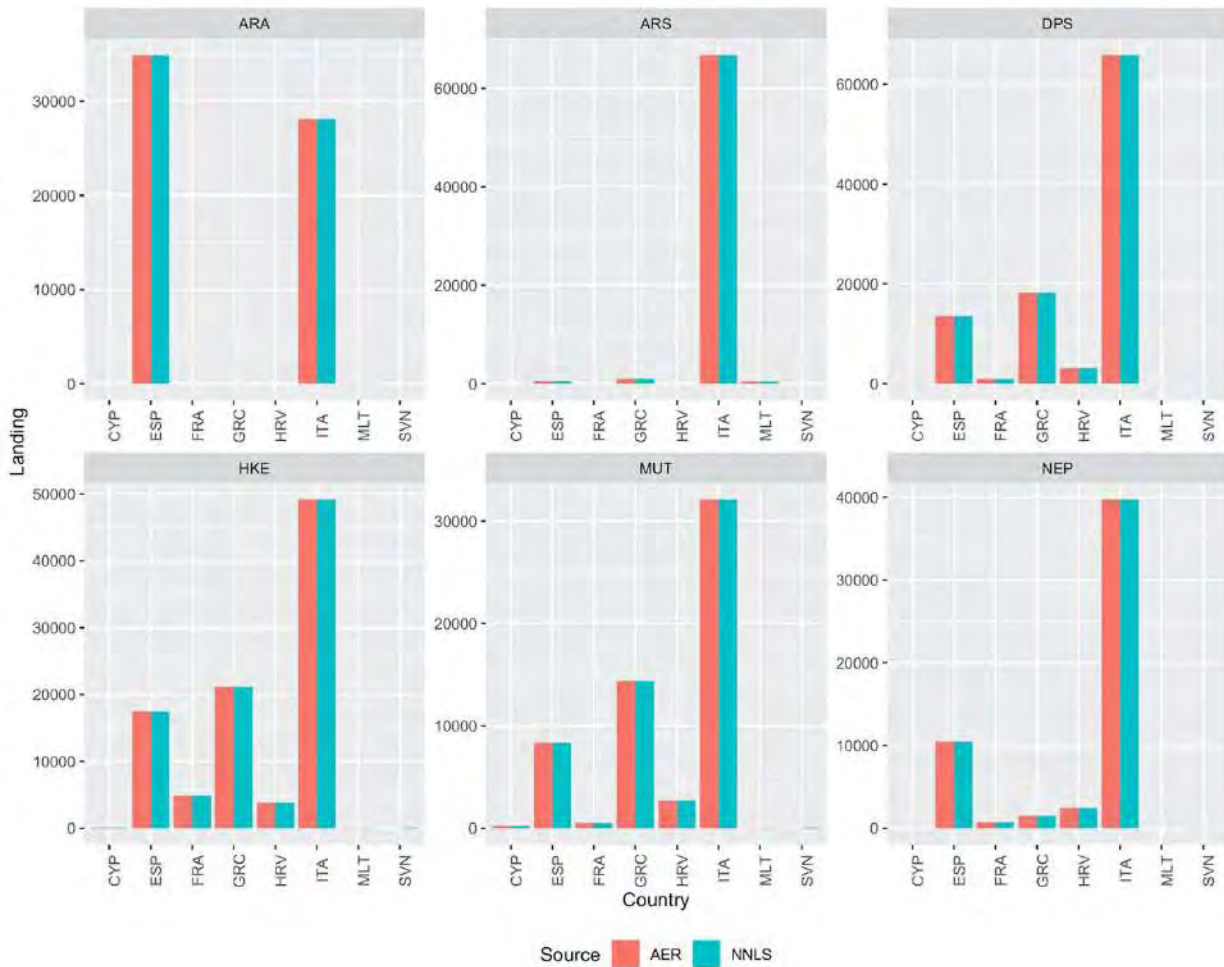


Figure 46 – Barplot comparing the total annual landings (2017), by species and countries, reported in the Annual Economic Report (AER) on the EU Fishing Fleet (STECF 19-06) with the corresponding values estimated using Non-negative least square (Russo et al., 2018).

At this aggregated level, model estimates fit very well the official data and negligible differences could be detected. In contrast, when the spatial origin of landings is considered (i.e. the GSA level), some differences are evident.

For instance, in the case of giant red shrimp *Aristeomorpha foliacea* (Fig. 47), the model suggests that a relevant portion of the Italian catch are coming from GSA12. Conversely, in the case of the blue and violet shrimp *Aristeus antennatus* (Fig. 48), the model indicates GSA16 is the main area of origin of Italian landings, whereas the relative importance of other GSAs is less relevant.

Aristaeomorpha foliacea

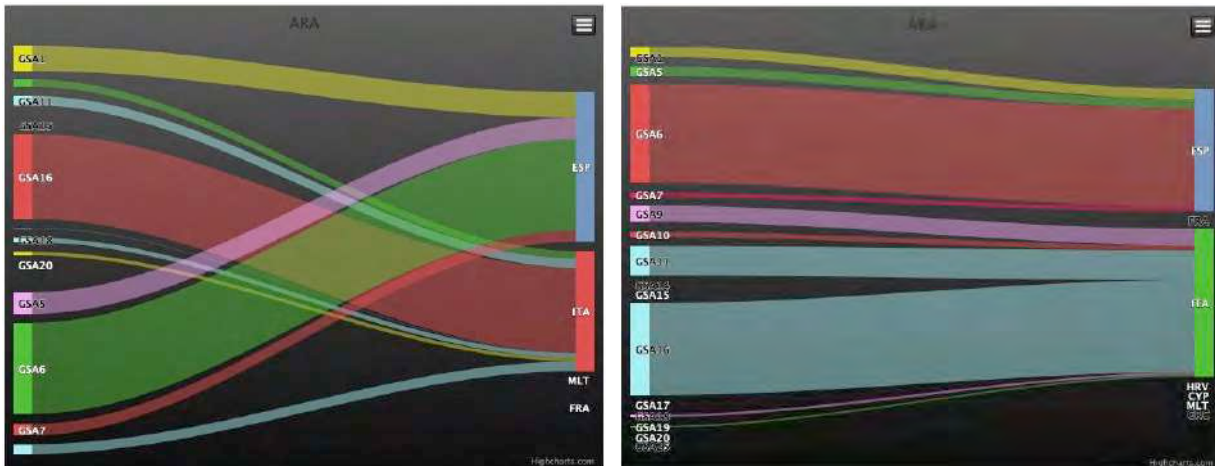


Figure 47 – Sankey plots comparing the official (left) and estimated (right) fluxes of landings from GSA to countries for the Giant Red Shrimp (*Aristaeomorpha foliacea* - ARS). Landings flow from area of origin (GSAs) to countries. Size of fluxes are proportional to landings in Tonnes.

Aristeus antennatus

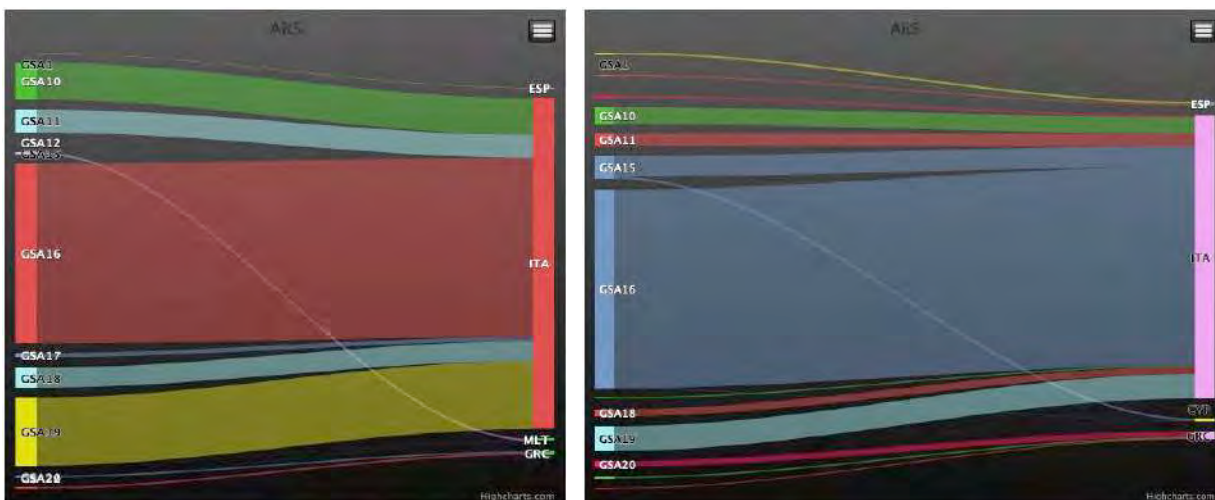


Figure 48 - Sankey plots comparing the official (left) and estimated (right) fluxes of landings from GSA to countries for the blue and violet shrimp (*Aristeus antennatus* - ARA). Landings flow from area of origin (GSAs) to countries. Size of fluxes are proportional to landings in Tonnes.

In the case of deep water rose shrimp *Parapenaeus longirostris* (Fig. 49), differences between model and official data occur for Italy (where the contribute of GSAs in the Tyrrhenian Sea estimated by the model is higher than the official ones) and for Greece (for which a complex of GSAs instead of the only GSA22 is identified as the origin of catch).

Parapenaeus longirostris

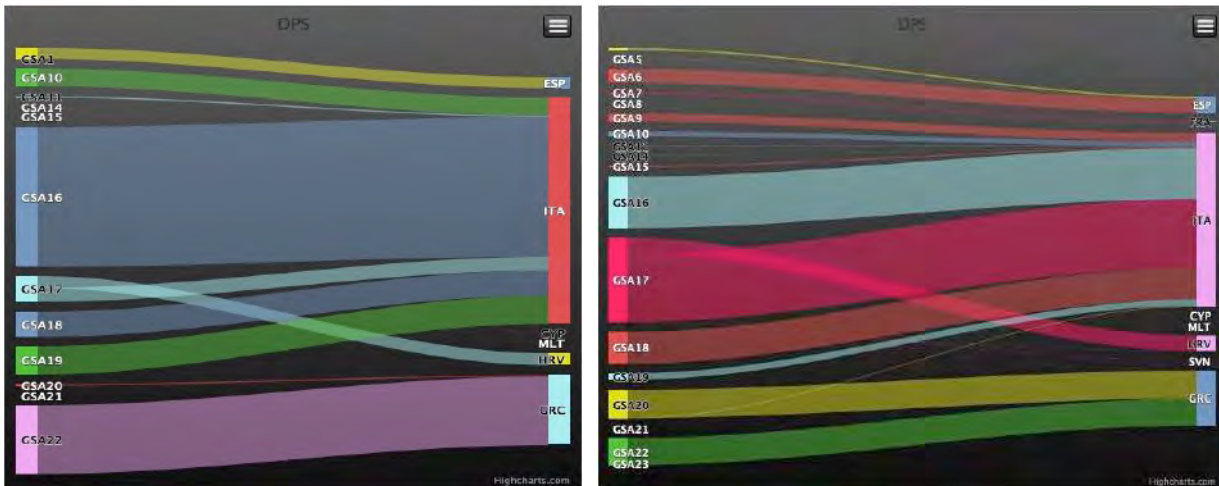


Figure 49 - Sankey plots comparing the official (left) and estimated (right) fluxes of landings from GSA to countries for the Deep water rose shrimp (*Parapenaeus longirostris* - DPS). Landings flow from area of origin (GSAs) to countries. Size of fluxes are proportional to landings in Tonnes.

In the case of European hake *Merluccius merluccius* (Fig. 50), the NNLS models attributes a crucial role, in terms of origin of catch, to the Adriatic GSA17-19 (for Italy), GSA22 (for Greece), and to GSA6 (for Spain).

Merluccius merluccius



Figure 50 - Sankey plots comparing the official (left) and estimated (right) fluxes of landings from GSA to countries for the European hake (*Merluccius merluccius* - HKE). Landings flow from area of origin (GSAs) to countries. Size of fluxes are proportional to landings in Tonnes.

GSA6 (for Spain) and GSA17 (for Italy) dominate the pattern for red mullet *Mullus barbatus* (Fig. 51) and for Norway lobster *Nephrops norvegicus* (Fig. 52).

However, this exercise was carried out under the assumption of homogeneity for several characteristics (efficiency, catchability of nets) of trawling fleets operating in the Mediterranean Sea were similar among countries. This means that this modelling approach was primarily driven by effort data.

Mullus barbatus

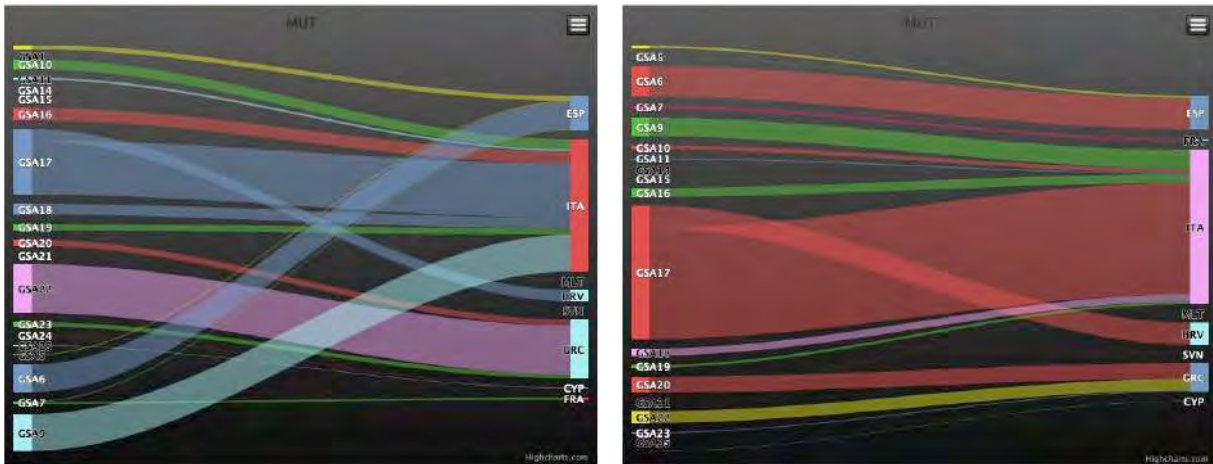


Figure 51 - Sankey plots comparing the official (left) and estimated (right) fluxes of landings from GSA to countries for the red mullet (*Mullus barbatus* - MUT). Landings flow from area of origin (GSAs) to countries. Size of fluxes are proportional to landings in Tonnes.

Nephrops norvegicus

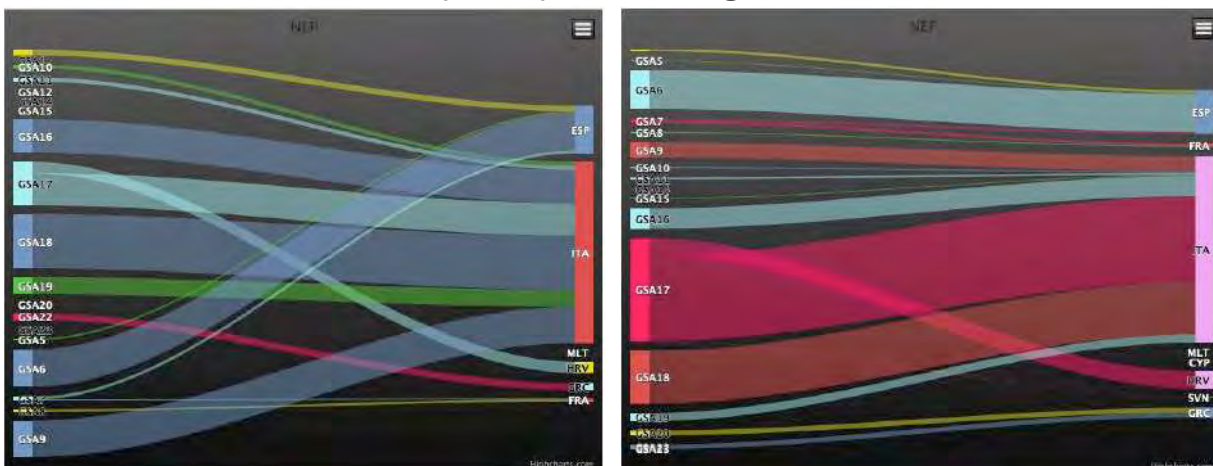


Figure 52 - Sankey plots comparing the official (left) and estimated (right) fluxes of landings from GSA to countries for the red mullet (*Nephrops norvegicus* - NEP). Landings flow from area of origin (GSAs) to countries. Size of fluxes are proportional to landings in Tonnes.

Discussion (or Conclusion)

The main goal of this part of the MED UNITS project was to estimate the fishing footprint for different fleet segments and countries operating in the Mediterranean Sea, even combining different data sources, processing and modelling approaches (including MCDA and ANN), and devoting part of the work to the quantitative comparison of obtained results. A series of maps were prepared to visualize the different fishing footprint returned by the application of the different methodologies in order to support a visual inspection and comparison of the outputs.

The first part of this report was devoted to the presentation of the results obtained from the application of two processing approaches (i.e. VMSBASE and CNR) to the AIS data for the Mediterranean vessels equipped with this tracking devices. The results of these methods were compared with each other and with the outputs of three external projects: the EMODNET project, the H2020 MINOUW project, and the Global Fishing Watch initiative.

These comparisons were carried out in two ways:

1. a quantitative analysis of fishing effort values detected in each cell;
2. a coverage analysis in terms of spatial extension of fishing activities (using the DCF ecological indicator 5).

The results of these comparisons could be summarized as follows:

- The two methodologies (i.e. VMSBASE and CNR) applied to process AIS data returned very consistent results in terms of structure and pattern of the fishing footprint. This is evident by looking at the maps and when post processing tool (i.e. Hot spot analysis) is applied;
- From a strictly quantitative point of view, that is in terms of total annual effort estimated in each cell, the two methodologies returned correlated by scaled results. In practice, the CNR method estimated higher level of effort than VMSBASE. An investigation about the reasons behind this discrepancy is beyond the scope of this deliverable, but it is worth noticing that previous experiences (ICES 2019) highlighted the opportunity to apply and compare different approaches to the estimation of fishing effort from AIS data;
- A crucial aspect is represented by the spatial extension of fishing activities (Fig. 32 of this report). This indicator stresses the coherence between the two main approaches (CNR and VMSBASE) and with the EMODNET outputs. In contrast, the analysis of Global Fishing Watch evidenced implausible values of this indicators, which could be determined by the coarse spatial resolution characterizing this method (Fig. 25), while the H2020 MINOUW project (in which AIS data were used in combination with VMS) estimated (in the GSA 5, 6, 9, 16, 20, 22, and 23 that is the only ones overlapping the MED UNITS project) higher values of this indicator in some GSAS (5, 6, 22, and 23) and lower values in others (9 and 16). The presence of higher values of this indicator when AIS and VMS data are combined could be easily interpreted and it is not an original finding (Russo et al. 2016; Shepperson

et al. 2018; ICES 2019). The presence of lower values could be related to difference in terms of fleet considered.

In summary, results reveal both promising findings and key limitations of estimating fishery indicators from AIS data. For monitoring fishing, AIS has some notable limitations: it can be easily turned off and broadcast incorrect identity information, its reception is poor in the Mediterranean southern areas and it is adopted by only a fraction of fishing vessels (above 15 m) that is not evenly distributed between Mediterranean regions making it difficult to compare activity. Despite these limitations, the use of available, high-frequency and seamless AIS data and performing algorithms proved useful in identifying and delimiting the more important fishing grounds in the Mediterranean basin and so responding to some of the specific aims of the project.

Thus, given that standardized methods to analyze and produce estimated of fishing intensity and fishing footprint, with particular emphasis on the presence of fishing activities in some areas, do not exist (ICES 2019), is fundamental to apply and compare different methods.

Another main topic of this Deliverable is represented by the modelling of fishing effort and fishing footprint for the fleets (i.e. those of the countries in the North Africa) without AIS.

To this end, the most influential components affecting trawl and small-scale fishing in terms of its distribution and intensity were identified and analyzed by means of a Multi-Criteria Decision Analysis (MCDA), while quite promising results on the potential fishing intensity of the fishing fleets were derived (Fig. 33, 34). Noticeably, based on the quality assessment of the Activity index (A_c), some weaknesses were revealed and therefore the estimation of the Fishing Pressure index (FPI) for bottom trawlers and small-scale fisheries based on a Multi-Criteria Decision Analysis (MCDA) has been influenced. In particular, the weaknesses in the representativeness of the fishing fleet characteristics by registration port and the different methodologies applied to estimate the fishing gear type by port specifically, for the southern Mediterranean countries, should be taken into account when using the estimated MCDA-FPI index in a management process. However, the strongest component for A_c estimation was the temporal correlation, underlining that the estimation of the FPI was based mainly on recent data obtained from all Mediterranean countries. Finally, in order to consider some of these data weaknesses in the MCDA and to improve the FPI outcomes for the entire Mediterranean Sea, a modelling framework based on simple regression and Generalized Additive Models has been used. This modeling framework was very useful for handling data gaps related to the fishing categories/gear-types, the GT of fishing vessels and the variability in the spatial quality of registration ports in the study area. It is also worth noting that as the data quality of fishing vessels characteristics per registration port is improved, both the quality and reliability of the MCDA outcomes will increase. In addition, the involvement of relevant experts from several areas in the Mediterranean into the overall MCDA process, could be very useful for verifying the model input data and for providing more details about those special fisheries characteristics that should

be taken into account in the model and, finally, to improve the estimated fishing pressure index.

Deviation from the Workplan

Remedial actions

References

- Carvalho N, Keatinge M, and Guillen J. 2019. The 2019 Annual Economic Report on the EU Fishing Fleet (STECF 19-06). : 496.
- EC. 2003. COMMISSION REGULATION (EC) No 2244/2003 of 18 December 2003 laying down detailed provisions regarding satellite-based Vessel Monitoring Systems.
- Fabi G. 2017. Challenge 5: Fishery management.
- Ferrà C, Tassetti AN, Grati F, et al. 2018. Mapping change in bottom trawling activity in the Mediterranean Sea through AIS data. *Marine Policy* **94**: 275–81.
- Galdelli A, Mancini A, Tassetti AN, et al. 2019. A Cloud Computing Architecture to Map Trawling Activities Using Positioning Data. In: Volume 9: 15th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications. Anaheim, California, USA: American Society of Mechanical Engineers.
- Gerritsen H and Lordan C. 2011. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES Journal of Marine Science* **68**: 245–52.
- Getis A and Ord JK. 2010. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* **24**: 189–206.
- Hastie T and Tsibaranani R. 1990. Generalized additive models. London: Chapman and Hall.
- Hidalgo M, Ligas A, Bellido JM, et al. 2020. Size-dependent survival of European hake juveniles in the Mediterranean Sea. *Sci Mar* **83**: 207.
- ICES. 2019. Working Group on Spatial Fisheries Data (WGSFD).
- James M, Mendo T, Jones EL, et al. 2018. AIS data to inform small scale fisheries management and marine spatial planning. *Marine Policy* **91**: 113–21.
- Jiang B. 1996. Fuzzy overlay analysis and visualization in geographic information systems.
- Kavadas S, Maina I, Damalas D, et al. 2015. Multi-Criteria Decision Analysis as a tool to extract fishing footprints and estimate fishing pressure: application to small scale coastal fisheries and implications for management in the context of the Maritime Spatial Planning Directive. *Mediterranean Marine Science* **16**: 294.
- Kroodsmas DA, Mayorga J, Hochberg T, et al. 2018. Tracking the global footprint of fisheries. *Science* **359**: 904–8.
- Lee J, South AB, and Jennings S. 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES Journal of Marine Science* **67**: 1260–71.
- Loo M van der, Laan J van der, Logan N, and Muir C. 2019. stringdist: Approximate String Matching and String Distance Functions.
- Mérigot B, Gaertner JC, Brind'Amour A, et al. 2020. Stability of the relationships among demersal fish assemblages and environmental-trawling drivers at large spatio-temporal scales in the northern Mediterranean Sea. *Sci Mar* **83**: 153.
- Natale F, Carvalho N, and Paulrud A. 2015. Defining small-scale fisheries in the EU on the basis of their operational range of activity The Swedish fleet as a case study. *Fisheries Research* **164**: 286–92.
- Pante E, Simon-Bouhet B, and Irisson J-O. 2019. marmap: Import, Plot and Analyze Bathymetric and Topographic Data.
- Petza D, Maina I, Koukourouli N, et al. 2017. Where not to fish – reviewing and mapping fisheries restricted areas in the Aegean Sea. *Medit Mar Sci* **18**: 310.
- Russo T, D'Andrea L, Franceschini S, et al. 2019a. Simulating the effects of alternative management measures of trawl fisheries in the Central Mediterranean Sea: application of a multi-species bio-economic modelling approach. *Front Mar Sci*.
- Russo T, D'Andrea L, Parisi A, et al. 2016. Assessing the fishing footprint using data

integrated from different tracking devices: Issues and opportunities. *Ecological Indicators* **69**: 818–27.

Russo T, D'Andrea L, Parisi A, and Cataudella S. 2014. VMSbase: An R-Package for VMS and Logbook Data Management and Analysis in Fisheries Ecology (M Castonguay, Ed). *PLoS ONE* **9**: e100195.

Russo T, Franceschini S, D'Andrea L, et al. 2019b. Predicting Fishing Footprint of Trawlers From Environmental and Fleet Data: An Application of Artificial Neural Networks. *Front Mar Sci* **6**: 670.

Russo T, Morello EB, Parisi A, et al. 2018. A model combining landings and VMS data to estimate landings by fishing ground and harbor. *Fisheries Research* **199**: 218–30.

Russo T, Parisi A, and Cataudella S. 2011a. New insights in interpolating fishing tracks from VMS data for different métiers. *Fisheries Research* **108**: 184–94.

Russo T, Parisi A, Prorgi M, et al. 2011b. When behaviour reveals activity: Assigning fishing effort to métiers based on VMS data using artificial neural networks. *Fisheries Research* **111**: 53–64.

Scott F, Prista N, and Reilly T. 2017. FecR: A Package For Calculating Fishing Effort.

Shepperson J, Hintzen N, Szostek C, et al. 2018. A comparison of VMS and AIS data: The effect of data coverage and vessel position recording frequency on estimates of fishing footprints. *ICES Journal of Marine Science* **75**: 988–98.

Taconet M, Kroodsmas D, and Fernandes J. 2019. Global Atlas of AIS-based fishing activity Challenges and opportunities. Rome: FAO.

Tasseti AN, Ferrà C, and Fabi G. 2019. Rating the effectiveness of fishery-regulated areas with AIS data. *Ocean & Coastal Management* **175**: 90–7.

Tasseti N, Ferrà C, Scarcella G, and Fabi N. 2016a. MEDSEA_CH5_Product_6 / Impact of fisheries on the bottom from AIS data combined with habitat vulnerability. EMODnet Medsea Checkpoint.

Tasseti N, Ferrà C, Scarcella G, and Fabi G. 2016b. MEDSEA_CH5_Product_7 / Change level of disturbance from AIS data combined with habitat vulnerability.

Thiessend A. 1911. DISTRICT No. 10, GR'EAT BASIN. *MONTHLY WEATHER REVIEW*: 3.

Weidema BP and Wesnæs MS. 1996. Data quality management for life cycle inventories—an example of using data quality indicators. *Journal of Cleaner Production* **4**: 167–74.

Wood S. 2006. Generalized Additive Models. An Introduction with R. Boca Raton, FL, USA.

Zimmermann H-J and Zysno P. 1980. Latent connectives in human decision making. *Fuzzy Sets and Systems* **4**: 37–51.

**Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/
SI2.793201 – SC03- Project MED_UNITS**

**“Study on Advancing fisheries assessment and management
advice in the Mediterranean by aligning biological and
management units of priority species”**

WP3-DELINEATE FISHING GROUNDS AND STOCK ASSESSMENT

**Task 3.2 - Combine the results on spatial
identification of fishing grounds with the spatial
distribution of the target species**

**Deliverable 3.2 – Maps combining the fishing
grounds and the spatial distribution of the target
species**

Responsible: Stefanos Kavadas, Irida Maina (HCMR)

Compiled by: Irida Maina, Stefanos Kavadas, Vinko Bandelj, Pierluigi Carbonara,
Lorenzo D'Andrea, Carmen Ferrà, Fabio Fiorentino, Simone Franceschini,
Alessandro Ligas, Claudia Musumeci, Tommaso Russo, Nora Tassetti, Paolo
Sartor, Mario Sbrana, Giuseppe Scarcella, Simone Libarato, George Tserpes,
Walter Zupa, Maria Teresa Spedicato

September 2020

Executive summary	3
Introduction	4
Materials and methods	5
Results	32
Discussion-Conclusions	44
References	46

Executive summary

The aim of Task 3.2 “Combine the results on spatial identification of fishing grounds with the spatial distribution of the target species” is to synthesize and combine the results of Task 3.1 “Fishing grounds delineation” on the identified fishing pressure with the information from the literature and previous projects on the spatial distribution of six target species studied in the project (European hake, *Merluccius merluccius*; Norway lobster, *Nephrops norvegicus*; deep-water rose shrimp, *Parapenaeus longirostris*; red mullet, *Mullus barbatus*; giant red shrimp, *Aristaeomorpha foliacea*; blue and red shrimp, *Aristeus antennatus*).

The goal of this synthesis is to delineate the fishing grounds by species or group of species. Those areas are characterized by both fishing activity and species abundance important for fisheries, including the target species and the fishing intensity from different sectors. Based on methods for analysing spatial patterns and mapping clusters, statistically significant hot and cold spots of the combined fishing grounds and species spatial distribution were studied. Aggregated outputs of the species under investigation provided spatial information about the number of exploited species in several fishing grounds.

A methodological framework based on spatial analytical techniques for exploring the patchy distribution of fishing effort and target species was used as a baseline for the investigation and mapping of fishing grounds in the Mediterranean Sea. The approach was based on combining the fishing effort of a certain fishery (bottom trawl, longlines, gillnets and trammel nets) and the potential habitat use of the investigated species. The steps followed for identifying and analyzing fishing grounds were: (i) estimation of the potential fishing grounds by species, (ii) Hot spot analysis and (iii) the production of aggregated hot and cold spots for the studied species.

Fishing effort for bottom trawlers, longlines and static nets with length category > 12 m, derived from Task 3.1, were based on the analysis of AIS data. In addition, a fishing pressure index for small scale fisheries <12 m (gillnets and trammel nets) and bottom trawl (particularly estimated for the south Mediterranean fishing fleets) were calculated based on a Multi-Criteria Decision Analysis methodology applied on environmental and fleet data in Task 3.1. The above datasets were used as input to the current analysis process. Spatial data of species distribution were also used and based on scientific literature and historical scientific projects (e.g., STOCKMED and MEDISEH). The estimations were performed for the entire Mediterranean basin, aiming to provide a more comprehensive view of the fishing grounds by target species.

In order to determine the potential fishing grounds by species, the spatial overlap (as a product) of two types of information was used: (i) the spatial distribution by species and (ii) the fishing effort. The produced outcomes, demonstrate the possibility for fishers to harvest a certain species in a specific area.

To perform hot spot analysis (estimation of hot and cold spots by species and métier), spatial clustering techniques based on the potential fishing grounds of each species were used. Maps of hot and cold spots for each species were based on the Getis-Ord G_i^* statistic.

Finally, the aggregated hot spots showed the areas over the Mediterranean with the highest number of commercial species for bottom trawling which are found in the Adriatic, Tyrrhenian, Strait of Sicily, Aegean, eastern Ionian, Balearic, Alboran, Libyan and Levantine Sea.

Introduction

A better understanding of fisheries dynamics is of utmost importance for proposing more effective management measures. Spatial analytical techniques, based on observation, pattern detection, experimentation and modelling have been proven to be powerful tools for the study of the complex nature and spatial patterns and trends in fisheries (Fortin and Dale, 2005). Such identification and analysis of the fisheries spatial structure and patterns can be very informative regarding the delineation of fishing grounds.

In that context, the aim of Task 3.2 “Combine the results on spatial identification of fishing grounds with the spatial distribution of the target species” is to synthesize and combine the results of Task 3.1 “Fishing grounds delineation” on the identified fishing pressure with the information from the literature and previous projects on the spatial distribution of the six target species considered in the project (i.e. European hake, *Merluccius merluccius*; Norway lobster, *Nephrops norvegicus*; deep-water rose shrimp, *Parapenaeus longirostris*; red mullet, *Mullus barbatus*; giant red shrimp, *Aristaeomorpha foliacea*; blue and red shrimp, *Aristeus antennatus*). The aim of this synthesis is to delineate the fishing grounds of a species or a group of species which is herein defined as “crucial areas characterized by both fishing activity and species abundance as a result of a strategy to maximize catches and economic gains” (Maina et al., 2016).

In addition, based on methods for analysing spatial patterns and mapping clusters, statistically significant hot and cold spots of the combined fishing grounds and species spatial distribution were studied. Aggregated outputs of all species under investigation provided spatial information about the total number of exploited species in several fishing grounds.

Managing fisheries in the Mediterranean Sea is quite complex, from an environmental, technical as well as from a geopolitical point of view. Besides the multispecies nature of several Mediterranean fisheries and the climate/environmental changes that affect, among other things, the rapid expansion of non-indigenous species, this basin includes wide areas of international waters and most commercial fish and shellfish stocks are shared among several coastal countries. Given that many of these countries are not part of the EU, they are having different research, data and management protocols, although the efforts put in place by GFCM and FAO regional projects are trying to uniform all these aspects at Mediterranean level with some important results achieved so far. All the above are influencing the effectiveness of fisheries management measures, since they are affecting fisheries yield and stocks status. The performance of the modelling approaches aiming to provide a better understanding of fisheries in the area is also influenced by these Mediterranean characteristics.

Under the overall framework of ecosystem-based marine spatial management, the results of this Task could be used to define areas important for fisheries, including the target species and the fishing intensity from different sectors. The spatial outcomes are also covering the entire Mediterranean basin, aiming to provide a more comprehensive view of the fishing grounds by target species.

Materials and methods

A methodological framework, based on spatial analytical techniques for exploring the patchy distribution of fishing effort and target species was used as a baseline for the investigation and mapping of fishing grounds in the Mediterranean Sea (Maina et al., 2016). The approach is based on combining the fishing effort of a certain fishery and the potential habitat use of the following target species:

- i) for the Mediterranean bottom trawl fisheries, all the six target species were considered in the analysis;
- ii) for the EU Mediterranean longlines and gillnets fisheries (> 12 m), only European hake was considered as a target species and used in the analysis;
- iii) for the Mediterranean small-scale fisheries using gillnets and trammel nets (< 12 m LOA), European hake and red mullet were considered as target species and used in the analysis.

The steps for identifying and analyzing fishing grounds are summarized in the flowchart of Figure 1 and are hereafter described in detail.

Several datasets on fishing effort and species spatial distribution were used, while the spatial data on species distribution were mainly based on previous studies performed in EU Mediterranean region (Tables 1; 2). For improving the final outcomes on the south-east Mediterranean Sea (non EU part), a literature review was also performed (Table 3).

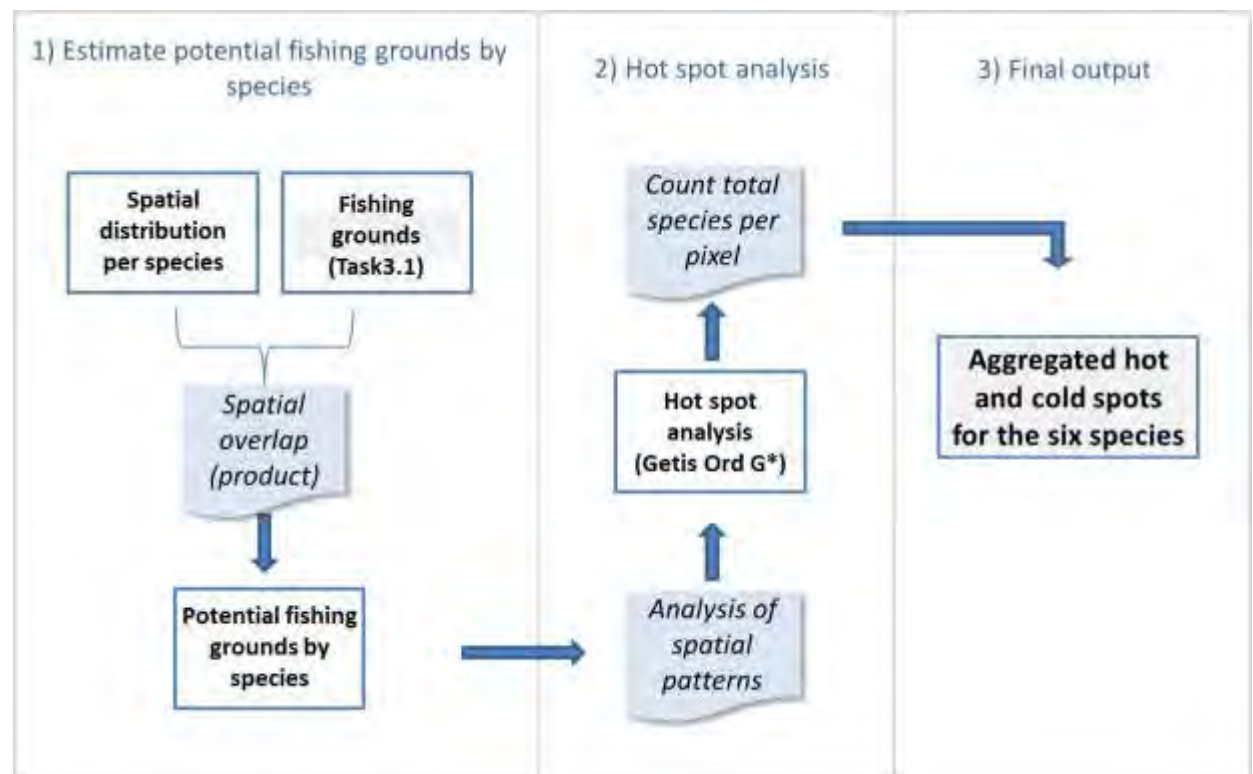


Figure 1: Flowchart of the methodological steps applied to identify and describe fishing grounds.

Data

Table 1: Spatial datasets used for the identification of fishing grounds for the north Mediterranean Sea (EU fishing fleets).

Spatial dataset for EU Med fishing fleets	Type of spatial dataset	Methods
Fishing Effort	Bottom trawl (>12 m) Longlines (>12 m) Nets (>12 m - mainly referring to set gillnets see details in Task 3.1 report)	Average of the following methods: (1) VMSbase platform (Russo et al., 2014; Russo et al., 2016, Russo et al., 2019a); (2) The procedure applied in Galdelli et al., 2019, Tasseti et al., 2019, Ferrà et al., 2018
	Small-scale fisheries <12m (gillnets and trammel nets)	Multi-Criteria Decision Analysis (MCDA) applied on environmental and fleet data (Kavadas et al., 2015) for small-scale components of the fleets;
Species distribution	<i>Merluccius merluccius</i> <i>Nephrops norvegicus</i> <i>Parapenaeus longirostris</i>	Spatial data used were based on Sbrana et al., 2019 and Sion et al., 2019 and validated through spatial visualization based on information of STOCKMED and MEDISEH projects.
	<i>Mullus barbatus</i> <i>Aristaeomorpha foliacea</i> <i>Aristeus antennatus</i>	Spatial data used were based on image re-processing (combined STOCKMED and Maina et al., 2016) and validated through spatial visualization based on information from MEDISEH project.

Table 2: Spatial datasets used for the identification of fishing grounds for the south-eastern Mediterranean Sea (non-EU fishing fleets).

Spatial dataset for non-EU Med fishing fleets	Type of spatial dataset	Methods
Fishing Effort	Bottom trawl (>12 m); Small-scale fisheries <12m (gillnets and trammel nets)	Multi-Criteria Decision Analysis (MCDA) applied on environmental and fleet data (Kavadas et al., 2015) for bottom trawl and small-scale fleets;
Species distribution	<i>Merluccius merluccius</i> <i>Nephrops norvegicus</i> <i>Parapenaeus longirostris</i> <i>Mullus barbatus</i>	Spatial data were estimated based on Generalized Additive Models. The models were conditioned with spatial data on species occurrence and abundance for EU Med countries (Table 1) and depth. Where it was possible, results were validated based

	<i>Aristaeomorpha foliacea</i> <i>Aristeus antennatus</i>	on existing literature (see table 3).
--	--	---------------------------------------

Table 3: Literature review on south-east (non-EU) Mediterranean studies about the target species and their fishing grounds. Given that the existing literature from these areas is limited, the works provided are based on scientific studies, grey literature and websites. The studies were categorized by area and target species.

Literature on non-EU Med fishing grounds	south west Med	south central Med	South and east Med
<i>Merluccius merluccius</i>	Boubekri, et al., 2018; Belhoucine et al., 2014 ; Orsi Relini et al., 2002.	Garofalo et al., 2008; UNEP-MAP-RAC/SPA. 2014; Khoufi et al., 2014.	Gucu and Bingel, 2011; Al-Absawy, 2010.
<i>Nephrops norvegicus</i>	http://www.fao.org/geonetwork/srv/en/main.home?uuid=fao-species-map-nep	http://www.fao.org/geonetwork/srv/en/main.home?uuid=fao-species-map-nep ; UNEP-MAP-RAC/SPA. 2014	Turan et al., 2016; http://www.fao.org/geonetwork/srv/en/main.home?uuid=fao-species-map-nep
<i>Parapenaeus longirostris</i>	Awadh and Aksissou, 2020; Nouar et al., 2016.	Rawag et al., 2004; Knittweis et al., 2013; UNEP-MAP-RAC/SPA. 2014; Milisenda et al., 2017; Quattrocchi et al., 2019; http://www.faomedsudmed.org/html/species/Parapenaeus%20longirostris.html	Hakkı and Mustafa, 2011; Abdel Razek et al., 2006.
<i>Mullus barbatus</i>	Talet et al., 2016	Garofalo et al., 2008; UNEP-MAP-RAC/SPA. 2014.	Mahmoud et al., 2017
<i>Aristaeomorpha foliacea</i>	http://www.faomedsudmed.org/html/species/Aristaeomorpha%20foliacea.html	UNEP-MAP-RAC/SPA. 2014; http://www.faomedsudmed.org/html/species/Aristaeomorpha%20foliacea.html	Bayhan et al., 2015
<i>Aristeus antennatus</i>	Nouar et al., 2016	http://www.fao.org/fishery/species/3422/en	http://www.fao.org/fishery/species/3422/en

Fishing effort spatial distribution

Fleets equipped with monitoring tracking devices (AIS/VMS)

The analysis of the fishing footprint of the fleets equipped with remote tracking devices (VMS, AIS or both) was based on AIS data and described in detail in Deliverable 3.1 (Fishing grounds delineation). A preliminary analysis of all the available AIS data for the year 2017 allowed to identify fishing vessels belonging to several Mediterranean countries, but the coverage of AIS data for non-EU countries is scant. The analysis of AIS data was based on two approaches:

- 1) the VMSbase platform (Russo et al. 2011a, b, 2014, 2016, 2019a). This allowed validating the output in some areas by comparing the pattern obtained using only AIS data with the one obtained integrating AIS and VMS data.
- 2) the procedure used in Ferrà et al. 2018; Tasseti et al. 2019; Galdelli et al. 2019 was also applied to AIS data.

In both cases, the outputs of these procedures were compared with an acknowledged reference dataset (i.e. the GlobalFishingWatch online repository). The output of these comparisons indicated that the level of detail obtained in Task 3.1 is high and the overall estimated footprint represents the best available reconstruction of the fishing effort. It is also worth noting that the outputs of the two procedures mentioned above are very coherent.

The spatial outcomes were represented by a series of grids coherent with the standard defined by the GFCM (<http://www.fao.org/gfcm/data/maps/grid/en/>), and populated with different metrics of fishing effort (i.e. fishing time in hours, fishing days, and nominal effort) and disaggregated by country, gear type (métier of level 4) and fleet segment. Herein, a spatial grid of ~ 1*1 nautical mile is used for the analysis purposes.

The distribution of the total fishing effort, expressed in fishing days, for the Mediterranean fleets operating with the bottom trawls, longlines and gillnets was based on AIS data and analyzed based on the two aforementioned approaches (see details in the Deliverable 3.1). An average of the spatial outcomes derived from the two methods was used as an input for the analysis performed in the present Deliverable and shown in figures 2-4.

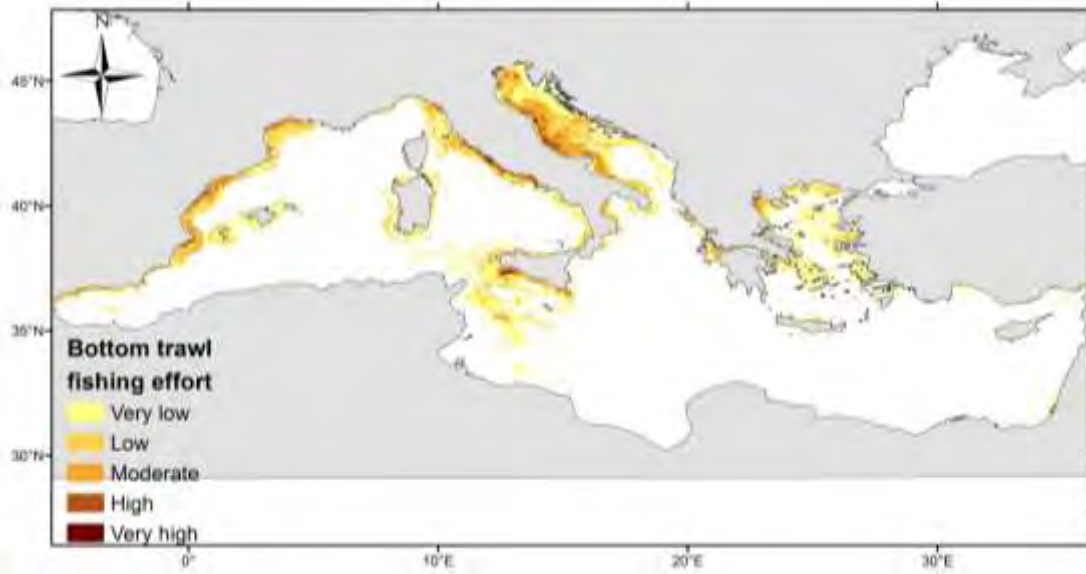


Figure 2: Spatial distribution of fishing effort of bottom trawlers based on Deliverable 3.1.

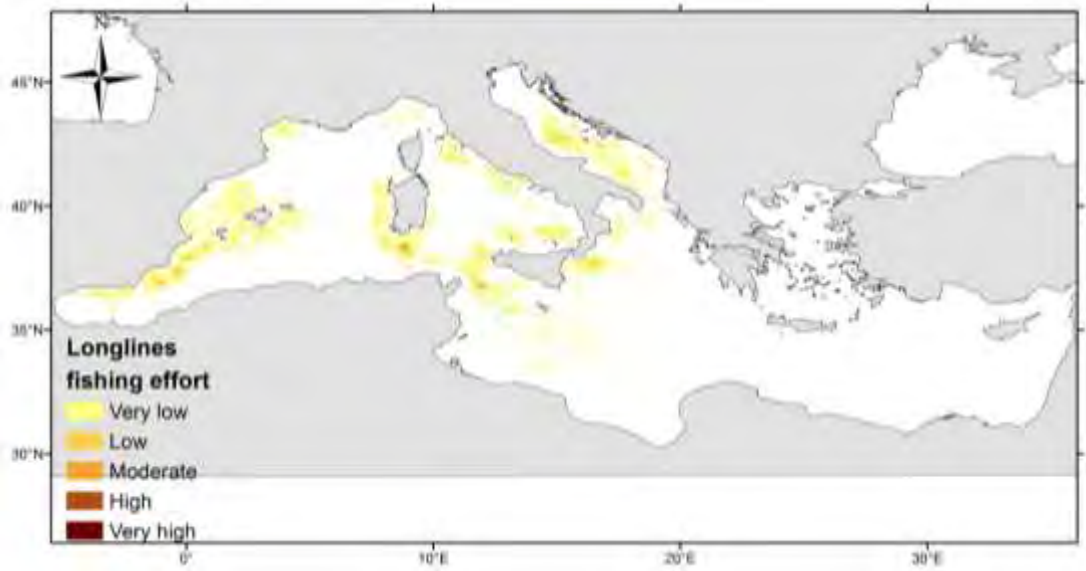


Figure 3: Spatial distribution of fishing effort of longlines based on Deliverable 3.1.

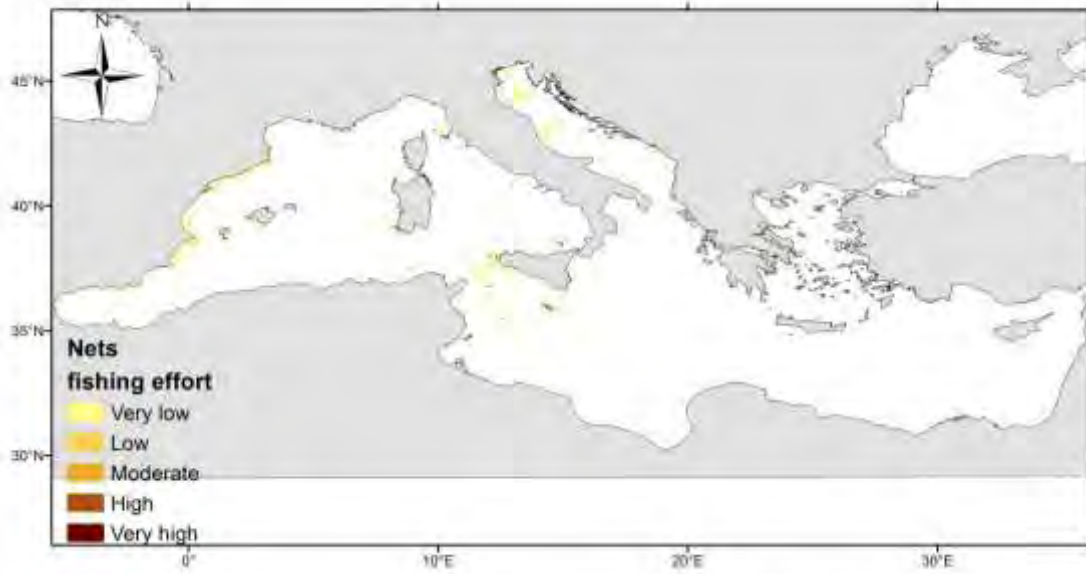


Figure 4: Spatial distribution of fishing effort of nets based on Deliverable 3.1.

Fleets not equipped with monitoring tracking devices (AIS/VMS)

The (Multi-Criteria Decision Analysis) MCDA method produces fisheries footprint by taking into consideration several interactions with other anthropogenic and environmental factors. The methodological approach and the application on Small-Scale Fisheries (SSF) are further described in Kavadas et al. 2015, while this method has been also applied to estimate a fishing pressure index (FPc) for bottom trawlers (see Hidalgo et al. 2020; Mérigot et al. 2020).

Several methods and processes such as the Analytic Hierarchy Process (AHP) and Fuzzy logic were applied to solve the multiple criteria problem. FPc was perceived as the fuzzy product of two indices: the fishery suitability index (Sc) and the activity index (Ac) based on the estimated spatial distribution of registered fishing vessels in the Mediterranean Sea (Data were based on the EU Fleet Register: https://webgate.ec.europa.eu/fleet-europa/search_en and on the General Fisheries Commission for the Mediterranean - GFCM fleet register: <http://www.fao.org/gfcm/data/fleet/register/en/>).

The MCDA was applied for the non-EU bottom trawl fleets (not equipped with VMS/AIS), and for small-scale fisheries (gillnets and trammel nets) of both EU and non-EU fleets with LOA <12m. The methodological approach and the data used for applying the model are described in detail in the Deliverable 3.1 and shown in figures 5 and 6.

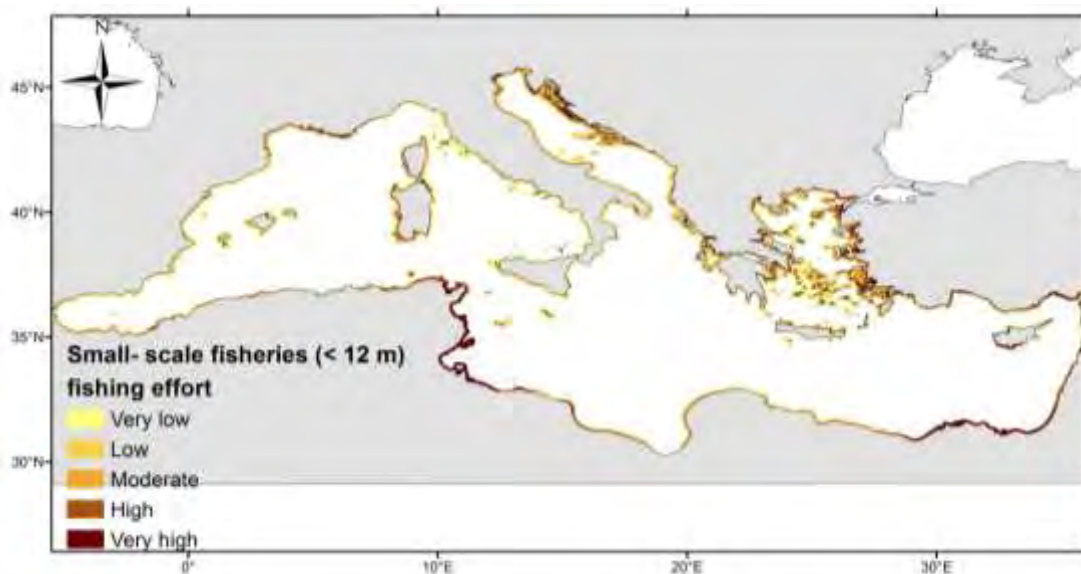


Figure 5: Spatial distribution of fishing effort of small-scale fishery vessels < 12m LOA based on Multi-Criteria Decision Analysis (see Deliverable 3.1)

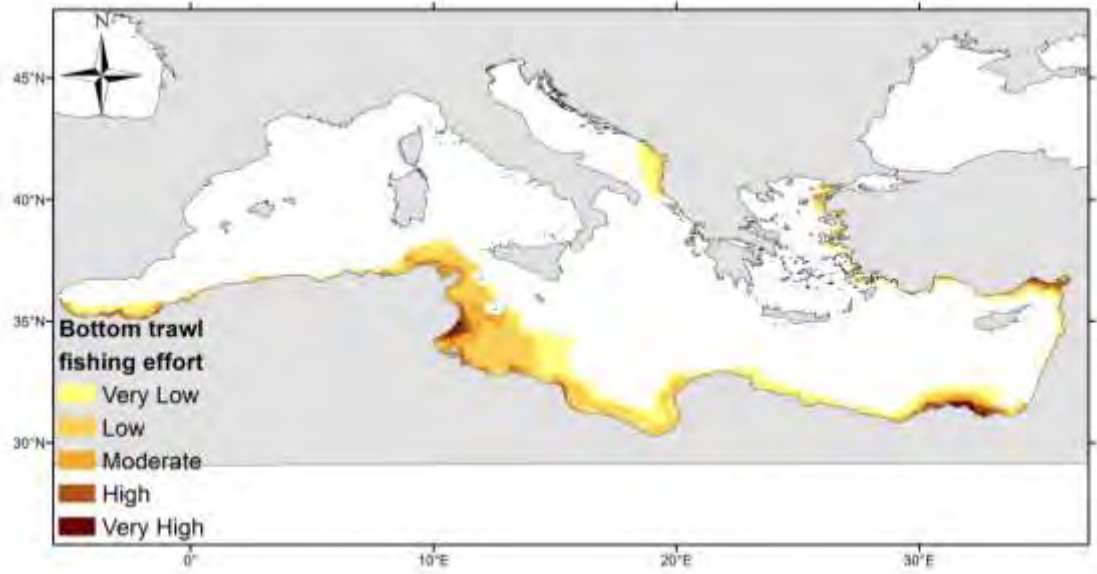


Figure 6: Spatial distribution of fishing effort of bottom trawlers (non-EU fleets) based on Multi-Criteria Decision Analysis (see Deliverable 3.1)

Species spatial distribution

In general, the spatial distribution of each species was obtained based on the following steps:

Step i- Information from literature (Sbrana et al., 2019; Sion et al., 2019; Maina et al., 2016) and previous projects (STOCKMED and MEDISEH) were combined and used for estimating the spatial distribution of each species covering the north Mediterranean Sea (EU part).

Step ii- Given that information on species spatial distribution were not available for the south-east Mediterranean coasts (non-EU part), Generalized Additive Models (GAMs) were used for estimating species spatial distribution. The models were conditioned with depth and information on species spatial occurrence and abundance based on Step i. A review was also carried out in the existing literature, aiming to retrieve of any information on species distribution or fishing grounds on these areas (Table 3). The maps were then adapted based on such review.

This procedure might differentiate by species. The analytical process of estimating the spatial datasets for each species is further described in the following subsections.

Merluccius merluccius

The spatial distribution of European hake, *Merluccius merluccius*, in the Mediterranean Sea, used as an input in this Deliverable 3.2., was based on the analysis made by Sion et al. (2019). A time series of data from the Mediterranean International bottom Trawl Surveys (MEDITS) from 1994 to 2015 was analysed. The use of GAMs, in which standardized biomass indices (kg km^{-2}) were analysed as a function of environmental variables (Sea Surface Temperature- SST, depth), explained how ecological factors could affect the spatio-temporal distribution of European hake biomass in the basin. High biomass levels predicted by the model were observed especially at 200 m depth and between 14°C and 18°C, highlighting the preference of the species for colder waters. In fact, the bottoms at the shelf break represent the main nursery areas of European hake (Bartolino et al., 2008; Ligas et al., 2015).

Predictions of the aforementioned modeling approach and particularly an average for the years 2014 – 2016 were used for this purpose (Fig. 7).

Spatial information from previous projects i.e. STOCKMED (Fig. 8) and MEDISEH (Fig. 9) was also taken into account only for interpretation purposes of the final results while those outcomes were not included directly to the estimation of fishing grounds for European hake. A comparison by visualization was performed based on the several data sources shown in Figures 7-9.

Furthermore, for estimating the spatial distribution of European hake in the south-eastern Mediterranean coasts (non-EU) the following approach was applied:

- 1) GAM was applied using as response variable the spatial occurrence of European hake (based on Maina et al. (2016) outcomes) and the depth as predictor variable,
- 2) GAM was applied using as response variable the spatial distribution of European hake's biomass (based on Sion et al. 2019 outcomes) and the depth as predictor variable.
- 3) The spatial outcomes of both models were then combined (through a spatial overlap procedure) to estimate the spatial distribution of this species in the south and eastern Mediterranean Sea (Fig. 7).

Although a few studies presenting spatially the fishing grounds by species in the south Mediterranean Sea (based on our literature search: e.g., Gucu and Bingel, 2011), it is reported that European hake is among the main target species of the entire Mediterranean demersal fishery that is fished in a wide bathymetric range mainly < 500 m (see references in Table 3).

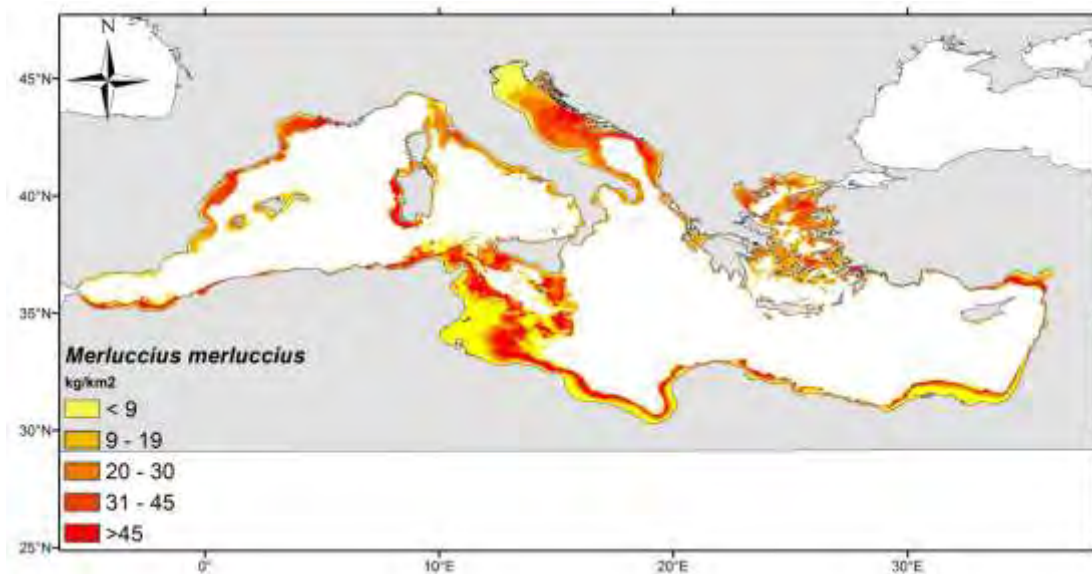


Figure 7: Spatial distribution of *Merluccius merluccius* biomass based on combined information from Sion et al. (2019) and Maina et al. (2016) (average for the years 2014 – 2016).

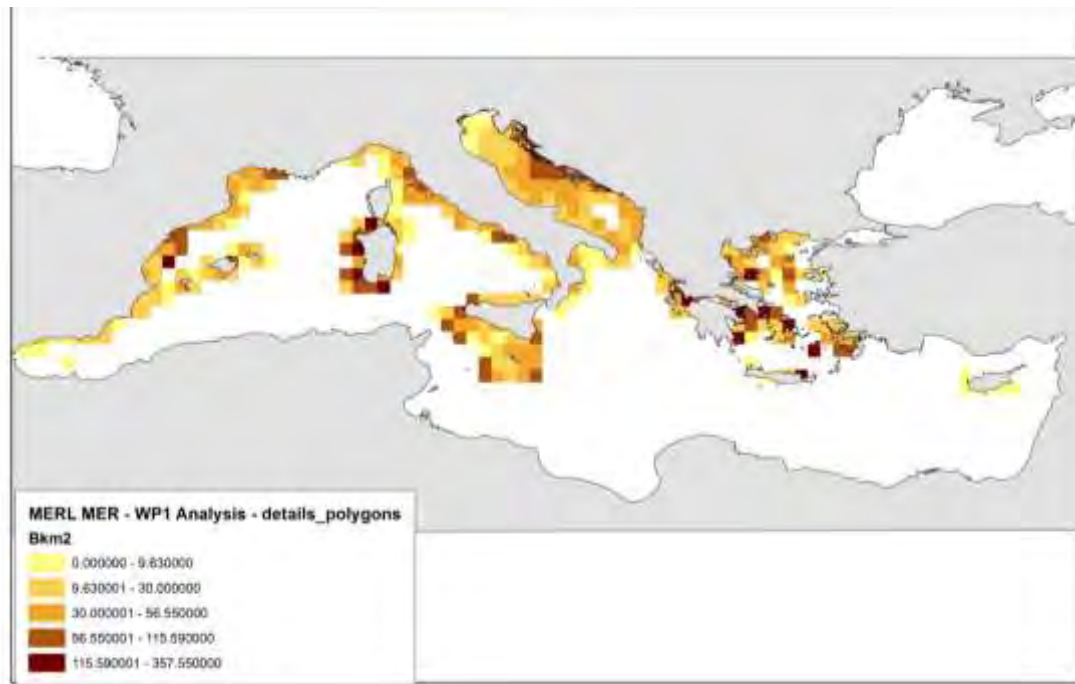


Figure 8: Spatial distribution of *Merluccius merluccius* biomass based on STOCKMED.

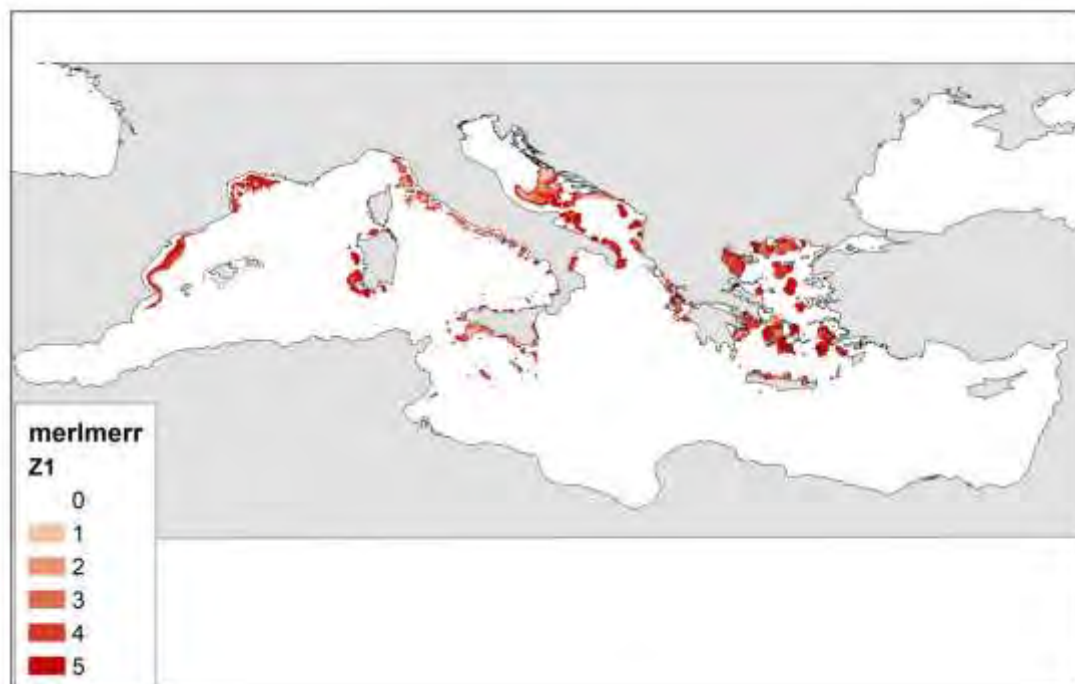


Figure 9: Spatial distribution of *Merluccius merluccius* nurseries based on MEDISEH. This map indicates the different levels of probability to find a nursery for this species.

Parapenaeus longirostris

The spatial distribution of deep-water rose shrimp, *Parapenaeus longirostris*, in the Mediterranean Sea, used as an input in this Deliverable, was based on the analysis made by Sbrana et al. (2019). A time series of data from the Mediterranean International bottom Trawl Surveys (MEDITS) from 1994 to 2015 was analysed. The use of GAMs, in which standardized biomass indices (kg km^{-2}) were analysed as a function of environmental variables (Sea Surface Temperature- SST, depth),

explained how ecological factors could affect the spatio-temporal distribution of deep-water rose shrimp biomass in the basin. Deep-water rose shrimp showed a widening bathymetric range in time, from 200-400 m in depth at the beginning of the time series to 150-600 m depth in the last few years. As regards temperature, the highest abundance of the species was associated with SSTs at around 10°C, while a second peak was observed at around 14°C. The map in Figure 11 presents the highest abundances of deep-water rose shrimp in the eastern Mediterranean (Sbrana et al., 2019).

Predictions of the aforementioned modeling approach and particularly an average for the years 2014 – 2016 were used for the present Deliverable (Fig. 10).

Spatial information from previous projects i.e. STOCKMED (Fig. 11) and MEDISEH (Figs. 12 and 13) was also taken into account for interpretation purposes of the final results while those outcomes were not included directly to the estimation of fishing grounds for deep-water rose shrimp. A comparison by visualization was performed based on the several data sources shown in Figures 10-13.

Furthermore, for estimating the spatial distribution of deep-water rose shrimp in the south-east Mediterranean coasts (non-EU) the following approach was applied:

- 1) GAM was applied using as response variable the spatial occurrence of deep-water rose shrimp (based on Maina et al. (2016) outcomes) and the depth as predictor variable,
- 2) GAM was applied using as response variable the spatial distribution of deep-water rose shrimp's biomass (based on Sbrana et al. 2019 outcomes) and the depth as predictor variable.
- 3) The spatial outcomes of both models were then combined (through a spatial overlap procedure) to estimate the spatial distribution of this species in the south and eastern Mediterranean Sea (Fig. 10).

The search in the existing literature revealed that the greatest abundance of *P. longirostris* are recorded between 100 and 300 m depth in the Mediterranean Sea. The Strait of Sicily, together with the seas around Greece, is the Mediterranean region with the greatest abundance of this species (see references in Table 3; <http://www.faomedsudmed.org/html/species/Parapenaeus%20longirostris.html>).

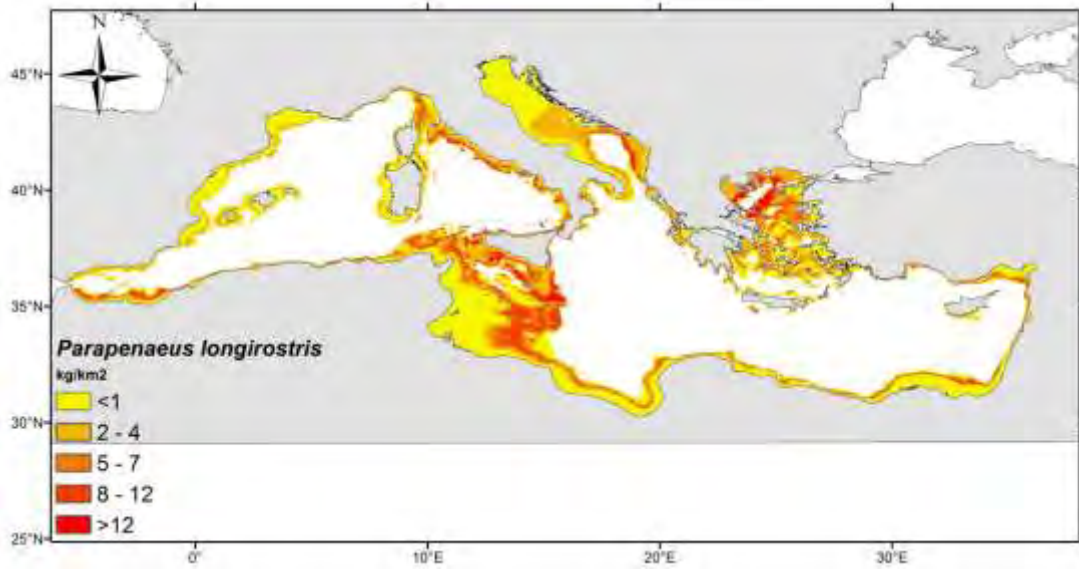


Fig 10. Spatial distribution of *Parapenaeus longirostris* biomass based on combined information from Sbrana et al. (2019) and Maina et al., (2016) (average for the years 2014 – 2016).

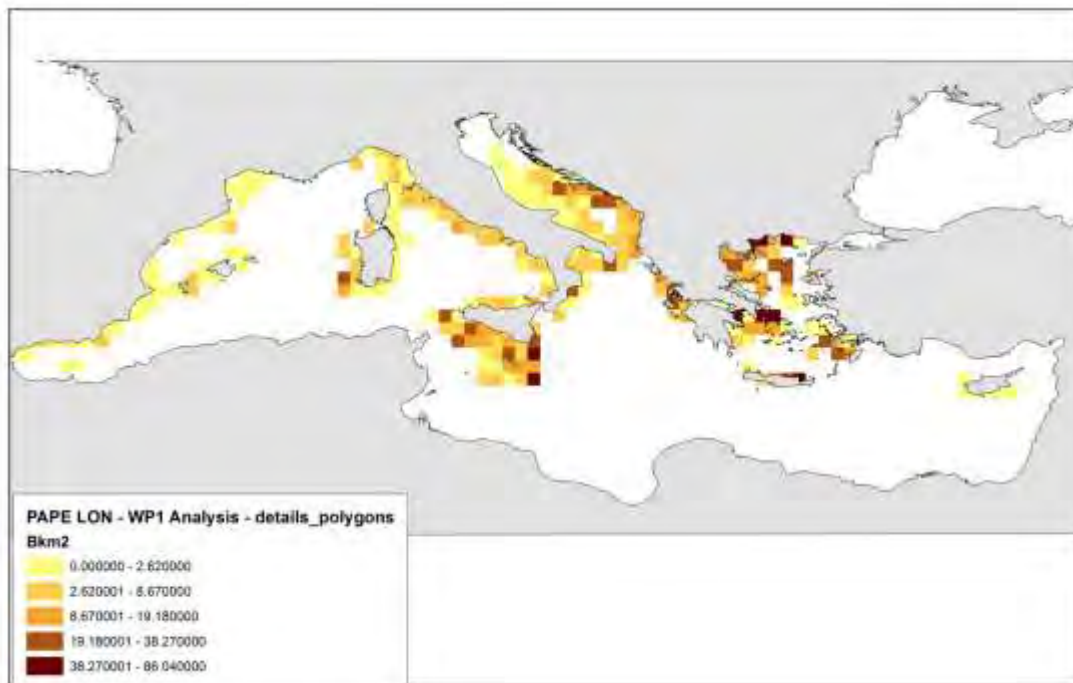


Fig 11. Spatial distribution of *Parapenaeus longirostris* biomass based on STOCKMED.

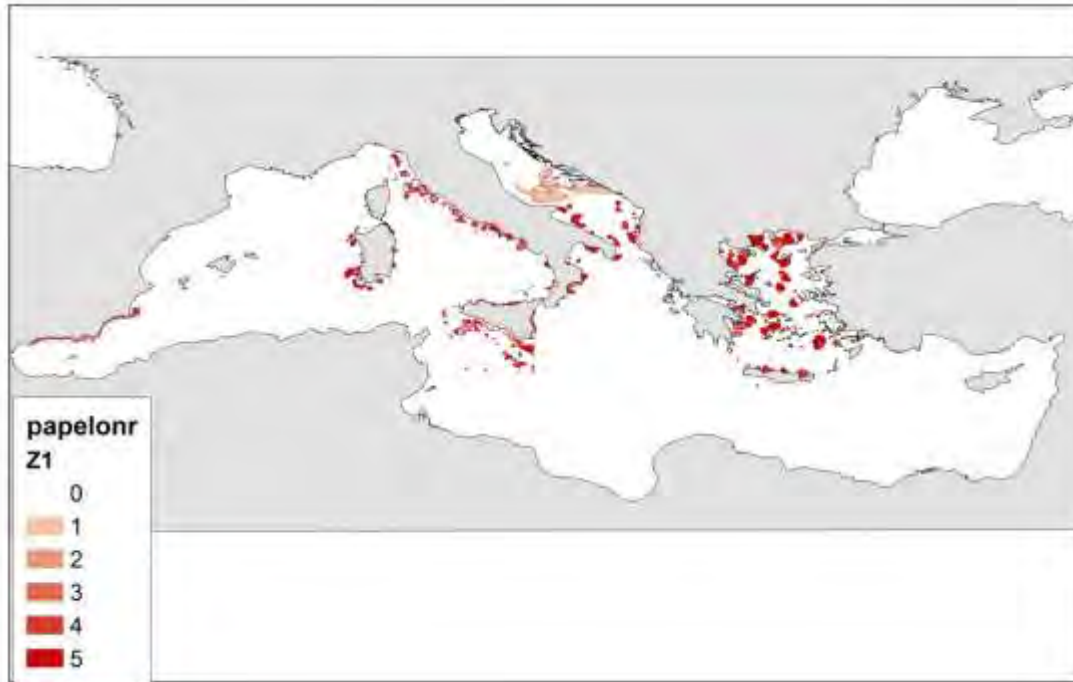


Fig 12. Spatial distribution of *Parapenaeus longirostris* nurseries based on MEDISEH. This map indicates the different levels of probability to find a nursery for this species.

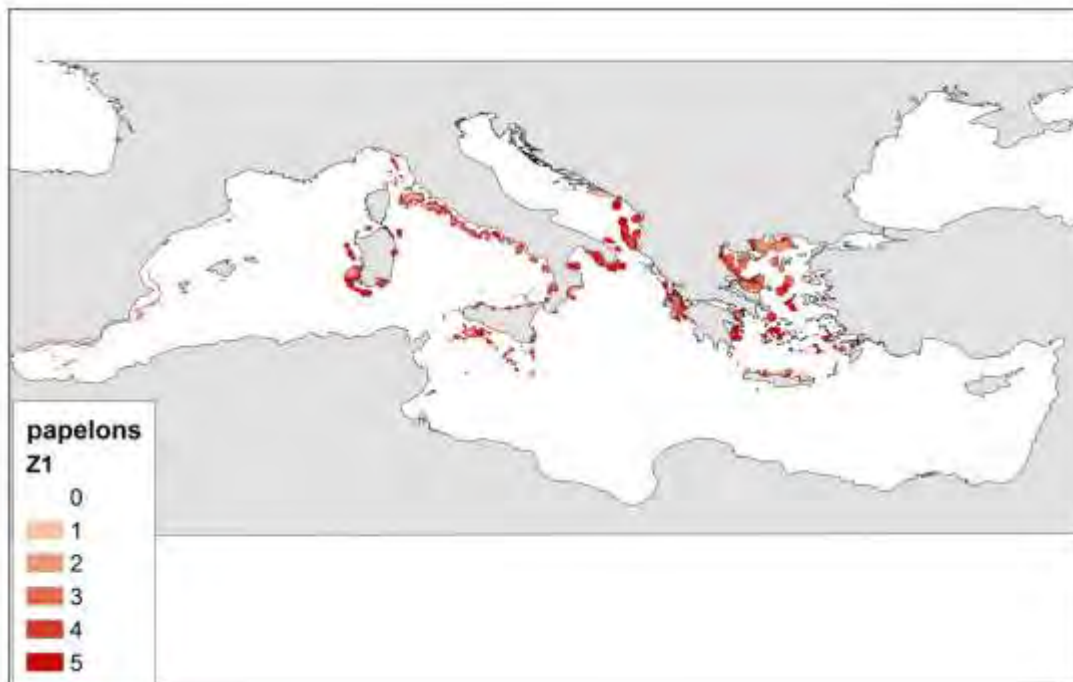


Fig 13. Spatial distribution of *Parapenaeus longirostris* spawners based on MEDISEH. This map indicates the different levels of probability to find a spawning hot spot for this species.

Nephrops norvegicus

The spatial distribution of Norway lobster, *Nephrops norvegicus*, in the Mediterranean Sea, used as an input in this Deliverable, was based on the analysis made by Sbrana et al. (2019). A time series of data from the Mediterranean International bottom Trawl Surveys (MEDITS) from 1994 to 2015 was analysed. The use of GAMs, in which standardized biomass indices (kg km^{-2}) were analysed as a function of environmental variables, explained how ecological factors (Sea Surface Temperature- SST, depth) could affect the spatio-temporal distribution of Norway lobster biomass in the basin. Norway lobster showed no particular temporal trend in bathymetric distribution over the years. There was a negative relationship between its biomass and SST. The highest biomass indexes of this species were associated with low temperatures, while biomass decreased with increasing SST (Sbrana et al., 2019).

Predictions of the aforementioned modeling approach and particularly an average for the years 2014 – 2016 were used for the purposes of this Deliverable (Fig. 14).

Spatial information from previous projects i.e. STOCMED (Fig. 15) and MEDISEH (Figs. 16 and 17) was also taken into account for interpretation purposes of the final results while those outcomes were not included directly to the estimation of fishing grounds for Norway lobster. A comparison by visualization was performed based on the several data sources shown in Figures 14-17.

Furthermore, for estimating the spatial distribution of Norway lobster in the south-eastern Mediterranean coasts (non-EU) the following approach was applied:

- 1) GAM was applied using as response variable the spatial occurrence of Norway lobster (based on Maina et al. 2016 outcomes) and the depth as predictor variable,
- 2) GAM was applied using as response variable the spatial distribution of Norway lobster's biomass (based on Sbrana et al. 2019 outcomes) and the depth as predictor variable.
- 3) The spatial outcomes of both models were then combined (through a spatial overlap procedure) to estimate the spatial distribution of this species in the south and eastern Mediterranean Sea (Fig. 14).

Based on the literature review, provided in Table 3, the *N. norvegicus* is not present in the eastern Mediterranean (Turan et al., 2016; <http://www.fao.org/geonetwork/srv/en/main.home?uuid=fao-species-map-nep>) and therefore eastern Mediterranean coasts were excluded from the spatial outcomes provided in Figure 14. Moreover, depth range of Norway lobster extends from 15 to 800 m, although it is typically found between 200 and 800 m in the Mediterranean Sea (<http://www.fao.org/geonetwork/srv/en/main.home?uuid=fao-species-map-nep>).

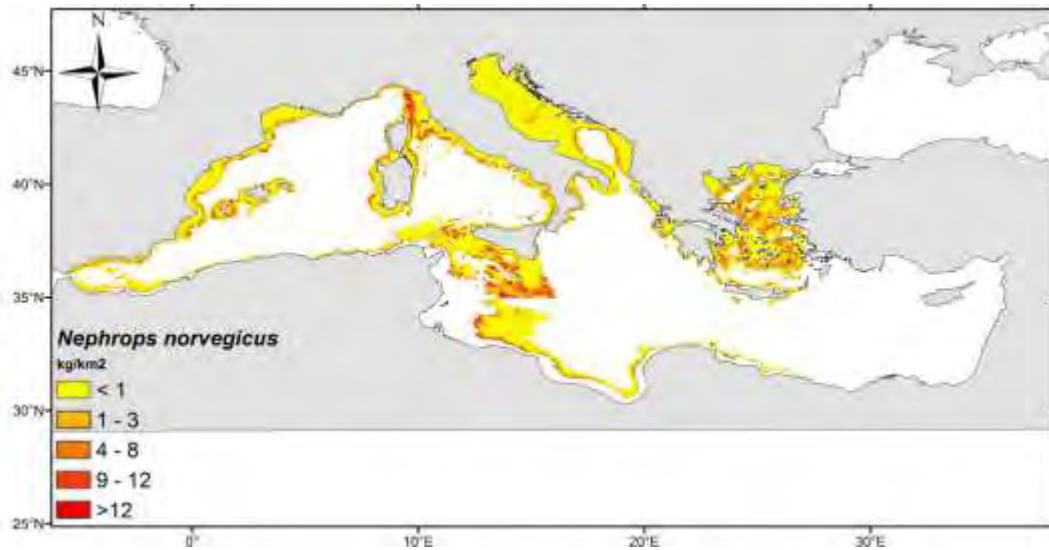


Fig 14. Spatial distribution of *Nephrops norvegicus* biomass based on combined information from Sbrana et al. (2019) and Maina et al., (2016) (average for the years 2014 – 2016).

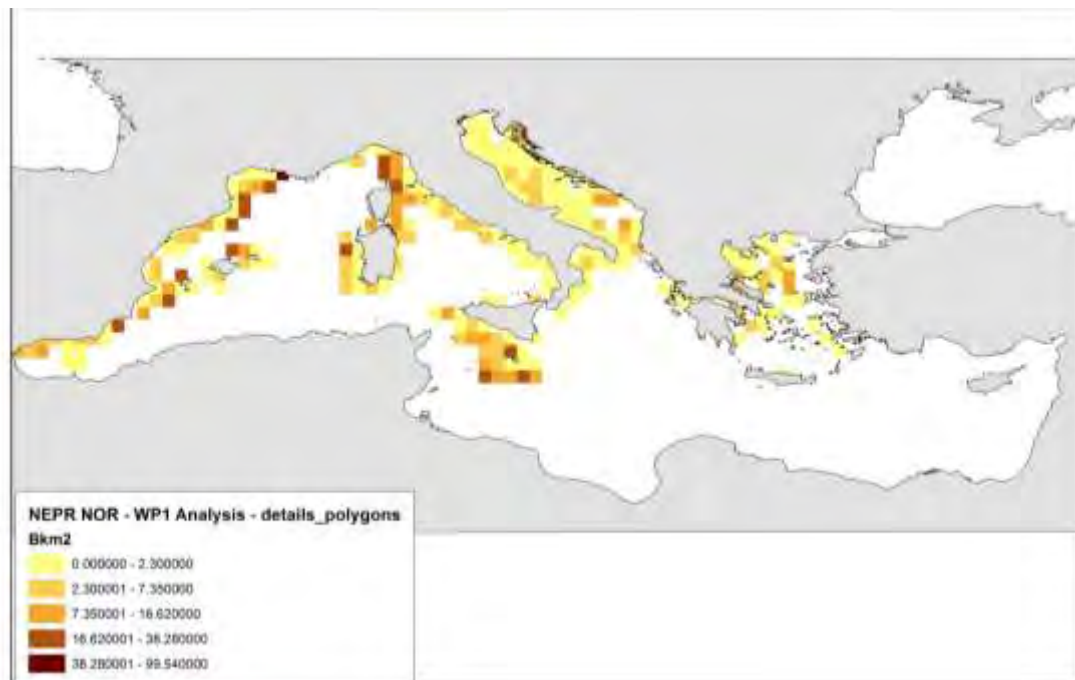


Fig 15. Spatial distribution of *Nephrops norvegicus* biomass based on STOCKMED.

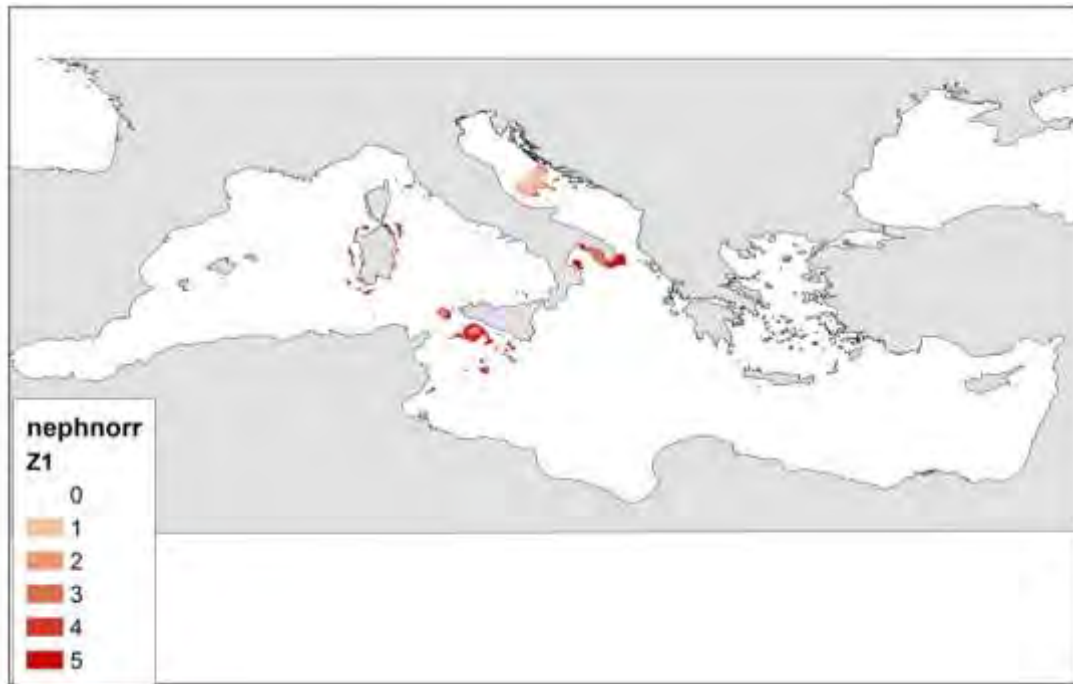


Fig 16. Spatial distribution of *Nephrops norvegicus* nurseries based on MEDISEH. This map indicates the different levels of probability to find a nursery for this species.

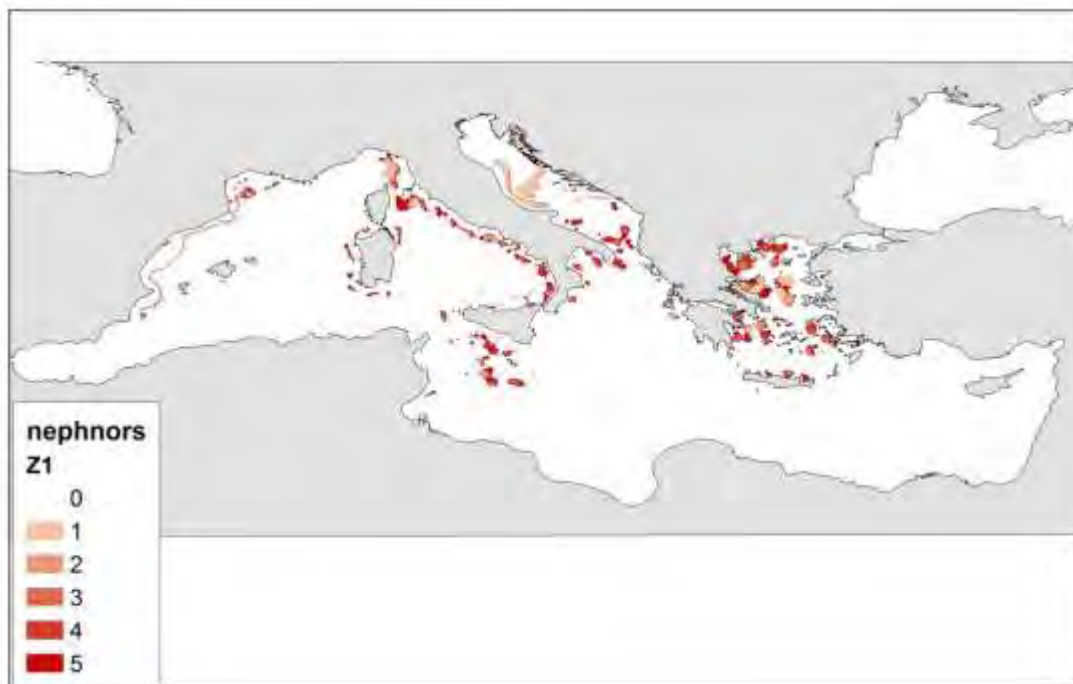


Fig 17. Spatial distribution of *Nephrops norvegicus* spawners based on MEDISEH. This map indicates the different levels of probability to find a spawning hot spot for this species.

Mullus barbatus

Based on the literature review provided in Table 3, red mullet is an important target species of the Mediterranean demersal fisheries that is fished along the west, central and eastern coasts.

Given that recent data based on literature were not available for the EU Mediterranean spatial scale for the *Mullus barbatus* spatial distribution, an analysis was performed based on the spatial information derived in the STOCKMED project supplemented by information on the species occurrence based on Maina et al. (2016). In particular, the following multi-step procedure was performed to improve the spatial coverage of STOCKMED project:

- 1) An IDW spatial interpolation method was applied to the spatial outcomes of STOCKMED project (Fig. 19) on the biomass of red mullet, *Mullus barbatus*, to improve the spatial resolution from 25*35 miles to 1*1 mile.
- 2) GAM was applied using as response variable the spatial distribution of red mullet's biomass (based on step 1 outcomes) and the depth as predictor variable.
- 3) The GAMs estimated by Maina et al. (2016) for modeling the red mullet occurrence in the Greek Seas were applied in a predictive mode for the entire Mediterranean Sea (by excluding the geographical coordinates and the swept area). Plots of the best fitting smooths showed a higher probability of finding red mullet in shallower waters, lower than 300m depth (Maina et al., 2016).
- 4) The outcomes of step 1, 2 and 3 were combined (based on a spatial overlap procedure) to estimate the abundance of red mullet in an improved spatial resolution (Fig 18).

Spatial information from MEDISEH (Fig. 20) was also taken into account only for interpretation purposes of the final results while those outcomes were not included directly to the estimation of fishing grounds for red mullet. A comparison by visualization was performed based on Figure 20. The period of the MEDITS survey, particularly foreseen by the protocol, is not suitable for monitoring the red mullet recruitment. Hence, the spatial distribution of red mullet recruits actually, was not considered in the estimations performed herein.

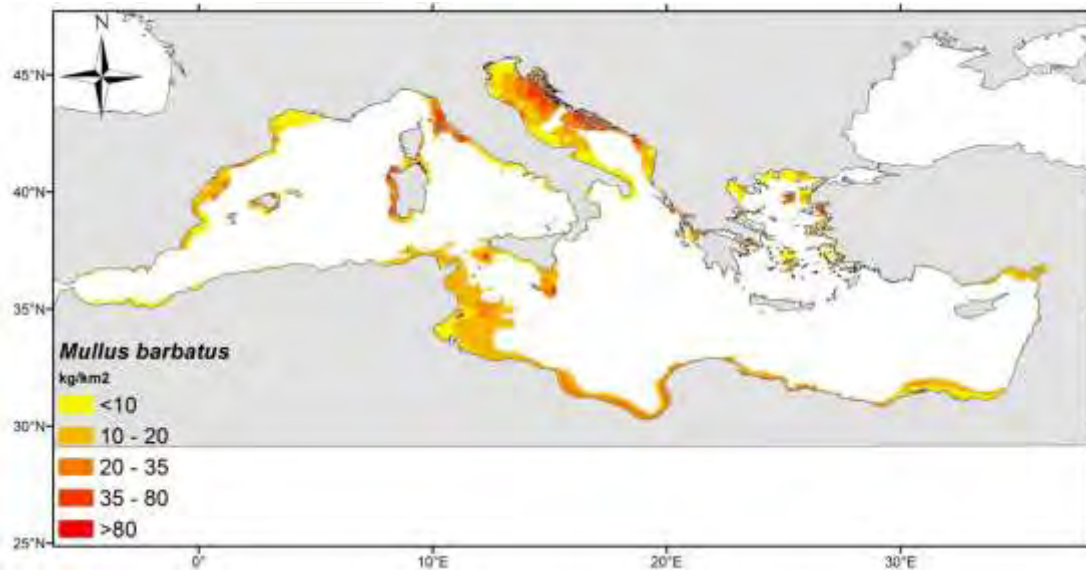


Fig 18. Spatial distribution of *Mullus barbatus* biomass based on combined information of STOCKMED and Maina et al. (2016).

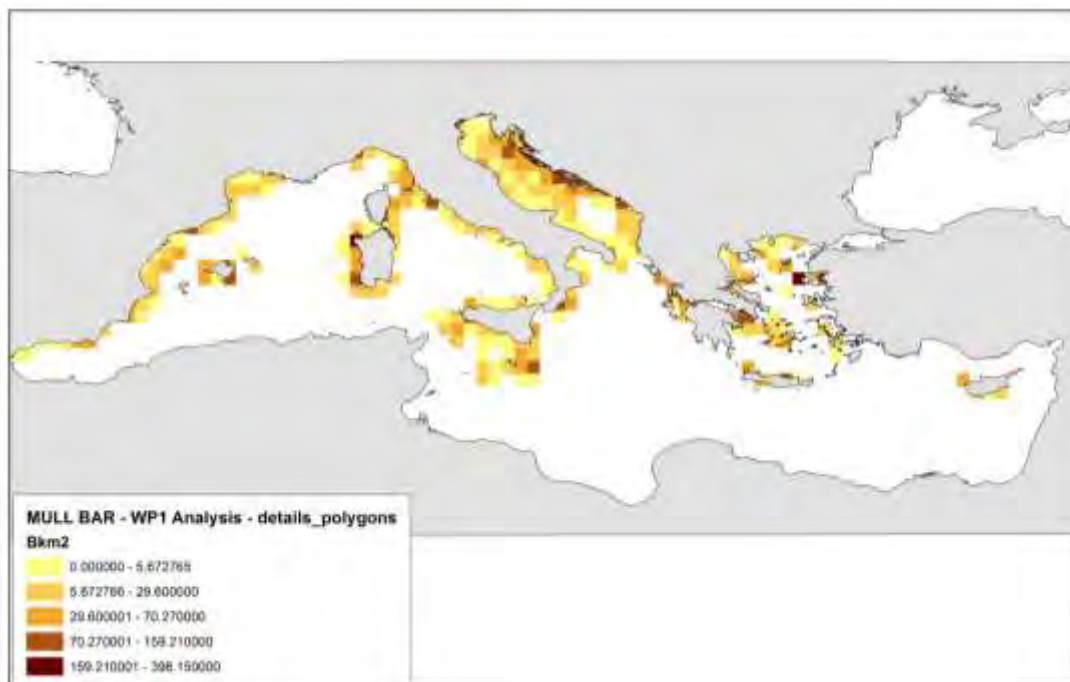


Fig 19. Spatial distribution of *Mullus barbatus* biomass based on STOCKMED.

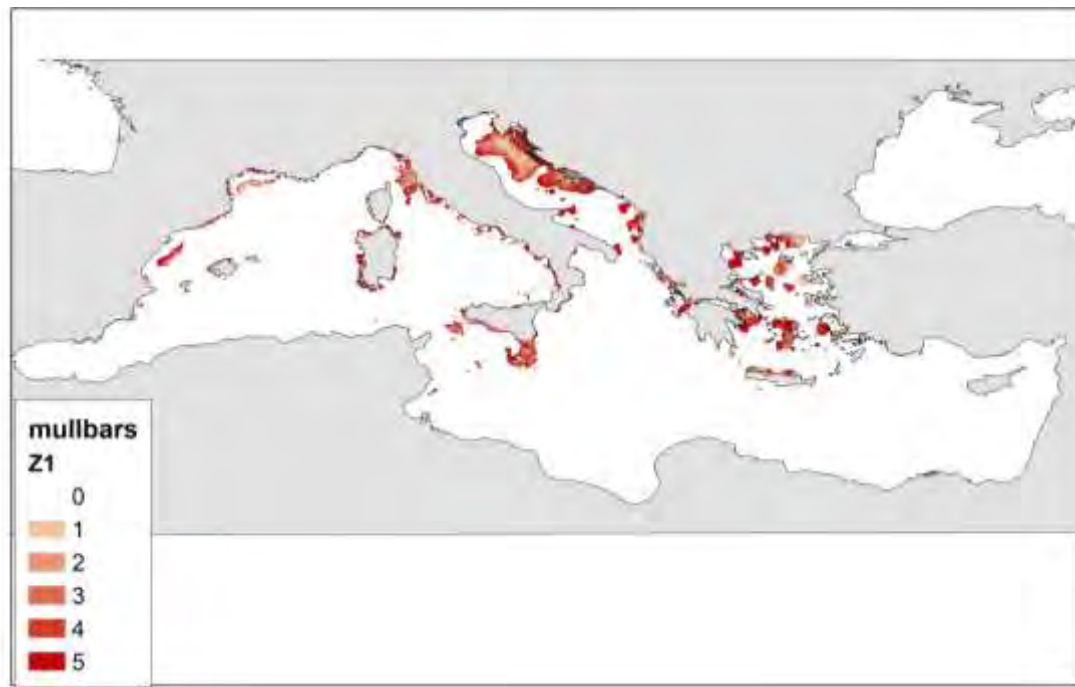


Fig 20. Spatial distribution of *Mullus barbatus* spawners based on MEDISEH. This map indicates the different levels of probability to find a spawning hot spot for this species.

Aristaeomorpha foliacea

Given that recent data based on literature were not available at the EU Mediterranean spatial scale for the spatial distribution of biomass for the giant red shrimp (*Aristaeomorpha foliacea*), an analysis was performed based on the spatial information derived in the STOCKMED project supplemented by information on the species occurrence based on Maina et al. (2016). In particular, the following multi-step procedure was performed to improve the spatial coverage of STOCKMED project:

- 1) An IDW spatial interpolation method was applied to the spatial outcomes of STOCKMED project (Fig. 22) on the abundance of giant red shrimp to improve the spatial resolution from 25*35 miles to 1*1 mile.
- 2) GAM was applied using as response variable the spatial distribution of giant red shrimp's biomass (based on step 1 outcomes) and the depth as predictor variable.
- 3) The GAMs estimated by Maina et al., 2016 for modeling the giant red shrimp in the Greek Seas were applied in a predictive mode for the entire Mediterranean Sea (by excluding the geographical coordinates and the swept area). Plots of the best fitting smooths showed a higher probability of finding giant red shrimp at depths > 450 m (Maina et al., 2016). The outcomes of step 1, 2 and 3 were combined (based on a spatial overlap procedure) to estimate the abundance of giant red shrimp in an improved spatial resolution (Fig. 21).

In the Mediterranean Sea, the distribution of giant red shrimp is patchy in nature, with the highest abundances found in the central-eastern basins. In the Central Mediterranean, there is a longitudinal segregation between the two species of red shrimp: *A. antennatus* decreases in abundance from the western to the eastern Mediterranean, whilst the opposite pattern applies to *A. foliacea* (Cau et al., 2002; Orsi Relini et al., 2013; Masnadi et al., 2018; Gujjarro et al., 2019). In Tunisian waters, the relative abundance of the two species has been reported to be in the ratio 1:1 (<http://www.faomedsudmed.org/html/species/Aristaeomorpha%20foliacea.html>; see Table 3 and references therein). The maps presenting the spatial distribution of giant red shrimps in the south-eastern Mediterranean Sea were adjusted (based on a correction factor), to show that the biomass increases from the western to the eastern areas. In particular, the spatial outcomes for the south-eastern Mediterranean were multiplied by the following correction factor i.e. 0.95 at west part, 1 at central part and 1.2 at eastern part.

Finally, spatial information from MEDISEH (Figs. 23 and 24) were also taken into account only for interpretation purposes of the final results while those outcomes were not included directly to the estimation of fishing grounds for *A. foliacea*.

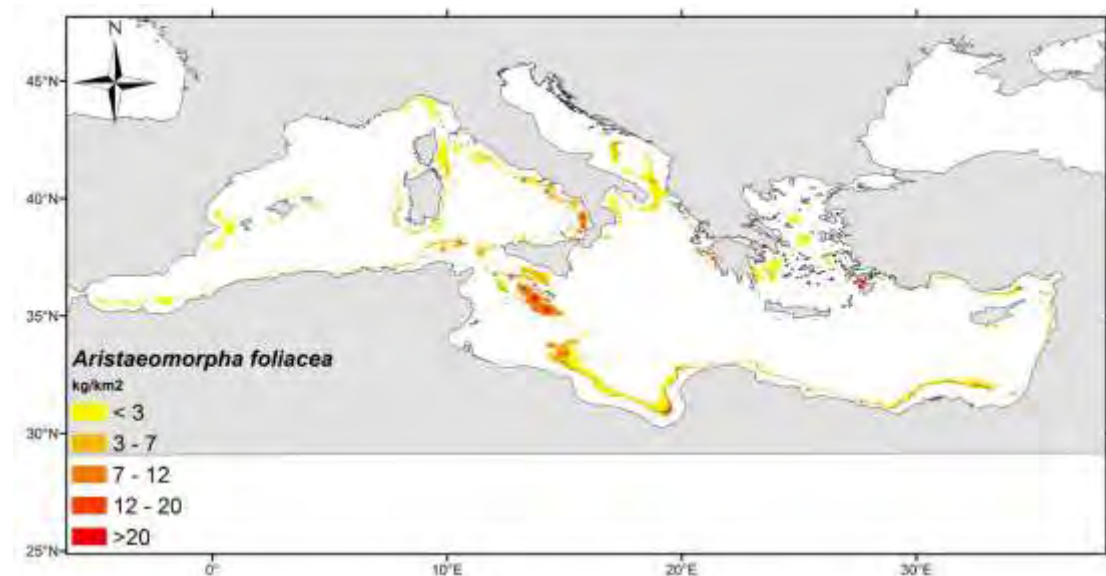


Fig 21. Spatial distribution of *Aristaeomorpha foliacea* biomass based on combined information of STOCKMED and Maina et al. (2016).

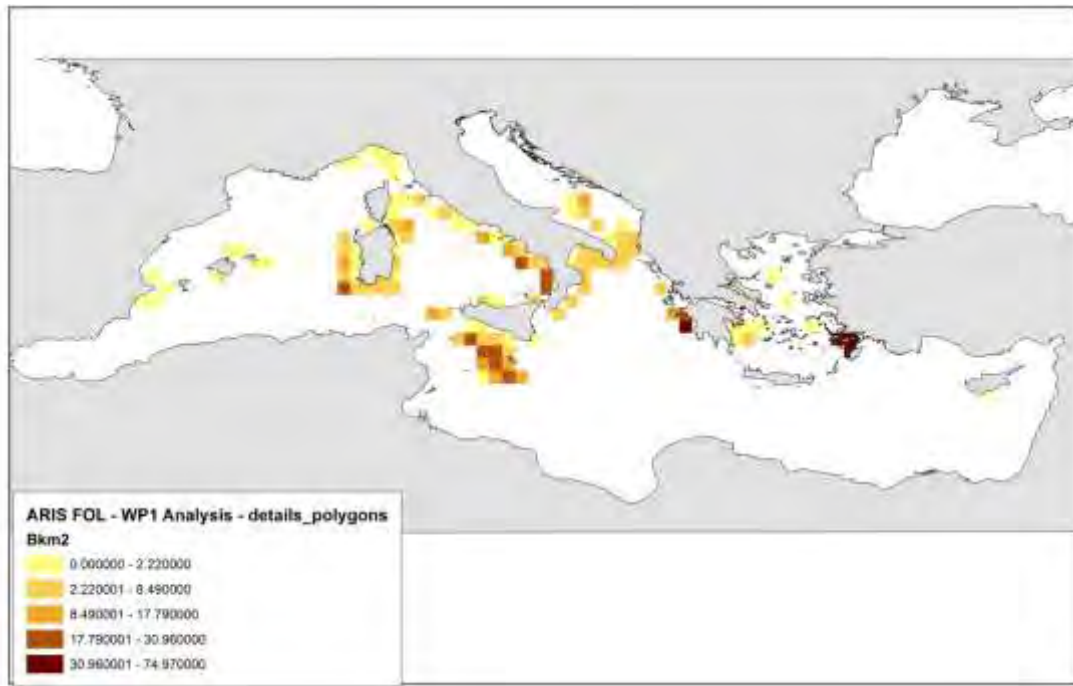


Fig 22. Spatial distribution of *Aristaeomorpha foliacea* biomass based on STOCKMED.

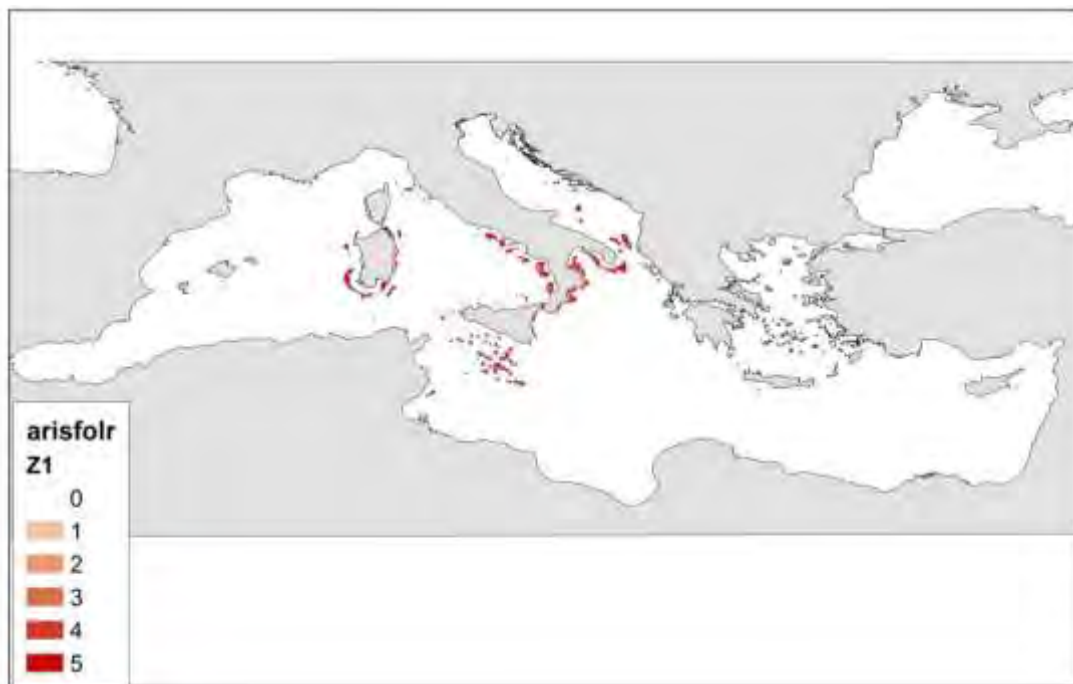


Fig 23. Spatial distribution of *Aristaeomorpha foliacea* nurseries based on MEDISEH. This map indicates the different levels of probability to find a nursery for this species.

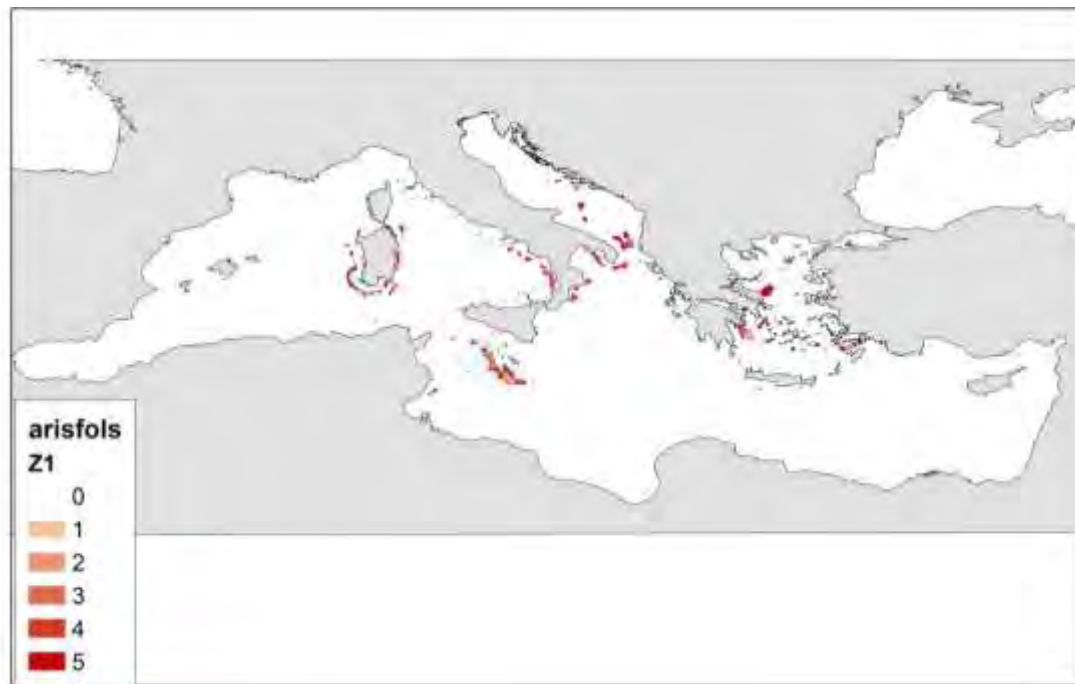


Fig 24. Spatial distribution of *Aristaeomorpha foliacea* spawners based on MEDISEH. This map indicates the different levels of probability to find a spawning hot spot for this species.

Aristeus antennatus

While recent data based on literature were not available for blue and red shrimp, *Aristeus antennatus*, spatial distribution of biomass and for the EU Mediterranean spatial scale, an analysis was performed based on the spatial information derived in the STOCKMED project supplemented by information on the species occurrence based on Maina et al. (2016). In particular, the following multi-step procedure was performed in order to improve the spatial coverage of the STOCKMED project:

- 1) An IDW spatial interpolation method was applied to the spatial outcomes of STOCKMED project (Fig. 26) on the abundance of *A. antennatus* to improve the spatial resolution 25*35 miles to 1*1 mile.
- 2) GAM was applied using as response variable the spatial distribution of *A. antennatus* biomass (based on step 1 outcomes) and the depth as predictor variable.
- 3) The GAMs estimated by Maina et al., 2016 for modeling the *A. antennatus* occurrence were applied in the Greek Seas were applied in a predictive mode for the entire Mediterranean Sea (by excluding the geographical coordinates and the swept area). Plots of the best fitting smooths showed a higher probability of finding *A. antennatus* at depths > 500 m (Maina et al., 2016).
- 4) The outcomes of step 1, 2 and 3 were combined (based on a spatial overlap procedure) to estimate the abundance of *A. antennatus* in an improved spatial resolution (Fig. 25).

The maps showing the spatial distribution of *A. antennatus* in the south-eastern Mediterranean Sea, were adjusted (based on a correction factor) to show that the biomass decreases from the western to the eastern areas (see references in Table 3). In particular, the spatial outcomes for the south-eastern Mediterranean were multiplied by the following correction factor i.e. 1.2 at west part, 1 at central part and 0.95 at eastern part.

Finally, spatial information from MEDISEH on spawners hot spots (Fig. 27), was also taken into account only for interpretation purposes of the final results while that information was not included directly to the estimation of fishing grounds for *A. antennatus*.

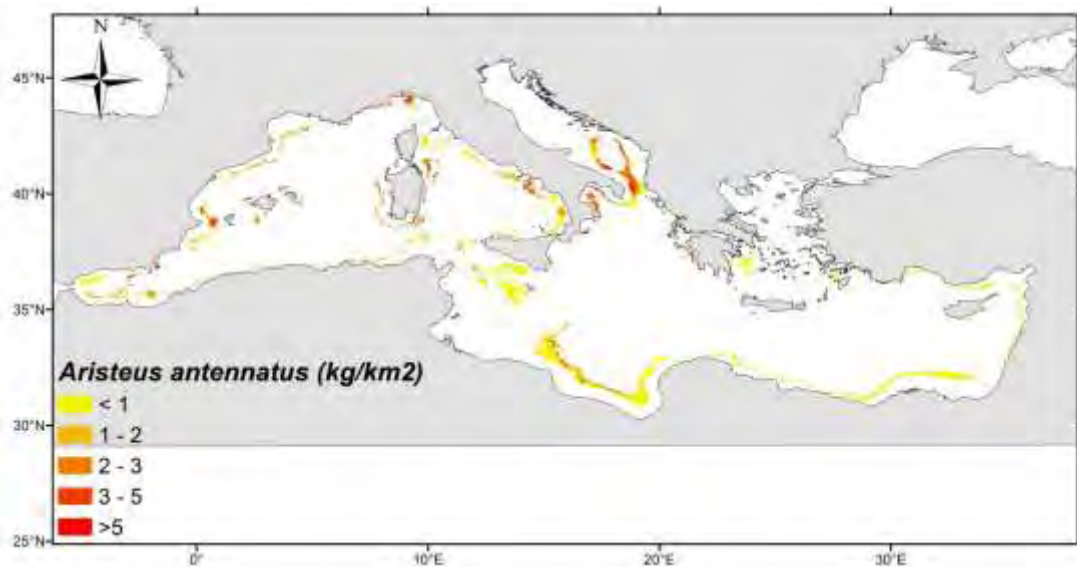


Fig 25. Spatial distribution of *Aristeus antennatus* biomass based on combined information of STOCKMED and Maina et al. (2016).

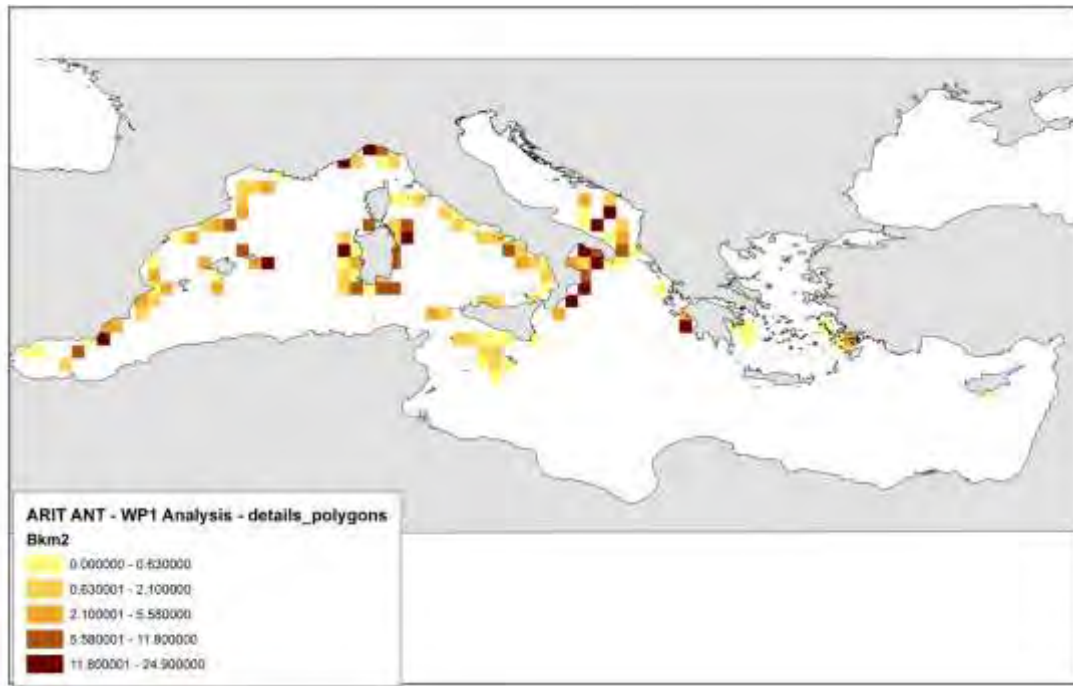


Fig 26. Spatial distribution of *Aristeus antennatus* biomass based on STOCKMED.

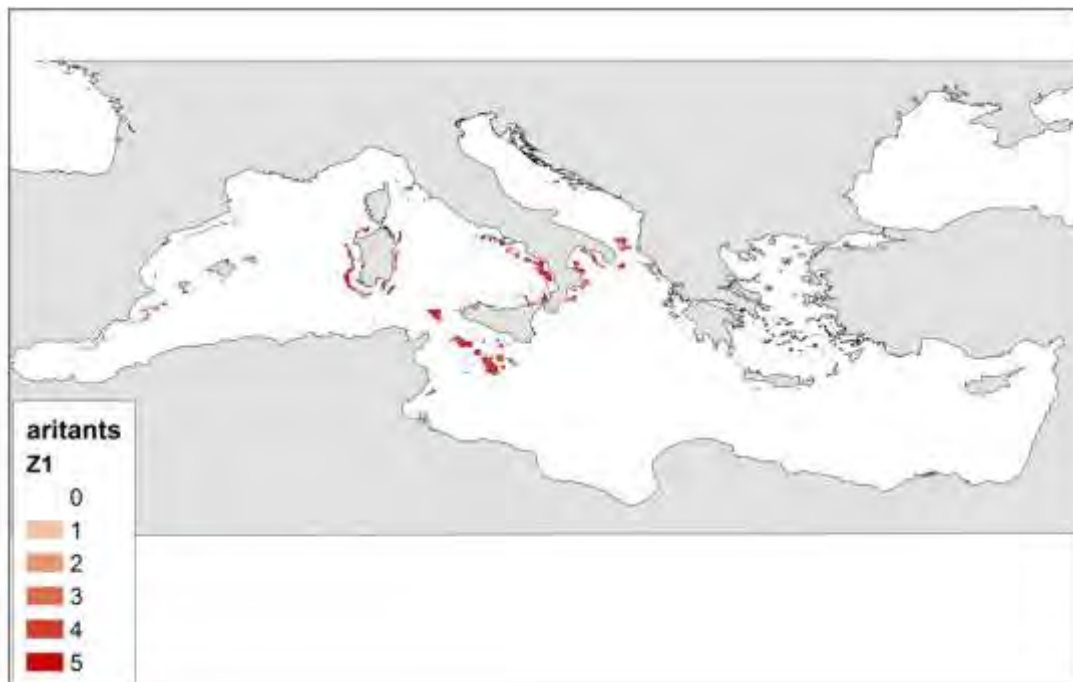


Fig 27. Spatial distribution of *Aristeus antennatus* spawners based on MEDISEH. This map indicates the different levels of probability to find a spawning hot spot for this species.

Analyzing spatial patterns - Mapping clusters

The potential fishing grounds of a species were determined by the spatial overlap of two types of layers:

- 1) the spatial distribution of each species (spatial data shown in Figures 7; 10; 14; 18; 21 and 25) and
- 2) the distribution of the fishing effort (spatial data shown in Figures 2-6).

This spatial overlap is the product of the two layers and demonstrates the possibility for a fisher to harvest a particular species in a specific area.

In the case of a high product, i.e. high probability of species abundance and high fishing effort, and also high values in all neighbouring cells, the specific site was considered as a hot spot. In contrast, in the case of a low (but positive) product, i.e. low probability of species abundance and low fishing effort, and also low values of all neighbouring cells, the specific site was considered as a cold spot.

Based on methods used for analyzing spatial patterns and mapping clusters (Maina et al., 2016), statistically significant hot and cold spots of the potential fishing grounds, for each species, were estimated. Additionally, hot and cold spots for the assemblages of all species under investigation were identified by aggregating hot and cold spots for each species.

The analysis of spatial patterns was based on the exploration of 'global measures', which reveal whether a pattern is 'dispersed', 'random' or 'clustered' in space and over what scale that clustering occurs (Fortheringham et al., 2000; Fortin and Dale, 2005). The global exploration of spatial patterns concerning the potential fishing grounds for each fishing strategy was based on the estimation on the Global Moran's I statistic for a series of increasing distances which measure the intensity of spatial clustering for each distance (Goodchild, 1986). Global Moran's I statistic was also used to evaluate the most appropriate scale to be used in Getis-Ord G_i^* statistic, which detects 'local pockets' of dependence that may not be revealed when using global statistics (Getis and Ord, 1992). For the exploration of global measures, 5 km was set as a starting distance for Global Moran's I statistic.

A Getis-Ord G_i^* statistic was applied to the product of fishing effort and the probability of species presence values, to identify statistically significant hot and cold spots of potential fishing grounds (Getis and Ord, 1992). Hot Spot Analysis calculates the Getis-Ord G_i^* statistic for each feature in a weighted set of features. The Getis-Ord G_i^* statistic shows whether features with high or low values tend to be clustered in the study area. This method works by looking at each feature within the context of neighbouring features.

For each species, areas classified as hot spots or cold spots were then aggregated in two single outputs in order to identify species richness in the most important fishing grounds for commercial species assemblages. The aim of this

Deliverable, was to provide a general picture of fishing grounds for the six species under investigation and not to highlight any differences or fishers' preferences between species (e.g. based on economic importance or else). To that end, aggregated hot/cold spots for these species were estimated on the basis of equally weighting between the different species.

Given that the spatial scale and extent is critical for the reliability of the results based on the analysis of "global and local" measures (i.e. Global Morans I. statistic, Getis ord G_i^* statistic), the analysis was performed separately for western, central and eastern Mediterranean Sea.

Results

Potential fishing grounds by species and métier (Spatial overlap)

The results of the spatial overlap which demonstrates the possibility for a fisher to harvest a certain species in a specific area are shown in Figures 28-37.

Merluccius merluccius

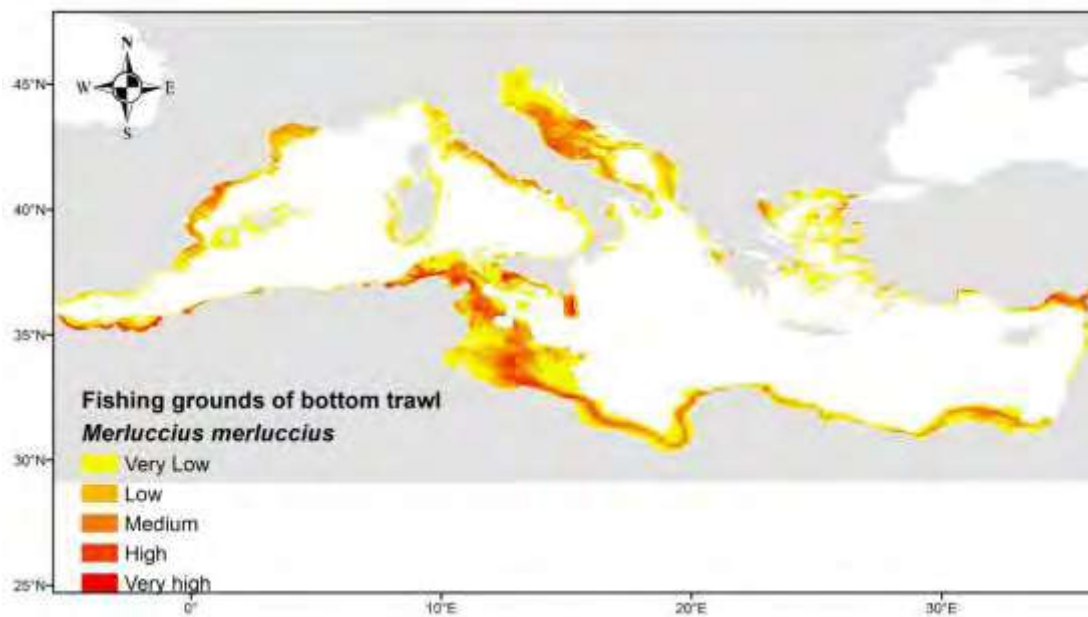


Figure 28. Potential fishing grounds of *Merluccius merluccius* for bottom trawling.

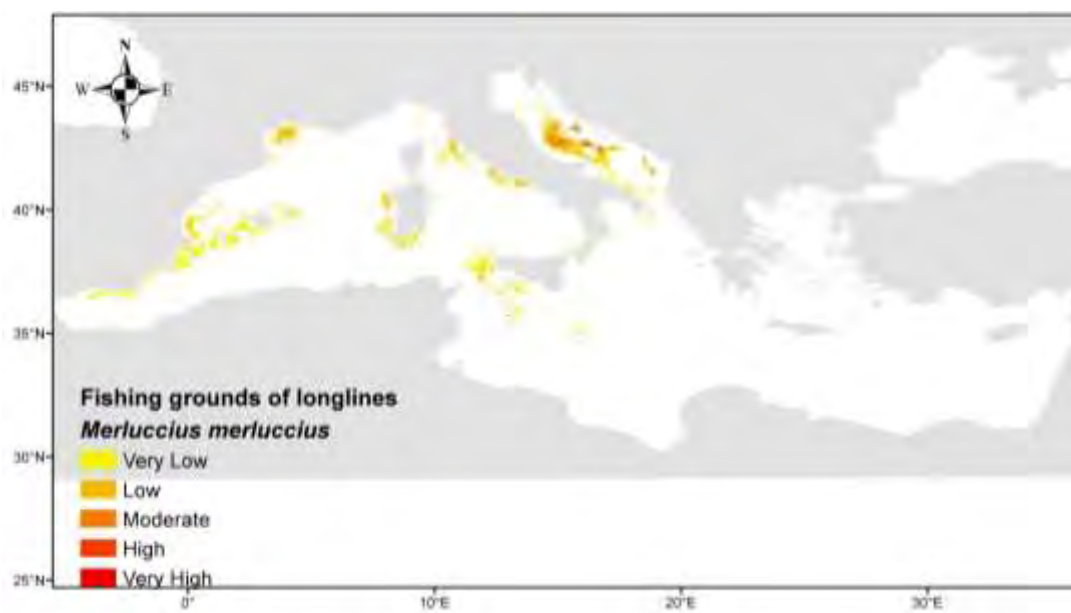


Figure 29 Potential fishing grounds of *Merluccius merluccius* for longlines (LOA > 12 m).

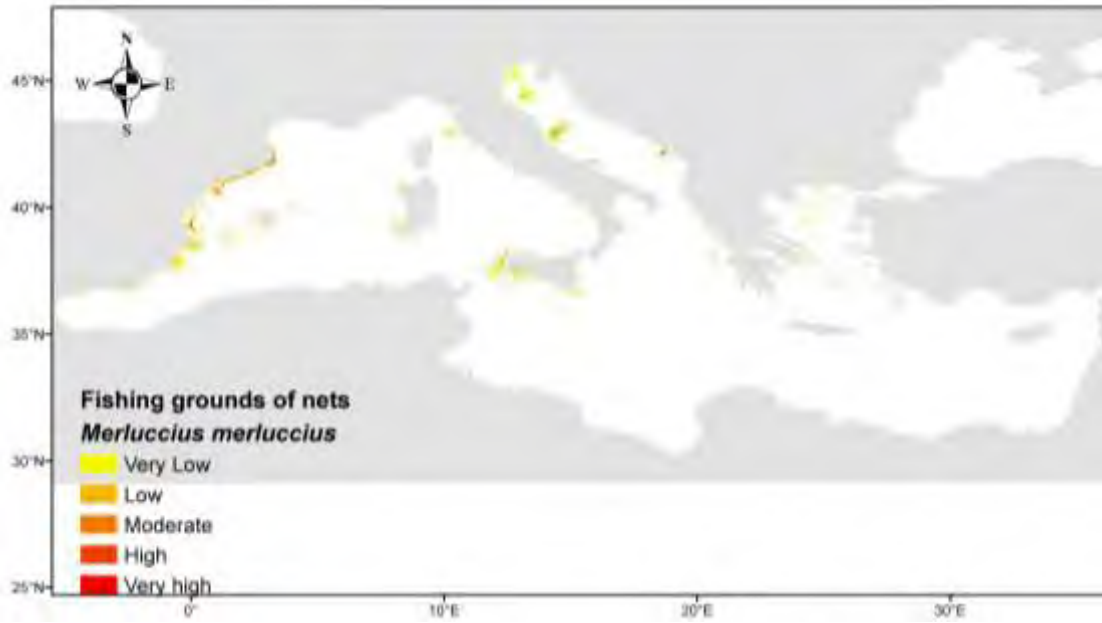


Figure 30. Potential fishing grounds of *Merluccius merluccius* for nets (LOA >12 m).

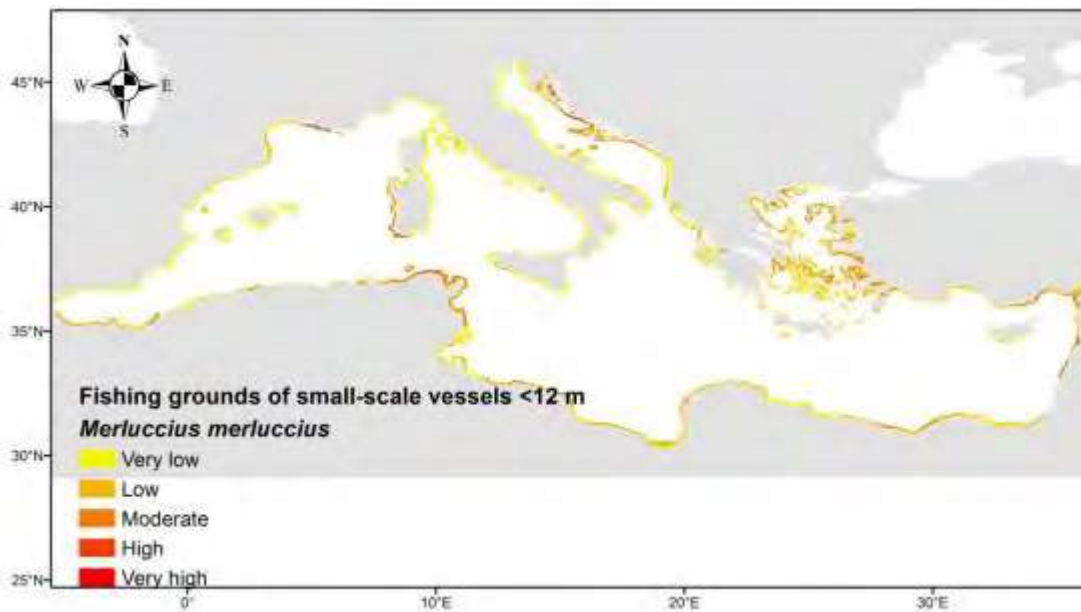


Figure 31. Potential fishing grounds of *Merluccius merluccius* for small-scale fisheries (LOA < 12 m).

Parapenaeus longirostris

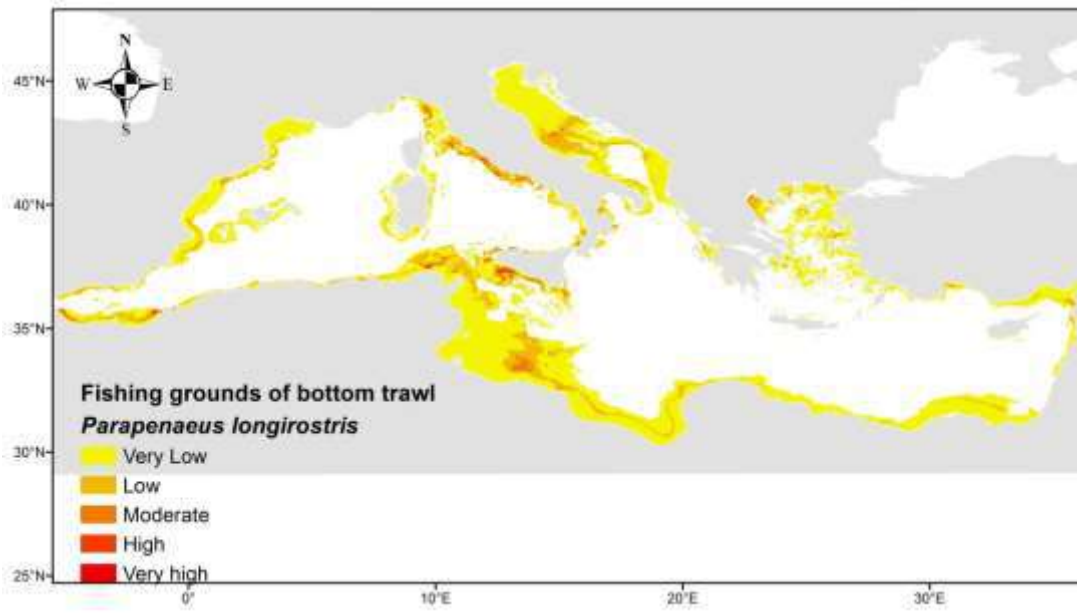


Figure 32. Potential fishing grounds of *Parapenaeus longirostris* for bottom trawling.

Nephrops norvegicus

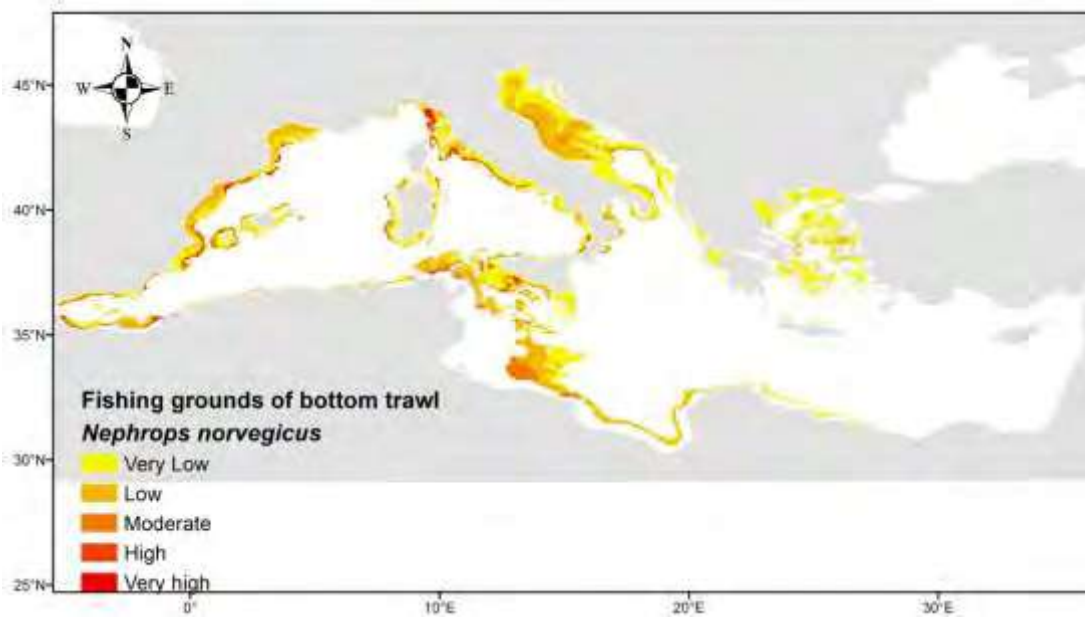


Figure 33. Potential fishing grounds of *Nephrops norvegicus* for bottom trawling.

Mullus barbatus

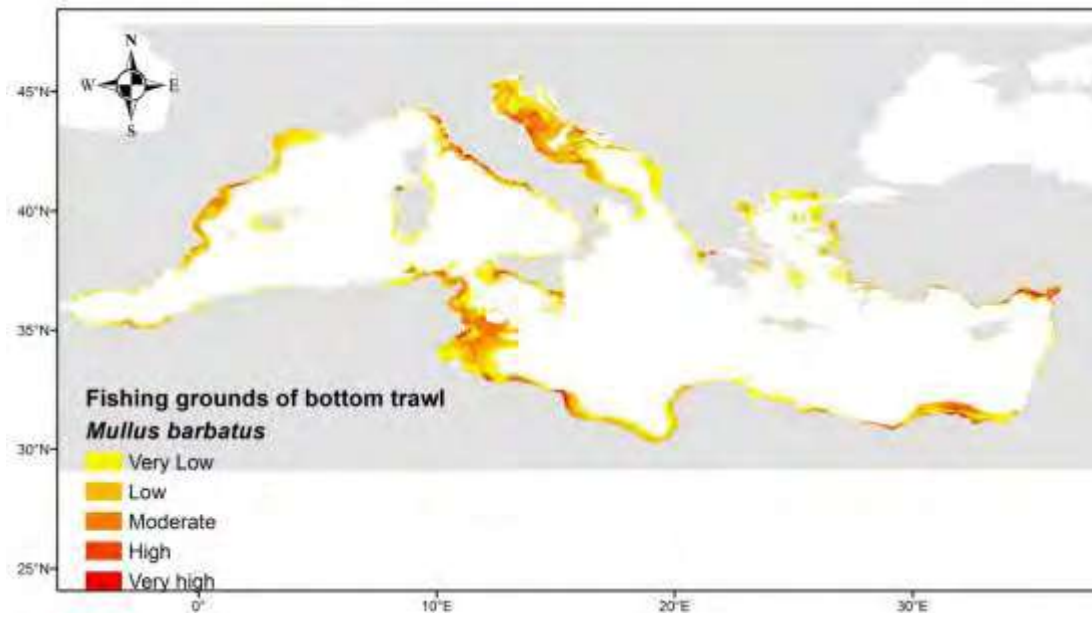


Figure 34. Potential fishing grounds of *Mullus barbatus* for bottom trawling.

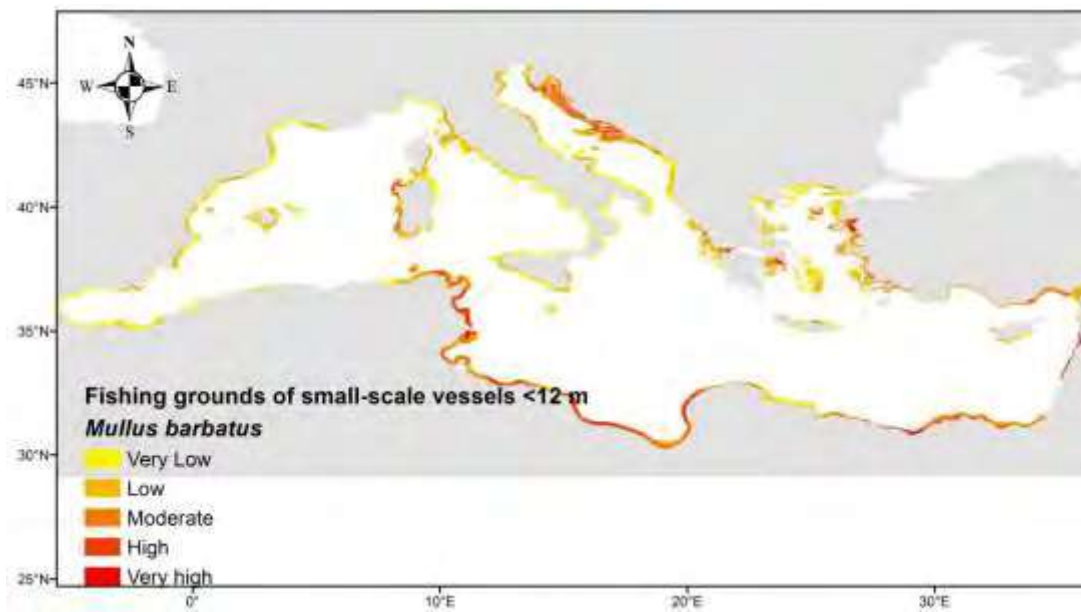


Figure 35. Potential fishing grounds of *Mullus barbatus* for small-scale fisheries (< 12 m LOA).

Aristaeomorpha foliacea

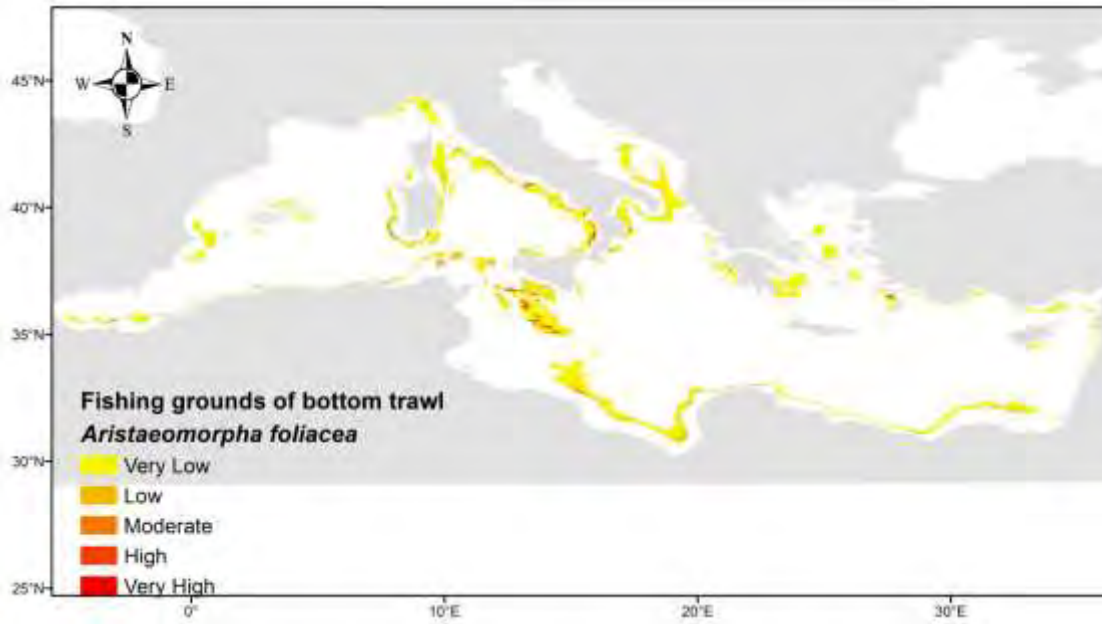


Figure 36. Potential fishing grounds of *Aristaeomorpha foliacea* for bottom trawling.

Aristeus antennatus

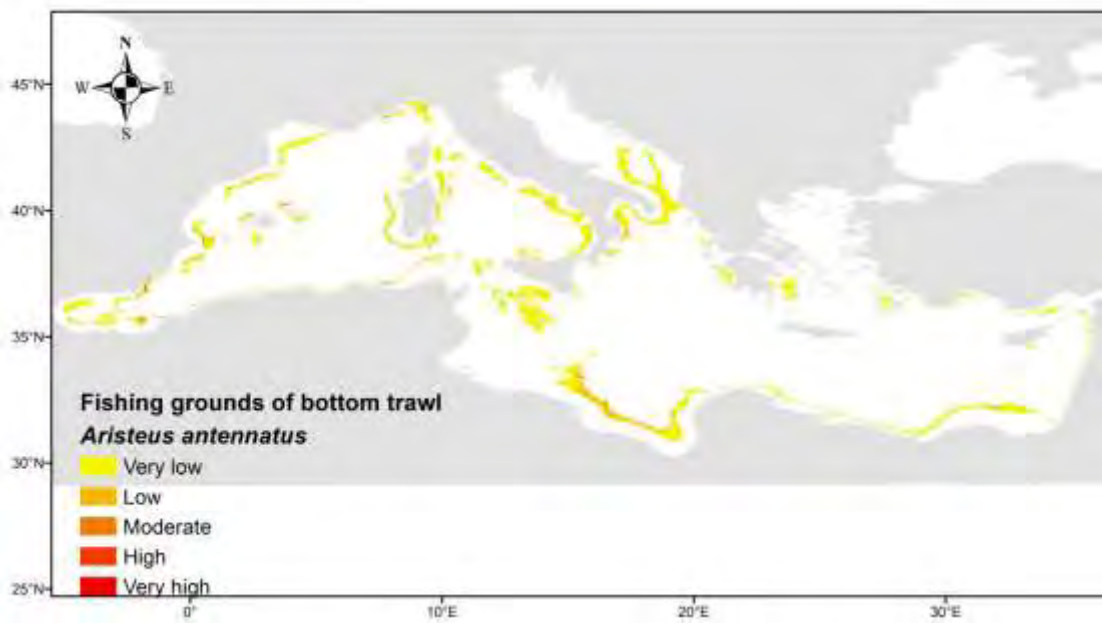


Figure 37. Potential fishing grounds of *Aristeus antennatus* for bottom trawling.

Hot spot analysis (estimate hot and cold spots by species and métier)

Spatial clustering was based on the potential fishing grounds of each species that were determined by the spatial overlap between the spatial distribution of each species abundance (based on literature and previous projects) and the distribution of fishing effort from bottom trawlers, longlines, nets and small-scale fisheries < 12 m LOA (Deliverable 3.1.). Maps of hot and cold spots for each species were based on the Getis-Ord G_i^* statistic. Such hot spots represent the most important fishing grounds for each species and are shown from figures 38 to 47.

Merluccius merluccius

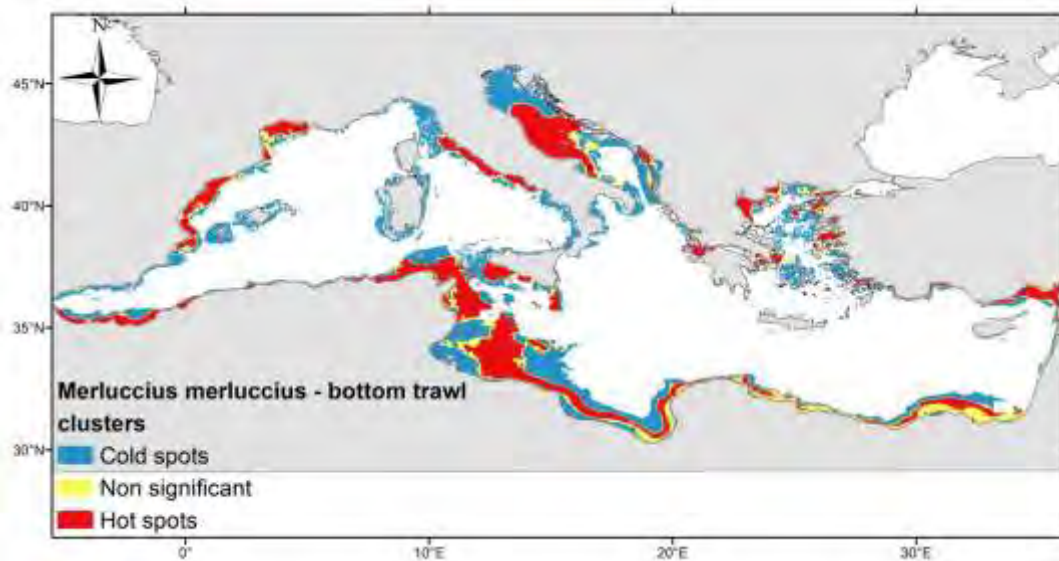


Figure 38. Estimated hot and cold spots of the *Merluccius merluccius* for bottom trawling. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate the absence of fishing activity or absence of species.

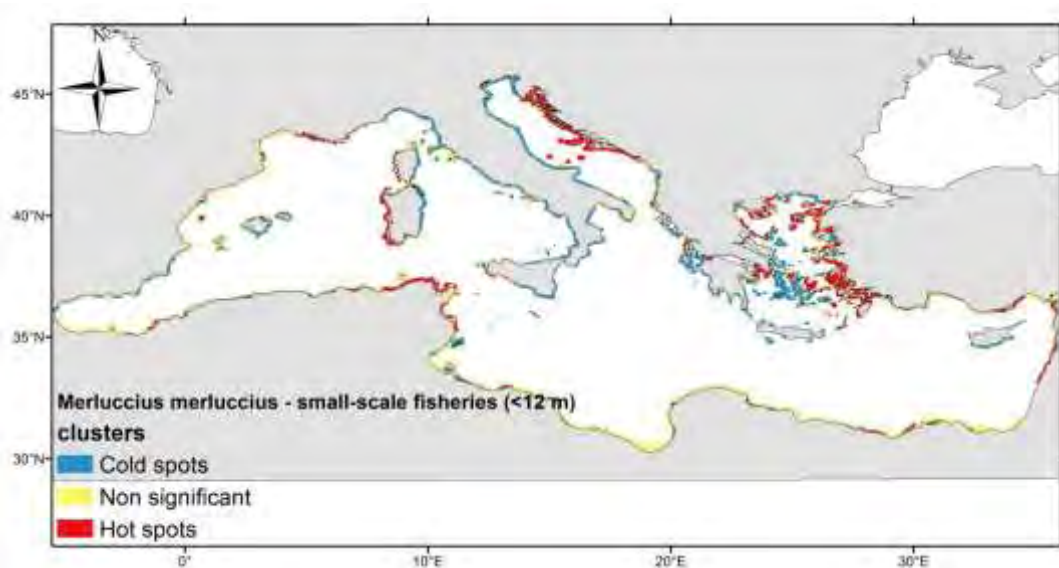


Figure 39. Estimated hot and cold spots of the *Merluccius merluccius* for small- scale fisheries (<12 m LOA). Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species.

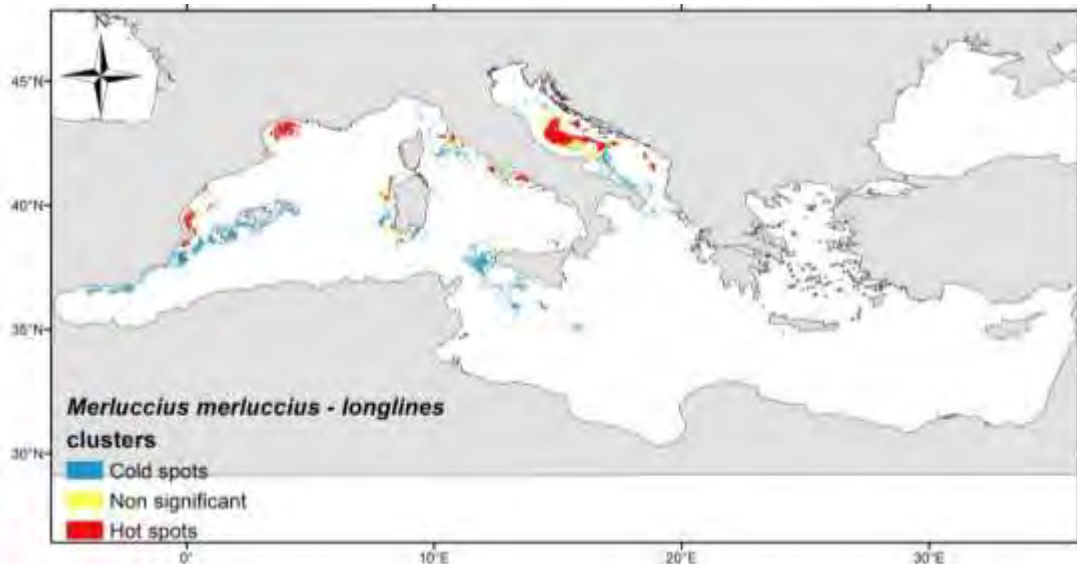


Figure 40. Estimated hot and cold spots of the *Merluccius merluccius* for longlines (LOA > 12 m). Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species.

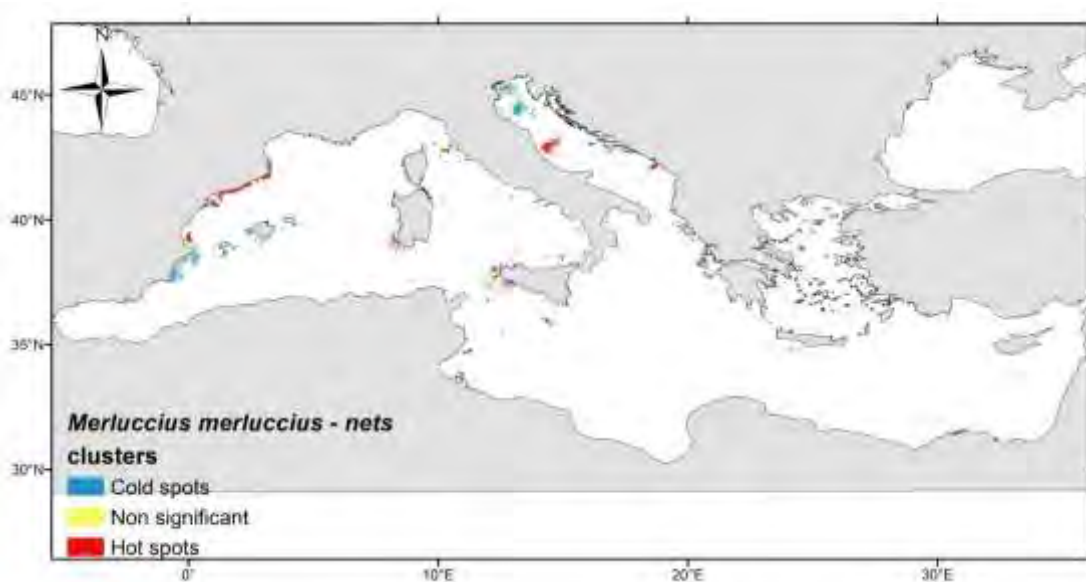


Figure 41. Estimated hot and cold spots of the *Merluccius merluccius* for nets (LOA > 12 m). Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species.

Parapenaeus longirostris

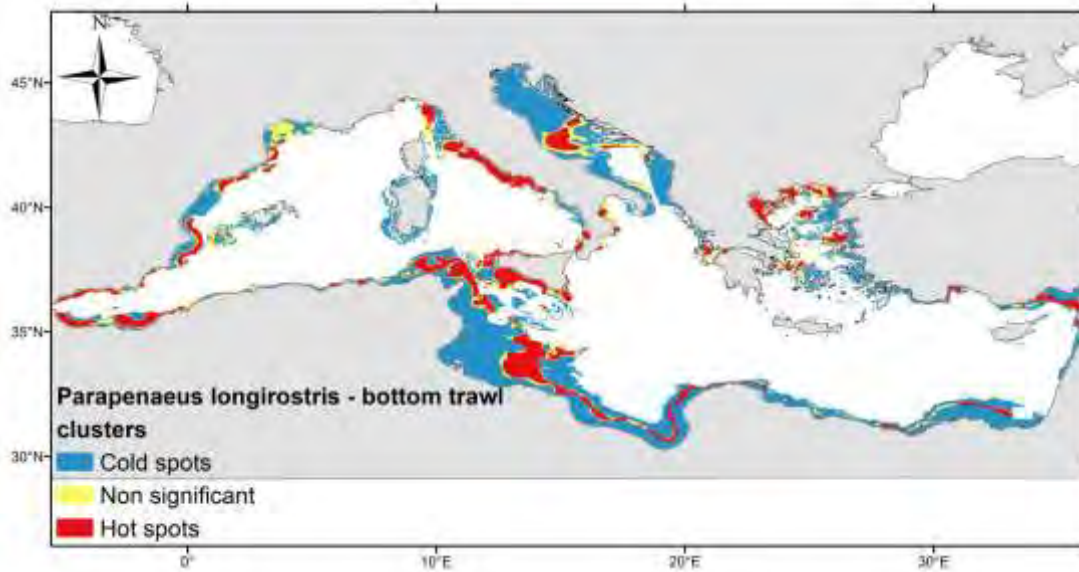


Figure 42. Estimated hot and cold spots of the *Parapenaeus longirostris* for bottom trawl. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species.

Nephrops norvegicus

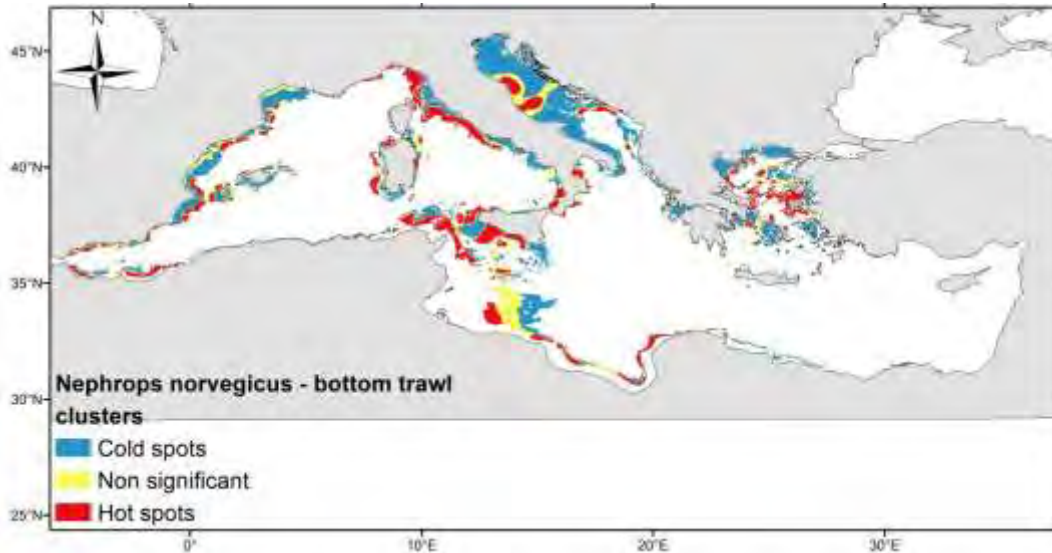


Figure 43. Estimated hot and cold spots of the *Nephrops norvegicus* for bottom trawl. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species

Mullus barbatus

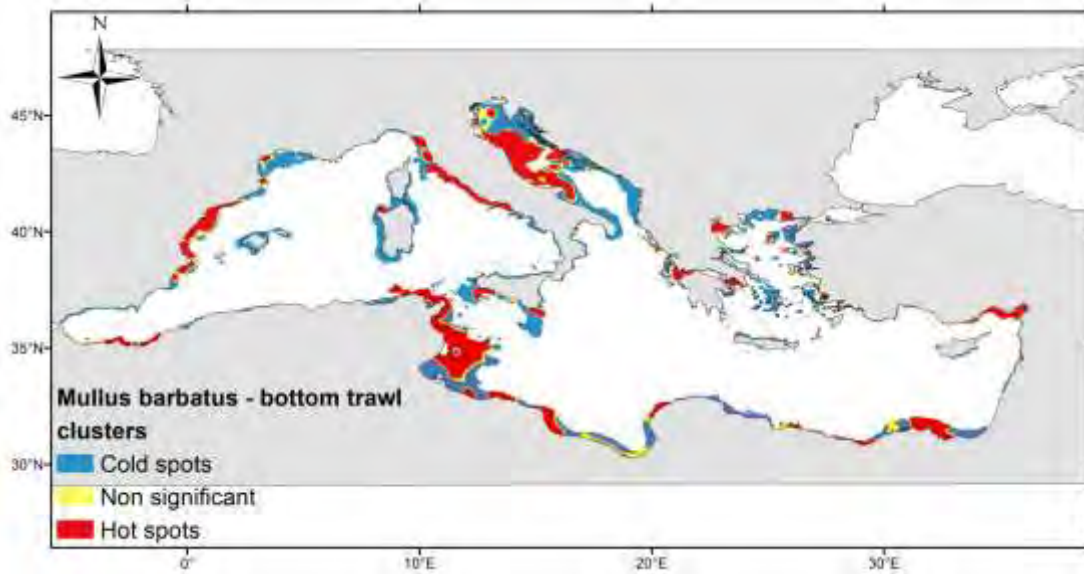


Figure 44. Estimated hot and cold spots of the *Mullus barbatus* for bottom trawling. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species

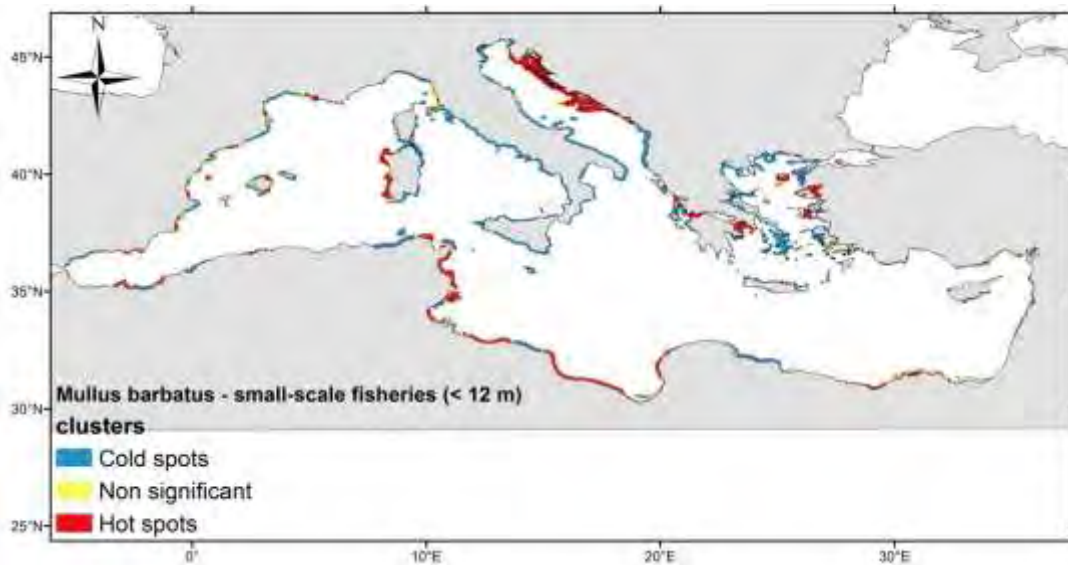


Figure 45. Estimated hot and cold spots of the *Mullus barbatus* for small-scale fisheries. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species

Aristaeomorpha foliacea

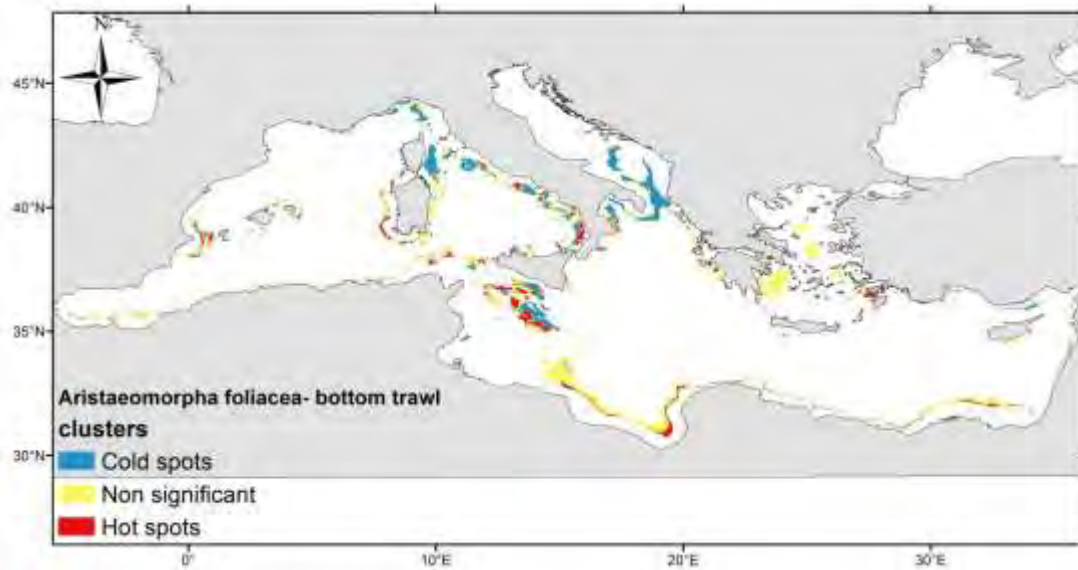


Figure 46. Estimated hot and cold spots of the *Aristaeomorpha foliacea* for bottom trawling. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species

Aristeus antennatus

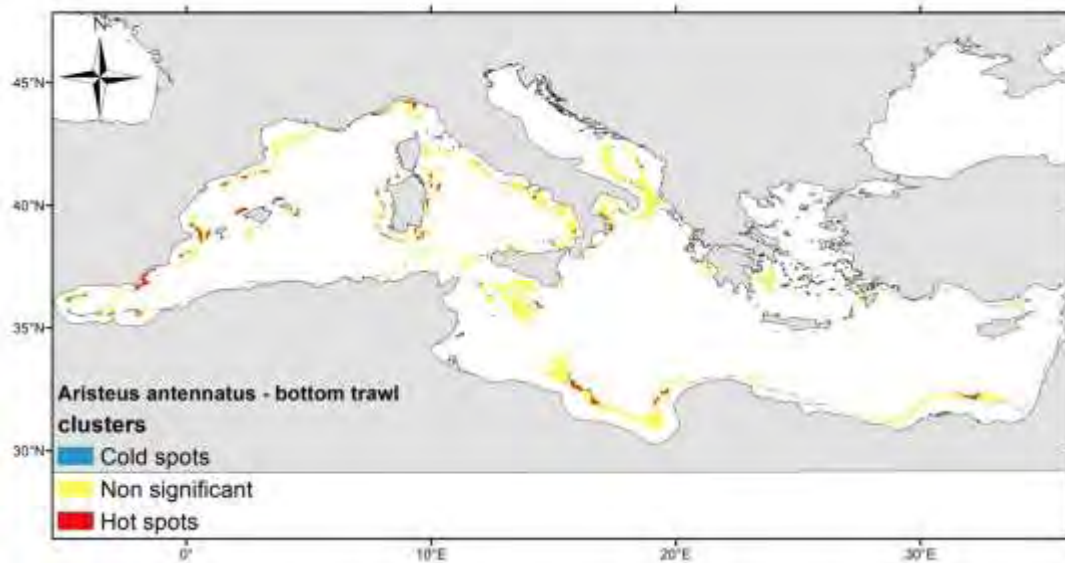


Figure 47. Estimated hot and cold spots of the *Aristeus antennatus* for bottom trawling. Non-significant values (at 95% significance level) are translated as no spatial clustering for the spatial overlap between the species abundance and fishing efforts. White coloured areas indicate absence of fishing activity or absence of species

Aggregated hot and cold spots

Aggregated hot spots showed that the main fishing grounds of the target species for bottom trawling (i.e. European hake, *Merluccius merluccius*; Norway lobster, *Nephrops norvegicus*; deep-water rose shrimp, *Parapenaeus longirostris*; red mullet, *Mullus barbatus*; giant red shrimp, *Aristaeomorpha foliacea*; blue and red shrimp, *Aristeus antennatus*) are located in the Adriatic, Tyrrhenian, Strait of Sicily, Aegean, eastern Ionian, Balearic, Alboran, Libyan and Levantine Sea (Fig. 48). For small-scale fisheries (<12 m LOA), hot spots for European hake and red mullet are located in the Aegean, central Ionian, eastern Adriatic, western Sardinia, Tunisia, Libyan and Levantine Sea (Fig. 49).

Aggregated cold spots of bottom trawl fisheries for the six target species were found mainly in the northern Adriatic Sea and Gulf of Lion (Fig. 50).

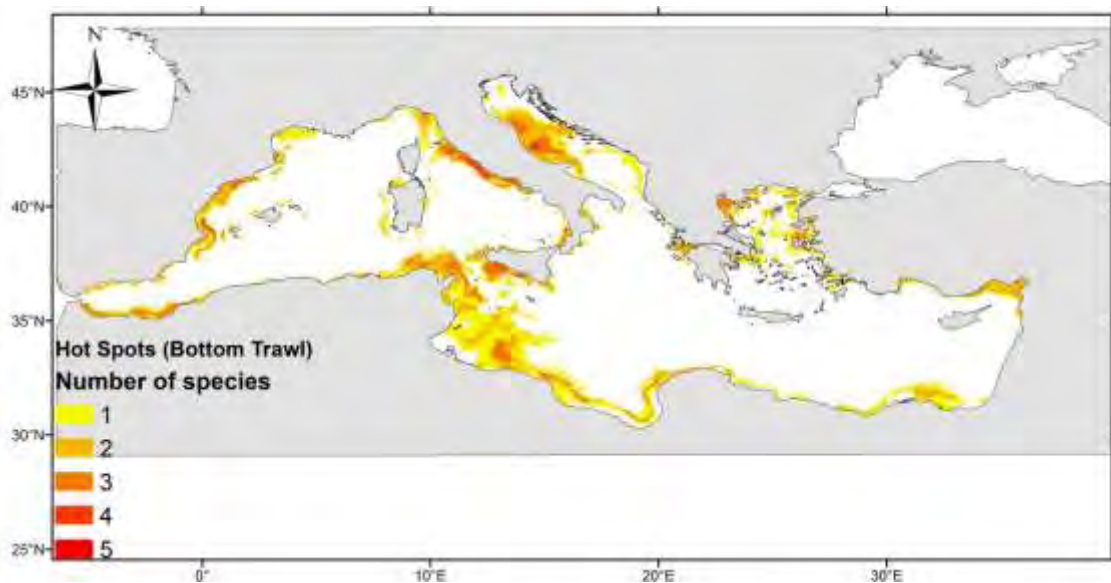


Fig 48. Bottom trawler fishing grounds - aggregated hot spots for the six species under investigation. The map indicates the number of species for which a specific cell has been found to be a hot spot.

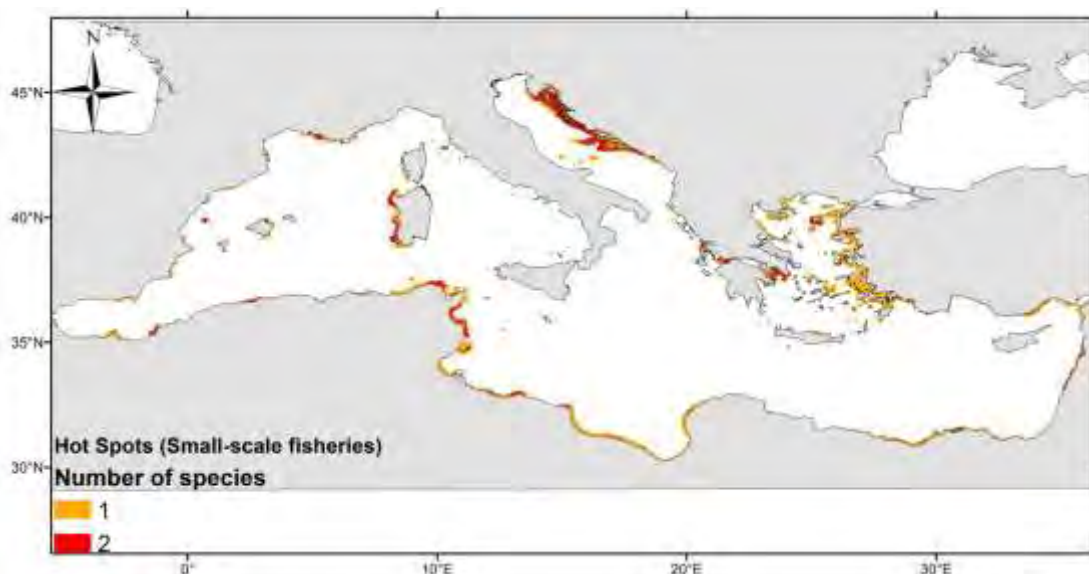


Fig. 49. Small-scale fisheries (< 12 m LOA) fishing grounds - aggregated hot spots for Hake and Red mullet. The map indicates the number of species for which a specific cell has been found to be a hot spot.

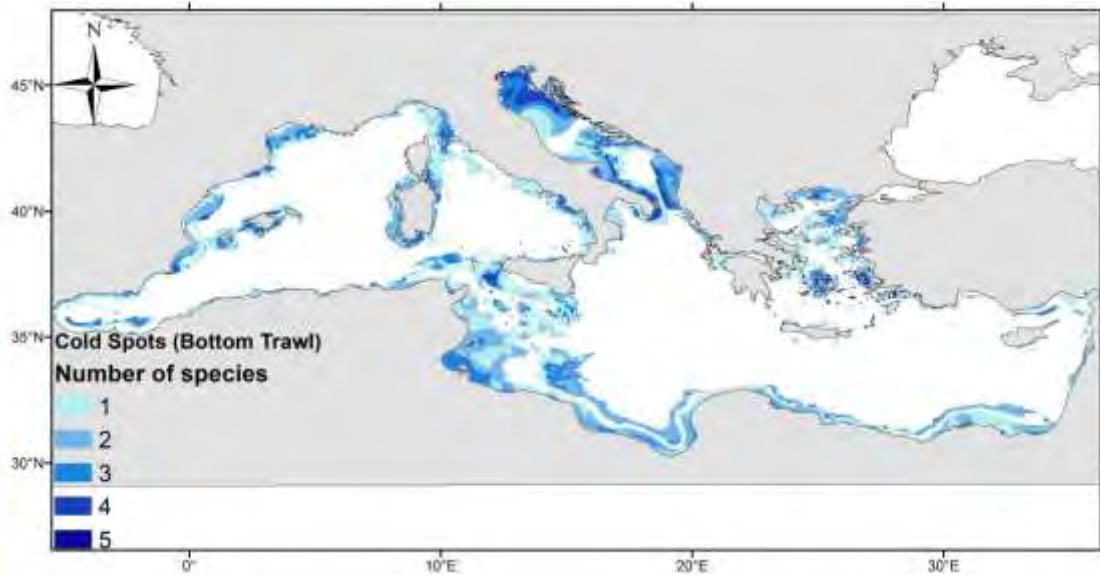


Fig. 50. Bottom trawler fishing grounds - aggregated cold spots for the six species under investigation. The map indicates the number of species for which a specific cell has been found to be a cold spot.

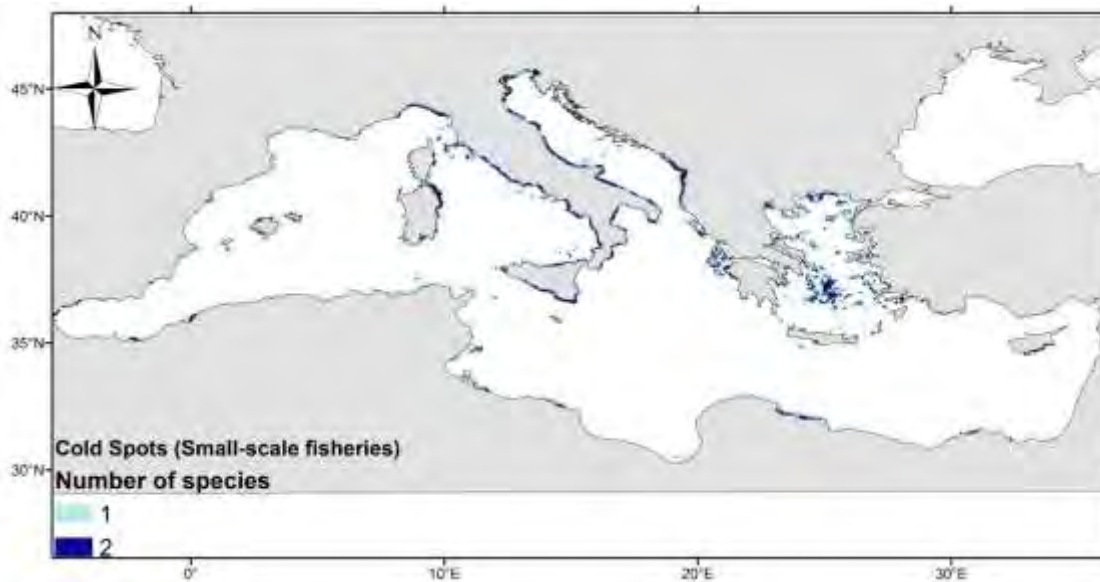


Fig. 51. Small-scale fisheries (< 12 m LOA) fishing grounds - aggregated cold spots for Hake and Red mullet. The map indicates the number of species for which a specific cell has been found to be a cold spot.

Discussion-Conclusions

The results presented in this Deliverable can be useful to ecosystem-based marine spatial management, an approach that recognizes the full array of interactions within an ecosystem, including human uses, rather than considering single issues, species, or ecosystem services in isolation. In that context, the mapping of fishing activities and the identification of fishing grounds is of utmost importance. It is vital to define areas that are important for fisheries in marine spatial planning and to estimate the cost of restricting fishing activities in specific areas.

The advantage of the approach followed in this Deliverable is the visualization of the spatial pattern of species richness in important fishing grounds covering the Mediterranean Sea. In particular, aggregated hot spots, for the six species studied, showed that the most important fishing grounds in terms of species biomass and fishing effort (fishing days) are found in the Adriatic, Aegean, eastern Ionian, Tyrrhenian, Strait of Sicily, Balearic Alboran, Libyan and Levantine Sea.

It should be also noted, that the information used for estimating the species distribution for *Merluccius merluccius*, *Nephrops norvegicus* and *Parapenaeus longirostris* was based on the most recent studies of Sbrana et al., (2019) and Sion et al., (2019). However, for the other species under investigation i.e. *Mullus barbatus*, *Aristaeomorpha foliacea* and *Aristeus antennatus* the spatial data used were based on image re-processing from previous projects (i.e. STOCKMED and validated through spatial visualization based on information from MEDISEH project) combined with information on occurrence based on the analysis particularly performed for the Greek Seas (Maina et al., 2016). Moreover, the analysis was also performed, under the assumption that the spatial and temporal patterns were not significantly changed among areas and years under investigation. Further scientific efforts on the analysis of the most recent data particularly for the *Mullus barbatus*, *Aristaeomorpha foliacea* and *Aristeus antennatus* will be undoubtedly very useful.

Presenting the fishing grounds outcomes (Deliverable 3.1. and this Deliverable) and the results based on MEDISEH project (where the probability of nursery and spawning grounds occurrence for the target species were investigated), would be also important for stock assessment modelling as well as for management advice on spatial closure of fishing activities.

Furthermore, at first sight, cold spots are considered unimportant fishing grounds for bottom trawlers and particularly for the six species studied, since these areas are characterized by relatively low fishing effort and species richness. However, this information is also essential and should be included as an input in a management process in several ways. Cold spots (aggregated or by species) can provide crucial information about limitations, non-preference and target species, and allow for a better understanding of fisheries dynamics (Murawski et al., 2005). Therefore, they can increase the knowledge on the possible effects of conservation measures, such as Marine Protected Areas (Stelzenmüller et al., 2008). Moreover, apart from the six

species studied, there are also other species that are driving the fishing effort patterns in the areas considered as cold spots that needs to be considered in future investigations and management processes.

In addition, the analysis of spatial patterns (including hot and cold spot estimations) was applied separately for western, central and eastern Mediterranean. In this context, the hot and cold spots presented here are indicating a relatively general scale of effort richness in the Mediterranean Sea. Further estimations at local level might be also useful to be considered in a management process of a particular area of interest.

Although it was possible to produce an estimate of the fishing footprint of fleets not equipped with VMS/AIS and especially for the northern African coasts, it is worth noting that the mapping of stocks was based to the data collected only by the EU member states. Given that the amount of data and the robustness of the whole set of analyses for the African coasts is generally limited, the estimation of fishing grounds for these areas was based on the existing information from EU Mediterranean countries and verified based on the existing literature. The merit of the methods applied herein, is that provide flexible options for handling differences in data availability, as is the case in the Mediterranean Sea. However, it is important to highlight the need of further investigating and validating these outcomes in future studies, at least based on some sets of observed data. Hence, the outcomes presented in this Deliverable could be used as a baseline on which future scientific efforts on the spatial identification of fishing grounds by species (or group of species) could be based.

References

Abdel Razek F.A. El-Sherief, S.S., Taha S.M., Muhamad, E.G. 2006. Some biological studies of *Parapenaeus Longirostris* (Lucas, 1846) (Crustacea, Decapoda) in the Mediterranean coast of Egypt. *Egyptian Journal of Aquatic Research* 1687-4285 vol. 32 no. 1, 2006: 385-400.

Al-Absawy, M. A. E. G. (2010). The reproductive biology and the histological and ultrastructural characteristics in ovaries of the female gadidae fish *Merluccius merluccius* from the Egyptian Mediterranean water. *African Journal of Biotechnology*, 9(17), 2544-2559.

Awadh, H., Aksissou, M. 2020, Population dynamics and stock assessment of deep water rose shrimp *Parapenaeus longirostris* (Lucas, 1846) from the Mediterranean Coast of Morocco. *Indian J. Fish.*, 67(2): 8-14, 2020 8 DOI: 10.21077/ijf.2019.67.2.93423-02

Bartolino, V., A. Ottavi, F. Colloca, G. D. Ardizzone, and G. Stefánsson. 2008, Bathymetric preferences of juvenile European hake (*Merluccius merluccius*). *ICES J. Mar. Sci.*, 65: 963-969. doi: 10.1093/icesjms/fsn079.

Bayhan, K.Y., Cartes, J.E., Fanelli E. 2015. Biological condition and trophic ecology of the deep-water shrimp *Aristaeomorpha foliacea* in the Levantine Sea (SW Turkey). *Mediterranean Marine Science* DOI: <http://dx.doi.org/10.12681/mms.867>

Belhoucine, F., Alioua, A., Bouhadiba, S., Boutiba, Z., 2014. Impact of some biotics and abiotics factors on the accumulation of heavy metals by a biological model *Merluccius merluccius* in the bay of oran in Algeria. *Journal of Biodiversity and Environmental Sciences* 2222-3045 Vol. 5, No. 6, p. 33-44

Boubekri, I., Caveen, A., Djebar, A., Amara, R., & Mazurek, H. (2018). Structure and spatio-temporal dynamics of the artisanal small-scale fisheries at the future MPA of “Taza” (Algerian coast, SW Mediterranean). *Mediterranean Marine Science*, 19(3), 555-571. doi:<https://doi.org/10.12681/mms.16192>

Cau A., Carbonell A., Follesa M.C., et al. 2002. MEDITS-based information on the deep-water red shrimps *Aristaeomorpha foliacea* and *Aristeus antennatus* (Crustacea: Decapoda: Aristeidae). *Sci. Mar.* 66(Suppl. 2): 103-124. <https://doi.org/10.3989/scimar.2002.66s2103>

Ferrà C, Tasseti AN, Grati F, et al. 2018. Mapping change in bottom trawling activity in the Mediterranean Sea through AIS data. *Marine Policy* 94: 275–81.

Galdelli A, Mancini A, Tasseti AN, et al. 2019. A Cloud Computing Architecture to Map Trawling Activities Using Positioning Data. In: Volume 9: 15th IEEE/ASME

International Conference on Mechatronic and Embedded Systems and Applications. Anaheim, California, USA: American Society of Mechanical Engineers.

Garofalo, G., Bel Hassen, M., Jarboui, O., Zgozi, S., Gristina, M., Fiorentino, F., Ragonese, S., Camilleri, M. 2008. Preliminary results on spatial distribution of abundance indices, nursery and spawning areas of *Merluccius merluccius* and *Mullus barbatus* in the central Mediterranean. GCP/RER/010/ITA/MSM-TD 19: 24 pp

Gucu A.C., Bingel, F. 2011. Hake, *Merluccius merluccius* L., in the northeastern Mediterranean Sea: a case of disappearance. *J. Appl. Ichthyol.* 27 , 1001–1012

Guijarro B., Bitetto I., D'Onghia G., Follesa M.C., Kapiris K., Mannini A., Marković O., Micallef R., Ragonese S., Skarvelis K., Cau A. 2019. Spatial and temporal patterns in the Mediterranean populations of *Aristaeomorpha foliacea* and *Aristeus antennatus* (Crustacea: Decapoda: Aristeidae) based on the MEDITS surveys. *Sci. Mar.* 83S1: 57-70. <https://doi.org/10.3989/scimar.05012.04A>

Hakkı D., Mustafa E., 2011. Spawning period and first maturity size of deep water rose shrimp (*Parapenaeus longirostris*) in the Aegean Sea. *African Journal of Biotechnology* Vol. 10(68), pp. 15407-15415, Available online at <http://www.academicjournals.org/AJB> DOI: 10.5897/AJB11.1113

Hidalgo, M., Ligas, A., Bellido, J., Bitetto, I., Carbonara, P., Carlucci, R., Guijarro, B., Jadaud, A., Lembo, G., Manfredi, C., Esteban, A., Garofalo, G., Ikica, Z., García, C., Gil de Sola, L., Kavadas, S., Maina, I., Sion, L., Vittori, S., & Vrgoc, N. 2019. Size-dependent survival of European hake juveniles in the Mediterranean Sea. *Scientia Marina*, 0. doi:<http://dx.doi.org/10.3989/scimar.04857.16A>

Kavadas, S., Maina, I., Damalas, D., Dokos, I., Pantazi, M., Vassilopoulou, V., 2015. Multi-criteria decision analysis as a tool to extract fishing footprints: Application to small scale fisheries and implications for management in the context of the maritime spatial planning directive. *Mediterranean Marine Science*, 16 (2), 294–304. <http://dx.doi.org/10.12681/mms.1087>

Khoufi, W. et al., 2014. Reproductive traits and seasonal variability of *Merluccius merluccius* from the Tunisian coast. *Journal of the Marine Biological Association of the United Kingdom*, 94(7), 1545–1556.

Knittweis, L., Arneri, E., Ben Meriem, S., Dimech, M., Fiorentino, F., Gancitano, V., Jarboui, O., Mbarek, K.B., Ceriola, L. 2013. Stock status and potential yield of deep water rose shrimp (*Parapenaeus longirostris*, Lucas 1846) in the south-central Mediterranean Sea. GCP/RER/010/ITA/MSM-TD-28. MedSudMed Technical Documents. No 28: 15 pp

Ligas A., Colloca F., Lundy M.G., Mannini A., Sartor P., Sbrana M., Voliani A., Belcari P. 2015, Modeling the growth of recruits of European hake (*Merluccius*

merluccius) in northwestern Mediterranean Sea with generalized additive models. Fishery Bulletin, 113: 69-81.

Mahmoud, H.H., Fahim, R.M., Srour, T.M., El-Bermawi, N., Ibrahim. M.A. 2017. Feeding ecology of *Mullus barbatus* and *Mullus surmuletus* off the Egyptian mediterranean coast. International Journal of Fisheries and Aquatic Studies 2017; 5(6): 321-325

Maina, I., Kavadas, S., Katsanevakis, S., Somarakis, S., Tserpes, G., Georgakarakos S. 2016. A methodological approach to identify fishing grounds: A case study on Greek trawlers. Fisheries Research. 183, 326-339, <http://dx.doi.org/10.1016/j.fishres.2016.06.021>

Masnadi F., Criscoli A., Lanteri L., Mannini A., Osio G.C., Sartor P., Sbrana M., Ligas A. 2018, Effects of environmental and anthropogenic drivers on the spatial distribution of deep-sea shrimps in the Ligurian and Tyrrhenian Seas (NW Mediterranean). Hydrobiologia, 816: 165-178.

Mérigot, B., Gaertner, J.C., Brind'Amour, A., Carbonara, P., Esteban, A., Garcia-Ruiz, C., Gristina, M., Imzilen, T., Jadaud, A., Joksimovic, A., Kavadas, S., Kolutari, J., Maina, I., Maiorano, P., Manfredi, C., Mifsud, R., Ordines, F., Peristeraki, P., Relini, G., Sbrana, M.,

Milisenda, G., Vitale, S., Massi, D., Enea, M., Gancitano, V., Giusto, G., Badalucco, C., Gristina, M., Garofalo, G., Fiorentino, F. (2017). Discard composition associated with the deep water rose shrimp fisheries (*Parapenaeus longirostris*, Lucas 1846) in the south-central Mediterranean Sea. Mediterranean Marine Science, 18(1), 53-63. doi:<https://doi.org/10.12681/mms.1787>

Murawski, S.A., Wigley, S.E., Fogarty, M.J., Rago, P.J., Mountain, D.G., 2005. Effort distribution and catch patterns adjacent to temperate MPAs. ICES J. Mar. Sci. 62(2), 1150–1167, <http://dx.doi.org/10.1016/j.icesjms.2005.04.005>.

Nouar, A., Kennouche, H., Ainouche N., Cartes. J.E. 2016. Temporal changes in the diet of deep-water Penaeoidean shrimp (*Parapenaeus longirostris* and *Aristeus antennatus*) off Algeria (southwestern Mediterranean). Scientia Marina 75(2). doi: 10.3989/scimar.2011.75n2279

Orsi Relini L., Mannini A., Relini G. 2013. Updating knowledge on growth, population dynamics and ecology of the blue and red shrimp, *Aristeus antennatus* (Risso, 1816), on the basis of the study of its instars. Mar. Ecol. 34: 90-102. <https://doi.org/10.1111/j.1439-0485.2012.00528.x>

Orsi Relini, L., Papaconstantinou, C., Jukic-Peladic, S., Souplet, A., Gil de Sola, L., Piccinetti, C., Kavadas, S., & Rossi, M. (2002). Distribution of the Mediterranean hake populations (*Merluccius merluccius smiridus* Rafinesque, 1810) (Osteichthyes: Gadiformes) based on six years monitoring by trawl-surveys: some implications for

management. *Scientia Marina*, 66(S2), 21-38.
doi:<http://dx.doi.org/10.3989/scimar.2002.66s221>

Quattrocchi, G., Sinerchia, M., Colloca, F., Fiorentino, F., Garofalo G., Cucco, A., 2019. Hydrodynamic controls on connectivity of the high commercial value shrimp *Parapenaeus longirostris* (Lucas, 1846) in the Mediterranean Sea. *Scientific Reports*, 9:16935 <https://doi.org/10.1038/s41598-019-53245-8>

Rawag A.A., Haddoud D.A., Zgozi S.W. (2004). Commercial demersal marine species of Libya. GCP/RER/010/ITA/MSM-TD-02. MedSudMed Technical Document, 2: 75-81.

Russo T, D'Andrea L, Franceschini S, et al. 2019a. Simulating the effects of alternative management measures of trawl fisheries in the Central Mediterranean Sea: application of a multi-species bio-economic modelling approach. *Front Mar Sci*.

Russo T, D'Andrea L, Parisi A, et al. 2016. Assessing the fishing footprint using data integrated from different tracking devices: Issues and opportunities. *Ecological Indicators* 69: 818–27.

Russo T, D'Andrea L, Parisi A, and Cataudella S. 2014. VMSbase: An R-Package for VMS and Logbook Data Management and Analysis in Fisheries Ecology (M Castonguay, Ed). *PLoS ONE* 9: e100195.

Russo T, Morello EB, Parisi A, et al. 2018. A model combining landings and VMS data to estimate landings by fishing ground and harbor. *Fisheries Research* 199: 218–30.

Russo T, Parisi A, and Cataudella S. 2011a. New insights in interpolating fishing tracks from VMS data for different métiers. *Fisheries Research* 108: 184–94.

Sbrana M., Zupa W., Ligas A., Capezzuto F., Chatzistryrou A., Follesa M.C., Gancitano V., Guijarro B., Isajlovic I., Jadaud A., Markovic O., Micallef R., Peristeraki P., Piccinetti C., Thasitis I., Carbonara P. 2019. Spatiotemporal abundance pattern of deep-water rose shrimp, *Parapenaeus longirostris*, and Norway lobster, *Nephrops norvegicus*, in European Mediterranean waters. *Sci. Mar.* 83S1: 71-80.
<https://doi.org/10.3989/scimar.04858.27A>

Sion L., Zupa W., Calculli C., Garofalo G., Hidalgo M., Jadaud A., Lefkaditou E., Ligas A., Peristeraki P., Bitetto I., Capezzuto F., Carlucci R., Esteban A., Follesa C., Guijarro B., Ikica Z., Isajlovic I., Lembo G., Manfredi C., Pérez J.L., Porcu C., Thasitis I., Tserpes G., Carbonara P. 2019. Spatial distribution pattern of European hake, *Merluccius merluccius* (Pisces: Merlucciidae), in the Mediterranean Sea. *Sci. Mar.* 83S1: 21-32. <https://doi.org/10.3989/scimar.04988.12A>

Stelzenmüller, V., Maynou, F., Bernard, G., Cadiou, G., Camilleri, M., Crec'hriou, R., Criquet, G., Dimech, M., Esparza, O., Higgins, R., Lenfant, P., Pérez-Ruzafa, Á.,

2008. Spatial assessment of fishing effort around European marine reserves: Implications for successful fisheries management. *Mar. Pollut. Bull.* 56 (12), 2018–2026. doi:10.1016/j.marpolbul.2008.08.006.

Talet, L.B., Talet, A.B., Boutiba, Z., 2016. Population dynamic parameters of the red mullet *Mullus barbatus* (Mullidae) in the Arzew. Gulf, Algeria. *Int. J. Aquat. Biol.* 4(1): 1-10

Tassetti AN, Ferrà C, and Fabi G. 2019. Rating the effectiveness of fishery-regulated areas with AIS data. *Ocean & Coastal Management* 175: 90–7.

Turan, C., Salihoğlu, B., Özgür Özbek, E., Öztürk, B. (Eds.) (2016). *The Turkish Part of the Mediterranean Sea; Marine Biodiversity, Fisheries, Conservation and Governance*. Turkish Marine Research Foundation (TUDAV), Publication No: 43, Istanbul, TURKEY.

UNEP-MAP-RAC/SPA. 2014. Status and conservation of fisheries in the Sicily Channel/ Tunisian Plateau. By H. Farrugio & Alen Soldo. Draft internal report for the purposes of the Mediterranean Regional Workshop to Facilitate the Description of Ecologically or Biologically Significant Marine Areas, Malaga, Spain, 7-11 April 2014.

Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-
Project MED_UNITS

*“Study on Advancing fisheries assessment and management
advice in the Mediterranean by aligning biological and
management units of priority species”*

WP3- DELINEATE FISHING GROUNDS AND STOCK ASSESSMENT

*Task 3.3 Updated stock assessments, retrospective comparisons,
bias estimations and implications for fisheries management
advice*

Deliverable 3.3 – Report on the stock assessments with the new stock configurations for the 6 target species of the study

Responsible: Giuseppe Scarcella (CNR) and Beatriz Guijarro (IEO)

Compiled by: Alessandro Ligas, Claudia Musumeci and Mario Sbrana (CIBM);
Isabella Bitetto and Andrea Pierucci (COISPA); Paola Pesci (CONISMA); Silvia
Angelini and Enrico Nicola Armelloni (CNR); George Tserpes and Danai
Mantopoulou-Palouka (HCMR)

July 2021 - Version: 2.0..

Contents

1. Introduction	6
2. Objectives	7
3. Materials and methods	8
4. Results	10
4.1 European hake in GSAs 1-12.	10
4.1.1 Input data	11
4.1.2 Model description.....	14
4.1.3 Outputs from the model.....	15
4.1.4 Comparison with other assessments with a different stock configuration.....	21
4.2 European hake in GSAs 22-27.	26
4.2.1 Input data	26
4.2.2 Model description.....	28
4.2.3 Outputs from the model.....	28
4.2.4 Comparison with other assessments with a different stock configuration.....	31
4.3 Red mullet in GSAs 1, 5-16.	33
4.3.1 Input data	33
4.3.2 Model description.....	37
4.3.3 Outputs from the model.....	38
4.3.4 Comparison with other assessments with a different stock configuration.....	45
4.4 Red mullet in GSAs 17-20.....	51
4.4.1 Input data	51
4.4.2 Model description.....	56
4.4.3 Outputs from the model.....	57
4.4.4 Comparison with other assessments with a different stock configuration.....	63
4.5 Red mullet in GSAs 22-25	68
4.5.1 Input data	68
4.5.2 Model description.....	71
4.5.3 Outputs from the model.....	71
4.5.4 Comparison with other assessments with a different stock configuration.....	76
4.6 Deep-water rose shrimp in GSAs 1, 5-16	77
4.6.1 Input data	77
4.6.2 Model description.....	80
4.6.3 Outputs from the model.....	80
4.6.4 Comparison with other assessments with a different stock configuration.....	86

4.7	Deep-water rose shrimp in GSAs 17-20	87
4.7.1	Input data	87
4.7.2	Model description.....	89
4.7.3	Outputs from the model.....	90
4.7.4	Comparison with other assessment with different stock configuration.....	98
4.8.	Norway lobster in GSAs 1, 2, 5-11	101
4.8.1	Input data	101
4.8.2	Model description.....	104
4.8.3	Outputs from the model.....	104
4.8.4	Comparison with other assessments with a different stock configuration.....	111
4.9.	Norway lobster in GSAs 17-19	116
4.9.1	Input data	116
4.9.2	Model description.....	119
4.9.3	Outputs from the model.....	120
4.9.4	Comparison with other assessments with a different stock configuration.....	123
4.10	Giant red shrimp in GSAs 5, 8-12	126
4.10.1	Input data	126
4.10.2	Model description.....	128
4.10.3	Outputs from the model.....	129
4.10.4	Comparison with other assessments with different stock configuration.....	135
4.11	Giant red shrimp in GSAs 15, 16, 18-20	136
4.11.1	Input data	136
4.11.2	Model description.....	138
4.11.3	Outputs from the model.....	139
4.11.4	Comparison with other assessments with different stock configuration.....	144
4.12	Blue and red shrimp in GSAs 1, 5, 6, 9-12	146
4.12.1	Input data	146
4.12.2	Model description.....	149
4.12.3	Outputs from the model.....	150
4.12.4	Comparison with other assessments with a different stock configuration.....	156
4.13	Blue red shrimp in GSAs 15, 16, 18-20	159
4.13.1	Input data	159
4.13.2	Model description.....	162
4.13.3	Outputs from the model.....	163
4.13.4	Comparison with other assessments with a different stock configuration.....	168
5.	Discussion and conclusion	169
6.	Deviation from the Workplan	172

7. Remedial actions	173
8. References.....	174

Executive summary

The present document summarizes and evaluates the new stock configurations explored applying the stock assessment methodologies routinely used for the six target stocks at single GSA or combination of GSAs. The outcomes of the new analyses are summarized in Table 1.

Table 1: Summary of the stock assessments' outputs.

N.	Species	GSAs combination	Method	Status according to Med_Units configuration
1	<i>M. merluccius</i>	GSAs 1 - 12	Age based (a4a)	In overfishing
2		GSAs 22 - 27	Production model (SPiCt)	In overfishing
3	<i>M. barbatus</i>	GSAs 1, 5 - 16	Age based (a4a)	In overfishing
4		GSAs 17-20	Age based (a4a)	Sustainably exploited
5		GSAs 23-25	Age based (a4a)	In overfishing
6	<i>P. longirostris</i>	GSAs 1, 5-16	Age based (a4a)	In overfishing
7		GSAs 17-20a	Age based (a4a)	In overfishing
8	<i>N. norvegicus</i>	GSAs 1, 5, 6, 9 – 11	Age based (a4a)	In overfishing
9		GSAs 17-19	Production model (SPiCt)	In overfishing
10	<i>A. foliacea</i>	GSAs 5, 8-12	Age based (a4a)	In overfishing
11		GSAs 13, 16, 18-20	Age based (a4a)	Sustainable exploited
12	<i>A. antennatus</i>	GSAs 1, 5, 6, 9-12	Age based (a4a)	In overfishing
13		GSAs 15, 16, 18-20	Age based (a4a)	In overfishing

1. Introduction

Delineating the stock structure of harvested resources is a major pre-requisite for the understanding of their population dynamics, conducting reliable stock assessment, and providing robust scientific advice for an effective management.

Nonetheless, stock structure uncertainty is pervasive in fisheries science. This problem is partly due to the intrinsically multifaceted concept of a stock as an ecological unit, and because the definition of stock can vary according to the objective of scientific research (Cadrin et al. 2005; Hammer and Zimmermann, 2005).

From a fisheries management perspective, these ecological units should be treated separately and are to be maintained at levels in which yield is maximized and sustainable use of the resource is promoted (Cope and Punt, 2009). Knowledge of ecological stock structure of a species is considered essential for effective management. The reason for such importance is related to the fact that most commonly used stock assessment methods rely on the assumption that stocks are discrete and self-recruiting units with homogeneous life history parameters (Begg et al., 1999; Begg and Waldman, 1999). However, these assumptions are frequently violated either because of insufficient information to delineate discrete stock units or because of the existence of greater stock complexity (e.g. occurrence of mixing between stocks or presence of different subunits within a single stock; Secor, 1999; Kerr et al., 2010; 2016; Ying et al., 2011). Thus, stock assessment uncertainty and bias can originate from two different types of error when defining a stock (Punt, 2003). The first type occurs when multiple stocks are assessed as a single unit. In this scenario, differences in life history parameters and susceptibility to fisheries might lead to a decrease in biocomplexity (Hilborn, 2003); local depletion; or, in the extreme, the loss of genetic diversity (Stephenson, 1999). The second type of error occurs when only a portion of a stock is assessed as a closed unit. Although this error is less commonly addressed in the literature, it can also lead to severe bias in stock assessment estimates (Butterworth and Geromont, 2001; Field et al., 2006).

In the Mediterranean, the lack of knowledge on stock structure can lead to both types of error, depending on the nature of species considered: for instance, species with territorial behavior, such as Norway lobster can be the case of the first type of error, while relatively mobile species, such as European hake can represent the reverse case.

The new stock configurations can provide a scientifically sound approach in the evaluations of the status of the six target species already analyzed under the auspices of STECF and GFCM-SAC, because these new configurations can probably reduce assessment uncertainty and bias origination from the two different types of error defined before.

The methodologies to be used for the new stock assessments are the same employed in 2019-2021 joint benchmarks of the scientific bodies in the Mediterranean (GFCM/SAC and STECF) for selected species and areas.

2. Objectives

The main objective of this task is to perform stock assessments taking into consideration the outputs of the different WPs and the synthesis carried out in WP4 to quantify the effects of stock structure uncertainty on the six target species.

The stock assessments are carried out in six case studies species, based on the updated list of stock units for the examined target species as emanating from the present study, and results are compared with previous assessments.

The number of assessments per species has been tailored in liaison with CINEA and DGMARE and a final number of 13 new assessments had been carried out.

It is important to stress that the new assessments carried out in the framework of the present task are not going to provide any scientific advice in term of status of the stocks compared with relevant reference points. Notwithstanding, the new assessments can represent a useful tool to explore the new stock configurations indicated by WP1 and WP2.

3. Materials and methods

The methodologies for each assessment were selected on a case by case basis depending on the available data and taking also into consideration, when possible, the outputs of task 3.1 regarding the delineation of fishing grounds by gear and fleet and their link with the spatial distribution of the target species (task 3.2).

In each case study, possible sources of bias in previous estimations of stock status is investigated and identified, together with the potential improvements brought by the new management units proposed by the Med_Units.

The results of WP1 and WP2 are summarized in Table 2 in term of GSAs distribution of each stock. Taking into account the data availability and in liaison with CINEA/DGMARE, the assessments that have been carried out and presented in the following sections are summarized in Table 3

Table 2: GSA distribution of target stocks according to Genetic (G) and Otolith (O) outputs.

Sea-Basin	W	W	W	W	W	W	W	W	W	W	W/C	W/C	W/C	C	C	C	C	E	E	E	E	E	
<i>M. merluccius G</i>	1	4	5	6	7	8	9	10	11	12		13		16	17	18	19	20	22	23	25	26	27
<i>M. merluccius O</i>	1	4	5	6	7	8	9	10	11	12		13		16	17	18	19	20	22	23	25	26	27
<i>M. barbatus G</i>	1	5	6	7	8		9	10	11	12		14		17	18	19	20		23	25	26	27	
<i>M. barbatus O</i>	1	5	6	7	8		9	10	11	12		16	17		18	19	20		23	25	26	27	
<i>A. foliacea</i>				5	8	9		10	11	12		13	16		18	19		20	22	23	25		26
<i>A. antennatus</i>	1	5			6	9	10	11	12		13	16	18	19			20	22	23			26	
<i>P. longirostris</i>			5	6	7	8		9	10	11	12	13	16	17	18	19	20a	20b	22	23	25		26
<i>N. norvegicus</i>	1	6	7	8	9	10	11						17	18	19			22					

Table 3: Stock configurations of the assessments carried out and methodologies employed.

N.	Species	GSAs combination	Method
1	<i>M. merluccius</i>	GSAs 1 - 12	Age based (a4a)
2		GSAs 22 - 27	Production model (SPiCt)
3	<i>M. barbatus</i>	GSAs 1, 5 - 16	Age based (a4a)
4		GSAs 17-20	Age based (a4a)
5		GSAs 23-25	Age based (a4a)
6	<i>P. longirostris</i>	GSAs 1, 5-16	Age based (a4a)
7		GSAs 17-20a	Age based (a4a)
8	<i>N. norvegicus</i>	GSAs 1, 5, 6, 9 – 11	Age based (a4a)
9		GSAs 17-19	Production model (SPiCt)
10	<i>A. foliacea</i>	GSAs 5, 8-12	Age based (a4a)
11		GSAs 13, 16, 18-20	Age based (a4a)
12	<i>A. antennatus</i>	GSAs 1, 5, 6, 9-12	Age based (a4a)
13		GSAs 15, 16, 18-20	Age based (a4a)

A set of analysis were implemented, when possible, to evaluate if the stock configurations envisaged in Med_Units and presented in Table 3, improved the fit of data to model assumption. Different aspects of the assessment process were explored through a step procedure:

- Cohort consistency: this step was aimed to evaluate and to compare the fit of data to the age structure assumption, as observed in the catches and in the MEDITS index. Data considered were the age structure (both MEDITS and catches) for the MED_UNITS stock configuration (Table 3) and for all the available assessments conducted on GSAs that were aggregated in the reference assessment

(sub-area stock). Data for the sub-area stock were collected from STECF EWG 20-09 report (STECF, 2020a, 2020b) and from the most recent stock assessment forms provided by the GFCM website¹. Methodology was based on the standard cohort consistency function adopted during STECF EWGs, used to obtain and to plot the Pearson correlation coefficients and scatterplots between age classes. Cohort consistency plots obtained were visually compared to check for improvements or degradations between cohort strength in the reference stock and in each of the sub-area stock. It should be acknowledged that assumptions such as slicing parameters, age classes considered and length of the time-series may have varied among the data used. As a consequence, the drivers of any hypothetical differences in the correlation strength would be difficult to unravel. Nevertheless, this qualitative comparison will serve to stimulate the discussion on the data consistency used in the different assessment. Due to the limitations stated above the present analysis was not carried out in all the cases.

- Stock trajectories: this step was aimed to evaluate the likelihood of the Med_Units stock configuration to be comprehensive of the sub-areas aggregated. Data considered were the SSB, the recruitment and the F_{BAR} trends for the MED_UNITS stock configuration (reference stock) and for all the available assessments conducted on GSAs that were aggregated in the reference assessment (sub-area stock). Prior to calculate the coefficients, input data were trimmed according to the minimum common size in respect to available years. A semi-quantitative analysis was implemented for SSB and Recruitment, the data for each of the available sub-area stock were summed, and plotted along with the results coming from the reference stock. The Pearson correlation coefficient was also calculated. The hypothesis 0 was that the sum of the SSB and Recruitment estimated by the sub-areas assessment should be equal to the reference assessment, as a result the trend should be correlated and values largely comparable. For the F_{BAR} and F/F_{01} trend, the data from the sub-areas assessment were all plotted along with the results coming from the reference stock. The Pearson correlation coefficient was calculated for each pairwise contrast between the sub-areas and the reference assessment. The hypothesis 0 was that all the individual F trends should be comparable and correlated to the reference stock. In addition, the terminal F, F_{01} and F/F_{01} of each sub-area were plotted along with the results from the reference stock, to permit a visual inspection of results coherence. An R script used to perform the analysis described above was made available to the group, in order to ensure replicability of the analysis and a standardized result visualization.
- Model Diagnostics: this step was aimed to compare the model diagnostics between the sub-areas assessment and the Med_Units configuration, Data considered were the standardized residuals and the retrospective patterns for the MED_UNITS stock configuration (reference stock) and for all the available assessments conducted on GSAs that were aggregated in the reference assessment (sub-area stock). Images for the sub-area stock were gathered from the STECF EWGs 2020 reports (STECF, 2020a, 2020b) and collated with the respective figure produced for the reference stock. Again, this qualitative comparison will serve to stimulate the discussion on the model diagnostics obtained in the different assessment taking also into account, as far as possible, the potential influence of other factors as model selection or settings in the assessment.

¹ <http://www.fao.org/gfcm/data/safs/es/>

4. Results

4.1 European hake in GSAs 1-12.

The stock assessment of hake in the Western Mediterranean Sea is here presented. Considering the WP1 and WP2 outputs, the GSAs from 1 to 12 were lumped together (Figure 1).

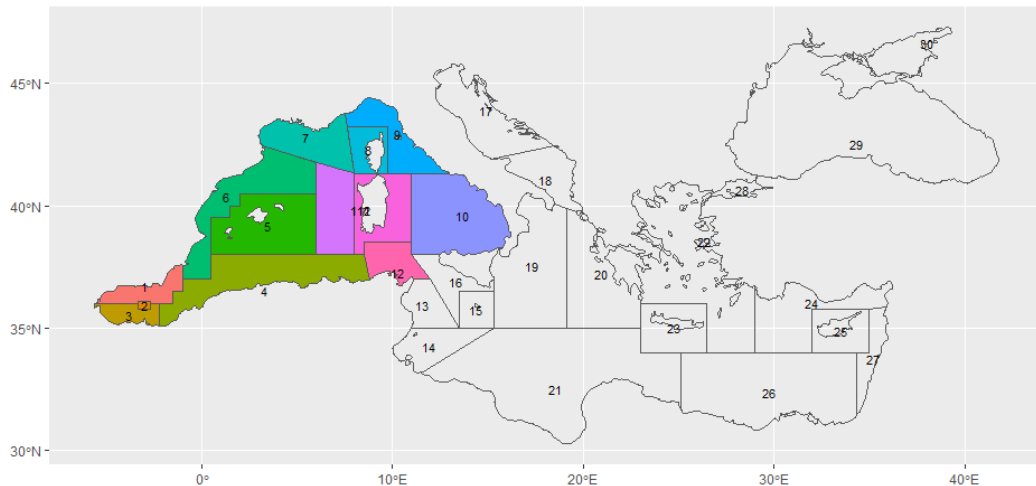


Figure 1: HKE in GSAs from 1 to 12 (colored GSAs): geographical location of the stock

Hake is distributed in the whole area between 10 and 800 m depth (Biagi et al., 2002; Colloca et al., 2003). Recruits peak in abundance between 150 and 250 m depth over the continental shelf-break and appear to move slightly deeper when they reach 10 cm total length. Migration from nurseries takes place when juveniles attained a critical size between 13 and 15.5 cm TL (Bartolino et al., 2008a, 2008b). Maturing hakes (15-35 cm TL) persist on the continental shelf with a preference for water of 70-100 m depth, while larger hakes can be found in a larger depth range from the shelf to the upper slope. Juveniles show a patchy distribution with some main density hot spots (i.e. nurseries areas) showing a high spatio-temporal persistence (Abella et al., 2005; Colloca et al., 2009) as also highlighted by the MEDISEH project in areas with frontal systems and other oceanographic structures that can enhance larval transport and retention (Abella et al., 2008).

European hake is a long lived fish mainly exploited by trawlers, especially on the continental shelves of the Gulfs (e.g. Gaeta, Salerno, Palermo) but also by artisanal fishers using fixed gears (gillnets, bottom long-line).

Trawl-survey data have evidenced highest biomass indices on the continental shelf of the GSA 10 (100-200 m; Spedicato and Lembo, 2011), where juveniles (less than 12 cm total length) are mainly concentrated. During autumn trawl surveys, one of the main recruitment pulses of this species is observed. Two main recruitment events (in spring and autumn; Spedicato and Lembo, 2011) are reported in GSA 10 as for other Mediterranean areas. European hake is considered fully recruited to the bottom trawlers at 10 cm TL (from SAMED, 2002). The length structures from trawl surveys are generally dominated by juveniles, while large size individuals are rare. This pattern might be also due to the different vulnerability of older fish beside the effect of high exploitation rates. The few large European hake caught during trawl surveys are generally females and inhabit deeper waters. The overall sex ratio ($\sim 0.41-0.47$) estimated from trawl survey data is slightly skewed towards males. The size at first maturity for females was recently estimated by Carbonara et al. (2019) at 33 cm, with a maturity range of 2.55 cm, and is in line with previous studies in the area (Recasens et al., 2008).

Spawning is taking place almost all year round, with a peak during winter–spring. Juveniles showed a patchy distribution with some main density hot spots (nurseries) showing a high spatio-temporal

persistence (Murenu et al., 2010) in western areas.

4.1.1 Input data

Fishery-Dependent data

Commercial data (catch data, LFDs and biological parameters) for the EU countries considered in the stock configuration (Spain, France and Italy) were collected from the EU DCR/DCF database. Catch statistics up to the year 2018 for non-EU countries (Morocco, Algeria and Tunisia) were retrieved from FAO-FishstatJ, while LFD and biological parameters for non-EU countries were not available (Table 4 and Table 5).

For GSA 8 (France) only landing data from 2010 to 2019 were available, data for years from 2005 to 2009 were assumed equal to those estimated during STECF 20-09 (STECF, 2020a). Also for GSA 2 (Spain), GSA 5 (France) years 2016 and 2017 and GSA 6 (France) years 2017 and 2018 only landing data were available. For these years and GSAs age data have been added using the SOP correction.

Considering the data available, this assessment was carried out considering years from 2005 to 2019.

Table 4: HKE in GSAs from 1 to 7: Catch data divided by country, GSA and year. Red values correspond to those estimated. Data for Morocco (MAR – GSA 3), Algeria (DZA – GSA 4) come from FAO-GFCM estimates.

Year	ESP - GSA 1	ESP - GSA 2	MAR - GSA 3	DZA - GSA 4	ESP - GSA 5	FRA - GSA 5	ESP - GSA 6	FRA - GSA 6	ESP - GSA 7
2005	378.75	2.17	157.00	12.00	109.91		3473.54		223.94
2006	384.70	1.46	547.00	44.00	124.79		3627.19		261.01
2007	340.37	1.17	194.00	193.00	184.94		2539.75		237.45
2008	343.66	2.10	132.00	96.00	128.45		3341.37		293.64
2009	639.95	2.42	197.00	307.00	82.85		3847.51		345.44
2010	591.20	1.44	146.00	288.00	118.58		2822.38		196.44
2011	688.99	1.14	227.00	383.00	91.04		3323.29		137.29
2012	483.62	1.13	233.00	515.00	67.69		2835.69		180.68
2013	384.90	2.26	159.00	670.00	121.76		3106.51		216.26
2014	313.28	1.50	146.00	888.00	123.82		2641.16		226.35
2015	206.87	0.72	463.00	905.00	102.20		1778.24		127.79
2016	200.66	0.66	260.00	674.40	69.61	0.05	1880.92		120.70
2017	320.60	1.88	309.00	620.00	72.84	0.03	1797.19	0.03	95.45
2018	437.46	3.61	318.00	511.00	143.49		2581.71	0.16	88.57
2019	300.53	1.71	295.67	601.80	124.40		1658.24		72.65

Table5: HKE in GSAs from 7 to 12: Catch data divided by country, GSA and year. Red values correspond to those estimated. Data for Tunisia (TUN – GSA 12) come from FAO-GFCM estimates.

Year	FRA - GSA 7	FRA - GSA 8	ITA - GSA 9	ITA - GSA 10	ITA - GSA 11	TUN - GSA 12
2005	1002.58	70.20	1859.98	1484.22	557.41	775.00
2006	1160.22	70.20	2281.69	1570.64	936.55	751.00
2007	1394.43	70.20	1733.03	1268.66	169.58	746.00
2008	2181.63	70.20	1321.13	1122.85	138.77	932.30
2009	2493.47	70.20	2005.74	1190.29	367.41	884.00
2010	2107.52	11.92	1583.52	1397.51	340.67	936.90
2011	1421.88	185.67	1879.53	1333.44	546.10	676.00
2012	1086.10	13.22	1185.75	1225.14	192.78	1046.00
2013	1608.12	4.31	1584.06	1087.82	228.06	1238.00

2014	1756.42	12.87	1550.79	1288.12	69.47	1671.00
2015	1027.84	12.19	1278.74	1073.15	322.88	1017.00
2016	922.84	39.85	1087.37	1080.32	367.48	872.00
2017	777.07	14.60	648.05	845.27	516.52	904.00
2018	825.36	21.15	719.69	544.54	503.80	880.00
2019	1072.23	18.00	921.87	638.24	449.49	885.33

Fishery-Independent data

LFDs (n/km^2) from the MEDITS surveys in GSAs 1, 5, 6, 7, 8, 9, 10 and 11 were sliced using the same set of growth parameters used for the commercial data. No data was available for GSAs 2, 3, 4, and 12.

Biological parameters (Table 6) are assumed equal to those used for the assessment of hake in GSAs 1, 5, 6 and 7, since this area produces the majority of the catch coming from the Western Mediterranean sea. Also natural mortality and proportion of matures (Table 7) are those used in the assessment of hake in GSAs 1, 5, 6 and 7.

Following the approach presented in most of the current assessments, sex combined parameters are used.

Table 6: HKE in GSAs from 1 to 12: VBGF and LW parameters.

Sex	L_{inf}	k	t_0
Combined	110	0.168	-0.005
	a		b
Combined	0.00677		3.0351

Table 7: HKE in GSAs from 1 to 12: natural mortality and maturity-at-age vectors.

Ages	Natural mortality	Proportion of matures (combined)
0	1.63	0.00
1	0.68	0.15
2	0.41	0.82
3	0.31	0.98
4	0.25	1.00
5+	0.22	1.00

The catch at age structure as well as the abundance at age resulting from the slinging are presented respectively in Figure 2 and Figure 3. Moreover the internal consistencies of such input data are presented respectively in Figure 4 and Figure 5.

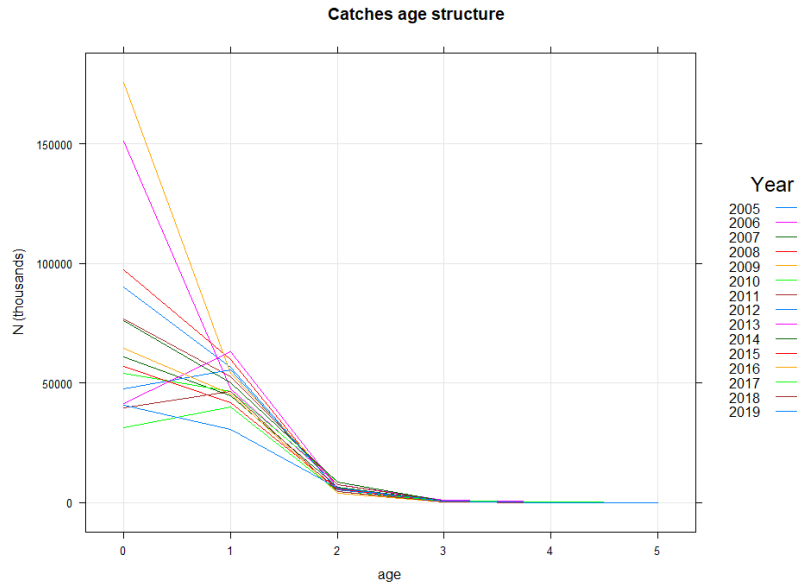


Figure 2: HKE in GSAs from 1 to 12: catch numbers-at-age

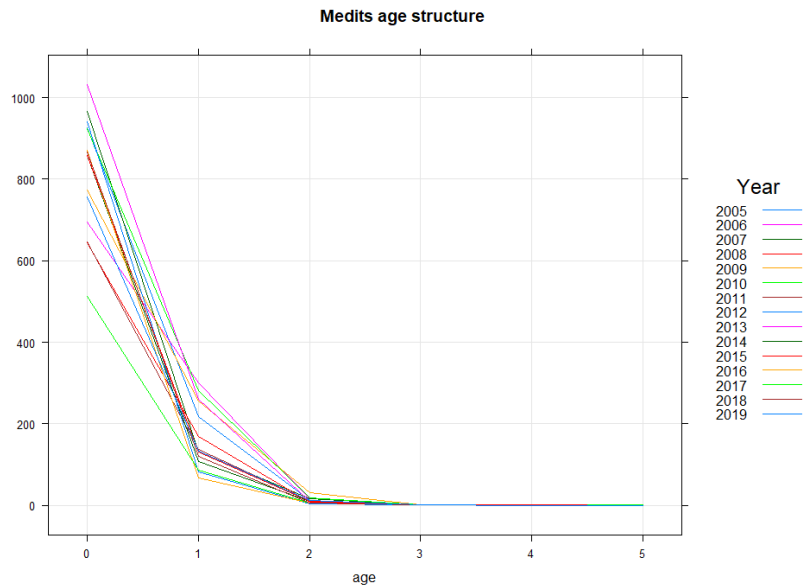


Figure 3: HKE in GSAs from 1 to 12: catch numbers-at-age in the MEDITS.

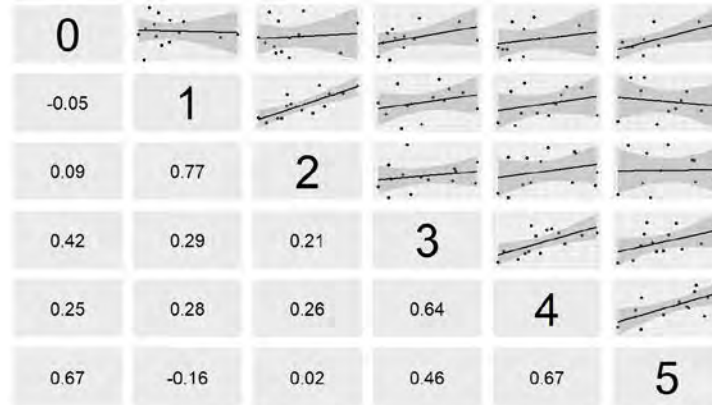


Figure 4: HKE in GSAs from 1 to 12: cohort consistency in the catch numbers-at-age.

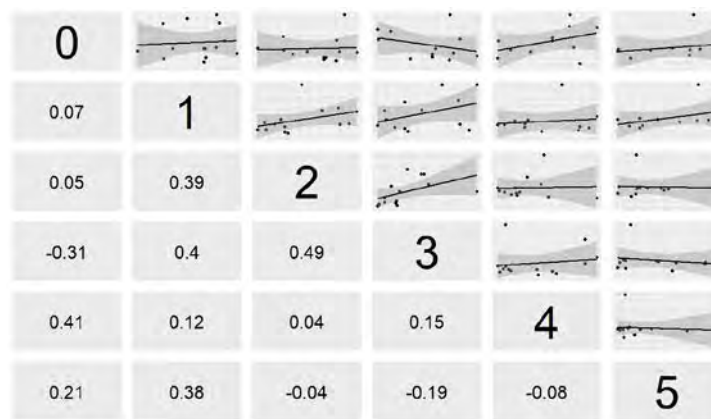


Figure 5: HKE in GSAs from 1 to 12: cohort consistency in the catch numbers-at-age in the MEDITS.

4.1.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment using a4a was carried out using as input data the period 2005-2019 for both the catch data the tuning file (MEDITS indices).

The natural mortality vector was estimated as an average of different methods (Gislason, Prodbiom revised version with unique solution, Chen & Watanabe, Brodziak (2011 and 2012), Lorenz and Gulland), consistently with the approach used in the GFCM benchmark assessment of hake in Adriatic Sea in 2019 (GFCM, 2019). Length-frequency distributions of commercial catches and surveys were transformed in age classes (plus group was set at age 5) using length-to-age slicing.

The number of individuals by age was SOP corrected [$SOP = \text{Landings} / \sum a$ (total catch numbers at age $a \times$ catch weight-at-age a)].

In the catches, a plus group at age 5 was set. A true age 5 was used in the survey.

F_{bar} range was fixed at 1-3, since most individuals belong to these age groups. Survey indices (N/km^2) from MEDITS survey were used.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $f_{\text{model}} = \text{factor}(\text{replace}(\text{age}, \text{age} > 4, 4)) + s(\text{year}, k=8)$

Catchability sub-model: $q_{\text{model}} = \text{list}(\sim \text{factor}(\text{replace}(\text{age}, \text{age} > 4, 4)))$

Model <- sca(hke.stk, FLIndices(hke.idx), fmodel, qmodel)

4.1.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the figures 6-11. The model estimates 39 parameters out of 180 observations; it is then around the threshold of 20% ratio between parameters and observations.

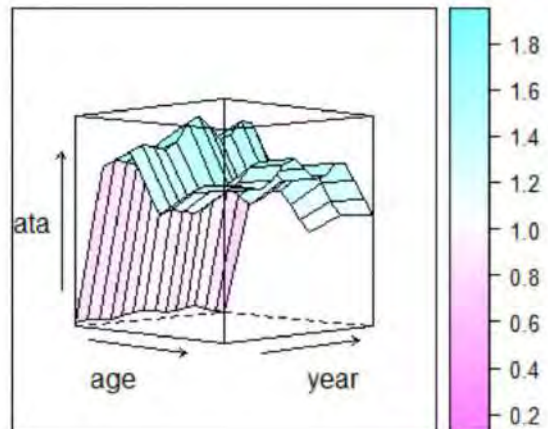


Figure 6: HKE in GSAs from 1 to 12: fishing mortality by age and year obtained from the a4a model (2005-2019).

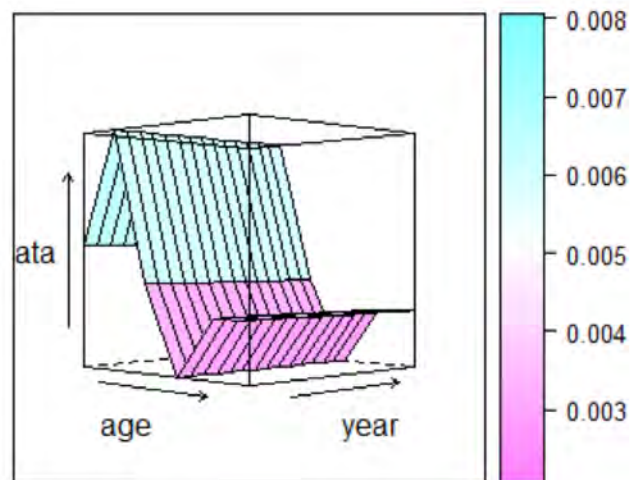


Figure 7: HKE in GSAs from 1 to 12: catchability of the survey by age and year obtained from the a4a model (2003-2019).

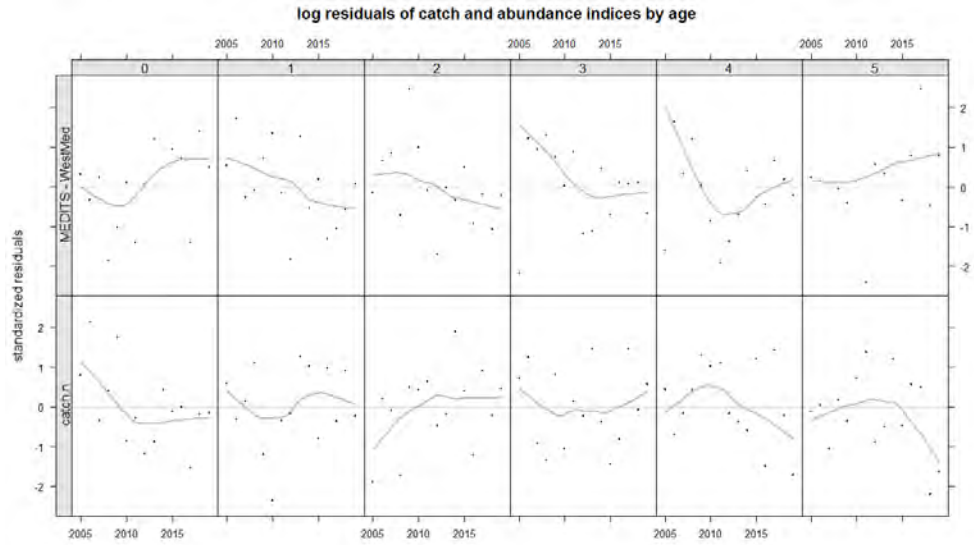


Figure 8: HKE in GSAs from 1 to 12: log residuals for the catch-at-age data of the fishery and the survey.

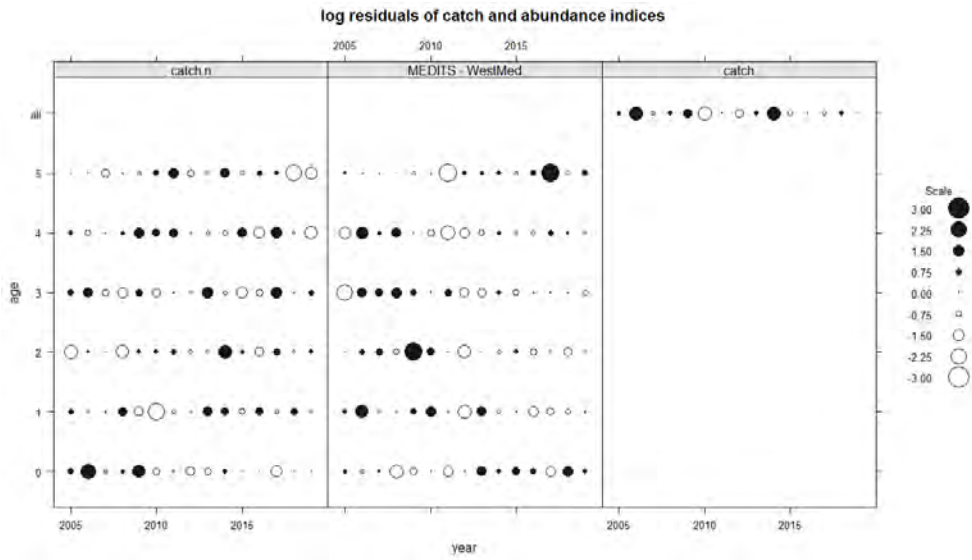


Figure 9: HKE in GSAs from 1 to 12: bubble plot of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

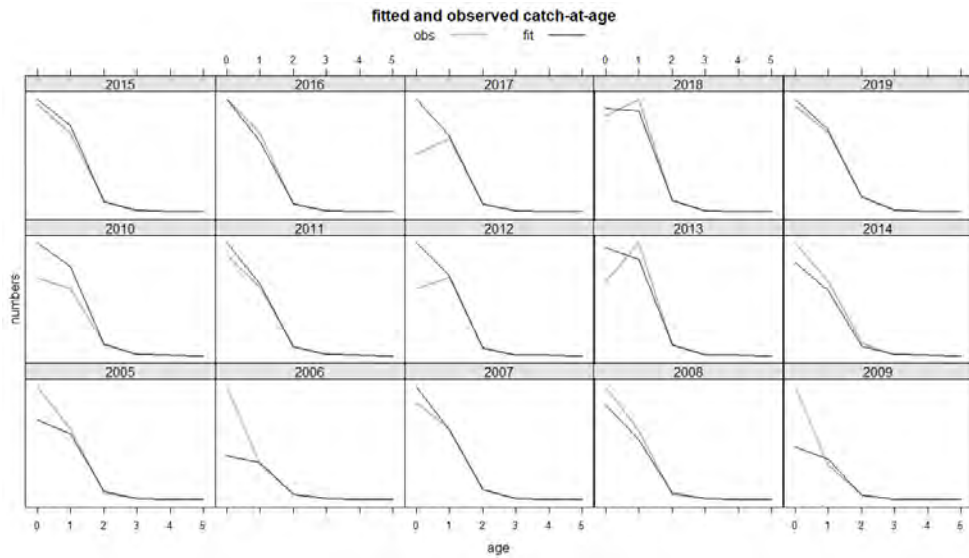


Figure 10: HKE in GSAs from 1 to 12: fitted vs observed values by age and year for the catches.

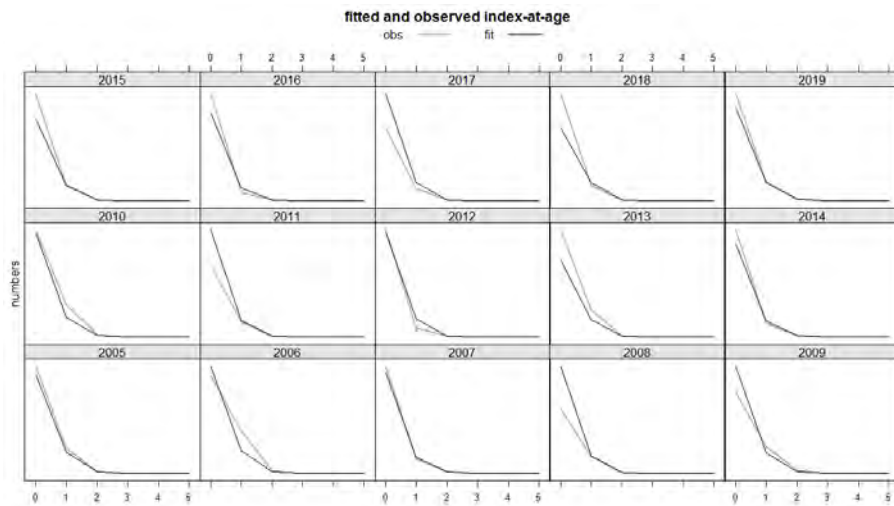


Figure 11: HKE in GSAs from 1 to 12: fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated and showed that biomass of the plus group (age 5+) is about 8% of the total SSB on average, with a peak in 2013 (12%) and the lowest value in 2019 (5%) (Figure 12).

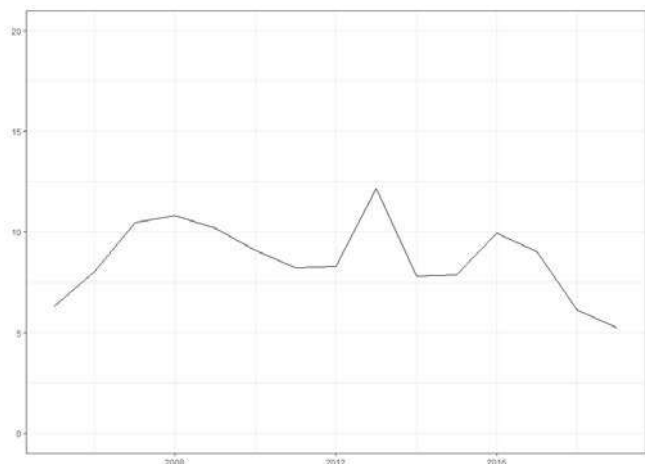


Figure 12: - HKE in GSAs from 1 to 12: cryptic biomass.

Figure 13 shows the final result of this assessment. The retrospective analysis shows a rather stable model (Figure 14). The Mohn's Rho for F and SSB are within the optimal range -0.2/0.2 (-0.001 for F, 0.096 for SSB, and -0.111 for recruitment; Hurtado-Ferro et al., 2014). Figure 15 shows the difference between the estimated and the observed catches: the trend appears reasonable with some differences at the beginning of the time series. Stock numbers-at-age, fishing mortality-at-age and summary results of the model are presented respectively in tables 8, 9 and 10, while figure 16 summaries the histograms of probability for $F_{0.1}$, F_{curr} and level of exploitation ($F_{curr}/F_{0.1}$ ratio) values.

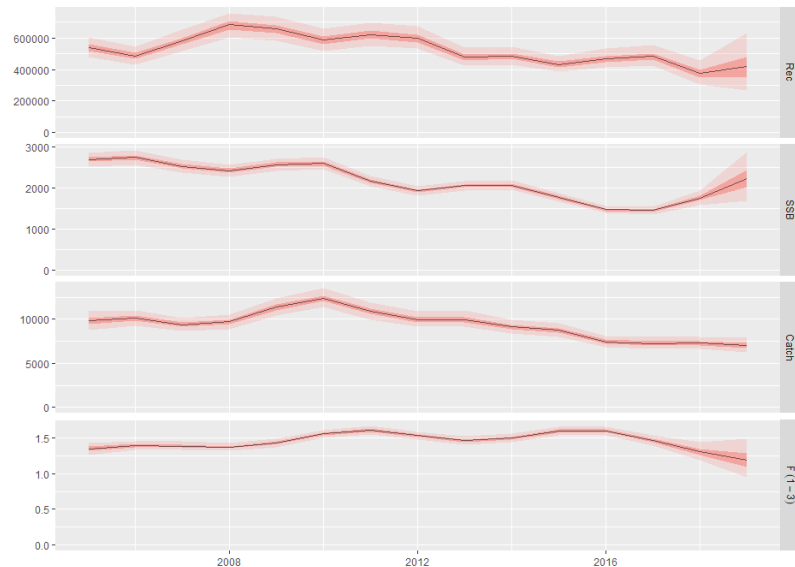


Figure 13 - HKE in GSAs from 1 to 12: output of the a4a model

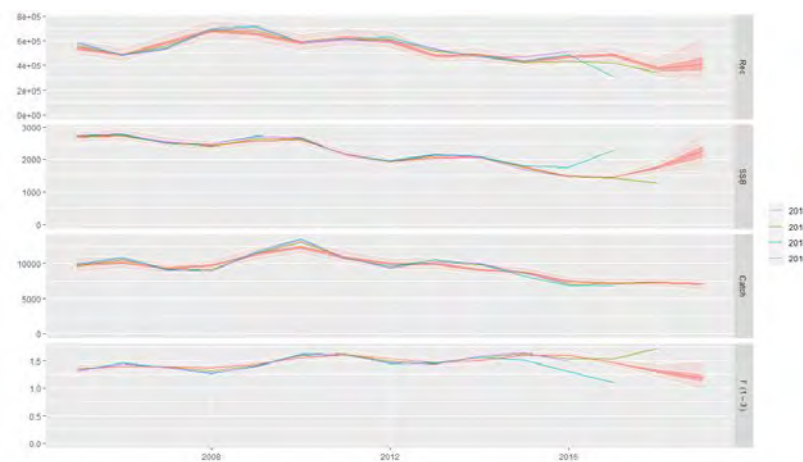


Figure 14: - HKE in GSAs from 1 to 12: retrospective analysis.

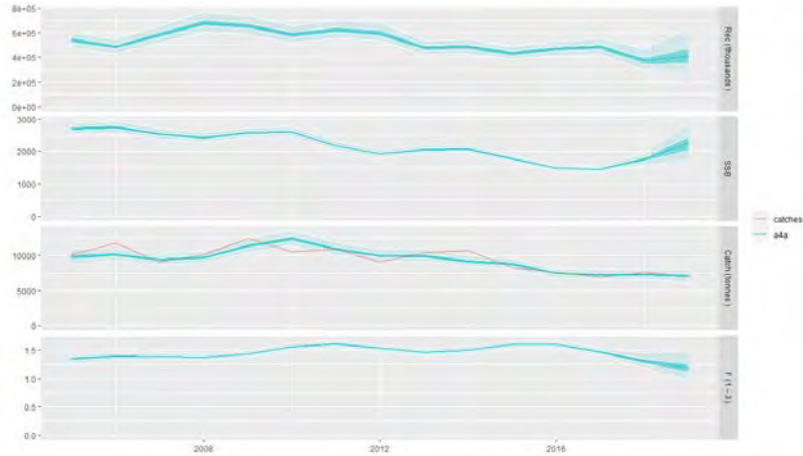


Figure 15: HKE in GSAs from 1 to 12: outputs of the a4a stock assessment model, with uncertainty; input catch data (red line) are plotted against the estimated catches

Table 8: HKE in GSAs from 1 to 12: Stock numbers-at-age (thousands).

Age/Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	539840.7	484775.8	581078.3	679891.5	655365.3	584480.8	621723.4	599086.1	477787.3
1	88258.61	81272.62	72321.2	86819.59	101901.3	97034.21	84467.1	88937.01	86960.41
2	10142.34	10642.25	9325.788	8367.38	10217.22	11220.91	9363.44	7709.911	8790.676
3	1453.206	1447.244	1440.047	1273.106	1163.276	1322.867	1261.288	991.664	889.21
4	284.179	365.513	350.803	351.212	314.46	273.44	281.805	257.776	215.065
5+	69.289	81.159	98.259	99.595	101.455	88.493	68.788	63.446	62.228
Age/Year	2014	2015	2016	2017	2018	2019			
0	484320.4	432972.2	468102.4	482608.6	373138.6	410617.3			
1	70352.81	70769.6	62058.08	67067.47	70995.17	56621.79			
2	9292.247	7209.794	6529.051	5713.449	7129.621	8937.677			
3	1102.131	1114.011	772.361	697.883	712.37	1065.405			
4	204.388	245.553	229.495	158.868	159.794	185.029			
5+	57.468	52.355	54.373	51.744	43.484	48.452			

Table 9: HKE in GSAs from 1 to 12: Fishing mortality-at-age.

Age/Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	0.263	0.273	0.271	0.268	0.280	0.304	0.315	0.300	0.286
1	1.435	1.485	1.477	1.460	1.526	1.658	1.714	1.634	1.556
2	1.537	1.590	1.581	1.563	1.634	1.776	1.835	1.750	1.666
3	1.070	1.107	1.101	1.088	1.138	1.236	1.278	1.218	1.160
4	1.227	1.270	1.263	1.248	1.305	1.418	1.465	1.397	1.331
5+	1.227	1.270	1.263	1.248	1.305	1.418	1.465	1.397	1.331
Age/Year	2014	2015	2016	2017	2018	2019			
0	0.293	0.313	0.313	0.287	0.256	0.233			
1	1.598	1.703	1.705	1.561	1.392	1.267			
2	1.711	1.824	1.826	1.672	1.491	1.357			
3	1.191	1.270	1.271	1.164	1.038	0.945			
4	1.366	1.456	1.458	1.335	1.190	1.083			
5+	1.366	1.456	1.458	1.335	1.190	1.083			

Table 10: HKE in GSAs from1 to 12: summary results of the a4a assessment.

Year	Catch (t)	SSB (t)	Rec (000)	F_{bar} (1-3)	Total biomass (t)
2005	9803	2696.9	539841	1.348	23073
2006	10101	2745.4	484776	1.394	22861
2007	9394.1	2539.9	581078	1.386	24261
2008	9705.8	2426.9	679892	1.370	25654
2009	11370.7	2577.1	655365	1.433	26464
2010	12381.6	2607.1	584481	1.557	26463
2011	10882.5	2168.2	621723	1.609	24620
2012	9958.2	1934.9	599086	1.534	23923
2013	9946.3	2059.3	477787	1.461	24313
2014	9147.4	2077.3	484320	1.500	19109
2015	8760.1	1780	432972	1.599	18345
2016	7465.1	1479.1	468102	1.601	17556
2017	7247.9	1450.3	482609	1.466	17906
2018	7268	1761.3	373139	1.307	17457
2019	7071.6	2240.3	410617	1.189	16476

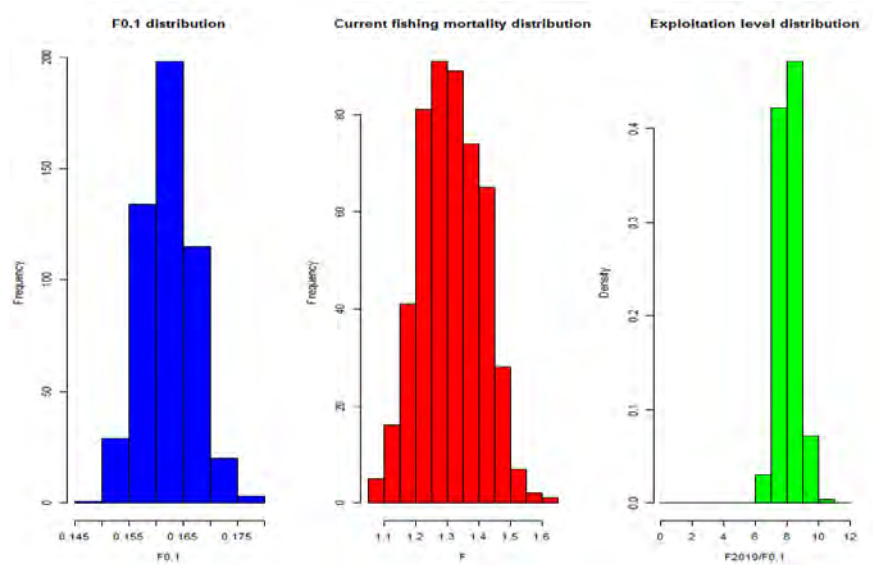


Figure 16: HKE in GSAs from1 to 12: Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F0.1 ratio) values.

Besides the model selection procedure, a further check was performed regarding the number of knots (k) of the smoother on year in the fmodel. A test based on AIC, BIC and GCV was performed on k ranging between 5 and 15 (Figure 17). All the model specifications highlight a consistent behavior in terms of main outcomes (Figure 18).

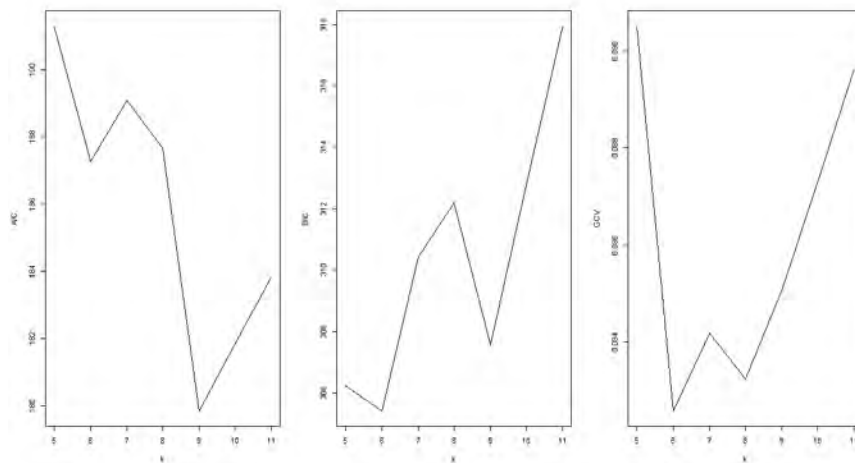


Figure 17: HKE in GSAs from 1 to 12: AIC, BIC and GCV values estimated on a range of k values of the smoother on year of the fmodel.

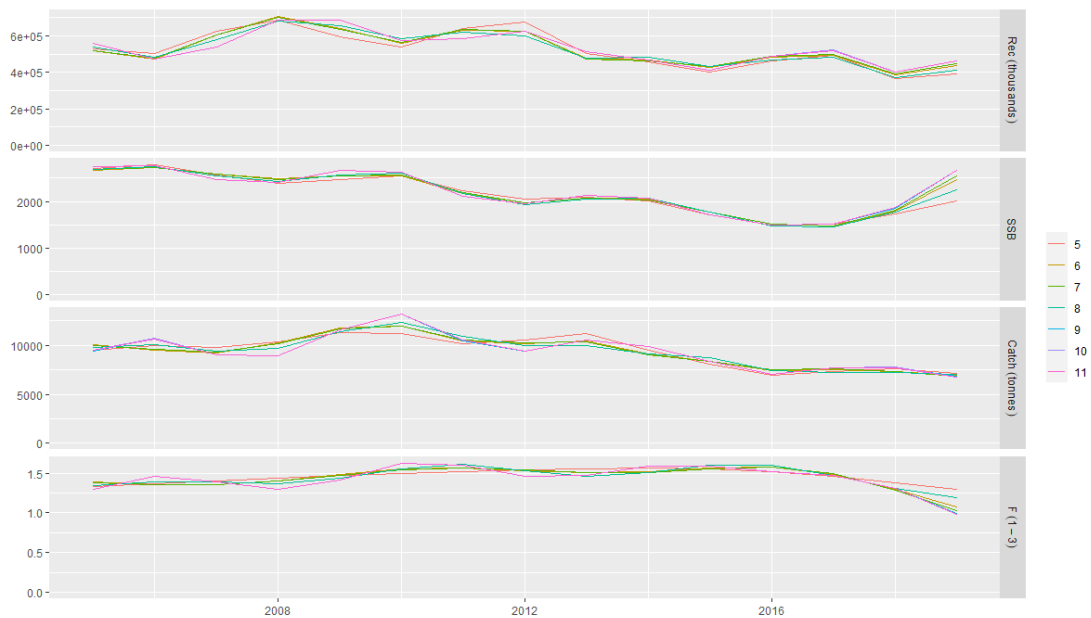


Figure 18: HKE in GSAs from 1 to 12: Outputs of model runs with different k values on the smoother on year in the fmodel.

4.1.4 Comparison with other assessments with a different stock configuration

Considering the output of the genetic and of the otolith microchemistry analyses, the stock of hake living in the Mediterranean basin seems to overlook the conventional GSAs boundaries used in the current assessments, resulting in only 3 sub-units: the Western Mediterranean stock (GSAs 1 – 12/13), the Central Mediterranean stock (GSAs 13, 16 – 20) and the Eastern Mediterranean stock (GSAs 22 – 27).

Here the genetic configuration has been followed and the assessment for the Western Mediterranean stock, including GSAs from 1 to 12, have been developed. Considering the current assessments developed within the STECF (STECF, 2020) and GFCM (GFCM, 2021) frameworks and the results of the hake benchmark assessments (GFCM, 2019), the assessment here presented has been compared with the those for GSAs 1-5-6-7 and for GSAs 8-9-10-11 (GFCM, 2021) in order to verify if there are evident improvement or deterioration of stock assessment quality, and to check the degree of coherence in the stock perception that can be found between the single GSA assessment and the MED-UNITS framework.

Model diagnostics

The diagnostics of the models for GSAs 1, 6, 7, 9 and 10 performed during the STECF EWG 20-09 (STECF; 2020) were plotted for comparison with the diagnostic of the model presented in this report (Figure 19). The diagnostics for single GSAs models and for the MED-UNITs model do not show severe issues. The magnitude and the trend of the residuals are comparable.

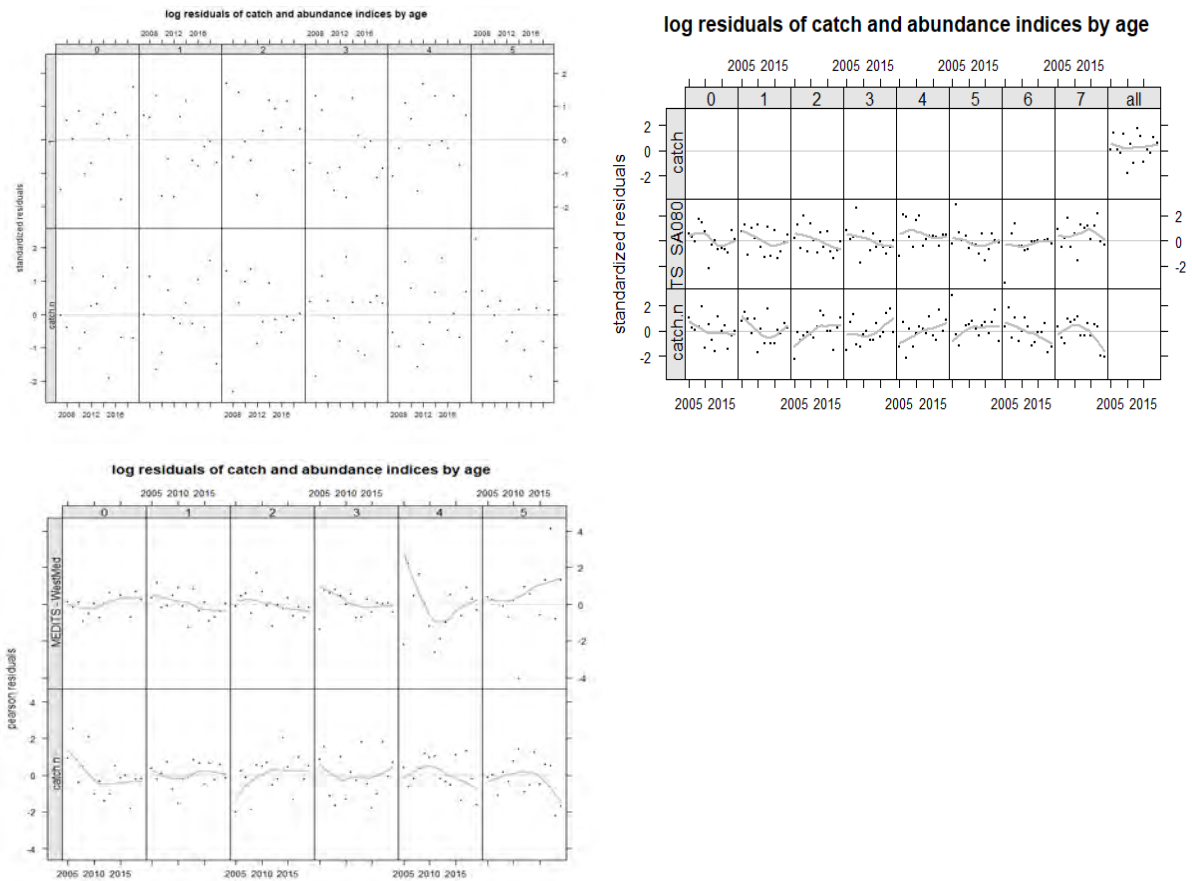


Figure 19: HKE in GSAs from 1 to 12: model diagnostic for the stock assessment of STECF EWG 20-09 for GSAs 1-5-6-7 (top left), GFCM assessment in GSAs 8-9-10-11 (top right); and on the MED-UNITs aggregation (bottom).

Retrospective patterns

The retrospective pattern observed for the MED-UNITs stock aggregation had a greater consistency than most of the single GSAs assessment (Figure 20). In particular, the models fitted on GSAs 1-5-6-7 is rather unstable on the SSB trend at the beginning of the time series, an issue that was not observed in the aggregated assessment.

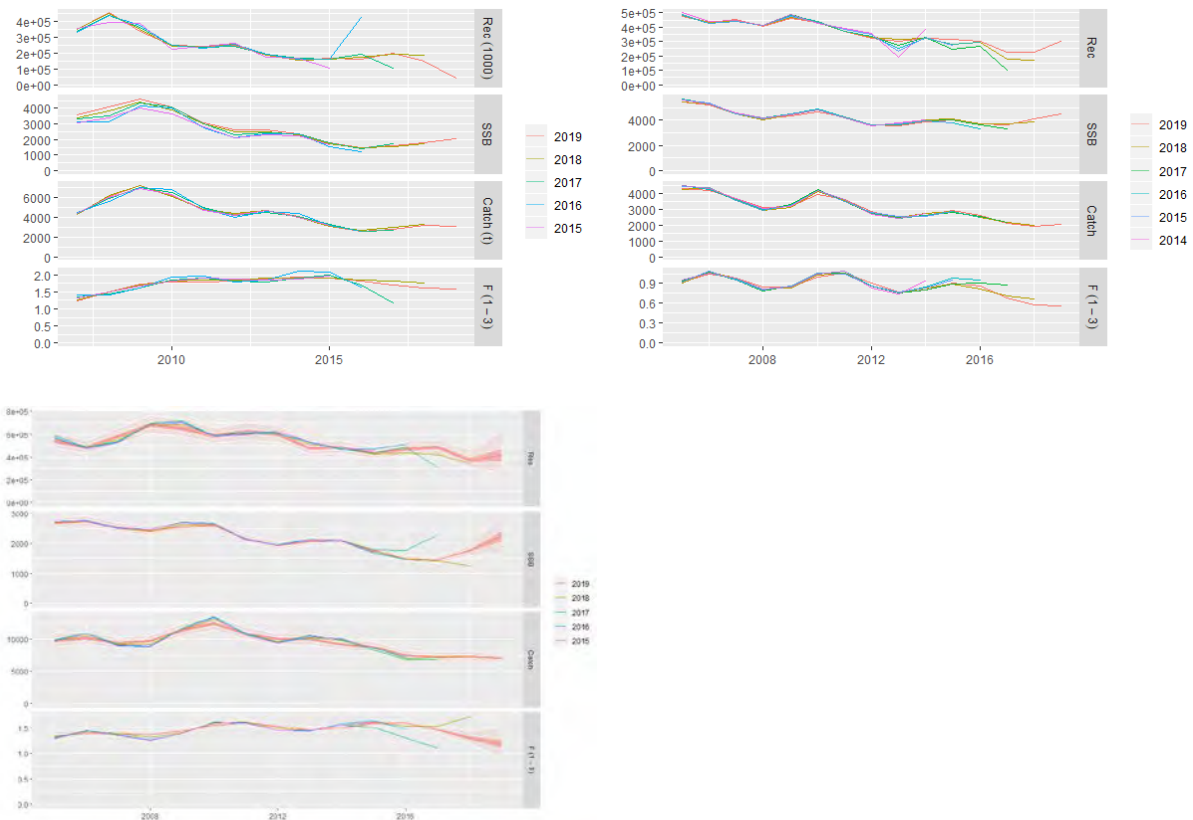


Figure 20: HKE in GSAs from 1 to 12: retrospective analysis for the stock assessment of STECF EWG 20-09 for GSAs 1-5-6-7 (top left), GFCM assessment in GSAs 8-9-10-11 (top right); and on the MED-UNITS aggregation (bottom).

Stock trajectories

The trajectories of the single GSA assessments (GSAs 1-5-6-7 and GSAs 8-9-10-11) have been compared to the MED-UNITs configuration to check the coherence for SSB and Recruitment (Rec 000) trend estimates, as well as for F_{bar} and F/F_{01} trends (Figure 21). Due to data availability (hake stock assessment for GSAs 1-5-6-7 considers only years from 2007 to 2019), the comparison was restricted to the time-frame 2007-2019. The comparisons for SSB and Recruitment were done by summing the contribution of single GSAs, then data were plotted and Pearson correlation coefficient were computed. SSB estimated by the two configurations show a similar trend at the end of the time series (2012 – 2019), whereas at the beginning of the time series the sum of the two GSAs assessment presents a peak in SSB. Also, the MED-UNITs configuration estimates lower SSB values. On the contrary, recruitment appears much more similar between the two considered configurations; thus the Pearson correlation coefficient is higher (0.892) compared to that one estimated for the SSB (0.533).

The F_{bar} and F/F_{01} trends were compared to the results obtained for single GSAs, including an estimation of the average annual value. The F_{bar} for the averaged single GSAs is quite similar both for the estimated values (except for the first year of the time series in which only the assessment of GSAs 8-9-10-11 is considered) and the trends. However, the F_{bar}/F_{01} results very different between the MED-UNITs configuration and the single GSAs assessments, with higher value for the MED-UNITs assessment.

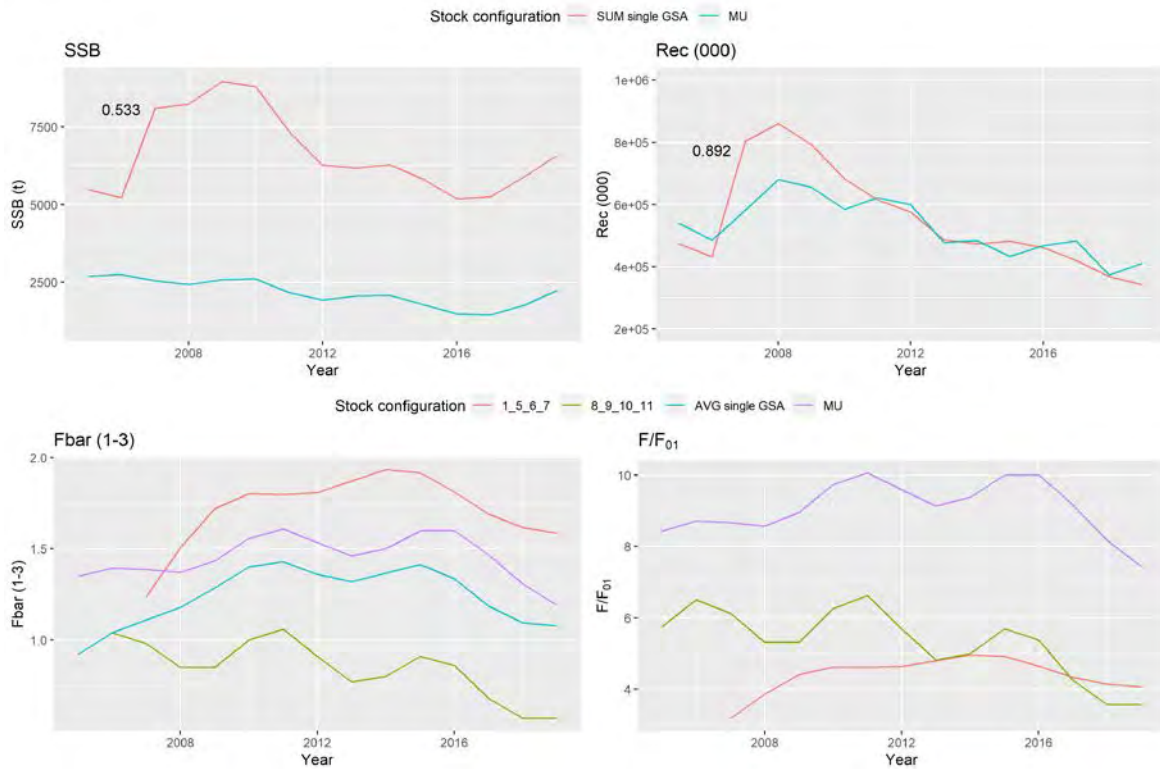


Figure 21: HKE in GSAs from 1 to 12: cohort consistency of the stock trajectories calculated on single GSAs 1, 6, 7, 9, 10, 13, 14, and 16 and on the MED-UNITS aggregation. MU= MED-UNITS aggregation.

F_{bar} (terminal year) and reference points

F_{bar}, F₀₁ and their ratio (F_{bar}/F₀₁) have been displayed for the last available year (2019) (Figure 22). The F_{bar} for 2019 for the MED-UNITS configuration rank between the F_{bar} estimated for the other two assessments, whereas the F₀₁ value for the MED-UNITS assessment result very similar to the one estimated by the assessment for GSAs 1-5-6-7. This results in a very high value for the ratio F/F₀₁ of the MED-UNITS configuration (~ 7) if compared to the other two assessments.

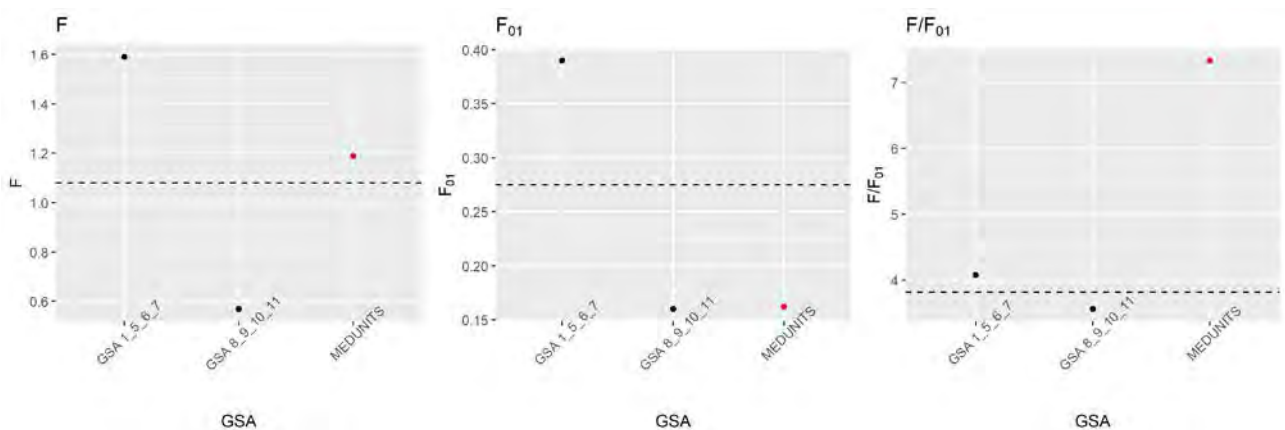


Figure 22: HKE in GSAs from 1 to 12: and on the MED-UNITS aggregation: F_{bar}, F₀₁ and their ratio calculated on single GSAs. Values for 1, 6, 7, 9, 10 refers to 2019, values for GSAs 13 and 14 refers to 2016 and values for GSA 16

Model diagnostics did not present large difference in the model fits, whereas a slight improvement in the retrospective patterns was observed in the MED-UNITS aggregation. The stock trajectories observe important differences between the SSB estimated for the MED-UNITS assessment and that one from the

single GSAs assessments, whereas recruitment and F_{bar} are more similar among the various assessment configurations. Finally the ratio between F_{bar} and F_{01} is very high for the MED-UNITs assessment compared to the current assessment for GSAs 1-5-6-7 and for GSAs 8-9-10-11. However, it has to be noted that similar ratio are showed in other hake assessments, e.g. the one for GSA 1, for GSAs 1 and 3, for GSA 6 (GFCM, 2021), that were not taken in account for comparison in this work since the hake benchmark assessment carried out in 2019 (GFCM, 2019) validated the assessments with multiple GSAs.

4.2 European hake in GSAs 22-27.

The stock assessment of hake in the Eastern Mediterranean Sea is here presented. Considering the WP1 and WP2 outputs, the GSAs from 22 to 27 were lumped together (Figure 23).

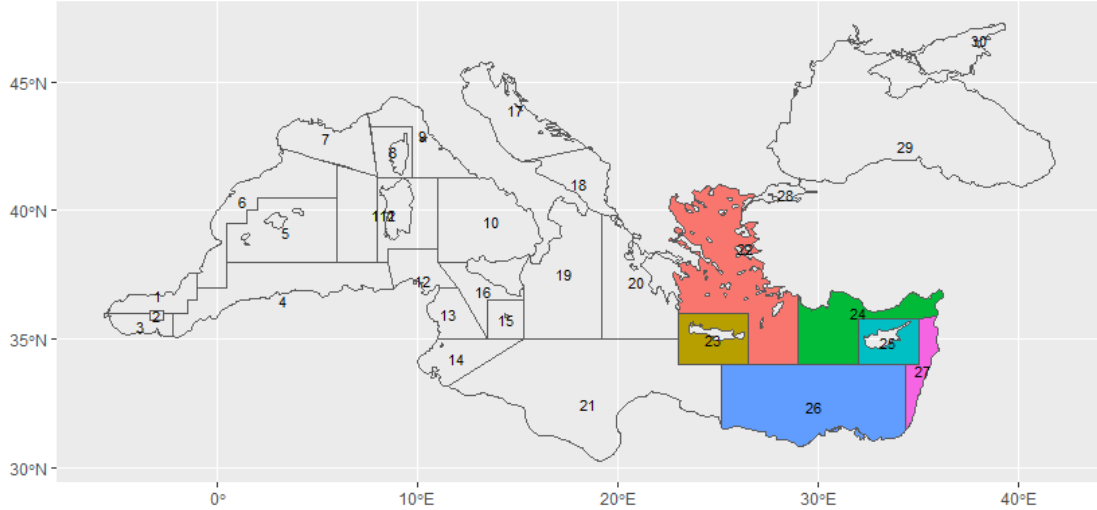


Figure 23: HKE in GSAs 22, 23, 24, 25, 26 and 27: geographical location of the stock

The genetic analysis conducted within the MED-UNITS project identify a unique stock for the Eastern Mediterranean Sea, including GSAs 22, 23, 24, 25, 26 and 27. Hake is one of the most important fish stocks in GSA 22 and it is mainly exploited by bottom trawlers, nets and longlines. The stock is distributed in depths between 50 and 600 m, with a peak in abundance between 200 and 300 m. The stock is mainly exploited by the Greek and Turkish fishing fleets but the landings of hake of the Turkish fleet are not available. Cyprus is also considered in this assessment, however its catches are very low compared to the Greek ones. The FAO/GFCM database reports also catches for the Syrian Arab Republic and Israel that are here considered.

4.2.1 Input data

Fishery-Dependent data

Considering the data available in the EU DCR/DCF for this group of GSAs only catch data have been considered. Specifically, the bulk of the hake fishery occurs in GSA 22 (Greece), followed by GSA 23 (Greece) and GSA 25 (Cyprus). In this assessment also data for Syria and Israel have been taken into account since they are reported in the FAO-GFCM database. Indeed, Turkey was not considered in this assessment since scarce and occasionally data are available from the FAO-GFCM database. The data included in the FAO-GFCM database are assumed as catch data (Table 11).

Considering the observations from the Greek experts contained in the last assessment (STECF, 2020b; GFCM, 2021), the time series considered for this assessment goes from 1997 to 2019.

Table 11: HKE in GSAs 22, 23, 24, 25, 26 and 27: Catch data divided by country, GSA and year. Red values correspond to those estimated. Data for Cyprus (CYP – GSA 25) from 1997 to 2012, for the Syrian Arab Republic and for Israel come from the FAO/GFCM database. Data for Greece in GSA 22 (GRC – GSA 22) are those used in STECF 20-15.

Year	CYP - GSA 25	GRC - GSA 22	GRC - GSA 23	Syrian Arab Republic	Israel
1997	4	3995	21.65	300	86
1998	2	3243	21.65	125	134
1999	5	3221	21.65	110	60
2000	6	3626	21.65	87	62

2001	8	2799	21.65	52	73
2002	3	2841	21.65	63	68
2003	11	3216	21.65	70	60
2004	10	3884	21.65	86	39
2005	28	3886	21.65	110	36
2006	23	4646	21.65	62	18
2007	4	5173	21.65	65	10
2008	5	5111	21.65	46	6
2009	2.60	5197	21.65	52	2
2010	3.00	4607	21.65	49	2
2011	4.90	4158	21.65	45	79
2012	7.30	4028	21.65	37	3
2013	5.50	4792	16.89	33	5
2014	2.70	3162	21.56	30	5
2015	2.70	2731	26.50	28	5
2016	2.66	2364	38.68	33	0
2017	3.29	3159	50.77	30	0
2018	2.14	3179	62.86	30	0
2019	3.03	3300	84.09	30	0

Fishery-Independent data

The MEDITS survey is carried out in GSAs 22, 23 and 25. Considering the differences in the time series, two survey indexes have been considered: one for GSAs 22 and 23 together (Figure 24) and one for GSA 25 only (Figure 25). The MEDITS survey for GSAs 22 and 23 has been performed since 1994, however in some years the survey was not carried out (2002, 2007, 2009-2012, 2015 and 2017). Whereas the MEDITS survey in GSA 25 was carried out since 2005, only in 2014 the survey was not executed.

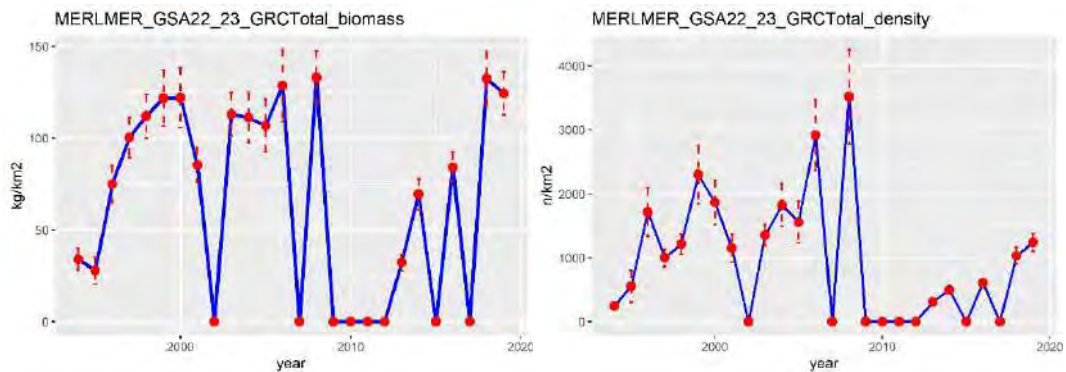


Figure 24 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Total biomass (left panel) and density (right panel) from the MEDITS survey for the stock of hake in GSAs 22 and 23

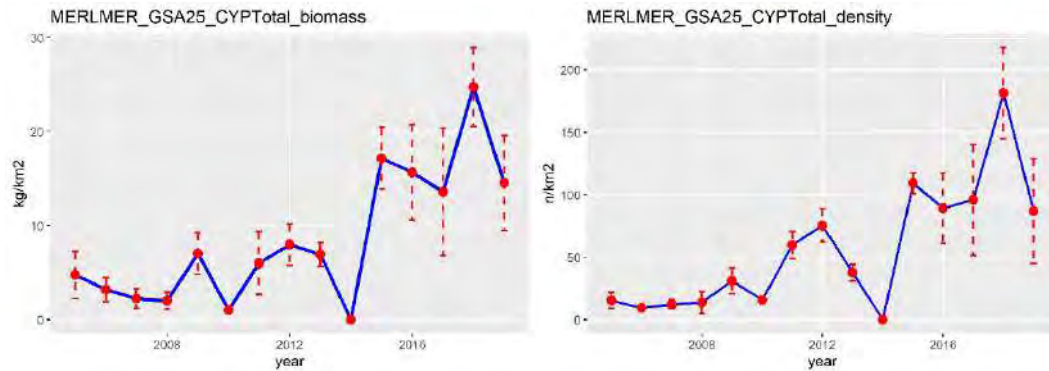


Figure 25 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Total biomass (left panel) and density (right panel) from the MEDITS survey for the stock of hake in GSAs 25.

4.2.2 Model description

The SPiCT assessment method is a state-space version of the Pella-Tomlinson surplus production model (Pella and Tomlinson 1969). SPiCT allows the inclusion of prior distributions for parameters that are difficult to estimate. The continuous time formulation of the model allows for arbitrary and irregular data sampling without a need for catch and index observations to match temporally. SPiCT requires a time series of landings or catches and one or more time series of commercial or survey CPUE indices. The expected output include all parameter estimates and the most interesting derived quantities are the F/F_{MSY} and B/B_{MSY} that quantify the stock status. More information on the SPiCT assessment method is described in Pedersen and Berg (2016). SPiCT is available as an R (R Core Team 2015) package in the github online repository: <https://github.com/mawp/spict>. For fast and efficient estimation, SPiCT uses the Template Model Builder package (TMB, Kristensen et al., 2016).

Two models have been developed:

- Option1 : catch data (Greece, Cyprus, Syria and Israel) from 1997 to 2019 and two survey indexes: i) MEDITS in GSA 22 and 23 from 1997 to 2019 and ii) MEDITS for GSA 25 from 2005 to 2019
- Option2 : catch data (Greece, Cyprus, Syria and Israel) from 1997 to 2019 and only the MEDITS survey for GSA 22 and 23 from 1997 to 2019

4.2.3 Outputs from the model

Option 1: this model has been developed considering catch data from 1997 to 2019 and two survey indexes: one considering the MEDITS index for GSAs 22 and 23 from 1997 to 2019 and one considering only the MEDITS index for GSA 25 from 2005 to 2019. Figure 26 summarizes the input data used for this option; the MEDITS survey is carried out mainly in spring/summer time and this was specified within the model.

Priors have been used in this option. Specifically, the initial biomass was assumed around the B_{MSY} value, whereas the initial depletion level was set as medium exploitation. A prior for r was also included, this was assumed equal to 0.53, as shown in Fishbase (Froese and Pauly, 2019) for hake.

The stock results in good status, with low fishing mortality and high biomass in the most recent years (Figure 27 – left panel and Table 12). Diagnostics appear quite good except for some autocorrelation in the residuals of the MEDITS index, for which significant lags are identified (Figure 27– right panel); also, the retrospective analysis shows a quite unstable model (Figure 28). Whereas, the initial values seem to not influence the parameter estimates.

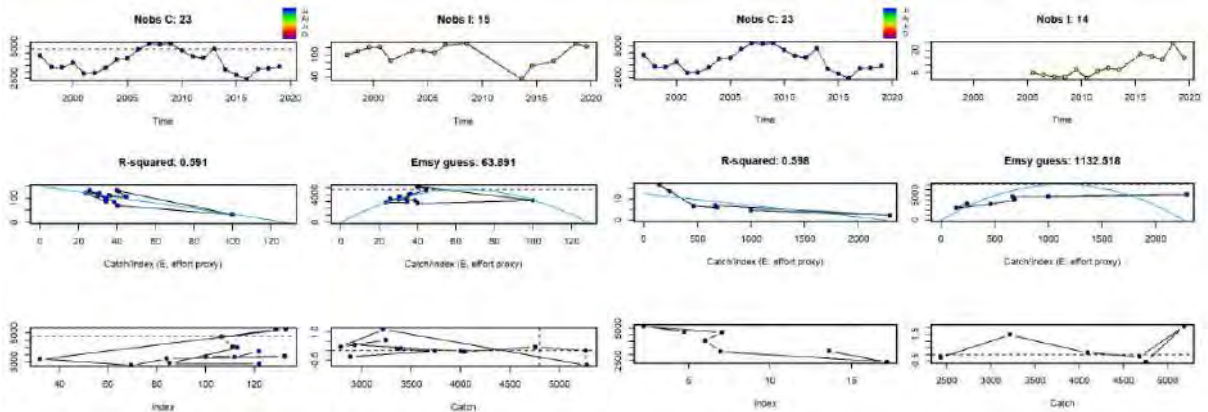


Figure 26 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Input data for the SPICT model. Option 1.

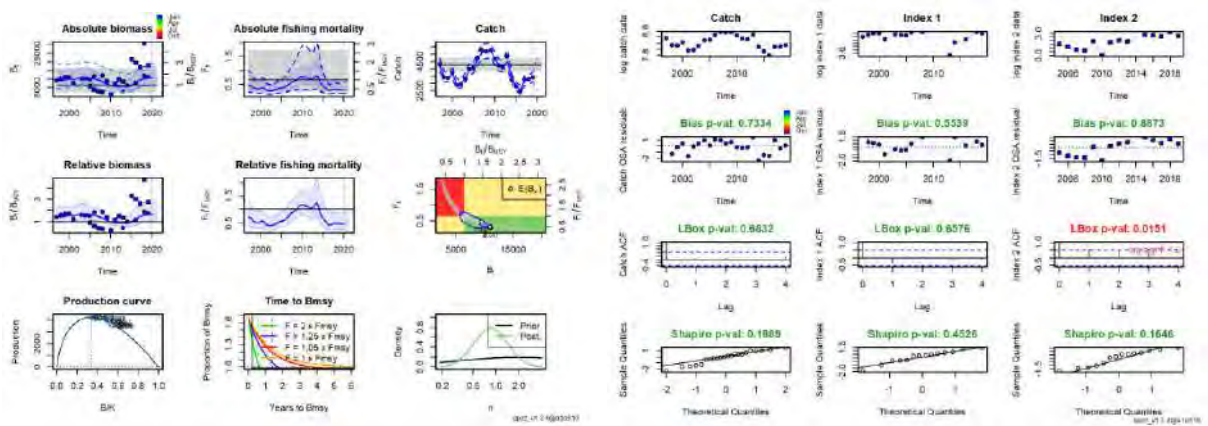


Figure 27 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Output (left panel) and diagnostics (right panel) of the SPICT model. Option 1.

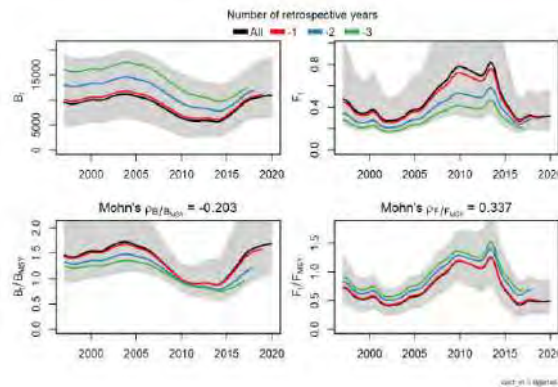


Figure 28 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Retrospective analysis. Option 1.

Table 12: HKE in GSAs 22, 23, 24, 25, 26 and 27. Estimated values from the SPICT model. Option 1.

K	r	MSY	Bmsy	Fmsy	BBmsy	FFmsy
19184	0.556	4247	6478	0.66	1.66	0.47

Option 2: this model has been developed considering catch data from 1997 to 2019 and the MEDITS index for only GSAs 22 and 23 from 1997 to 2019, being this survey the most representative one for the this stock. Figure 29 summarizes the input data used for this option. The MEDITS survey is carried out mainly in spring/summer time and this was specified within the model (Figure 29, left panel).

Priors have been used in this option and these have been set as in the previous runs.

As in the previous runs, this option describes a stock in good status, with low fishing mortality and high

biomass in the most recent years (Figure 30 – left panel and Table 13). Diagnostics appear good (Figure 30 – right panel) and the retrospective shows a rather stable model (Figure 31). The other parameters are in line with the SPiCT guidelines (Mildenberger et al., 2019), thus suggesting a good setting for this model.

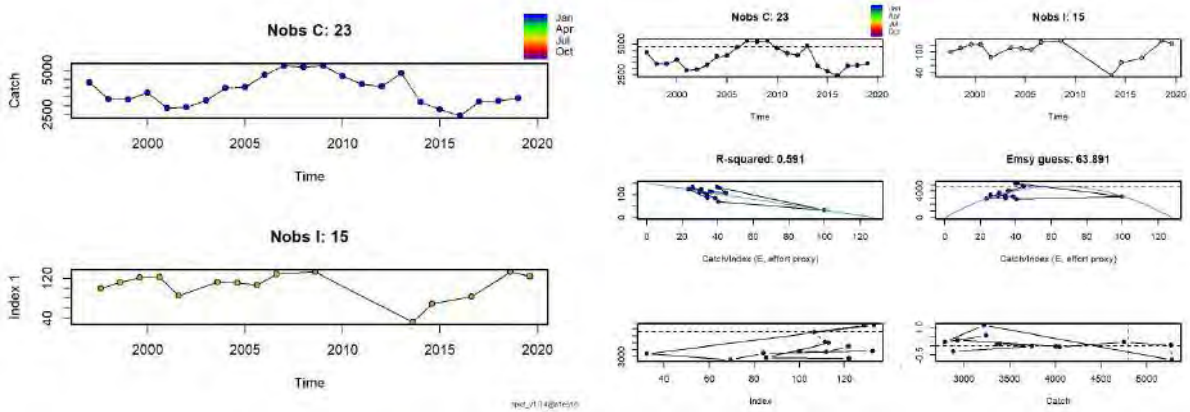


Figure 29 – HKE in GSAs 22, 23, 24, 25, 26 and 27. Input data for the SPiCT model. Option 2.

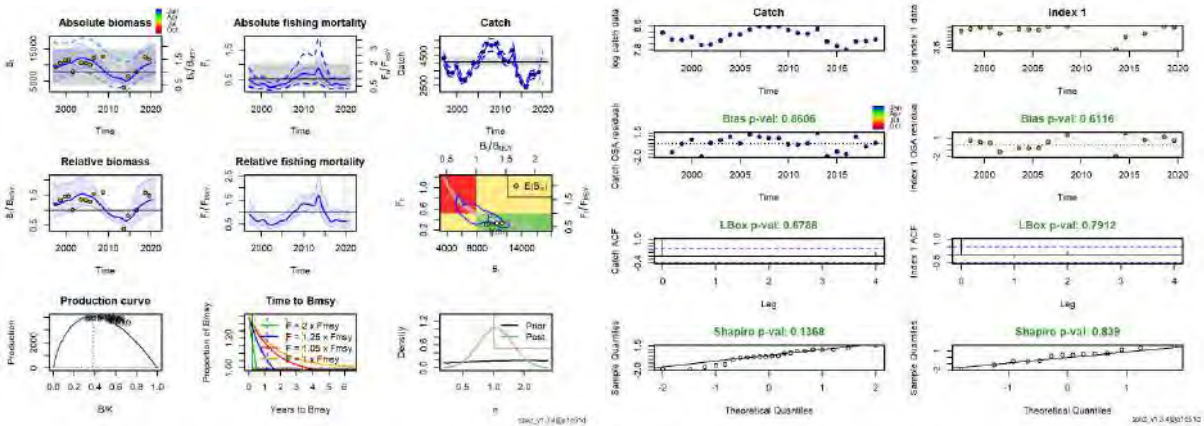


Figure 30 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Output (left panel) and diagnostics (right panel) of the SPiCT model. Option 2.

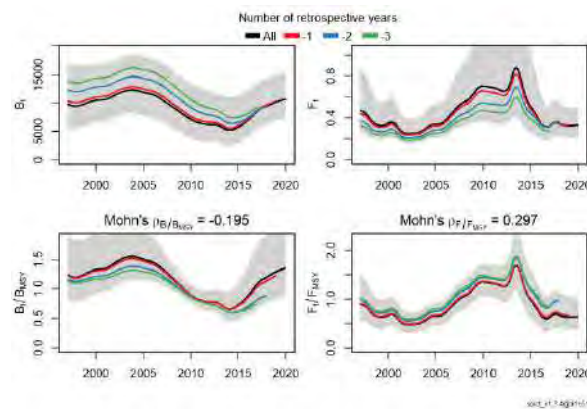


Figure 31 - HKE in GSAs 22, 23, 24, 25, 26 and 27. Retrospective analysis. Option 2.

Table 13: HKE in GSAs 22, 23, 24, 25, 26 and 27. Estimated values from the SPiCT model. Option 2.

K	r	MSY	Bmsy	Fmsy	BBmsy	FFmsy
20954	0.545	4094	7897	0.52	1.29	0.63

Another run including only one MEDITS index grouping all GSAs (22, 23 and 25) has been tested. This configuration did not converge and thus was not reported in this document.

The two options presented in this document show similar outcomes for this stock. However option two showed the best diagnostics and thus it was selected for comparison with the current assessments.

4.2.4 Comparison with other assessments with a different stock configuration

Considering the output of the genetic and of the otolith microchemistry analysis, the stock of hake living in the Mediterranean basin seems to overlook the conventional GSAs boundaries used in the current assessments, resulting in only 3 sub-units: the Western Mediterranean stock (GSAs 1 – 12/13), the Central Mediterranean stock (GSAs 13, 16 -20) and the Eastern Mediterranean stock (GSAs 22 – 27).

Here the genetic configuration has been followed and the assessment for the Eastern Mediterranean stock, including GSAs from 22 to 27, has been developed. Considering the current assessments developed within the STECF and GFCM frameworks, in the Eastern Mediterranean only the stock of hake living in GSA 22 is performed and thus used for comparison with the model developed for the MED-UNITS project.

The assessment of hake in GSA 22 is carried out considering also the age structure and thus both a SPiCT model and a4a model have been developed. However, none of these assessments have been considered valid for giving advice (STECF, 2020b). Considering the data available within the MED-UNITS project, here only a SPiCT model was developed (see the previous paragraph).

Model diagnostics

The diagnostics of the SPiCT model for GSA 22 performed during the STECF EWG 20-15 (STECF, 2020) were plotted for comparison with the diagnostic of the model presented in this report (Figure 32). The diagnostics for the MED-UNITS configuration result better than those showed by the single GSA model, with the Shapiro test for catch improved in the new configuration.

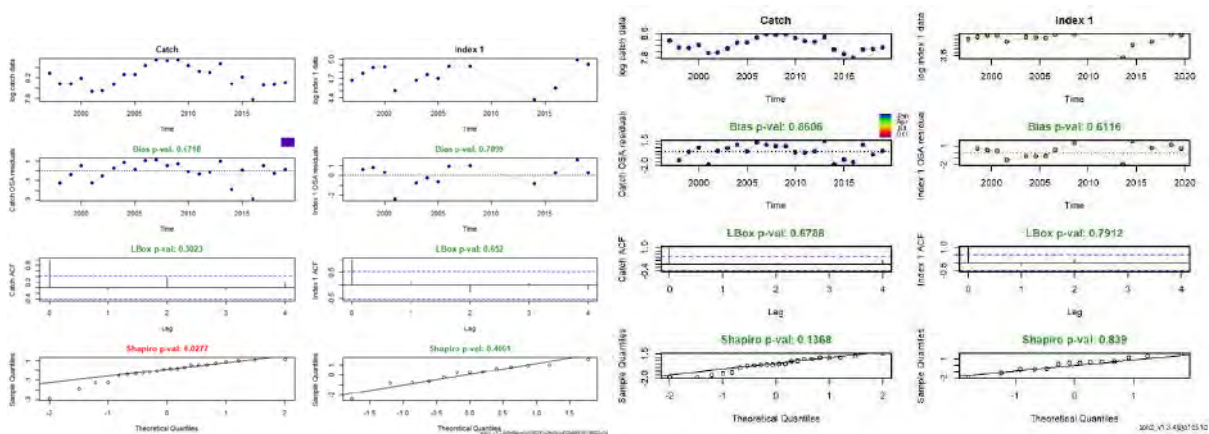


Figure 32: HKE in GSAs 22, 23, 24, 25, 26 and 27: model diagnostic for the stock assessment of STECF EWG 20-15 for GSA 22 (left panel) and on the MED_UNITS aggregation (right panel).

Retrospective patterns

The retrospective pattern observed for the MED-UNITS stock aggregation is improved compared to the single GSA assessment (Figure 33). Specifically the single GSA assessment results particularly unstable, whereas the MED-UNITS aggregation seems rather stable, except for some differences in the stock biomass. The Monhn’s rho for the retrospective of the MED-UNITS aggregation is in agreement with the condition expressed in Hurtado-Ferro et al. (2014).

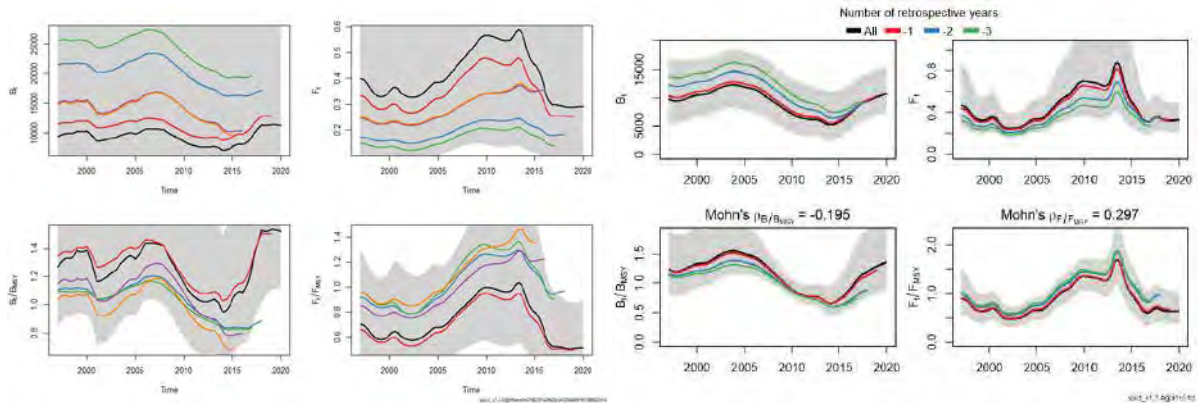


Figure 33: HKE in GSAs 22, 23, 24, 25, 26 and 27: retrospective analysis for the stock assessment STECF EWG 20-15 for GSA 22 (left panel) and on the MED_UNITS aggregation (right panel).

F_{bar} (terminal year) and reference points

F_{bar} (F), F₀₁ and their ratio (F_{bar}/F₀₁) have been displayed for the last available year (2019) (Figure 34). The F_{bar} estimated for the MED-UNITS configuration is higher compared to the one estimated from the single GSA assessment, whereas the F_{0.1} of the MED-UNITS aggregation results lower than the one from the single GSA assessment. Consequently, the ratio F_{bar}/F_{0.1} is higher for the MED-UNITS configuration compared to the single GSA assessment, however they are comparable.

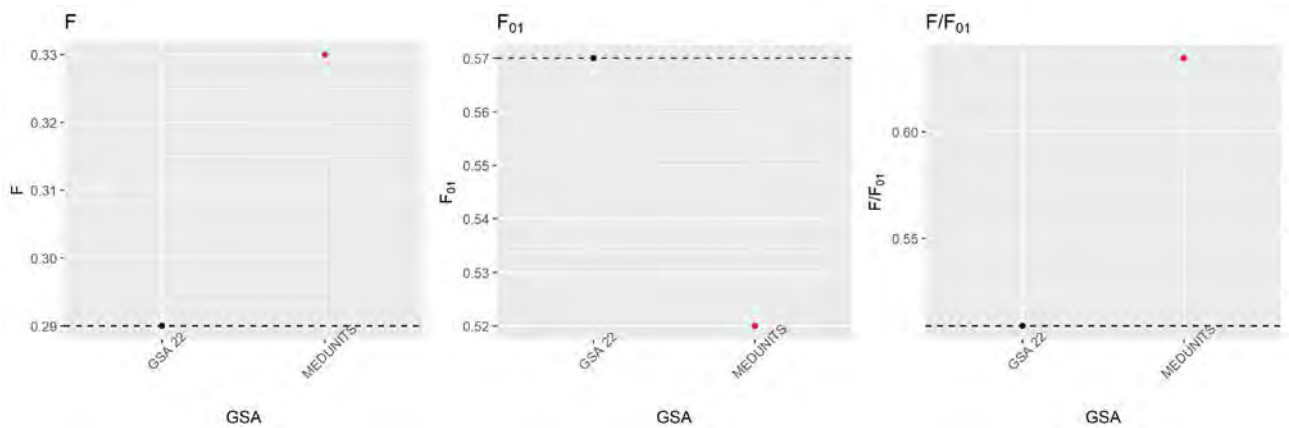


Figure 34: HKE in GSA 22 and on the MED-UNITS aggregation: F_{bar} (F, left panel), F₀₁ (middle panel) and their ratio (F/F₀₁, right panel) calculated on single GSAs.

The comparison of the model diagnostics indicates an improved situation for the MED-UNITS aggregation; this is also showed by the retrospective pattern that presents a higher stability for the MED-UNITS configuration. The F_{bar} and F_{0.1} are different between the two models; however, they are comparable. The ratio between F_{bar} and F_{0.1} is higher for the MED-UNITS aggregation compared to the single GSA assessment, however the two assessments describe a similar situation for the stock of hake living in Aegean and Levantine waters that results being under exploited with high biomass.

4.3 Red mullet in GSAs 1, 5-16.

In the western Mediterranean, red mullet (MUT), *Mullus barbatus*, represents a target resource for the trawling fleet operating on the continental shelf, mainly up to 200m depth. According to the outcomes of the genetic analysis, as well as of the otoliths shape and microchemistry analysis, performed under the Med_Units project, a stock configuration including GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12 was identified. Otolith shape and microchemistry included also GSA 16 in the cluster. Considering that GSAs 13 and 15 are lacking samples, while they are in strict geographical proximity with some of the GSAs aggregated by the otolith analysis, these areas were included in the cluster. For the same reason, GSA 14 was included notwithstanding the genetic analysis, which set this area apart from the other GSAs. Therefore, the stock configuration used to perform an assessment included GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16 (Figure 35).

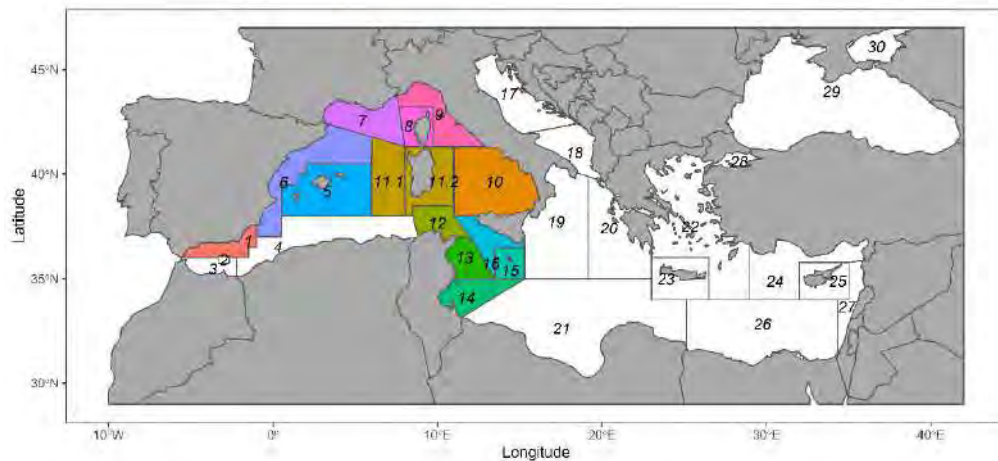


Figure 35: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: geographical location of the stock

4.3.1 Input data

Fishery-Dependent data

Commercial data (catch data, LFDs and biological parameters) for the EU countries considered in the stock configuration (Spain, France, Italy and Malta) were collected from the EU DCR/DCF database. Catch statistics up to the year 2018 for non-EU countries (Tunisia) were retrieved from FAO-FishstatJ, while LFD and biological parameters for non-EU countries were not available.

Corrections were necessary to align EU DCR/DCF data to stock configurations used in the most recent available stock assessment within STECF and GFCM context:

- Catch data and LFDs for GSA 16 for the years 2006-2019 were not matching those used within the most recent stock assessments. The updated information was kindly provided by local experts.
- Landing data for GSA 15 included in the landings.csv file was not matching those provided in the catch.csv file. The latter information was the same used in the most recent stock assessment and was therefore used to build the stock object.
- LFD (landings) for GSA 16: for the years 2002-2005 LFD information was taken from EU DCR/DCF, considering only statistics for the gear OTB, to align the information with data provided by local experts.

Reconstructions were also needed to make up for some scattered information. Landing data were reconstructed only for those areas which mean contribution to the total landings was > 10%:

- Landing data for GSA 9: missing information for the year 2002 was provided by local experts.

- Landing data for GSA 16: missing information for the years 2002-2005 was obtained by calculating the mean proportion GSA 16 out to the total landings for the years 2006-2008, and applying it to the annual catches (2002-2005).
- Landing data for GSAs 12, 13 and 14: landing data for these areas were combined, and the missing information for the year 2019 was obtained by calculating the mean proportion of GSAs 12-14 out of total landings for the years 2016-2018, and applying it to the 2019 catches.
- LFD (landings) for GSA 9: missing information for the year 2002 was filled by raising the 2002 landings to the mean LFD for the years 2003-2005.
- Discard data: discard data was reconstructed at the GSA level by calculating the ratio discard/landings of the first three available years, and multiplying this value for the annual landing data. The year 2006 for GSA 11 represented an outlier and was excluded from the calculation of the mean landing/discard value.
- LFD (discards): missing discards LFD were filled by raising the annual discards to the mean LFD of the three available years.

In addition, some data were excluded:

- GSA 8 (catches and LFD): data were scattered and represented a negligible amount.
- GSA 16 discards: decision taken according to the assumptions done in the most recent stock assessment.
- GSA 5 and 15 discards: data were scattered and represented a negligible amount.
- GSA 5 and 15 LFD (landings): data were scattered and negligible.

Figure 36 summarize the FDI data used in the assessment.

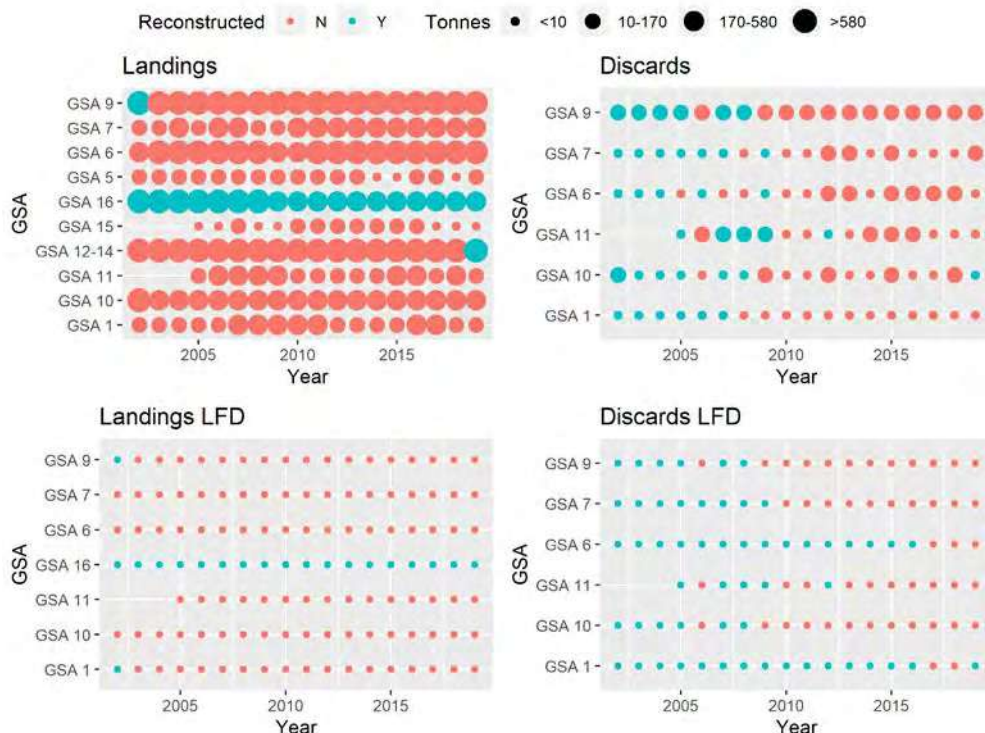


Figure 36: summary of the data used to build the stock object. Blue color indicates reconstructed data, size (only for Landings and Discards) indicates the amount (tons).

LFDs were sliced (deterministic slicing by sex) to obtain age structure using the growth parameters available under EU DCR/DCF for the GSA 10 (Table 14). The choice of GSA 10 was driven by the observation conducted during STECF EWG 20-09, where it confirmed the consistency of these parameters with the most recent age reading protocol for *Mullus barbatus* (Carbonara et al., 2018). The natural mortality vector was estimated using the Chen and Watanabe model (Table 15). Mean weight-at-age were calculated using the LW parameters available under EU DCR/DCF for the GSA 10 (selected for coherence with growth parameters) and applied to the whole time series. The maturity vector was set according to STECF EWG 20-09 for GSA 10. The Plus group was set to 4 according to the majority of the available stock assessments for this species.

No LFDs were available for GSAs 12-14, and LFDs for GSAs 5 and 15 were not used for age slicing. However, landings from these areas were added to the stock object, and the catch numbers-at-age was raised to those additional landings.

Fishery-Independent data

LFDs (n/km^2) from the MEDITS surveys in GSAs 1, 6, 7, 8, 9, 10, 11, 15, and 16 were sliced (deterministic age slicing by sex) using the same set of growth parameters used for the commercial data. GSA 5 data were available from 2007 onward, and they represented a low abundance value, therefore they were excluded. No data was available for GSAs 12, 13 and 14.

Table 14: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: VBGF and LW parameters by sex.

Sex	Linf	k	t0
F	30.11	0.239	-0.680
M	25.56	0.252	-0.911
	a		b
F	0.011		18.086
M	0.011		16.491

Table 15: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: natural mortality and maturity-at-age vectors.

Age	Natural mortality (F)	Natural mortality (M)	Proportion of matures (combined)
0	0.97	0.84	0.00
1	0.59	0.55	1.00
2	0.45	0.44	1.00
3	0.38	0.38	1.00
4+	0.34	0.34	1.00

The catch at age structure as well as the abundance at age resulting from the slicing are presented in Figure 37 and Figure 38. Moreover, the internal consistencies of such input data is presented in Figure 39 and Figure 40.

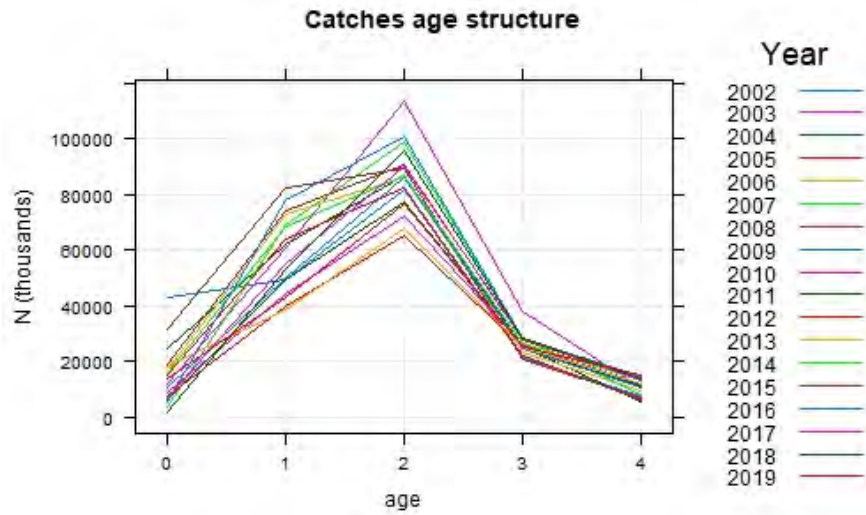


Figure 37: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: catch numbers-at-age

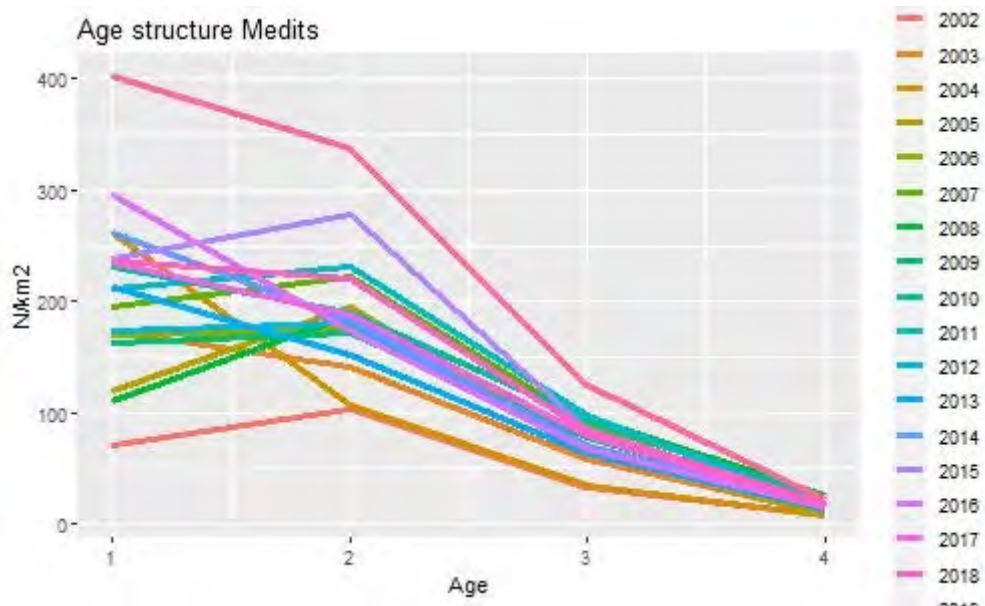


Figure 38: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: catch numbers-at-age in the MEDITS.

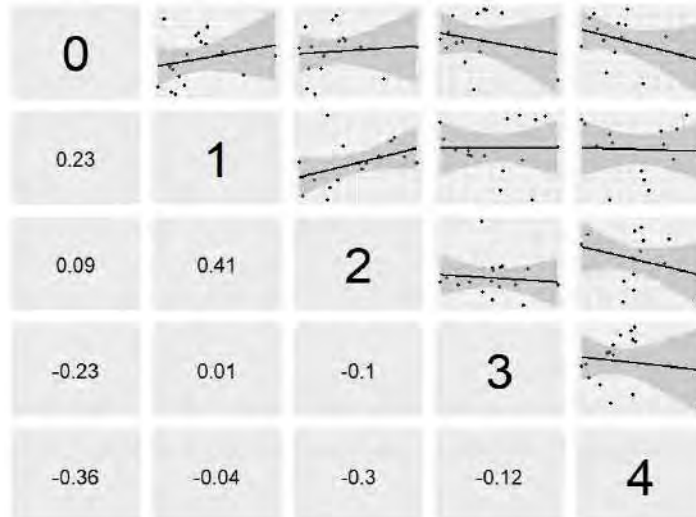


Figure 39: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: cohort consistency in the catch numbers-at-age.

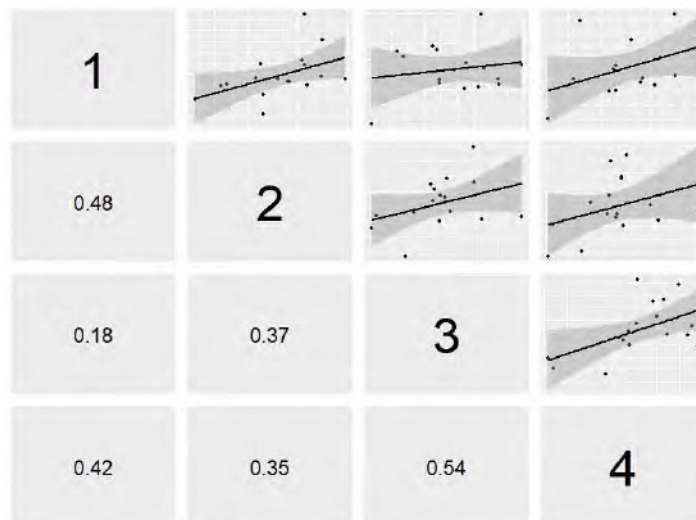


Figure 40: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: cohort consistency in the catch numbers-at-age in the MEDITS.

4.3.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment using a4a was carried out using as input data the period 2002-2019 for the catch data and 2002-2019 for the tuning file (MEDITS indices).

A natural mortality vector computed using Chen and Watanabe model was used in the assessment. Length-frequency distributions of commercial catches and surveys were split by sex and then transformed in age classes (plus group was set at age 4) using length-to-age slicing with different growth parameters by sex, derived from those adopted in the last assessment on red mullet in GSA 10. No correction on t_0 was adopted for consistency with the most recent stock assessment of red mullet in GSA 10.

The number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers at age $a \times$ catch weight-at-age a)].

In the catches, a plus group at age 4 was set. A true age 4 was used in the survey.

F_{bar} range was fixed at 1-3.

The assessment was performed by sex combined. Given that the landings were composed mainly of individuals between 1 and 3 years, these ages were selected as F_{bar} range. Survey indices (density by age) from MEDITS were used considering that spring surveys are not designed to detect recruitment of red mullet. Recruitment (age class 0) was detected just in some years when surveys were carried out in late summer or autumn. For that reason, age 0 class was not included in the tuning indices used for the assessment (Figure 38).

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $f_{\text{model}} = \sim s(\text{replace}(\text{age}, \text{age} > 2, 2), k = 3) + s(\text{year}, k = 10)$

Catchability sub-model: $q_{\text{model}} = \text{list}(\sim \text{factor}(\text{age}))$

SR sub-model: $\text{srmod} \sim \text{geomean}(\text{CV}=0.1)$

Model <- sca(stock = stk, indices = idx, fmodel, qmodel, srmod)

The n1model and vmodel used in the final fit are the default ones.

4.3.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the following figures. The model estimated 42 parameters out of 162 observations; it is then around the threshold of 25% ratio between parameters and observations (Figures 41-46).

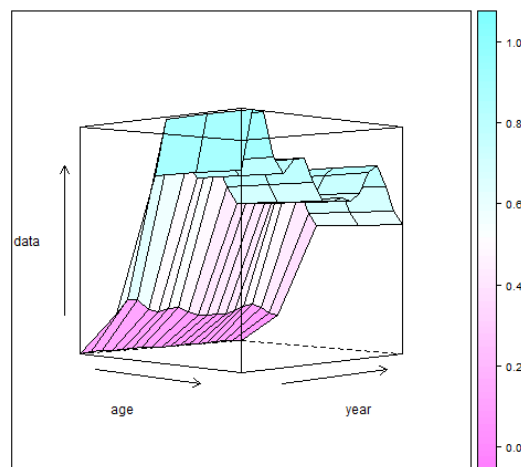


Figure 41: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: fishing mortality by age and year obtained from the a4a model (2002-2019).

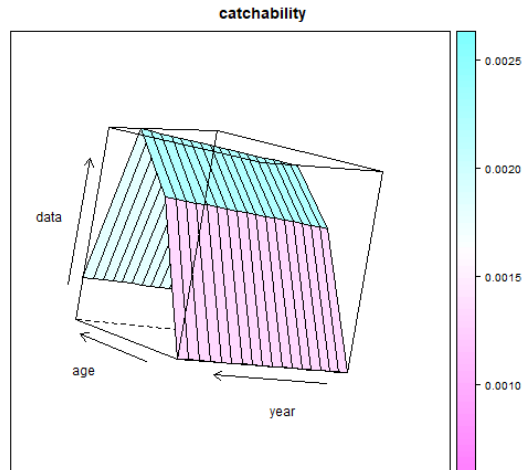


Figure 42: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: catchability of the survey by age and year obtained from the a4a model (2003-2019).

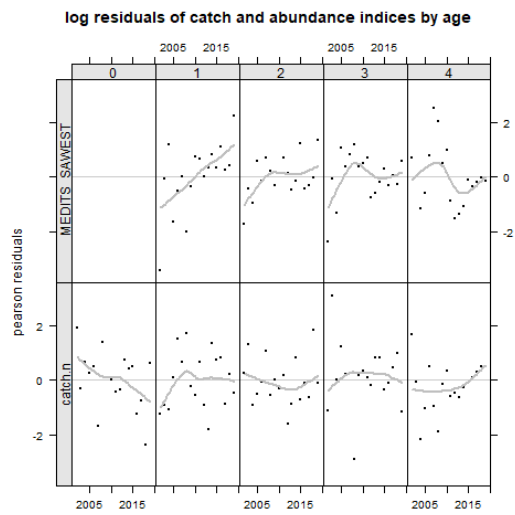


Figure 43: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: log residuals for the catch-at-age data of the fishery and the survey, and the catches.

log residuals of catch and abundance indices

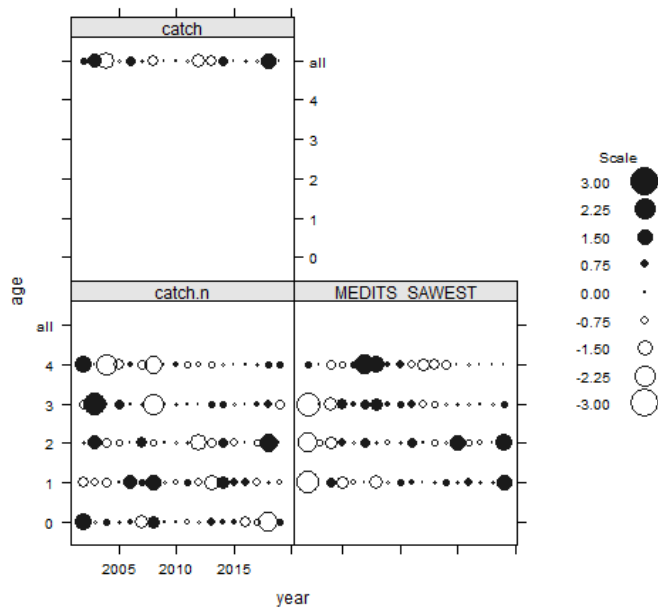


Figure 44: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: bubble plot of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

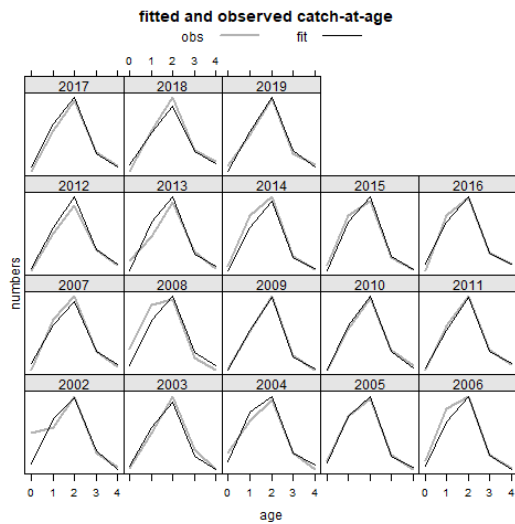


Figure 45: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: fitted vs observed values by age and year for the catches.

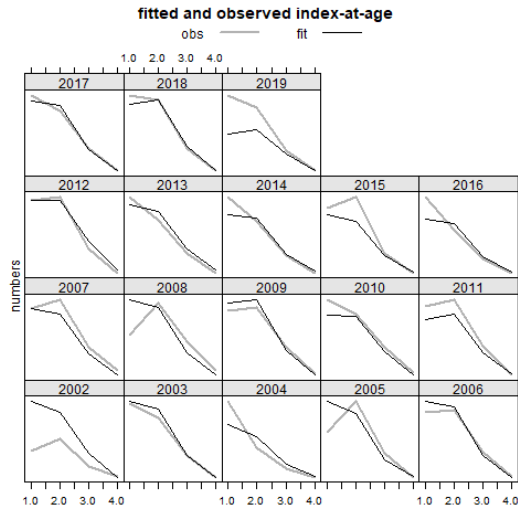


Figure 46: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated and showed that biomass of the plus group (age 4+) was about 12% of the total SSB on average, with spikes just above 15% (Figure 47).

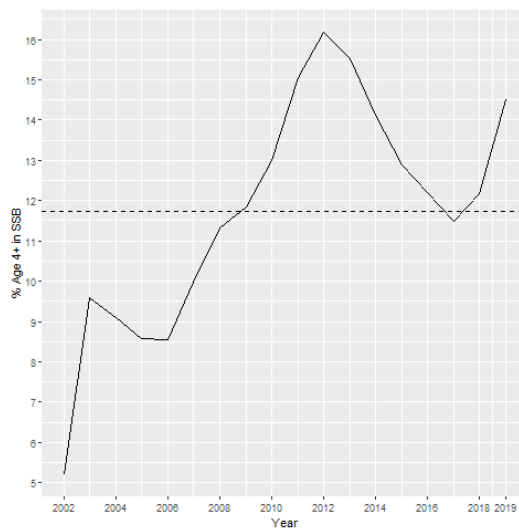


Figure 47: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: cryptic biomass.

The retrospective analysis shows some minor effect in F , SSB and recruitment, indicating that the assessment model is stable (Figure 48). The Mohn's Rho for F and SSB are within the optimal range $-0.2/0.2$ (0.096 for F , -0.092 for SSB, and -0.082 for recruitment).

Stock numbers-at-age, fishing mortality-at-age and summary results of the model are presented respectively in tables 16, 17 and 18 and in figure 49, while Figure 50 summarizes the histograms of probability for $F_{0.1}$, F_{curr} and level of exploitation ($F_{curr}/F_{0.1}$ ratio) values.

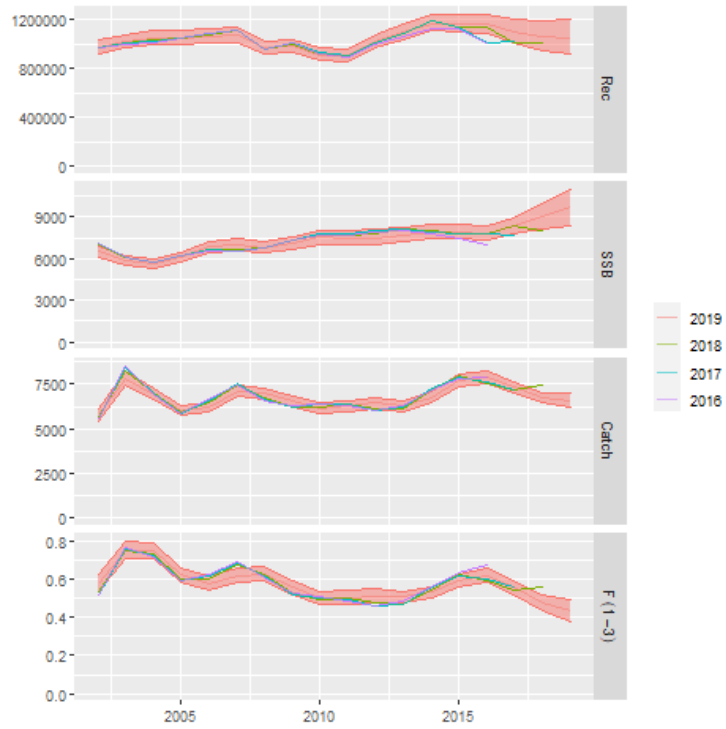


Figure 48: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: retrospective analysis.

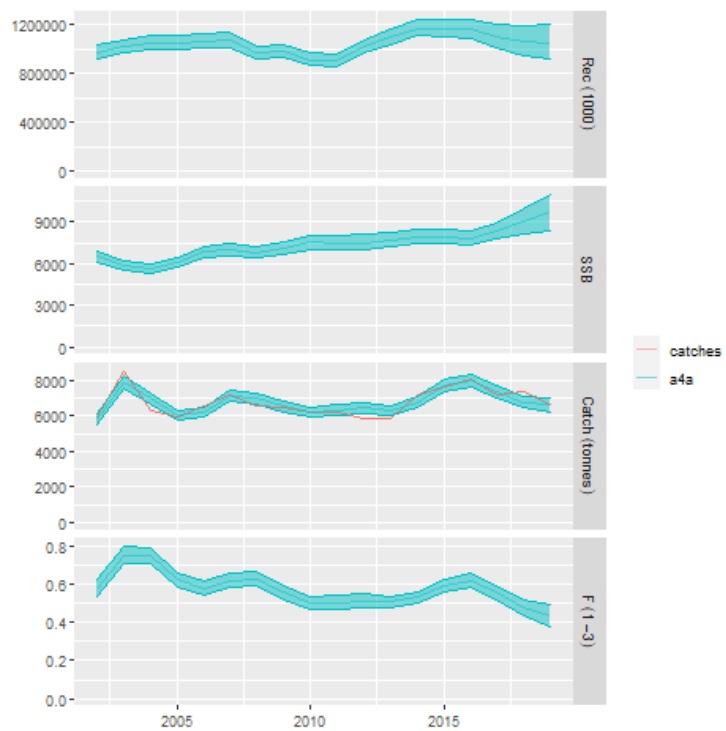


Figure 49: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: outputs of the a4a stock assessment model, with uncertainty; input catch data (red line) are plotted against the estimated catches

Table 16: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: Stock numbers-at-age (thousands).

Age/Year	2002	2003	2004	2005	2006	2007	2008	2009	2010
0	976232.3	1020212	1049471	1049095	1062418	1071117	966960.9	980279	911820.4
1	424835.9	396577.2	414904.6	425569.5	426676.6	428879.9	434142.8	389301.8	395982.7
2	180652.7	197138.6	173311.3	181580.9	194841.8	198170.5	196208	197995.2	181955.9
3	51660.94	53492.05	46216.71	40854	50713.08	57913.7	55572.94	54520.91	60515.43
4+	9935.099	19622.43	18525.21	16500.22	17326.06	21829.06	24164.97	23955.79	25954.16
Age/Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	909540.7	1018775	1097284	1167584	1167540	1162307	1101664	1058661	1056411
1	369766.4	370186.3	412657.6	441876.3	471307.6	470535.9	469679.6	445857.5	435626.2
2	188561.5	175902.2	175757.1	196314.9	208350.3	217339.5	215097.4	219311.8	213692.2
3	59738.26	61823.25	56986.01	57416.59	62109.02	60425.17	61016.36	65766.33	74396.94
4+	30727.97	32142.68	33008.74	31901.07	30634.16	29126.74	27214.56	29191.09	34851.84

Table 17: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: Fishing mortality-at-age.

Age/Year	2002	2003	2004	2005	2006	2007	2008	2009	2010
0	0.017	0.022	0.022	0.018	0.017	0.018	0.018	0.016	0.015
1	0.198	0.258	0.256	0.213	0.197	0.212	0.214	0.190	0.171
2	0.773	1.007	1.002	0.832	0.770	0.828	0.838	0.742	0.670
3	0.773	1.007	1.002	0.832	0.770	0.828	0.838	0.742	0.670
4+	0.773	1.007	1.002	0.832	0.770	0.828	0.838	0.742	0.670
Age/Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	0.015	0.015	0.015	0.015	0.017	0.018	0.016	0.014	0.013
1	0.172	0.175	0.173	0.181	0.203	0.212	0.190	0.163	0.148
2	0.671	0.683	0.675	0.707	0.794	0.827	0.741	0.638	0.578
3	0.671	0.683	0.675	0.707	0.794	0.827	0.741	0.638	0.578
4+	0.671	0.683	0.675	0.707	0.794	0.827	0.741	0.638	0.578

Table 18: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: summary results of the a4a assessment.

Year	Catch (t)	SSB (t)	Rec (000)	F_{bar} (1-3)	Total biomass (t)
2002	5783.958	6496.6	976232.3	0.582	8651.3
2003	7849.099	5886.574	1020212	0.757	8786.3
2004	6948.712	5630.997	1049471	0.753	8446.7
2005	6029.773	6086.789	1049095	0.625	8952.8
2006	6216.208	6836.804	1062418	0.579	9020.0
2007	7111.231	6994.668	1071117	0.623	10201.1
2008	6967.196	6810.777	966960.9	0.630	9167.8
2009	6527.364	7133.46	980279	0.558	9409.6

2010	6158.054	7533.207	911820.4	0.504	10199.1
2011	6298.985	7518.375	909540.7	0.505	10271.3
2012	6434.058	7503.063	1018775	0.514	10243.8
2013	6267.729	7720.538	1097284	0.507	10004.8
2014	6763.379	7967.79	1167584	0.532	11032.3
2015	7697.222	7984.624	1167540	0.597	10718.5
2016	7921.408	7846.96	1162307	0.622	11253.3
2017	7324.48	8338.452	1101664	0.558	11283.7
2018	6740.711	9027.691	1058661	0.480	12223.6
2019	6617.944	9637.566	1056411	0.434	12764.9

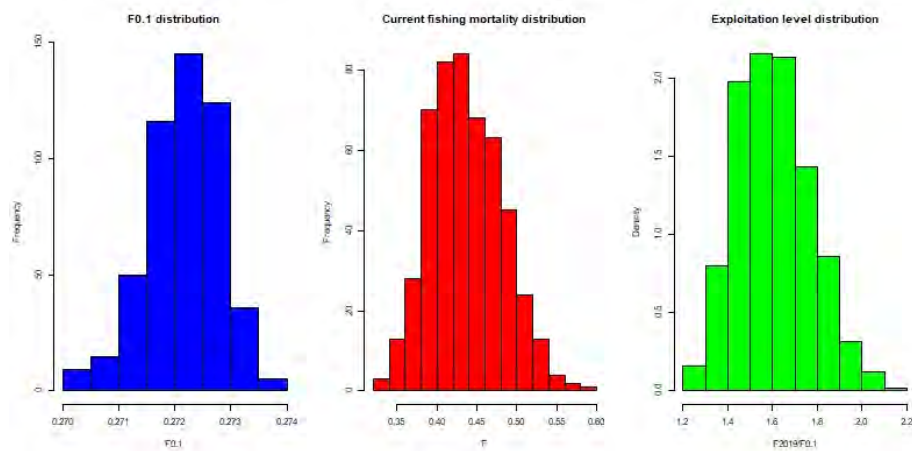


Figure 50: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F0.1 ratio) values.

Besides the model selection procedure, a further check was performed regarding the number of knots (k) of the smoother on year in the f model. A test based on AIC, BIC and GCV was performed on k ranging between 5 and 15 (Figure 51). All the model specifications highlight a consistent behavior in terms of main outcomes (Figure 52). A k value of 10 was retained as best balancing AIC/BIC reduction and GCV augmentation.

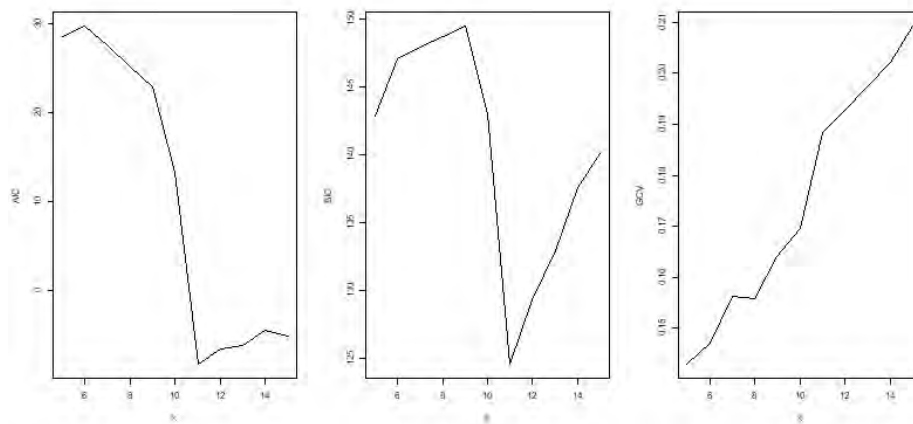


Figure 51: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: AIC, BIC and GCV values estimated on a range of k values of the smoother on year of the fmodel.

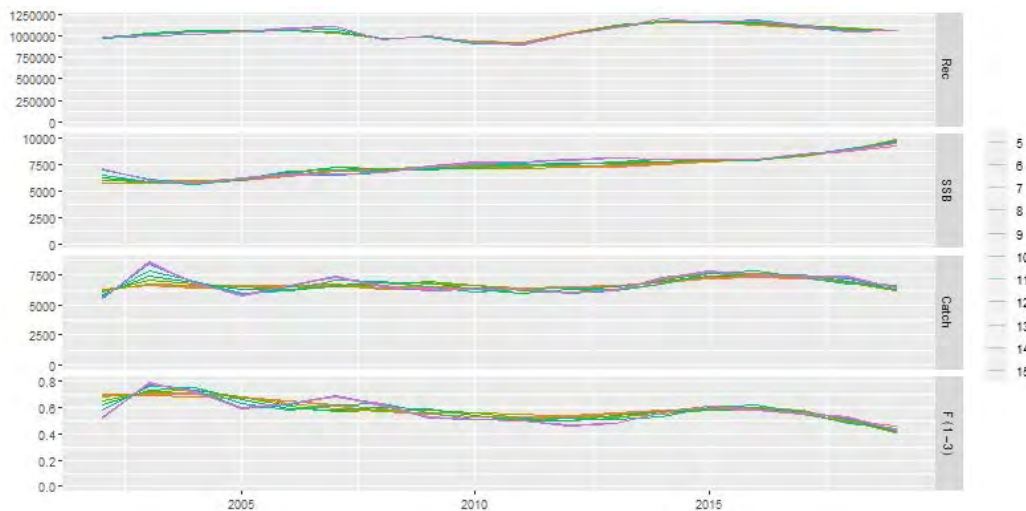


Figure 52: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: Outputs of model runs with different k values on the smoother on year in the fmodel.

4.3.4 Comparison with other assessments with a different stock configuration

Considering the output of the genetic and of the otolith microchemistry analysis, estimation of stock status of red mullet based on single GSAs is prone to overlook for stock boundaries. Red mullet stock assessment can be framed into the “second type of error” defined in the report introduction (section 2): in particular, this type of error occurs when only a portion of a stock is assessed as a closed unit. Ignoring complex population structure and stock connectivity can lead to misperception of the magnitude of fish productivity, which can translate to suboptimal utilization of the resource (Kerr *et al.*, 2017). Basing on this consideration, a broad stock unit is likely to accommodate for the assessment of the entire population structure, potentially improving the understanding of the stock dynamic. However, performing a stock assessment on such a large area implies several constraints, such as difficulty to accommodate local variations in the growth parameters and failing to observe local depletions.

In this paragraph are showed a set of comparisons of the new stock configuration to the GSAs based assessment for red mullet, to verify if there are evident improvement or deterioration of stock assessment quality, and to check the degree of coherence in the stock perception that can be found between the

single GSA assessment and the MED_UNITS framework. A brief discussion of the results is also provided.

Cohort consistency

To evaluate and compare the fit of data to biological assumptions, cohort consistency of stock configurations (for both MEDITS indices and FDI) were plotted for a visual comparison. For the present evaluation were available stock objects for GSAs 6, 7, 9 and 10 (retrieved from STECF EWG 20-09 report; STECF, 2020a), which were therefore used in the comparison. Figure 53 reports the cohort consistency plot calculated on the MEDITS indices. Different age classes were considered in the assessments, in particular the age 0 was excluded in the cases of GSA 9 and in the MED_UNITS aggregation. The overall cohort consistency of the MED_UNITS aggregation was slightly lower than GSAs assessment of areas 6 and 7, in particular for the age classes 2-3.

Figure 54 reports the cohort consistency plot calculated on the FDI information (catch at age), from which was possible to observe that correlation strength of the MED_UNITS aggregation was among the highest for the age classes 1-2, while the strength deteriorated for the oldest age classes. To sum up the cohort consistency comparison results, a little degradation of cohort consistency was observed in the MED_UNITS aggregation regarding both the MEDITS and FDI information.

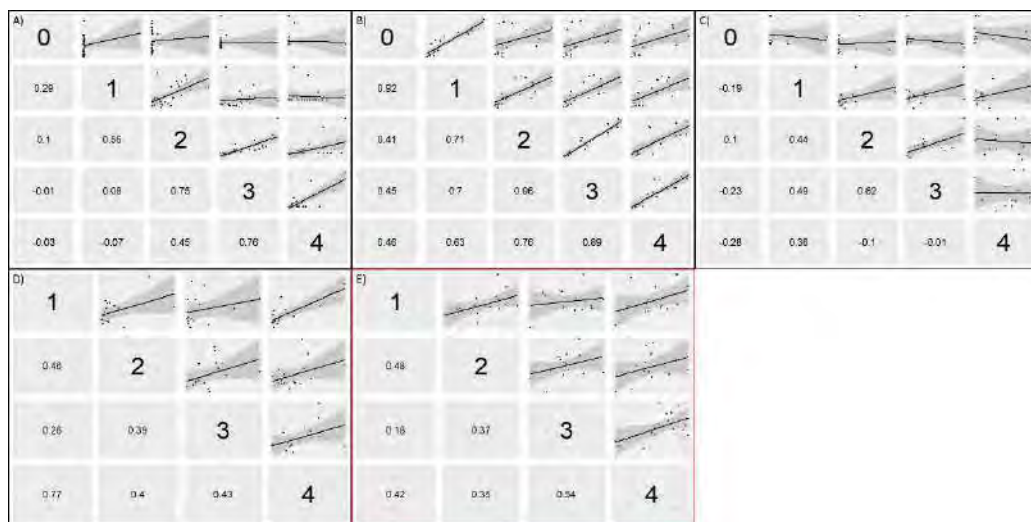


Figure 53: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: cohort consistency of the MEDITS index calculated on GSA 6 (A); 7 (B); 9(C); 10 (D); and on the MED_UNITS aggregation (E, red box).

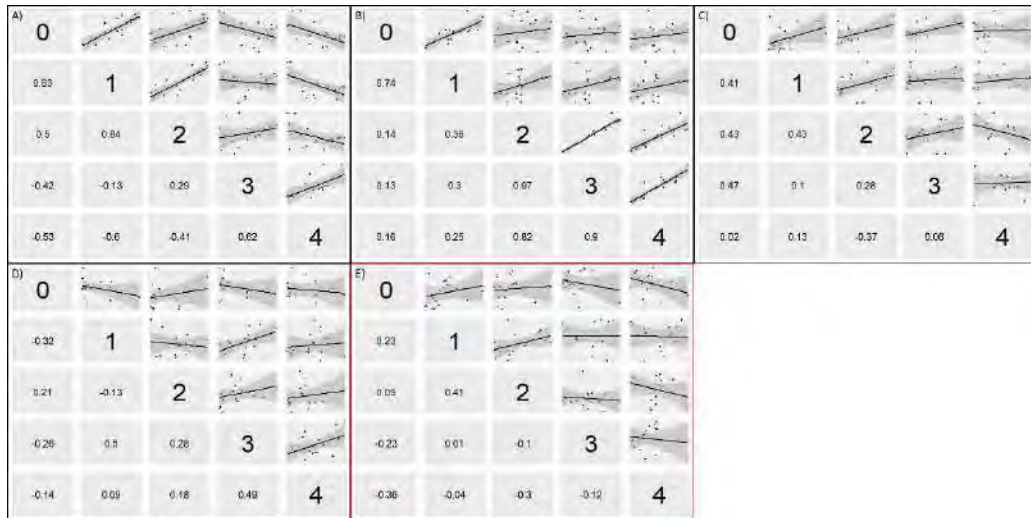


Figure 54: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: cohort consistency of the FDI calculated on GSA 6 (A); 7 (B); 9(C); 10 (D); and on the MED_UNITS aggregation (E; red box).

Model diagnostics

The diagnostics of the models for GSAs 1, 6, 7, 9 and 10 performed during the STECF EWG 20-09 (STECF, 2020a) were plotted for comparison with the diagnostic of the model presented in this report (Figure 55). In overall, neither the diagnostics for single GSAs models nor the diagnostics for MED_UNITS model showed severe issues. The magnitude of the residuals was comparable, and the only systematic deviation from the mean was observed for GSA 10. The major issue observed in the MED_UNITS diagnostic was the slight rising pattern in the Age class 1 for the MEDITS index, a pattern that was nevertheless also detected in the stock assessment of GSA 7.

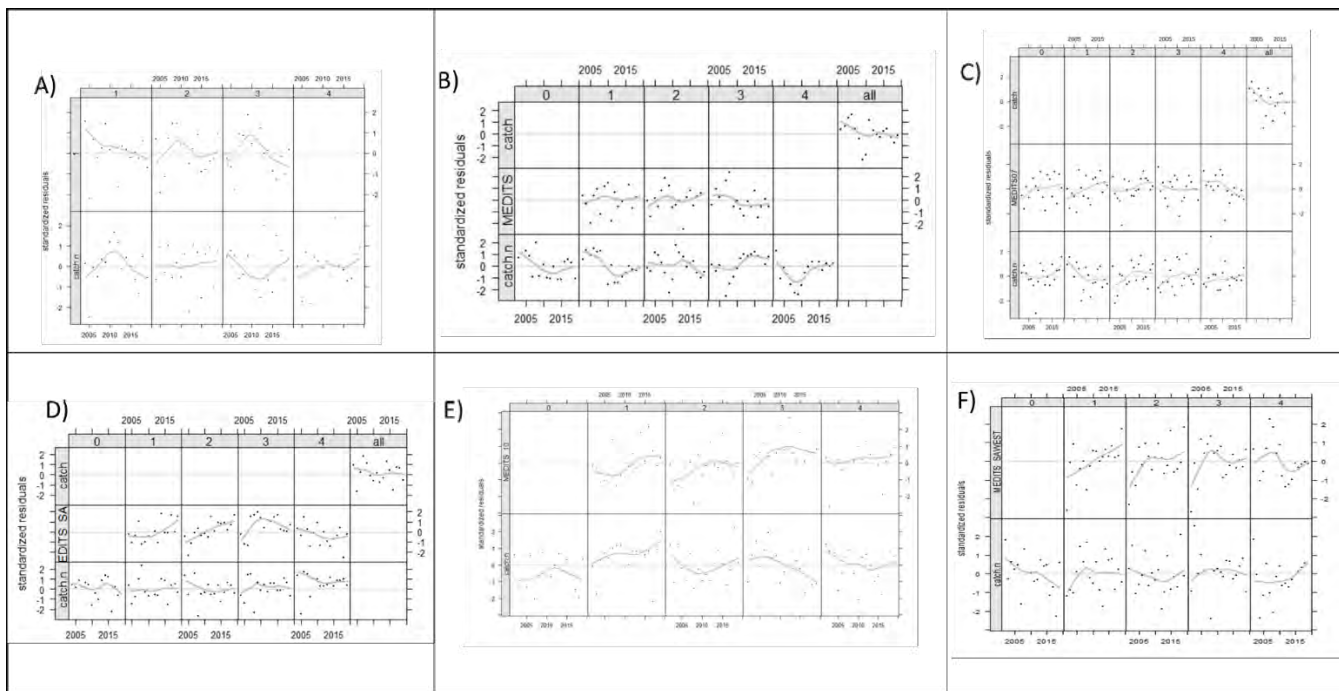


Figure 55: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: model diagnostic for the stock assessment of STECF EWG 20-09 for GSAs 1 (A), 6 (B), 7 (C), 9(D) and 10 (E); and on the MED_UNITS aggregation (F).

Retrospective patterns

The retrospective pattern observed for the MED_UNITS stock aggregation had a greater consistency than most of the single GSAs assessment (Figure 56). In particular, the models fitted on GSAs 6, 7, 9 and 10 were quite unstable on the recruitment trends, an issue that was not observed in the aggregated assessment. Also, the F trend showed instability for GSAs 1, 6 and 10. SSB and catch trend did not show a large issue in any of the models analyzed.

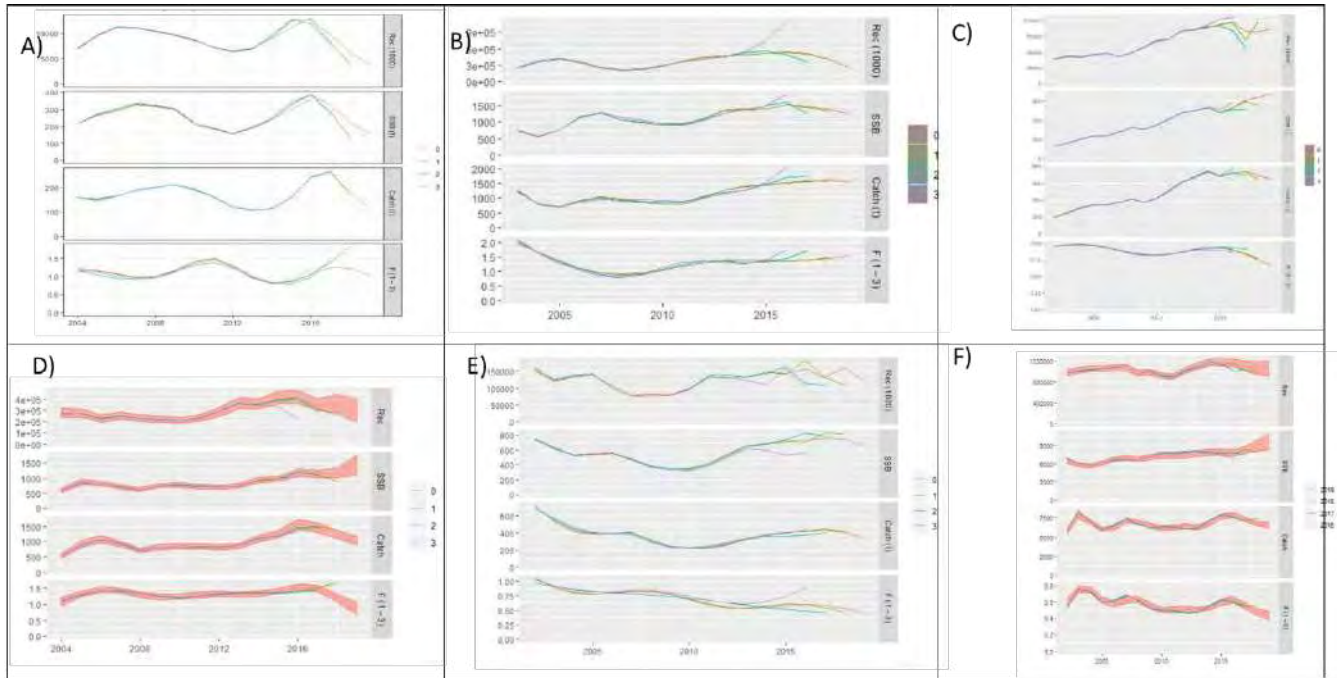


Figure 56: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: retrospective analysis for the stock assessment of STECF EWG 20-09 for GSAs 1 (A), 6 (B), 7 (C), 9(D) and 10 (E); and on the MED_UNITS aggregation (F).

Stock trajectories

The trajectories of most of the single GSA assessment (GSAs 1, 6, 7, 9, 10, 13, 14, and 16; > 90% average landings) were semi-quantitatively compared to the MED_UNITS aggregation to check the coherence for SSB and Recruitment trend estimates, as well as for Fbar and F/F01 trends (Figure 57). The comparison was restricted to the time-frame 2008-2016 due to data availability. SSB estimates were consistent in terms of magnitude, while the trend obtained by summing the single GSAs was increasing faster in the period 2013-2016. As a result, trends were moderately consistent, showing a Pearson correlation coefficient of 0.536. Recruitment estimates were 39% larger on average by considering the sum of the single GSAs assessment, while the trends were highly correlated, showing a Pearson correlation coefficient of 0.939. The Fbar and F/F01 trends were compared to the results obtained for single GSAs, including an estimation of the average annual value. Fbar figures in the period 2008-2016 were heterogeneous both in term of trend and magnitude: it was observed a slightly decreasing trend for GSAs 10, 13-14 and 16, a more stable trend for GSAs 1, and 7, and an increasing trend for GSAs 6, 9 with annual values ranging from 1.82 (GSAs 13-14 in 2012) to 0.36 (GSA 16 in 2016). The trend estimated for the MED_UNITS aggregation was slightly increasing and showed the largest correlation with the trend of GSA 9 (Pearson correlation coefficient of 0.59). The mean of the single GSAs Fbar estimates was on average 87% larger than the trend estimated for the MED_UNITS aggregation, which experienced one of the lowest values among the displayed trends. Nevertheless, when scaling the annual Fbar values by the most recent F01 estimate of each GSA, the F/F01 value estimated within the MED_UNITS framework took a value falling roughly on the mean of the single GSAs estimates.

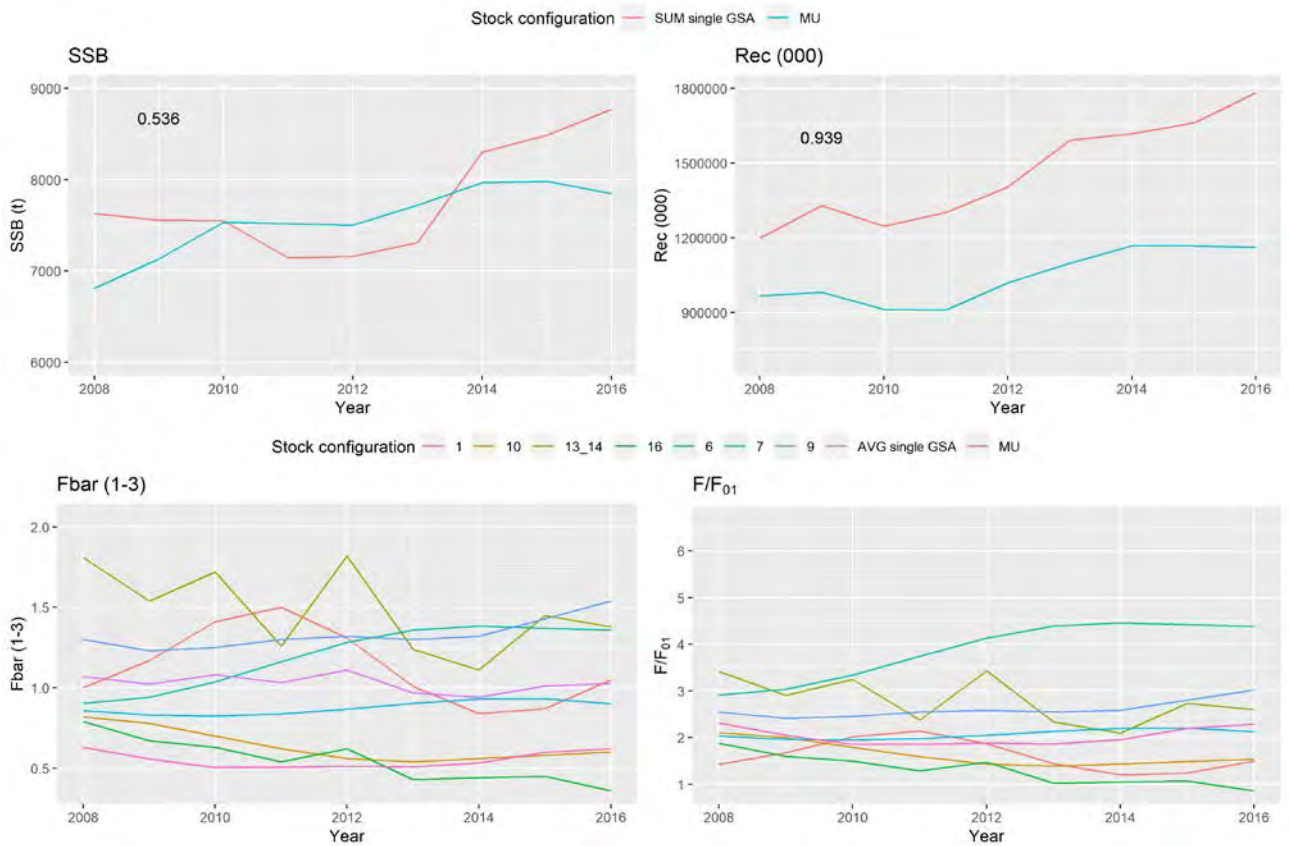


Figure 57: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16: cohort consistency of the stock trajectories calculated on single GSAs 1, 6, 7, 9, 10, 13, 14, and 16 and on the MED_UNITS aggregation. MU= MED_UNITS aggregation.

Fbar (terminal year) and reference points

F bar, F01 and their ratio were displayed for the last available year, namely 2019 for GSAs 1 to 10 (as well as for the MED_UNITS aggregation), 2017 for GSA 16 and 2016 for GSAs 13-14 (Figure 58). Like in Figure 21, where the data were limited to 2016, the Fbar estimated within the MED_UNITS framework was one of the lowest among the stock configuration considered. Nevertheless, also the F01 estimate was far below the mean. The resulting F/F01 final result was therefore close to the mean estimated from the single stock GSAs, where GSA 6 behaved as an outlier ($F/F01 \sim 5$). By taking out GSA 6 estimated from the mean value calculation (red dashed line in Figure 22), it results that MED_UNITS estimate for F/F01 was falling on the average of the single GSAs stock assessments.

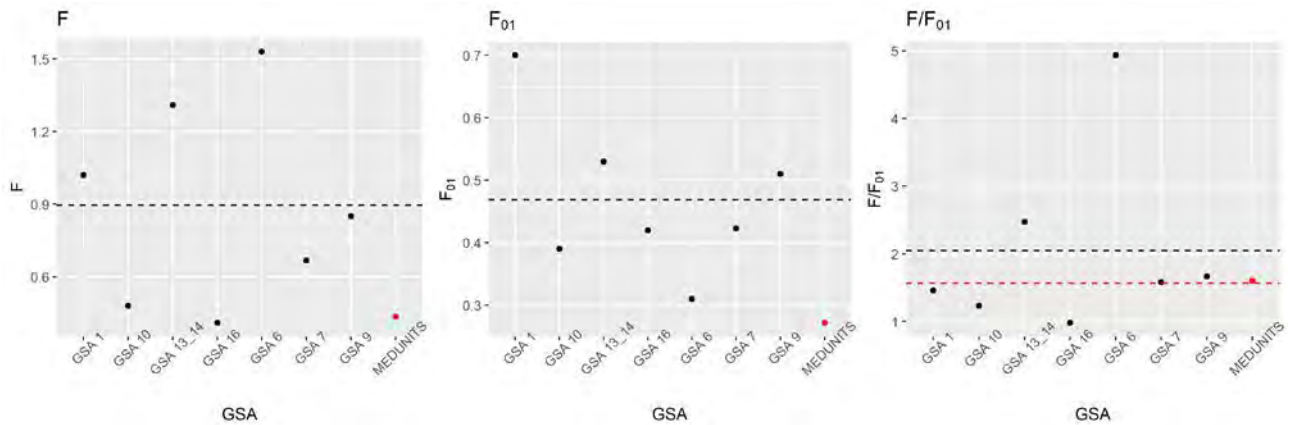


Figure 58: MUT in GSAs 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16 and on the MED_UNITS aggregation: Fbar, F01 and their ratio calculated on single GSAs. Values for 1, 6, 7, 9, 10 refers to 2019, values for GSAs 13 and 14 refers to 2016 and values for GSA 16

The comparison of cohort consistency indicated that few or no improvement in the age structure of the MUT stock in the MEDITS index was obtained by aggregating the western Mediterranean stocks, and deterioration was observed in the FDI information. Nevertheless, it is difficult to unravel the sources of heterogeneity that caused a poor age structure. Potential factors can be both biological and operational. Among the biological factors, there is the possibility that using a single set of growth parameters for the entire western Mediterranean population may not be suited for accounting for local variations, especially for a short-living species such as *Mullus barbatus* for which a great heterogeneity in the growth parameters used in the single GSAs stock assessments was observed. Regarding the operational factors, the most likely to undermine the aggregation of FDI information may be some degree of heterogeneity in the sampling effort over the considered areas, including the fact that FDI information was not available for some important GSAs. Indeed, a lack of cohort consistency was also observed for the single GSAs, and it is not possible to exclude that coupling some slightly biased FDI to unbiased FDI can undermine the overall result.

Regarding model diagnostics, no large difference was observed in the model fits, while a slight improvement in the retrospective patterns was observed in the MED_UNITS aggregation. In particular, recruitment and F trends were more stable than most of the single areas assessments. The observation on the stock trajectories prompted arguments both in favor and against the accuracy of the MED_UNITS aggregation assessment. The main argument in support of the MED_UNITS aggregation regarded the recruitment estimate: the pattern derived from the sum of single GSAs assessment was highly correlated with the MED_UNITS estimates, although it was overestimated by nearly 40%. Considering the instability observed in the retrospective pattern of single GSAs models, especially for the recruitment trend, it may be possible that this figure originated from the sum of a slight bias observed in most single GSAs assessment. As such, aggregating the western population may have enabled us to estimate the overall recruitment process with more precision.

The main observation against the MED_UNITS aggregation was the heterogeneity in the F pattern values and trajectories. On one hand, the MED_UNITS F/F₀₁ trend was well representing the overall perception coming from the single assessment, and this observation may support the hypothesis to consider the western population behaving as a single stock unit. On the other hand, the single assessment may be more suited to give insights on the local exploitation, a piece of information that can easily be lost by giving a unique trend for the entire area. For example, the F trends estimated for the Sicily channel (GSAs 13-14 and 16) were declining, while the trends estimated for the westernmost part of the basin (GSAs 1 and 6) were cycling or even increasing. As a result, the aggregated assessment was potentially underestimating the risk of local depletion.

4.4 Red mullet in GSAs 17-20.

In the central Mediterranean, red mullet (MUT), *Mullus barbatus*, represents one of the most valuable fish species for the trawling fleet operating on the continental shelf, mainly below 200m depth. According to the outcomes of the genetic analysis, as well as of the otoliths shape and microchemistry analysis, performed under the Med_Units project, a stock configuration including GSAs 17, 18, 19 and 20 was identified and used to perform an assessment (Figure 59).

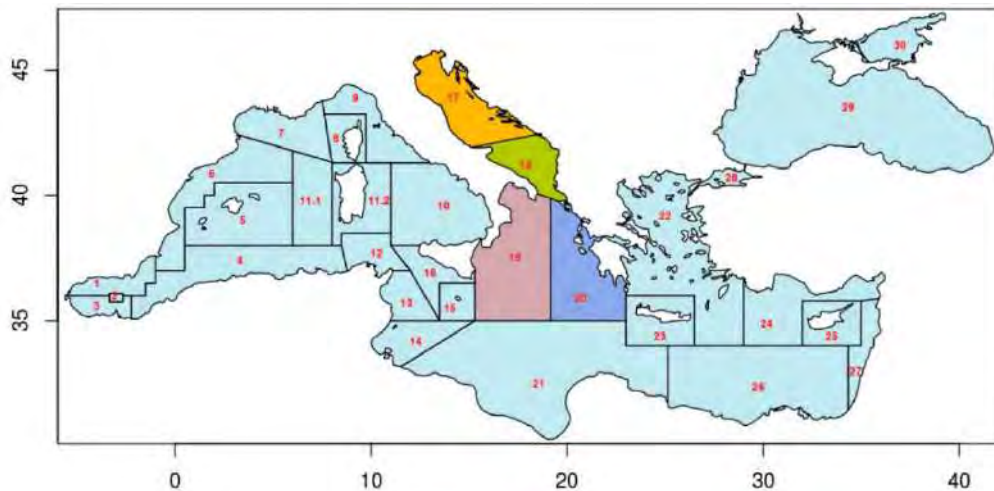


Figure 59: MUT in GSAs 17, 18, 19 and 20. Geographical location of the stock

4.4.1 Input data

Fishery-Dependent data

Commercial data (landing and discard data, LFDs and biological parameters) for the EU countries considered in the stock configuration (Italy, Croatia, Slovenia) were collected from the EU DCR/DCF and from STECF EWG 20-15 and GFCM WGSAD 2020 reports. Croatia commercial data (2006-2012) not present in DRC/DCF were collected instead from Recfish project database. Albanian and Montenegrin commercial data were available from STECF EWG 20-15 (STECF, 2020b). Greek commercial data (2006-2019) were collected from HCMR database (Table 20). In Table 21 proportion of landing by countries and GSAs are reported.

LFDs were sliced (deterministic slicing by sex) to obtain age structure using the growth parameters of Carbonara et al. (2018) (Table 19). The choice was driven by the consistency of these parameters with the most recent age reading protocol for *Mullus barbatus*.

Table 19: MUT in GSAs 17, 18, 19 and 20. VBGF and LW parameters by sex.

Sex	Linf	k	t0
F	30	0.244	-0.61
M	26	0.264	-0.73
	a		b
F	0.0089		3.098
M	0.0086		3.106

Table 2019: MUT in GSAs 17, 18, 19 and 20. Summary of the commercial data used to build the stock object. Green color indicates year data from Recfish project database for Croatia summed up with DCF data for Slovenia. Cream color indicates reconstructed data for ALB. The values indicates catch amount in (tons).

Year	17_HRV_SVN	17_ITA	17_18_ALB_MNE	18_ITA	19_ITA	20_GRC
2002				4911.1	2272.7	
2003				2369.5	2451.8	376.2
2004				2063.6	951.0	396.7
2005				1449.5	1113.7	309.8
2006	808.43	3101	395	1934.4	887.4	231.0
2007	958.18	3298	395	1802.2	541.1	289.3
2008	829.56	3158	397	960.8	447.8	278.5
2009	848.25	2433	395	1046.0	539.5	268.9
2010	794.75	1979	393	681.2	538.1	175.8
2011	1110.11	2693	390	551.1	588.4	354.2
2012	1292.98	1850	414	2530.3	486.6	236.1
2013	1089.75	2270	408	1269.2	274.5	223.7
2014	1157.28	2845	362	1391.8	252.5	209.5
2015	1132.38	3130	428	1676.7	303.7	207.3
2016	956.75	2541	436	1535.5	278.0	257.1
2017	992.43	3740	432	632.7	298.8	235.0
2018	850.73	4550	335	1186.7	552.1	406.0
2019	751.38	2329	415	972.7	462.1	345.0

Table 2120: MUT in GSAs 17, 18, 19 and 20. Proportion of total catch in the stk object by each country and GSA. From red high to green low values.

17_HRV_SVN	17_ITA	18_ALB_MNE	18_ITA	19_ITA	20_GRC
0.11	0.42	0.05	0.26	0.12	0.03
0.13	0.45	0.05	0.25	0.07	0.04
0.14	0.52	0.07	0.16	0.07	0.05
0.15	0.44	0.07	0.19	0.10	0.05
0.17	0.43	0.09	0.15	0.12	0.04
0.20	0.47	0.07	0.10	0.10	0.06
0.19	0.27	0.06	0.37	0.07	0.03
0.20	0.41	0.07	0.23	0.05	0.04
0.19	0.46	0.06	0.22	0.04	0.03
0.16	0.46	0.06	0.24	0.04	0.03

0.16	0.42	0.07	0.26	0.05	0.04
0.16	0.59	0.07	0.10	0.05	0.04
0.11	0.58	0.04	0.15	0.07	0.05
0.14	0.44	0.08	0.18	0.09	0.07

Fishery-Independent data

LFDs (n/km²) from the MEDITS surveys in GSAs 17, 18, 19 and 20, available from MedUnits datacall, were sliced (deterministic age slicing by sex) using the same set of growth parameters used for the commercial data. GSA 20 data were available for 2006, 2008, 2014, 2018 and 2019 only (See Table 22).

Greek LFDs available under EU DCR/DCF present some missing years. These missing years were filled in with LFDs of close years. Natural mortality vector was estimated using the Chen and Watanabe model (Table 23) for consistency with the last GSA 17, 18, 19 and 20 red mullet stock assessments. The maturity vector used in the same stock assessments was also used). Plus group was set to 4 + according to the majority of the available stock assessments for this species.

The catch at age structure as well as the abundance at age resulting from the slicing are presented respectively in Figure 61 and Figure 62. Moreover, the internal consistencies of such input data are presented respectively in Figure 63 and Figure 64.

Table 22: MUT in GSAs 17, 18, 19 and 20. Available Medits data. Green values from 2006 - 2019 have been used to build the index object

year	GSA_17_ITA		GSA_17_HRV		GSA_17_SVN		GSA_18_ITA_ALB_MNE		GSA_19_ITA		GSA_20_GRC	
	density	biomass	density	biomass	density	biomass	density	biomass	density	biomass	density	biomass
1994	215.6	7.1					21.3	1.0	264.6	7.9	303.9	14.7
1995	375.6	10.1					9.4	0.5	694.6	17.9	211.7	9.1
1996	239.9	7.5					151.9	2.9	281.5	7.9	261.7	7.5
1997	232.8	7.9					72.1	2.7	106.5	3.6	867.9	13.6
1998	212.8	7.5					106.5	3.6	260.3	7.0	1115.0	31.8
1999	2684.3	45.3			137.8807	12.9	995.8	7.8	61.0	2.5	867.4	29.0
2000	602.7	10.7			23.908	1.2	101.0	5.0	147.6	6.0	386.2	16.0
2001	407.2	13.7			NA	NA	154.9	6.2	302.7	11.6	1303.0	50.3
2002	689.6	15.5	597.4	25.3	518.735	12.2	123.4	3.7	432.0	14.2	NA	NA
2003	295.9	8.5	993.6	39.6	542.2232	9.6	90.1	4.0	262.1	8.2	1120.2	31.2
2004	319.6	7.4	1579.0	54.4	24.39682	0.0	90.0	3.5	610.5	14.6	1592.6	47.3

2005	311.8	8.0	1261.2	47.8	1786.65	71.2	520.1	9.2	428.4	13.6	729.0	27.8
2006	653.0	14.6	1475.3	54.9	NA	NA	117.6	5.2	459.5	14.1	961.9	34.9
2007	312.4	6.8	822.7	31.7	27.2285	0.4	365.5	10.5	4876.4	32.5	NA	NA
2008	362.8	10.9	1881.2	65.6	NA	NA	227.7	11.1	2679.3	75.7	1359.8	49.9
2009	282.6	9.3	1371.4	48.5	35.74586	0.1	301.8	11.4	340.8	10.0	NA	NA
2010	314.8	8.7	1735.0	61.9	11.89602	0.0	289.9	9.1	987.1	22.3	NA	NA
2011	697.6	13.5	1219.8	38.4	NA	NA	638.2	13.3	483.6	14.3	NA	NA
2012	1123.3	23.1	3228.4	87.8	35.27619	1.2	4919.2	36.1	531.2	14.0	NA	NA
2013	1946.0	37.4	4870.5	121.2	NA	NA	2734.3	40.6	1511.0	33.0	NA	NA
2014	3717.6	76.5	5621.8	160.0	885.763	25.3	5644.5	36.7	5808.3	54.7	863.7	25.9
2015	1733.3	29.0	5519.1	145.1	222.0888	5.9	1699.3	28.9	1634.1	36.2	NA	NA
2016	5676.8	68.1	3162.9	97.4	342.2774	0.7	3649.4	31.1	563.9	19.5	1271.6	49.3
2017	6604.8	100.7	4759.2	130.3	70.21576	1.3	8089.4	99.6	5325.8	71.2	NA	NA
2018	8681.5	93.5	4637.5	104.9	279.9961	7.4	2841.0	82.1	3822.7	61.4	2626.2	80.9
2019	3957.4	57.5	4669.9	132.6	1466.431	49.5	431.5	12.8	2806.4	31.5	2128.7	78.9

Table 21: MUT in GSAs 17, 18, 19 and 20. Natural mortality and maturity-at-age vectors

Age	Natural mortality (combined)	Proportion of matures (combined)
0	0.94	0.00
1	0.62	1.00
2	0.5	1.00
3	0.43	1.00
4+	0.39	1.00

MUT GSA 17 18 19 20 ITA ALB MNE HRV SVN GRC Catches age structure

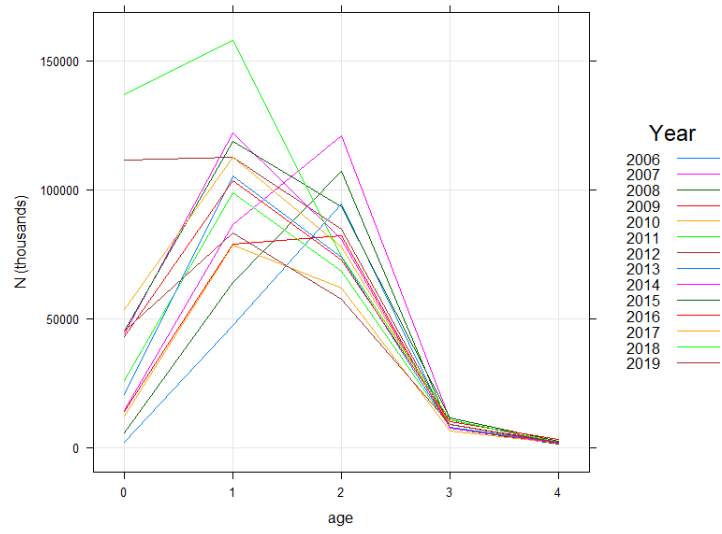


Figure 60: MUT in GSAs 17, 18, 19 and 20. Catch numbers-at-age

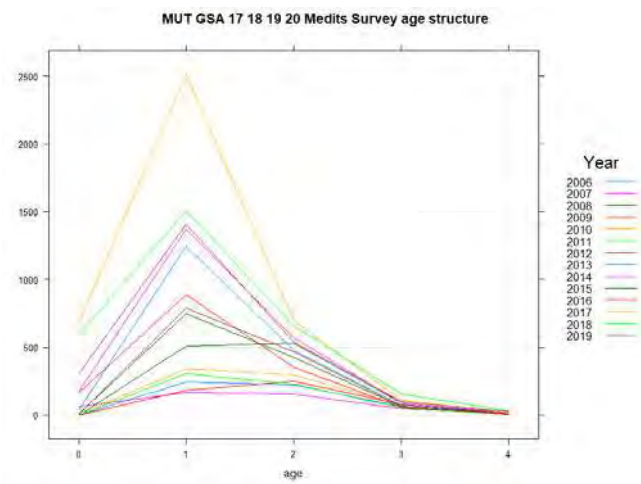


Figure 61: MUT in GSAs 17, 18, 19, and 20: catch numbers-at-age in the MEDITS

MUT GSA 17 18 19 20 ITA ALB MNE HRV SVN GRC Cohorts consistency in the catch

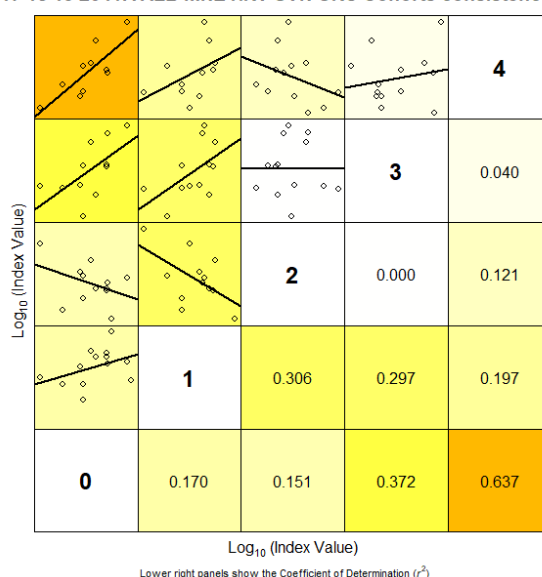


Figure 62: MUT in GSAs 17, 18, 19 and 20. Cohort consistency in the catch numbers-at-age.

MUT GSA 17 18 19 20 Cohorts consistency in Medits survey

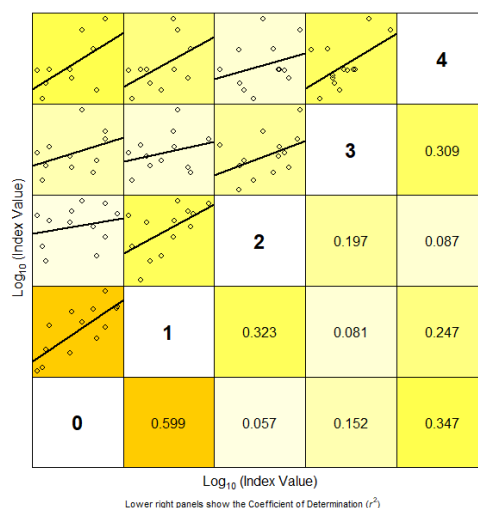


Figure 63: MUT in GSAs 17 18 19 and 20. Cohort consistency in the catch numbers-at-age in the MEDITS

4.4.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment using a4a was carried out using as input data the period 2006-2019 for the catch data and 2006-2019 for the tuning file (MEDITS indices).

Length-frequency distributions of commercial catches and surveys were split by sex and then transformed in age classes (plus group was set at age 4) using length-to-age slicing with different growth parameters by sex. Then, the assessment was performed by sex combined. No correction on t_0 was adopted for consistency with the most recent stock assessment of red mullet in GSA 17 and 18 (STECF EWG 20-15; STECF, 2020b) and in GSA 19 (GFCM, 2021).

After the slicing of the reconstructed LFDs, the number of individuals by age was SOP corrected [SOP = Landings / $\sum a$ (total catch numbers at age x x catch weight-at-age a)].

In the catches, a plus group at age 4 was set. A true age 4 was used in the survey.

Given that the landings were composed mainly of individuals between 1 and 3 years, these ages were selected as F_{bar} range.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $f_{\text{model}} = \sim s(\text{replace}(\text{age}, \text{age} > 3, 3), k = 3) + s(\text{year}, k = 7)$

Catchability sub-model: $q_{\text{model}} = \text{list}(\sim s(\text{replace}(\text{age}, \text{age} > 2, 2), k = 3, \text{by} = \text{breakpts}(\text{year}, 2012)))$

SR sub-model: $\text{srmod} \sim \text{geomean}(\text{CV} = 0.2)$

The $n1_{\text{model}}$ and v_{model} used in the final fit are the default ones.

4.4.3 Outputs from the model

The diagnostics and outputs of the best run are presented in the following figures. The model estimated 36 parameters out of 140 observations; it is then below the threshold of 25% ratio between parameters and observations (Figures 64-69).

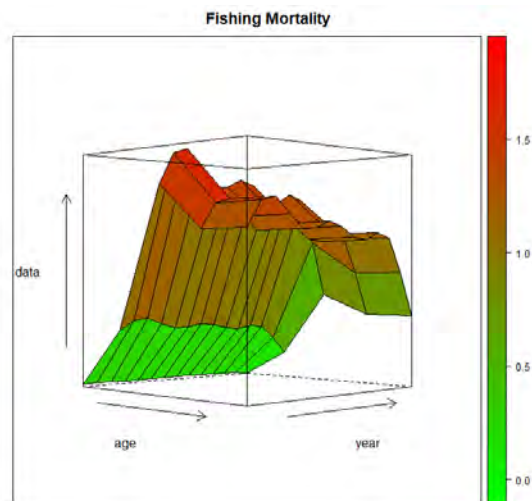


Figure 64: MUT in GSAs 17, 18, 19 and 20. Fishing mortality by age and year obtained from the a4a model (2006-2019).

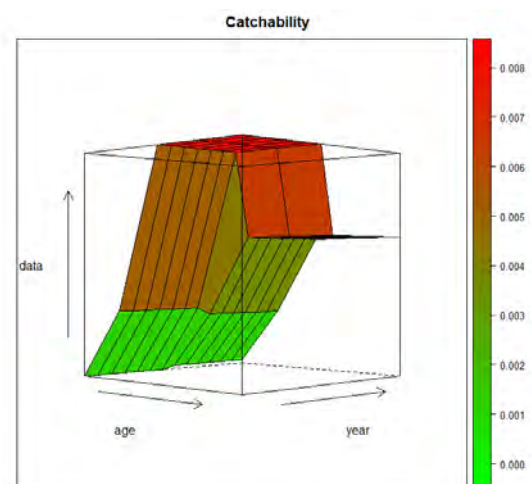


Figure 65: MUT in GSAs 17, 18, 19 and 20. Catchability of the survey by age and year obtained from the a4a model (2006-2019).

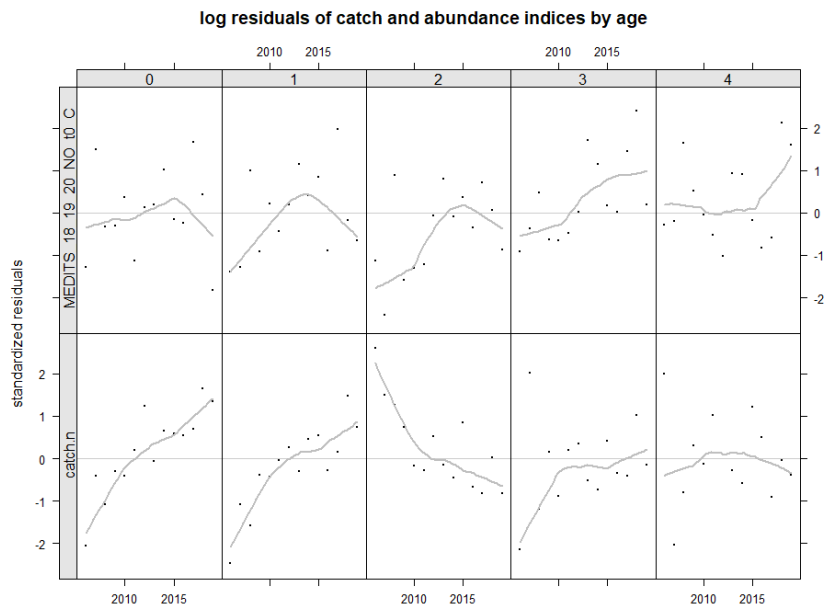


Figure 66: MUT in GSAs 17, 18, 19 and 20. Log residuals for the catch-at-age data of the fishery and the survey.

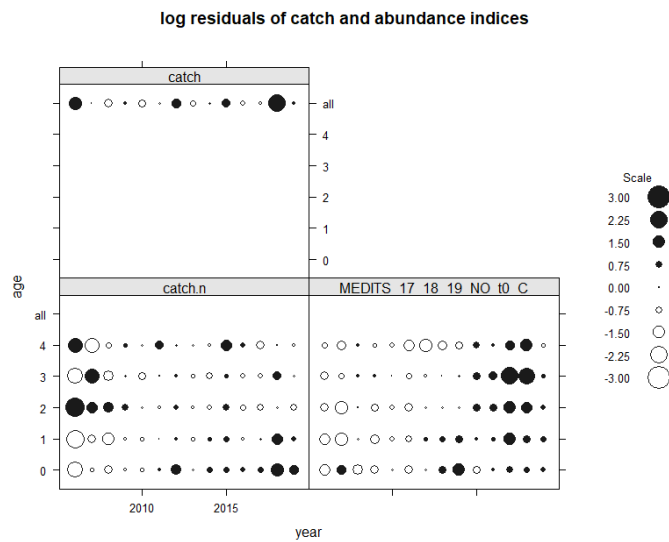


Figure 67: MUT in GSAs 17, 18, 19 and 20. Bubble plot of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

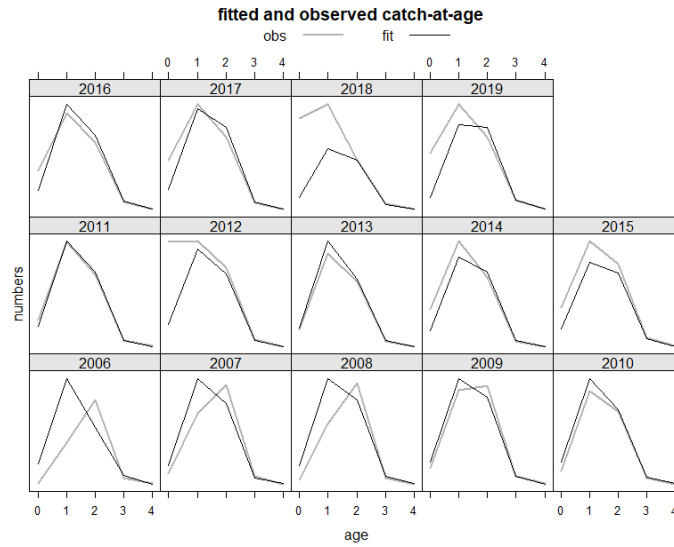


Figure 68: MUT in GSAs 17, 18, 19 and 20. Fitted vs observed values by age and year for the catches.

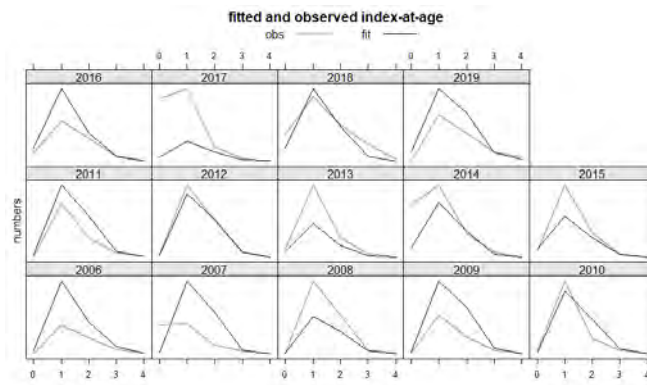


Figure 69: MUT in GSAs 17, 18, 19 and 20. Fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated and showed that biomass of the plus group (age 4+) was about 4.8% of the total SSB on average (Table 24).

Table 22: MUT in GSAs 17, 18, 19 and 20. Cryptic biomass estimates by year.

year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
age														
4+	3.72	3.58	4.22	4.96	3.74	4.86	3.05	4.05	4.84	5.65	6.83	6.31	5.69	5.85

The retrospective analysis shows that the model is quite stable with some small comb effect in F, SSB and recruitment (Figure 70). The Mohn's Rho for F and SSB and recruitment are within the optimal range - 0.2/0.2 (-0.094 for F, 0.15 for SSB, and -0.064 for recruitment).

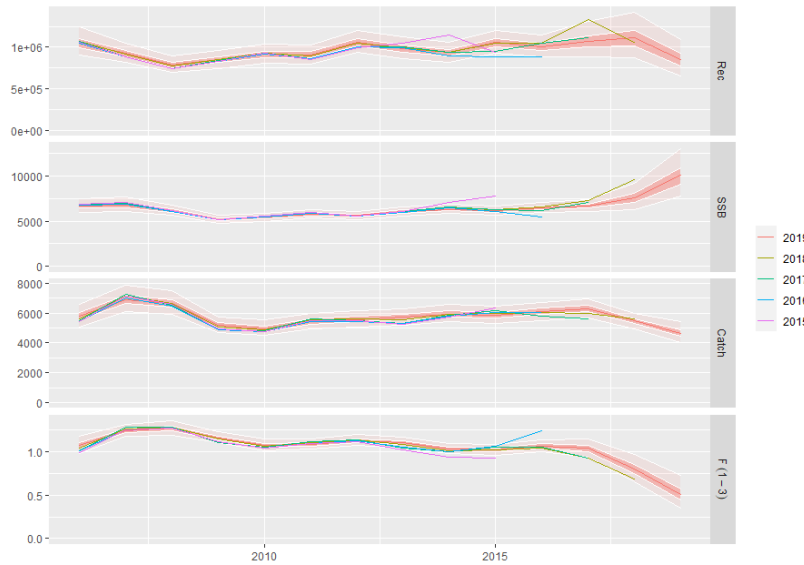


Figure 70: MUT in GSAs 17, 18, 19 and 20. Retrospective analysis.

Stock numbers-at-age, fishing mortality-at-age and summary results of the model are presented respectively in tables 25, 26 and 27 and in Figure 71, while Figure 72 summaries the histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F01 ratio) values.

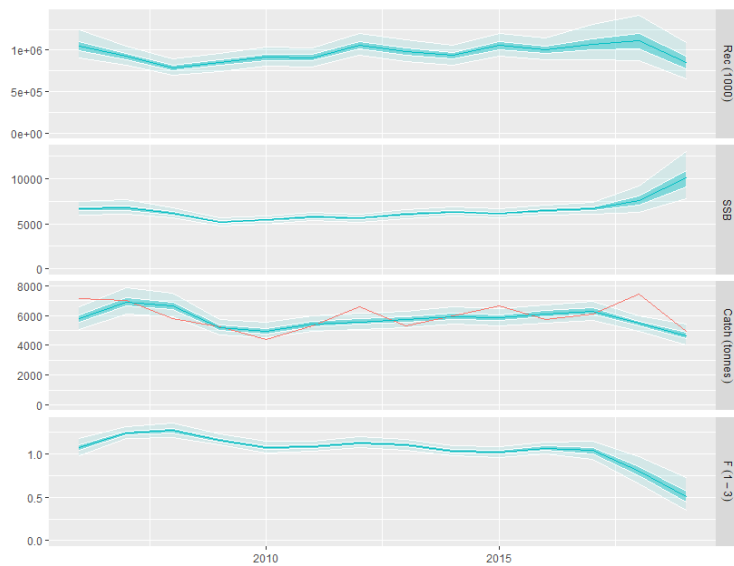


Figure 71: MUT in GSAs 17, 18, 19 and 20. Outputs of the a4a stock assessment model, with uncertainty; input catch data (red line) are plotted against the estimated catches.

Table 23: MUT in GSAs 17, 18, 19 and 20. Stock numbers-at-age (thousands).

age/year	2006	2007	2008	2009	2010	2011	2012	2013
0	1052779	925524.3	783222.6	840824.2	918633.4	931883.9	1071778	975398.9
1	415757.6	398739.8	348642.7	294842.5	317722.7	348075.3	352883.1	405263.9
2	95652.8	139507.2	124335.5	107746.2	95879	107206.2	116497.3	115794.9
3	17237.7	12232.9	14022.9	12136.2	12430.8	12487.8	13595.3	13845.2
4+	2433.2	3679.3	2464.9	2487.5	2523.4	2842.8	2854.6	2907

age/year	2014	2015	2016	2017	2018	2019
0	920275.4	1057785	1032493	1119858	1217335	1071040
1	369241.3	349316.4	401739.2	391399.8	424970.3	465854.7
2	135043.4	127610.6	121647.3	136405.6	134816.5	163934
3	14474.4	19028.3	18436.2	16172.7	19010.3	27253.4
4+	3082.3	3554.9	4661.9	4468.9	4149.1	6257.4

Table 24: MUT in GSAs 17, 18, 19 and 20. Fishing mortality-at-age

age/year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.01
1	0.48	0.55	0.56	0.51	0.47	0.48	0.50	0.48	0.45	0.44	0.46	0.45	0.34	0.20
2	1.56	1.80	1.83	1.66	1.54	1.57	1.64	1.58	1.46	1.44	1.52	1.48	1.10	0.66
3	1.25	1.44	1.46	1.33	1.23	1.25	1.31	1.27	1.17	1.15	1.22	1.18	0.88	0.53
4+	1.25	1.44	1.46	1.33	1.23	1.25	1.31	1.27	1.17	1.15	1.22	1.18	0.88	0.53

Table 25: MUT in GSAs 17, 18, 19 and 20. Summary results of the a4a assessment

Year	Catch	Recruitment	SSB	Fbar	Total Biomass
2006	5808	1052779.0	6698.775	1.10	21666.0
2007	7002	925524.3	6743.501	1.26	14575.1
2008	6621	783222.6	6165.202	1.28	19604.5
2009	5141	840824.2	5185.755	1.17	10950.9
2010	4933	918633.4	5392.162	1.08	18078.5
2011	5450	931883.9	5785.578	1.10	19113.4
2012	5667	1071777.7	5701.755	1.15	20043.1
2013	5810	975398.9	6125.81	1.11	20010.2
2014	5921	920275.4	6332.013	1.03	20022.4
2015	5791	1057785.3	6169.604	1.01	20294.5
2016	6094	1032493.2	6529.836	1.07	21264.4
2017	6338	1119857.8	6857.685	1.03	22745.3
2018	5508	1217335.4	8014.014	0.77	23834.8
2019	4683	1071039.9	11094.63	0.46	26235.2

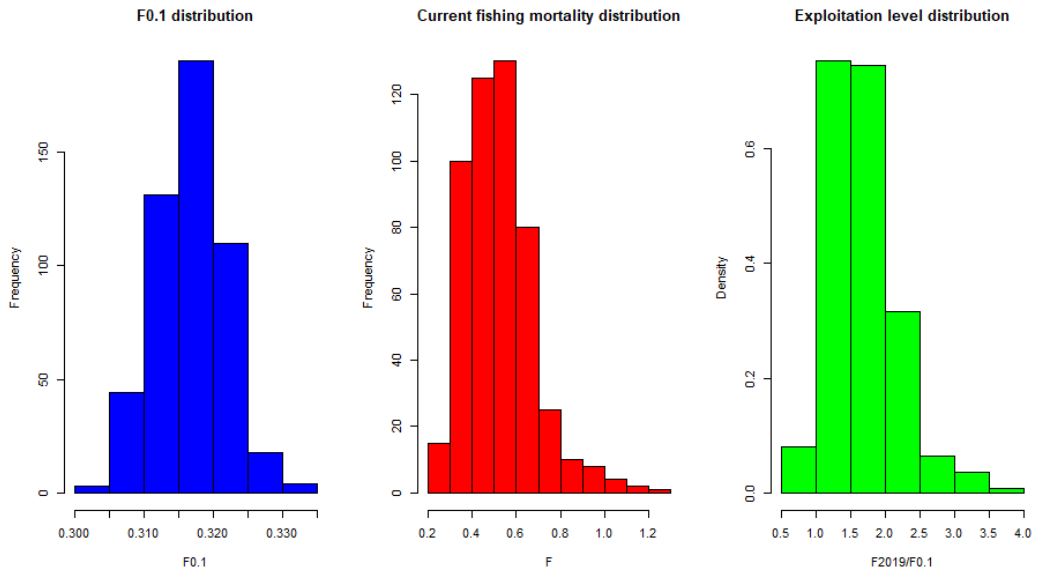


Figure 72: MUT in GSAs 17, 18, 19 and 20. Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F01 ratio) values.

Besides the model selection procedure, a further check was performed regarding the number of knots (k) of the smoother on year in the fmodel. A test based on AIC, BIC and GCV was performed on k ranging between 5 and 10 (Figure 73 and Figure 74). All the model specifications highlight a consistent behavior in terms of main outcomes (Figure 74). A k value of 7 was retained as the one providing the best retrospective.

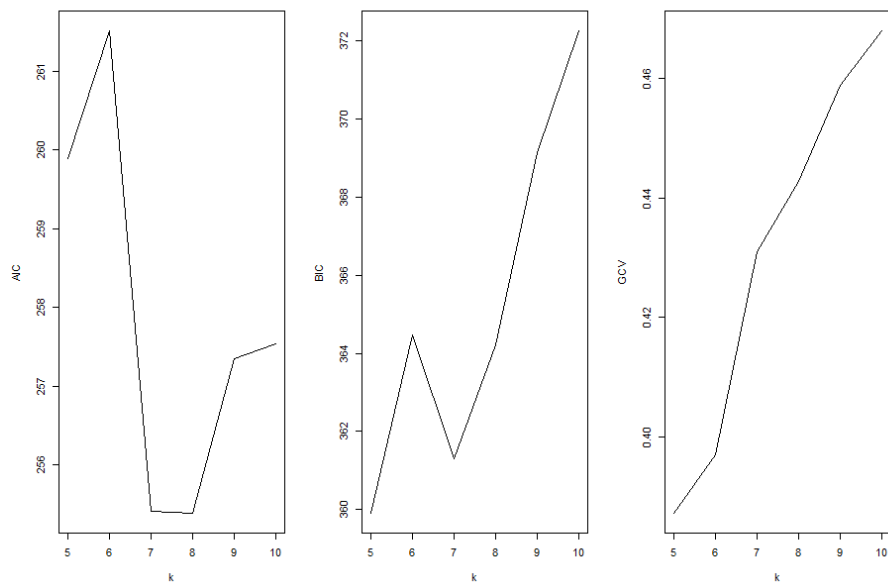


Figure 73: MUT in GSAs 17, 18, 19 and 20. AIC, BIC and GCV values estimated on a range of k values of the smoother on year of the fmodel.

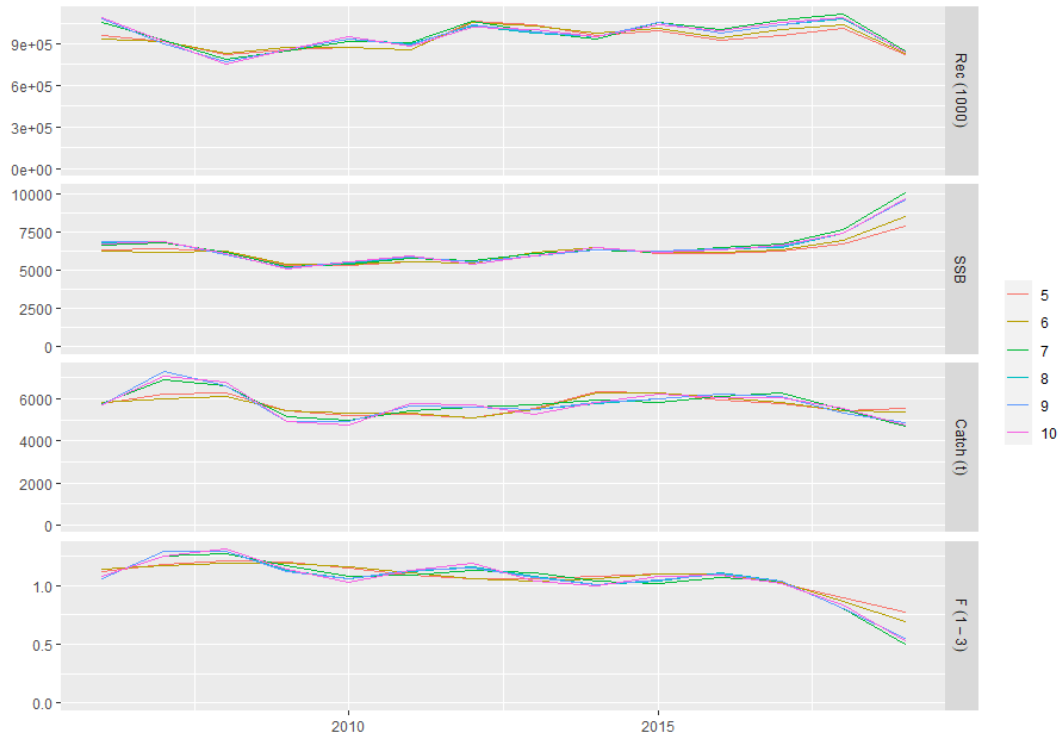


Figure 74: MUT in GSAs 17, 18, 19 and 20. Outputs of model runs with different k values on the smoother on year in the fmodel.

Reference point were calculated in FLBRP library. Considering the F current of 0.46 estimated for 2019, the fishing mortality level is above the reference point F0.1 of 0.32.

4.4.4 Comparison with other assessments with a different stock configuration

Considering the output of the genetic and of the otolith microchemistry analysis, estimation of stock status of red mullet based on single GSAs is prone to overlook for stock boundaries. Red mullet stock assessment can be framed into the “second type of error” defined in the report introduction (section 2): in particular, this type of error occurs when only a portion of a stock is assessed as a closed unit. Ignoring complex population structure and stock connectivity can lead to misperception of the magnitude of fish productivity, which can translate to suboptimal utilization of the resource (Kerr *et al.*, 2017). Basing on this consideration, a broad stock unit is likely to accommodate for the assessment of the entire population structure, potentially improving the understanding of the stock dynamic. However, performing a stock assessment on such a large area implies several constraints, such as difficulty to accommodate local variations in the exploitation pattern and failing to observe local depletions.

In this paragraph are showed a set of comparisons of the new stock configuration to the GSAs based assessment for red mullet, to verify if there are differences in the stock assessment quality, and to check the degree of coherence in the stock perception that can be found between the single GSA assessment and the MED_UNITS framework. A brief discussion of the results is also provided.

Cohort consistency

To evaluate and compare the fit of data to the different assumptions on stock boundaries, cohort consistency of stock configurations (for both MEDITS indices and commercial data) were plotted and statistically compared with a t-test. Hypothesis 0 was that the mean of the Pearsons correlation coefficient calculated on the single GSAs was not different from the value calculated on the MED_UNITS aggregation.

Figure 75 reports the cohort consistency plot calculated on the MEDITS indices, from which it is possible to observe that overall cohort consistency of the MED_UNITS aggregation was slightly lower than single GSAs 17_18 and GSA 19 assessment. The statistical test based on the entire set of correlations on the MEDITS indices are showed in Table 28.

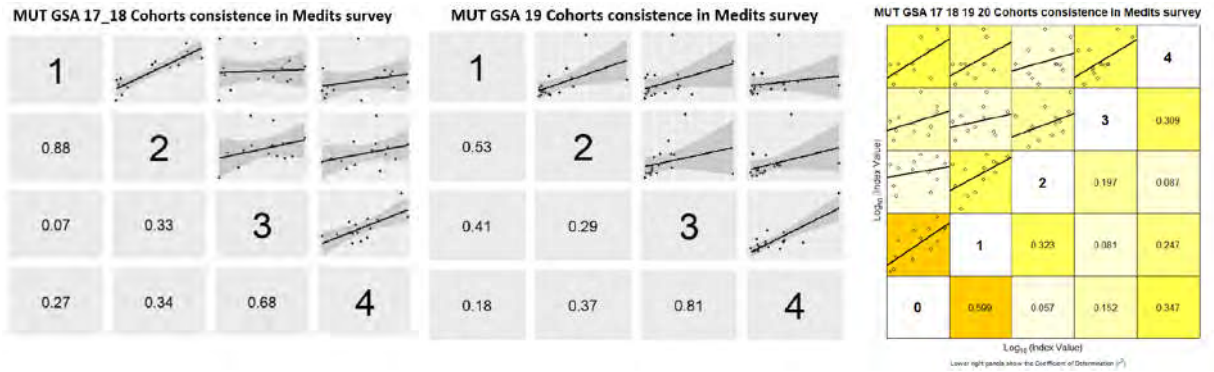


Figure 75: MUT in GSAs 17, 18, 19 and 20: cohort consistency of the MEDITS index calculated on GSA 17 + 18 (left), GSA 19 (central) and on the MED_UNITS aggregation (right).

Table 26: MUT in GSAs 17, 18, 19 and 20. Statistical comparison of cohort consistency in the MEDITS, between MED_UNITS configuration and those calculated on GSA 17 18 and GSA 19. MU= MED_UNITS

GSA	Mean R	t_test (p-val)
17-18	0.295	0.059
19	0.229	0.005
MU	0.632	-

Figure 76 reports the cohort consistency plot calculated on the commercial information (catch at age), from which was possible to observe a lower consistency of the MED_UNITS aggregation compared to single GSAs especially between age 1-2. The statistical test does not show any relevant difference between the three configurations analyzed (Table 29).

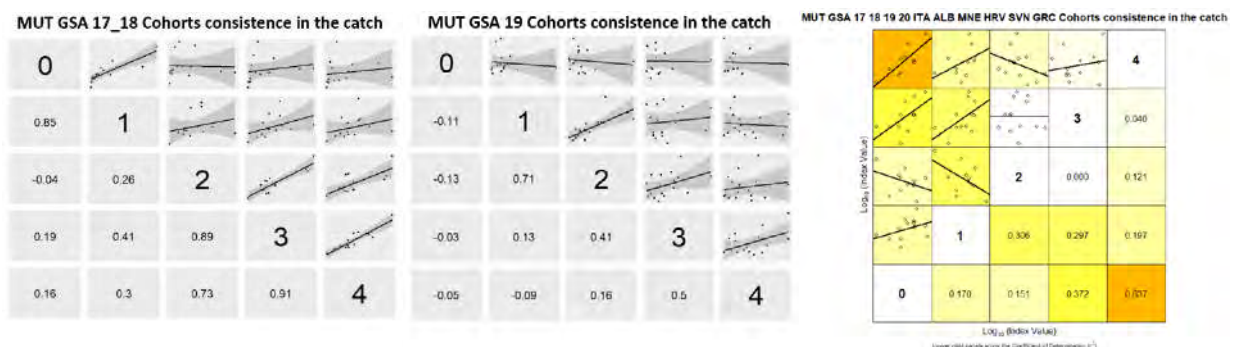


Figure 76: MUT in GSAs 17, 18, 19 and 20: cohort consistency of the catch calculated on GSA 17 + 18 (left), GSA 19 (central) and on the MED_UNITS aggregation (right).

Table 27: MUT in GSAs 17, 18, 19 and 20. Statistical comparison of cohort consistency in the catch, between MED_UNITS configuration and those calculated on GSA 17 and 18. MU= MED_UNITS

GSA	Mean R	t_test (p-val)
17_18	0.465	0.184
19	0.352	0.486
MU	0.181	-

Model diagnostics

The diagnostics of the models for GSAs 17 and 18 (performed during the STECF EWG 20-15; STECF, 2020b) and for GSAs 19 (performed during the GFCM WGSAD 2020) were plotted for comparison with the diagnostic of the model presented in this report (Figure 19). In overall, the diagnostics catch for GSAs 17_18 models have less trend than MED_UNITS model. By contrast Medits index present more trend in GSA 17_18 than in MED_UNITS model. The magnitude of the residuals was comparable. No relevant patterns are presented in GSA 19 (Medits log residual bubbles plot, Figure 77).

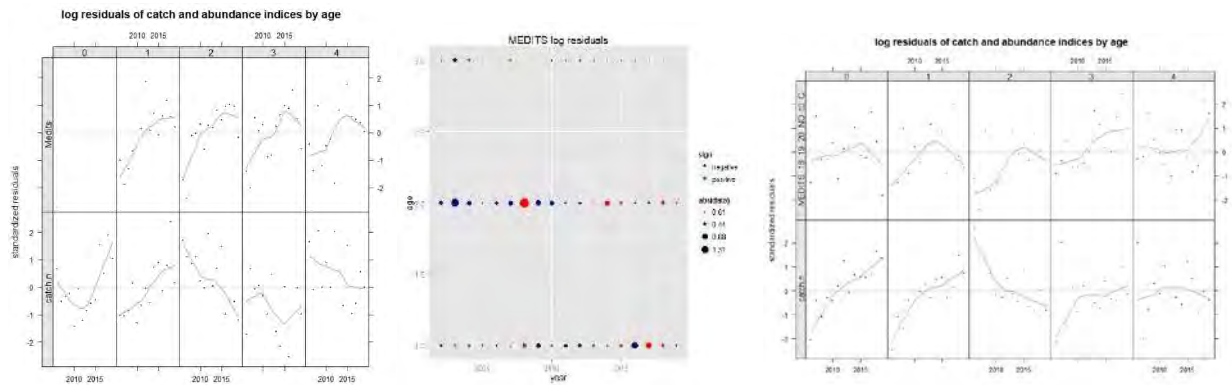
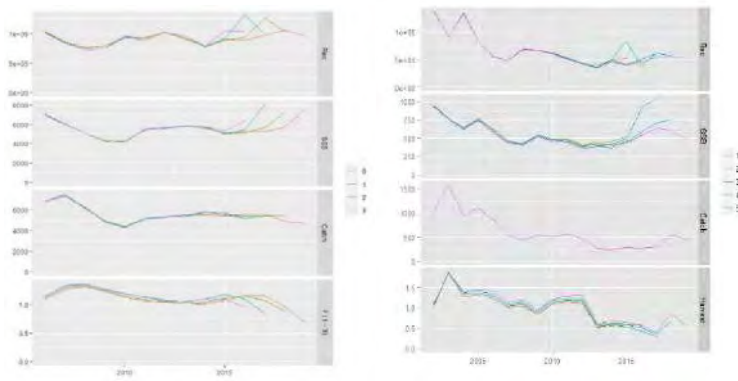


Figure 77: MUT in GSAs 17, 18, 19 and 20: model diagnostic for the stock assessment of STECF EWG 20-09 for GSAs 17 18 (left), GSA 19 (central) and on the MED_UNITS aggregation (right).

Retrospective patterns

The retrospective analysis observed for the MED_UNITS stock aggregation shows more consistency and stability than in GSAs 17_18 and in GSA 19 assessments (Figure 78). In particular, the models fitted on GSAs 17_18 and in GSAs 19 were quite unstable on the recruitment trends, SSB and F.



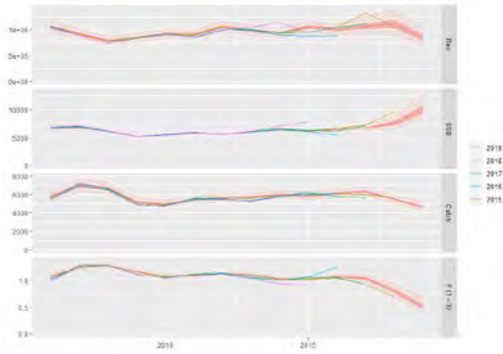


Figure 78: MUT in GSAs 17, 18, 19 and 20. Retrospective analysis for the stock assessment of STECF EWG 20-09 for GSAs 17 18 (left), GSA 19 (central) and on the MED_UNITS aggregation (right).

Stock trajectories

The comparison was restricted to the time-frame 2006-2019 due to data availability. The trajectories of F_{bar} in GSA 17_18 and aggregated GSAs are quite similar. The trajectory of F_{bar} in GSAs 19 is lower in the period from 2013-2017 compared to the MED_UNITS aggregation and GSAs 17_18 (Figure 79). The trajectories of F/F_{01} in the MED_UNITS aggregation are higher than in single GSAs assessments. The comparisons for SSB and Recruitment were done by summing the contribution of single GSAs, then data were plotted and Pearson correlation coefficient was computed. SSB estimates were consistent in terms of magnitude. On an overall basis, the trends are consistent, showing a Pearson correlation coefficient of 0.798. Recruitment estimates are larger in the period 2014-2019 by considering the sum of the single GSAs assessment, while the trends are highly correlated, showing a Pearson correlation coefficient of 0.835.

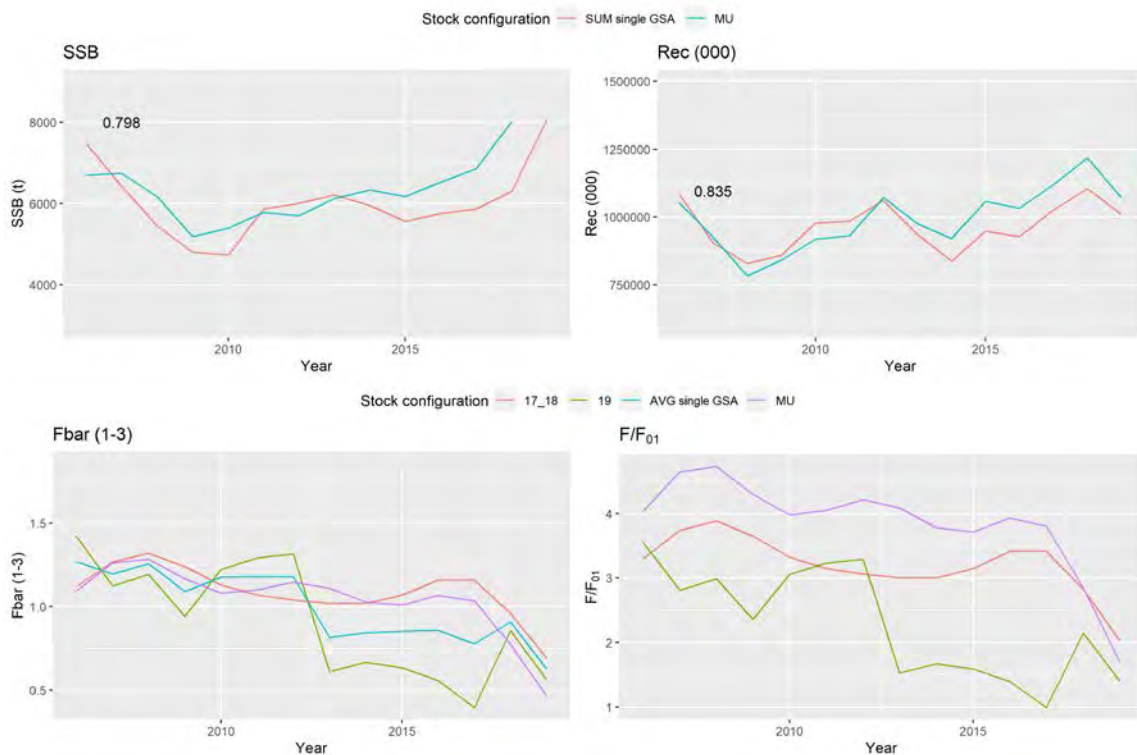


Figure 79: MUT in GSAs 17, 18, 19 and 20. Cohort consistency of the stock trajectories calculated on GSAs 17 and 18 and on the MED_UNITS aggregation. MU= MED_UNITS aggregation.

Fbar (terminal year) and reference points

F bar, F01 and their ratio were displayed for the last available year (2019) for GSAs 17_18, GSA 19 and Med_Units aggregation. F bar, F01 and their ratio are all higher in the GSAs 17 + 18 compared to GSA 19 and Med_Units aggregation (Figure 80).

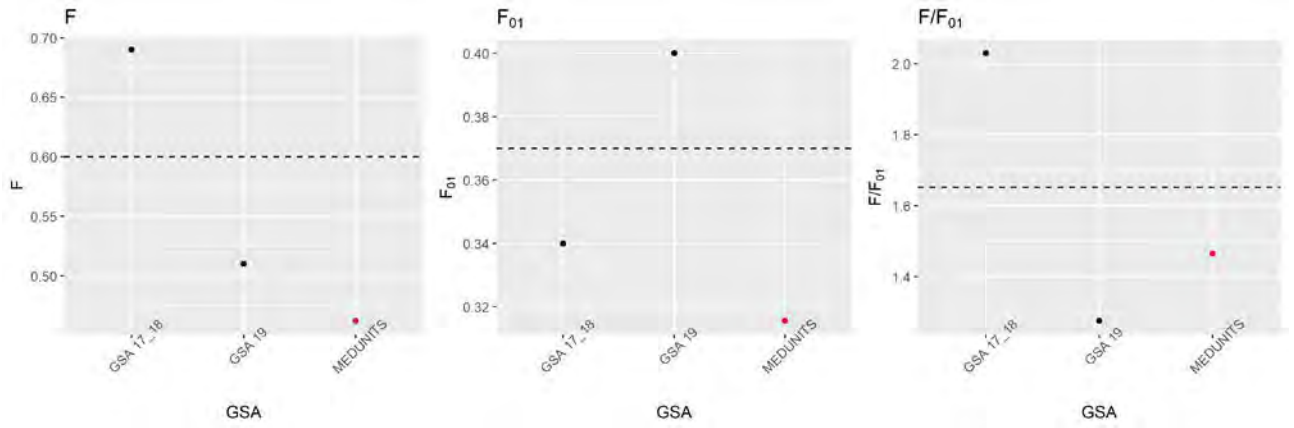


Figure 80: MUT in GSAs 17, 18, 19 and 20 and on the MED_UNITS aggregation: Fbar, F01 and their ratio calculated on single GSAs.

4.5 Red mullet in GSAs 22-25

In the eastern Mediterranean, red mullet (MUT), *M. barbatus*, represents a target resource for the trawling fleet operating on the continental shelf, mainly below 200m depth. Based on D3.2 including a synthesis of the various analyses carried out under WP1 and WP2 of the project a stock configuration including the eastern GSAs (22-25) was as identified (Figure 81).

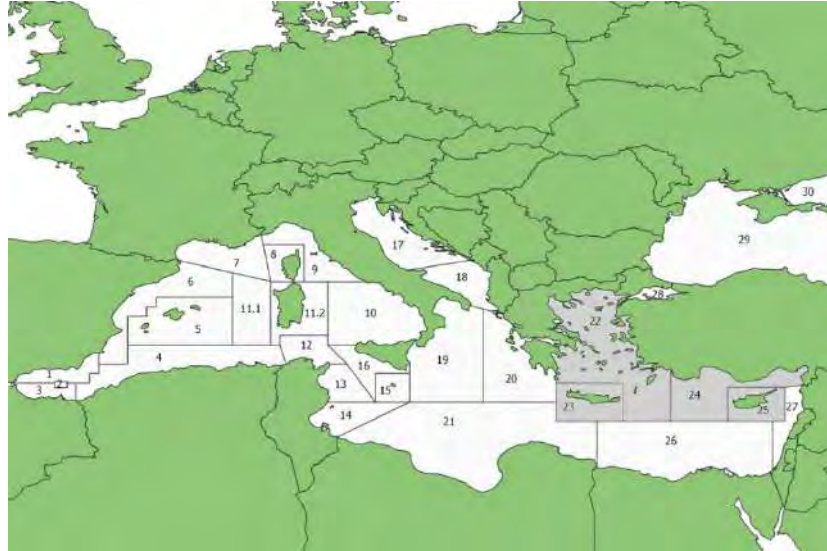


Figure 81: MUT in GSAs 22-25. Geographical location of the stock

4.5.1 Input data

Fishery-Dependent data

Commercial data (catch data, LFDs and biological parameters) for the EU countries considered in the stock configuration (Greece, Cyprus) were collected from the EU DCR/DCF database. Greek fleets exploit GSAs 22 and 23, while Cypriot fleets exploit GSA 25. It should be noted that due to large gaps in the Greek data (missing years: 2007, 2009-2013, 2015, 2017), owing to non-continuous implementation of the DCF, landing estimates for GSAs 22 and 23 were based on FAO statistics and the size composition of the catch in GSAs 22 and 23 for the missing years was assumed to be similar to that of GSA 25. Regarding non EU countries, the total landings of Turkey in GSAs 22 and 24 were considered, as retrieved from the FAO database. In this case it was also assumed that the size composition of the catch follows the same pattern observed in the DCF data of the EU member states. The discontinuity of catch-at-size data in GSAs 22 and 23, as well as the inclusion of catches from non EU countries resulted to a large inflation of the available catch-at-size data to reflect the total catch levels. No discards were assumed given that relevant information was scarce.

LFDs were sliced (deterministic slicing) to obtain age structure using the growth parameters employed for assessing GSA 22 during the STECF EWG 20-15 (STECF, 2020b; Table 1) and correcting the t_0 value to account for a birth date at the middle of the year. The choice of GSA 22 was driven by the fact that it has by far the highest landings (~70% of the total in the latest years) among the considered GSAs. Similarly, the assumed natural mortality and maturity vectors, as well as the length-weight relationship, were in accordance to those used in the STECF EWG 20-15 assessment of GSA 22 (Table 30 and Table 31).

Table 28: MUT in GSAs 22-25. Von Bertalanffy growth (VBGF) and length-weight relationship parameters.

	VBGF			Length/weight	
	Linf	k	t0	a	b
All sexes	326	0.17	-1.78	0.00885	3.07

Table 29: MUT in GSAs 22-25. Proportion of mature and natural mortality (M) at age

Age	1	2	3	4	5
Maturity	0.72	0.89	0.98	1	1
M	0.61	0.54	0.50	0.50	0.50

Figure 82 illustrates the estimated age composition of the total catch by year.

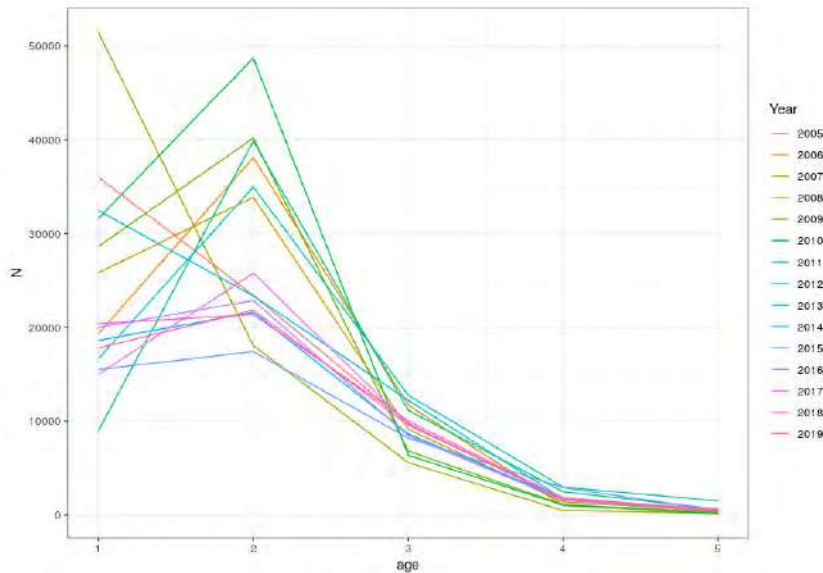


Figure 82: MUT in GSAs 22-25. Catch numbers-at-age

Fishery-Independent data

LFDs (n/km^2) from the MEDITS surveys in GSAs 22, 23 and 25 were sliced using the same growth parameters used for the commercial data. The data used covered the period 2005-2019 but similar gaps with those observed in the fishery dependent data existed for GSAs 22 and 23. Figure 83 illustrates the estimated age composition of the MEDITS survey data by year.

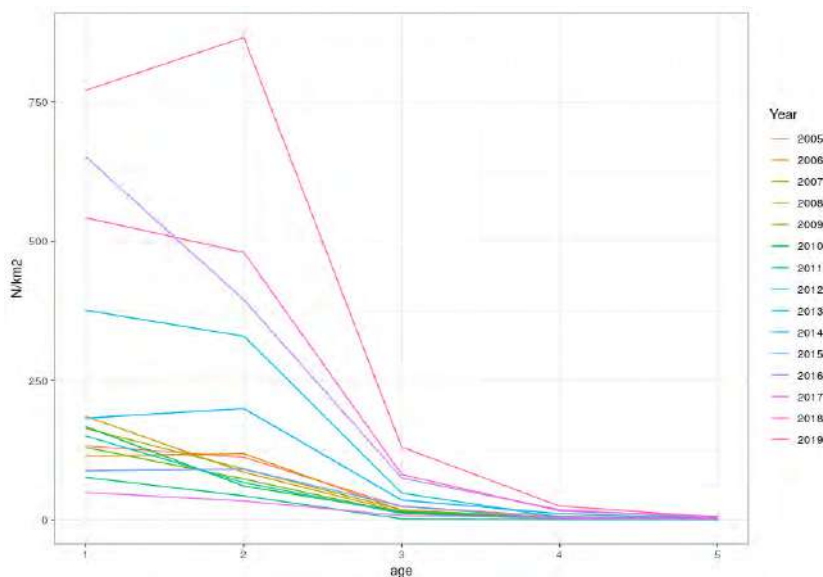


Figure 83: MUT in GSAs 22-25. Catch numbers-at-age in the MEDITS.

Cohort consistency plots for the fishery and survey data are shown in Figure 84 and Figure 85 respectively.

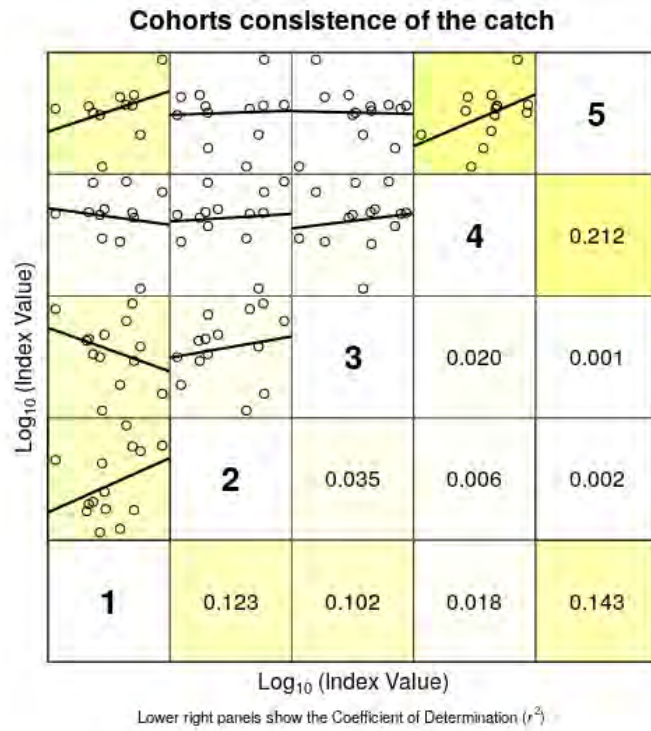


Figure 84: MUT in GSAs 22-25: cohort consistency in the catch numbers-at-age.

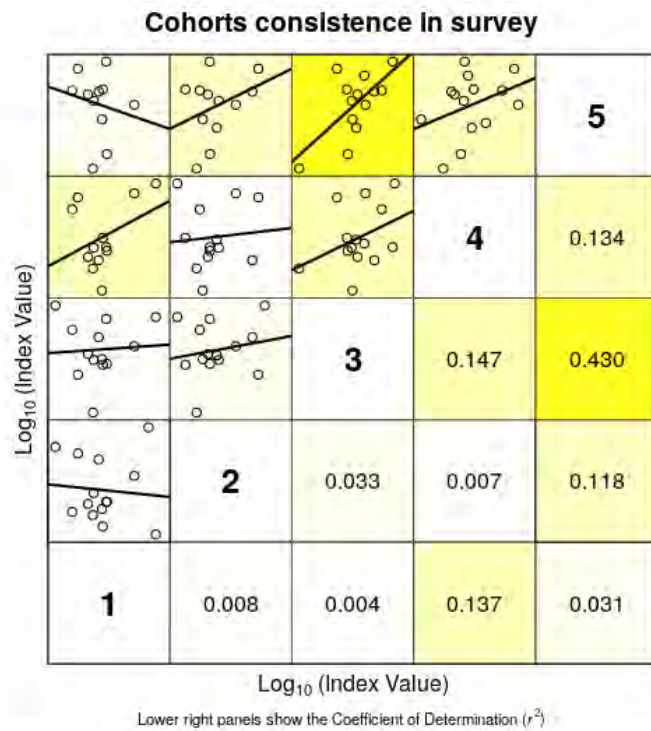


Figure 85: MUT in GSAs 22-25: cohort consistency in the catch numbers-at-age in the MEDITS.

4.5.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment using fisheries and survey data of the 2005-2019 period. A plus group of age 5 was set for the fisheries data, while the number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers at age $a \times$ catch weight-at-age a)], to account for the missing catch-at-size data. A true age 5 was used in the survey and F_{bar} range was fixed at 1-3.

The final model settings showing the best diagnostics outputs were:

Fishing mortality sub-model: $fmodel = \sim te(age, year, k=c(4,5)) + s(year, k=3) + s(age, k=3)$

Catchability sub-model: $qmodel = -list(\sim s(replace(age, age > 3, 3), k=3, by = breakpts(year, 2009)) + s(year, k=3))$

SR sub-model: $srmodel \sim geomean(CV=0.2)$

4.5.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in figures 86-91.

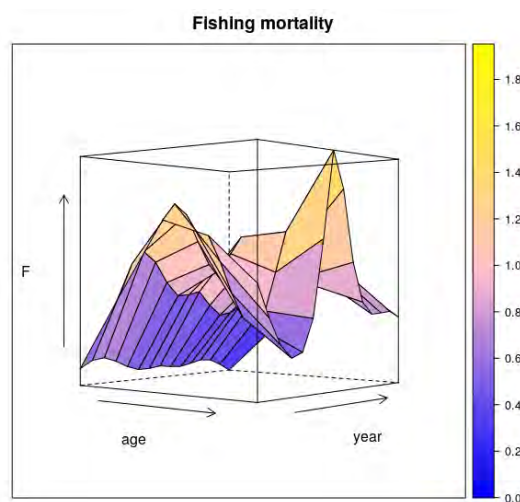


Figure 86: MUT in GSAs 22-25: fishing mortality by age and year obtained from the a4a model (2005-2019).

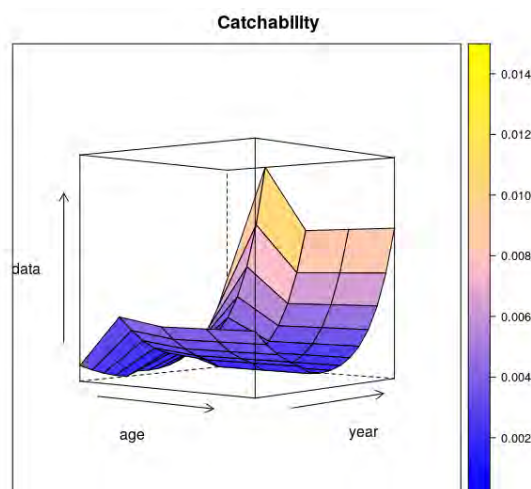


Figure 87: MUT in GSAs 22-25: catchability of the survey by age and year obtained from the a4a model (2005-2019).

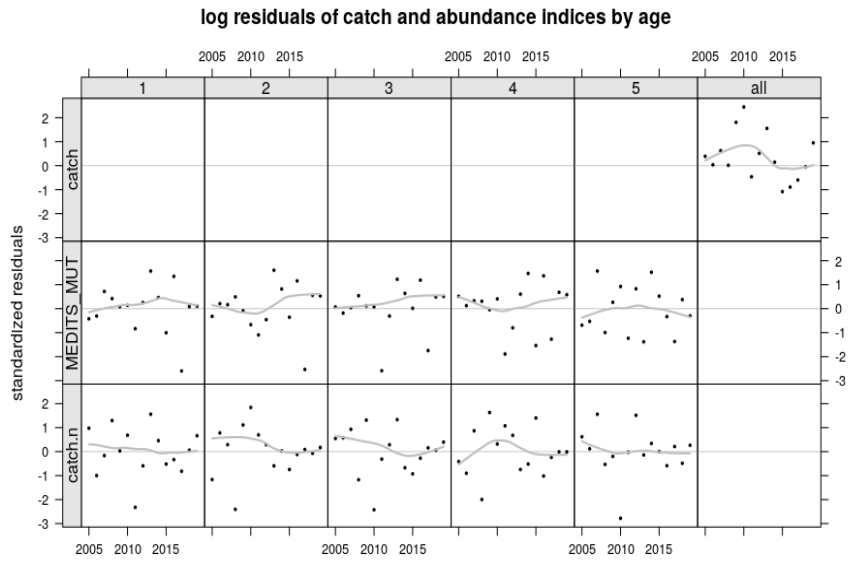


Figure 88: MUT in GSAs 22-25: log residuals for the catch-at-age fishery and survey data.

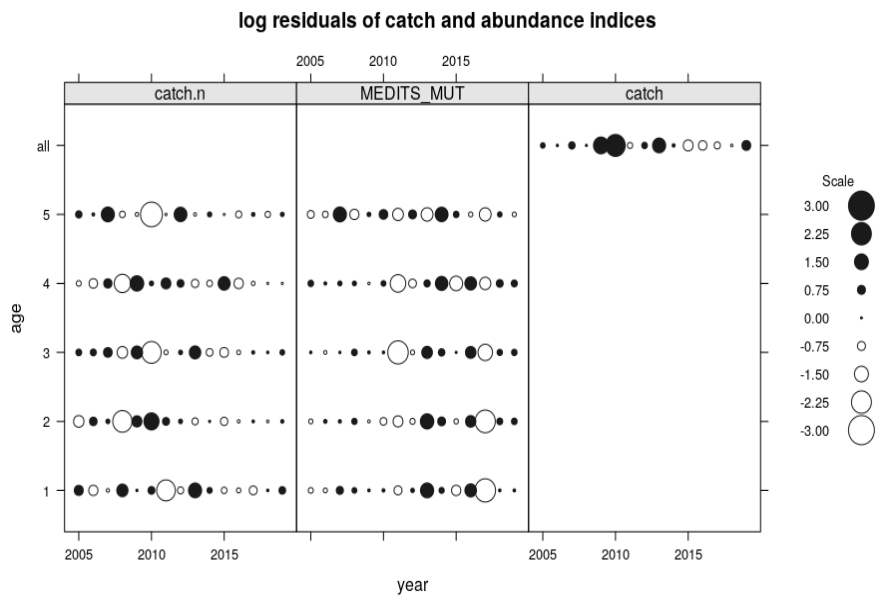


Figure 89: MUT in GSAs 22-25: bubble plot of the log residuals for the catch-at-age fishery and survey data.

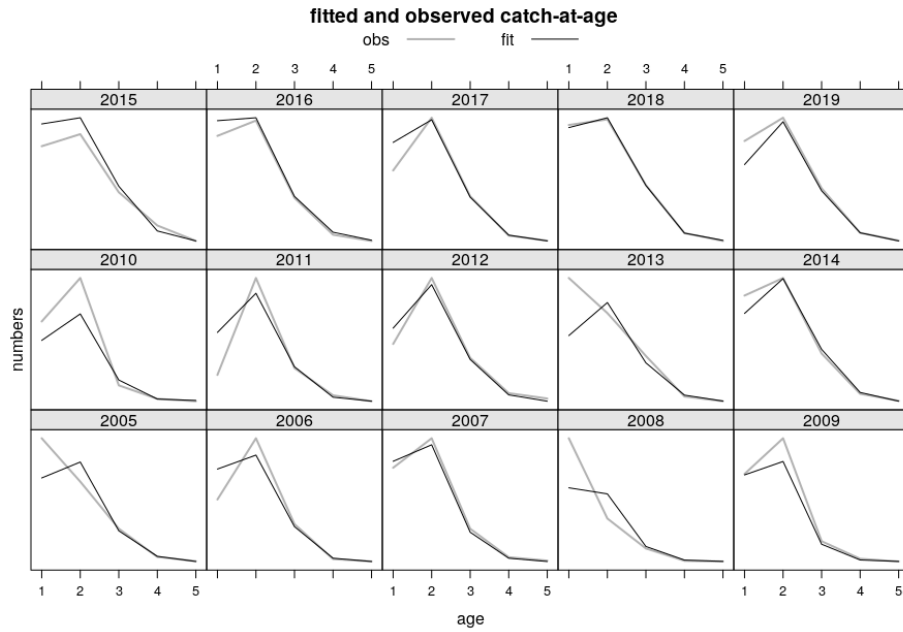


Figure 90: MUT in GSAs 22-25: fitted vs observed values by age and year for the catches.

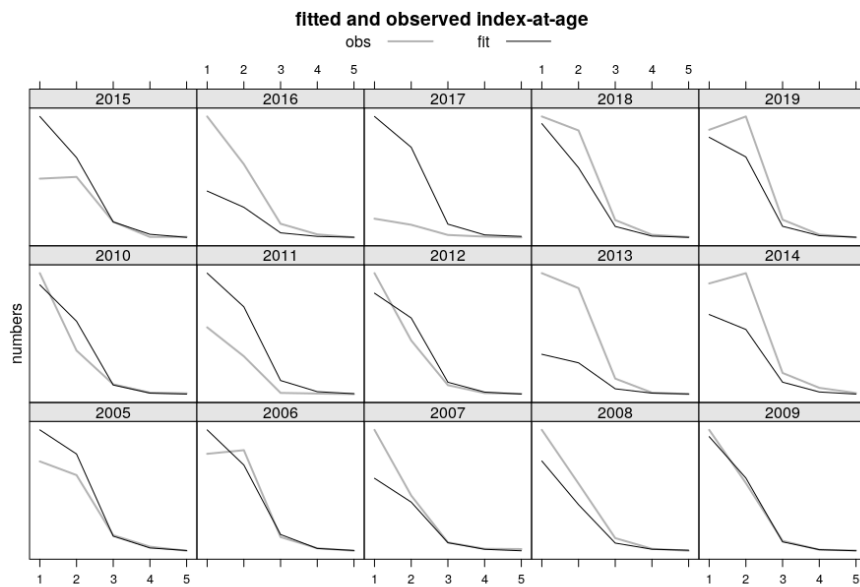


Figure 91: MUT in GSAs 22-25: fitted vs observed values by age and year for the survey.

The retrospective analysis (Figure 12) shows that the assessment model is moderately stable, and the catch estimates obtained by the a4a assessment are fitting well the observed catches. The Mohn's Rho for SSB and recruitment are falling within the optimal range-0.2- 0.2, while the value for F (-0.37) is outside.

Final assessment outcomes are given in figure 93, and tables 32-34. The value of F_{0.1} calculated by the FLBRP package on the a4a assessment results is equal to 0.4. The current F value (2019), as calculated by the a4a, is 0.48 indicating that the stock is under moderate overexploitation ($F/F_{0.1} = 1.2$; Figure 94).

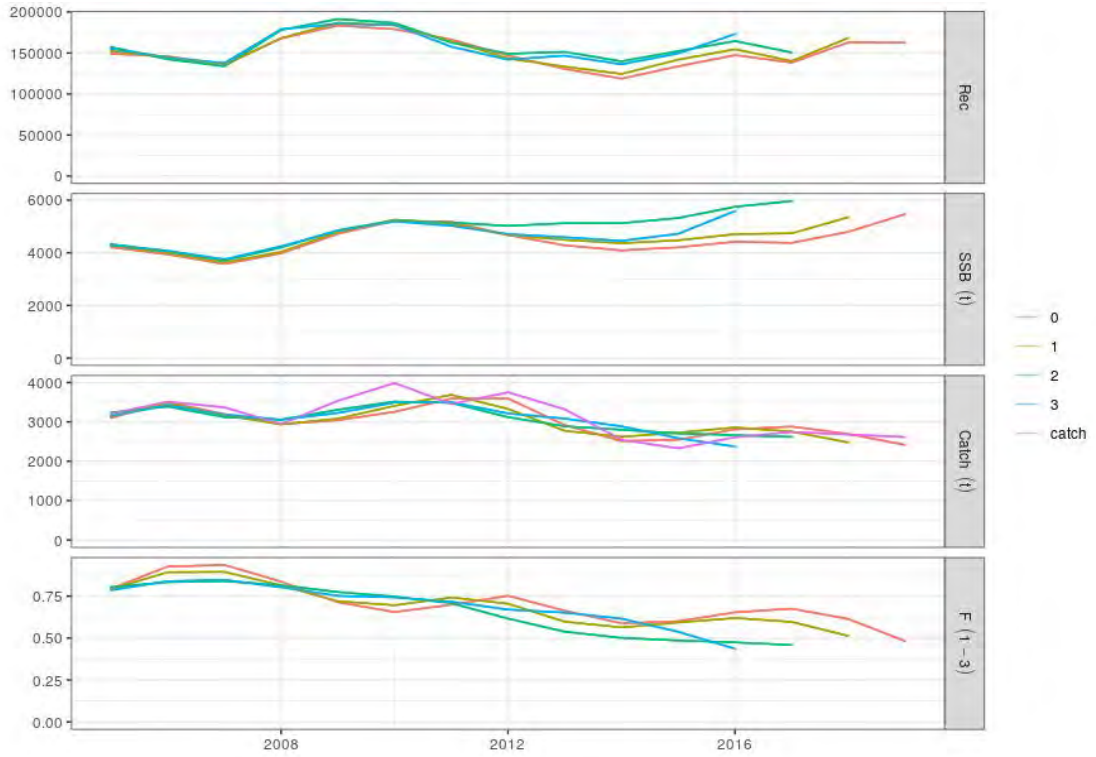


Figure 92: MUT in GSAs 22-25. Retrospective analysis.

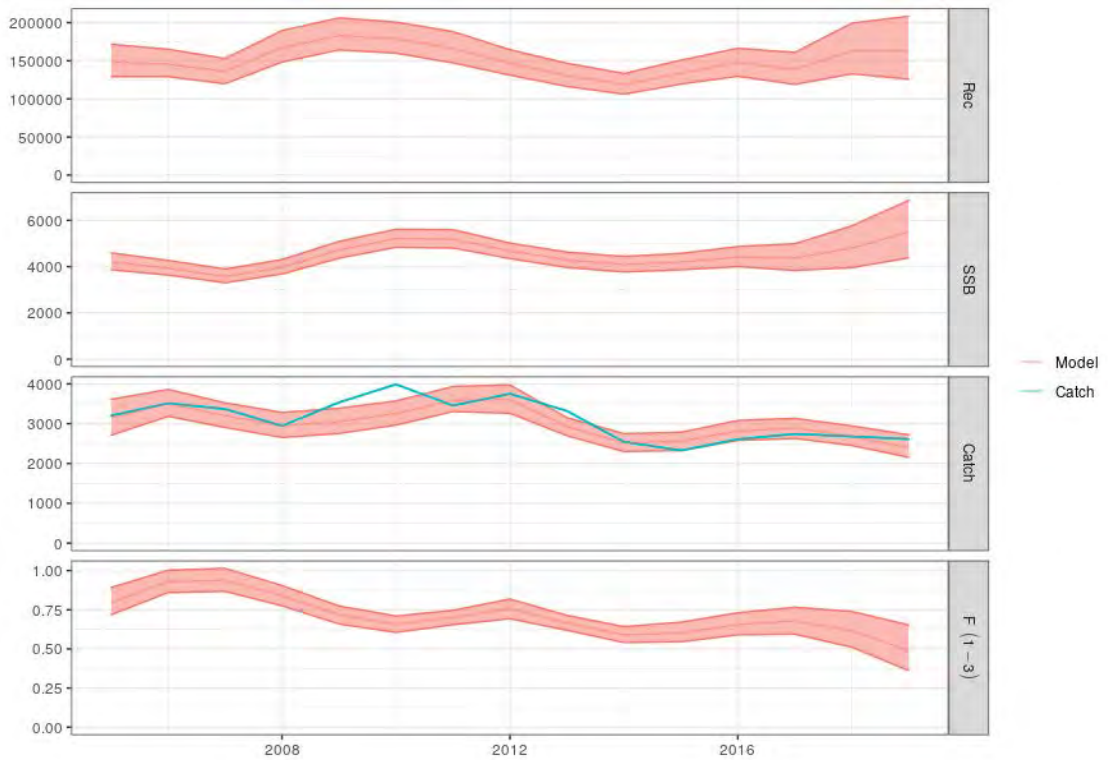


Figure 93: MUT in GSAs 22-25. Outputs of the a4a stock assessment model, with uncertainty; input catch data (light blue line) are plotted against the estimated catches

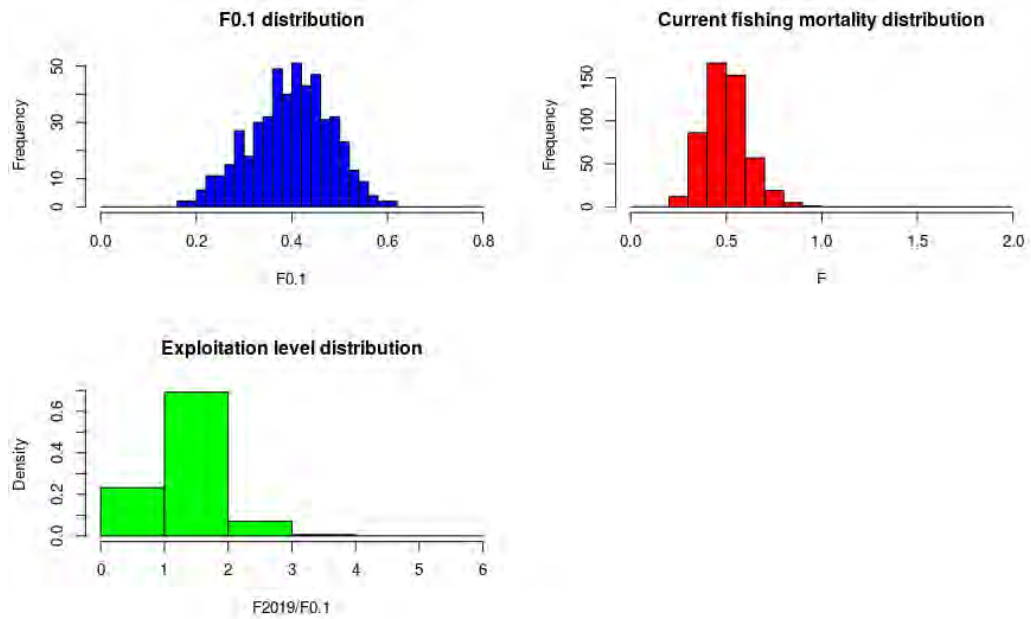


Figure 94: MUT in GSAs 22-25. Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F0.1 ratio) values.

Table 30: MUT in GSAs 22-25. Stock numbers-at-age (thousands).

Year	Age 1	Age 2	Age 3	Age 4	Age 5
2005	149028	65872	15008	3017	535
2006	145857	63411	17188	2401	676
2007	135822	58773	13205	2445	725
2008	167728	54089	11188	2018	927
2009	183476	69058	11119	2065	1023
2010	178980	79375	16540	2423	1101
2011	166451	79923	21075	3753	1088
2012	146987	74248	21086	4237	1066
2013	130536	64755	19167	3729	830
2014	118792	58324	18660	3912	823
2015	133745	53382	18203	4380	1106
2016	147414	58993	16424	4258	1442
2017	138277	63709	17265	3511	1509
2018	162946	60258	18579	3443	1281
2019	162541	74179	19212	3902	1183

Table 31: MUT in GSAs 22-25. F-at-age.

Year	Age 1	Age 2	Age 3	Age 4	Age 5
2005	0.24	0.8	1.33	1.21	0.9
2006	0.3	1.03	1.45	1.04	0.67
2007	0.31	1.12	1.38	0.82	0.48
2008	0.28	1.04	1.19	0.65	0.39
2009	0.23	0.89	1.02	0.59	0.42
2010	0.2	0.79	0.98	0.69	0.65
2011	0.2	0.79	1.1	0.97	1.2
2012	0.21	0.81	1.23	1.26	1.83
2013	0.2	0.7	1.09	1.15	1.56
2014	0.19	0.62	0.95	0.94	1.04
2015	0.21	0.64	0.95	0.86	0.75
2016	0.23	0.69	1.04	0.9	0.65
2017	0.22	0.69	1.11	0.97	0.66
2018	0.18	0.6	1.06	0.98	0.67
2019	0.12	0.45	0.88	0.88	0.62

Table 32: MUT in GSAs 22-25. Summary results of the a4a assessment

Year	Recruitment (000's	SSB (t)	Fbar	Catch (t)
2005	149028	4213	0.79	3100
2006	145857	3939	0.93	3502
2007	135822	3573	0.94	3198
2008	167728	3982	0.84	2940
2009	183476	4719	0.71	3043
2010	178980	5220	0.66	3255
2011	166451	5183	0.7	3591
2012	146987	4672	0.75	3593
2013	130536	4284	0.66	2919
2014	118792	4092	0.59	2513
2015	133745	4207	0.6	2546
2016	147414	4419	0.65	2811
2017	138277	4378	0.68	2883
2018	162946	4798	0.61	2692
2019	162541	5472	0.48	2416

4.5.4 Comparison with other assessments with a different stock configuration

So far, red mullet assessments have been performed only for GSAs 22 and 25, separately. For GSA 22 sustainable exploitation has been reported, while overexploitation was found for GSA 25. In the current case, four GSAs were considered jointly but there were several gaps and missing information on catch-at-age data. As a consequence, several substitutions were made and in fact LFDs from GSA 25 were considered as representative of the total catch in several years. Given that catches from GSA 25 represent less than 1% of the total catch, such an assumption may have biased our findings. In this context, it is normal to expect that previous “local” runs will show better cohort consistency than the current joint assessment. In general, direct comparisons, either on diagnostics or findings, among the previous and the current assessment can be misleading given the assumptions currently made regarding missing catch-at-size data.

4.6 Deep-water rose shrimp in GSAs 1, 5-16

In the western Mediterranean, deep-water rose shrimp (DPS), *Parapenaeus longirostris*, is an important commercial demersal species, targeted mainly by bottom trawlers. According to the outcomes of the genetic analysis performed under the MED_UNITS project, a stock configuration including GSAs 1, 5-16 was identified and used to perform a statistical catch-at-age stock assessment (Figure 95).

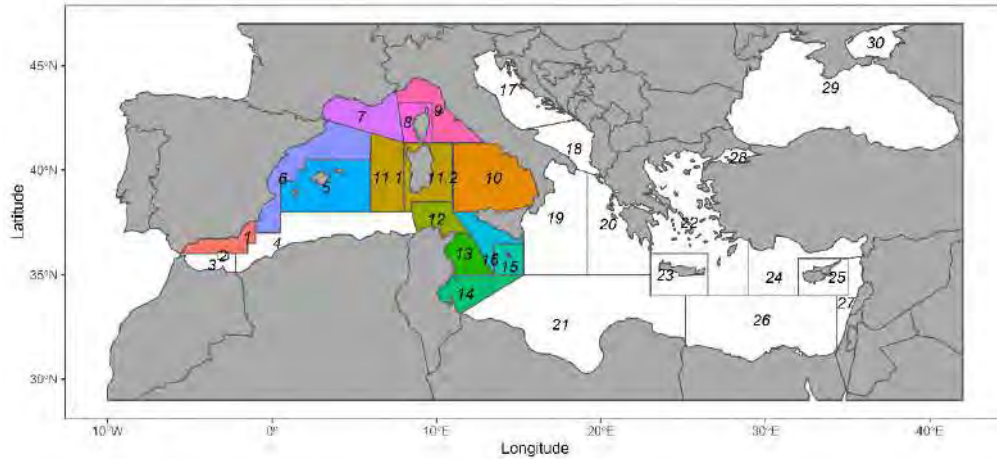


Figure 95: DPS in GSAs 1, 5-16. Geographical location of the stock

4.6.1 Input data

Fishery-Dependent data

Commercial data (catch data, LFDs and biological parameters) for the EU countries considered in the stock configuration (Spain, France, Italy, Malta) were collected from the EU DCR/DCF database. In addition, data from the Tunisian fleets exploiting GSAs 12-14 were obtained from the GFCM/FAO database. The commercial fisheries data covered the period 2009-2019. Thus fisheries data covered GSAs 1, and 5-16.

LFDs were sliced (deterministic slicing) to obtain age structure using the growth parameters employed for assessing GSAs 12-16 during the latest (2021) GFCM assessment (Table 35) and correcting the t_0 value to account for a birth date at the middle of the year. The choice of those parameters was driven by the fact that this GSA group has by far the highest landings (~70% of the total in the latest years) among the considered GSAs. Similarly, the assumed natural mortality and maturity vectors, as well as the length-weight relationship, were in accordance to those used in that GFCM assessment (Table 35 and Table 36).

Table 33: DPS in GSAs 1, 5-16. Von Bertalanffy growth (VBGF) and length-weight relationship parameters.

	VBGF			Length/weight	
	Loo	k	t_0	a	b
All sexes	44.59	0.6	-0.118	0.0033	2.46

Table 34: DPS in GSAs 1, 5-16. Proportion of mature and natural mortality (M) at age.

Age	0	1	2	3	4
Maturity	0.03	0.98	1	1	1
M	1.42	1.09	1.05	1.03	1.03

Figure 96 illustrates the estimated age composition of the total catch by year.

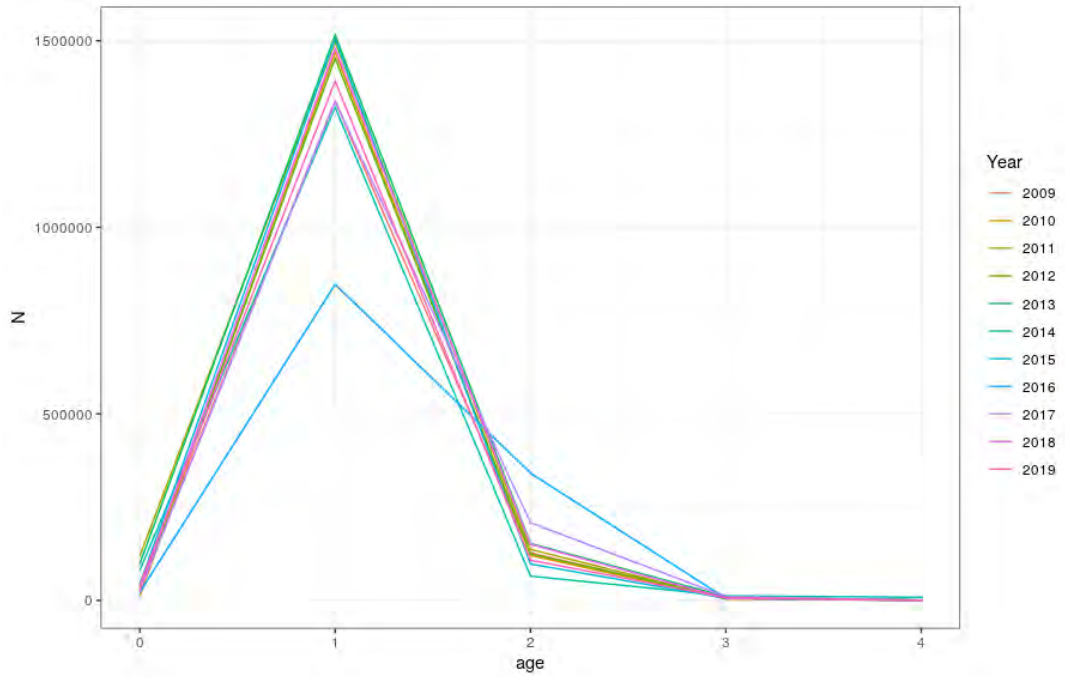


Figure 96: DPS in GSAs 1, 5-16. Catch numbers-at-age

Fishery-Independent data

LFDs (n/km^2) from the MEDITS surveys in the European GSAs (1, 3-11, 15-16) were sliced using the same growth parameters used for the commercial data. The data used covered the period 2009-2019. Figure 97 illustrates the estimated age composition of the MEDITS survey data by year.

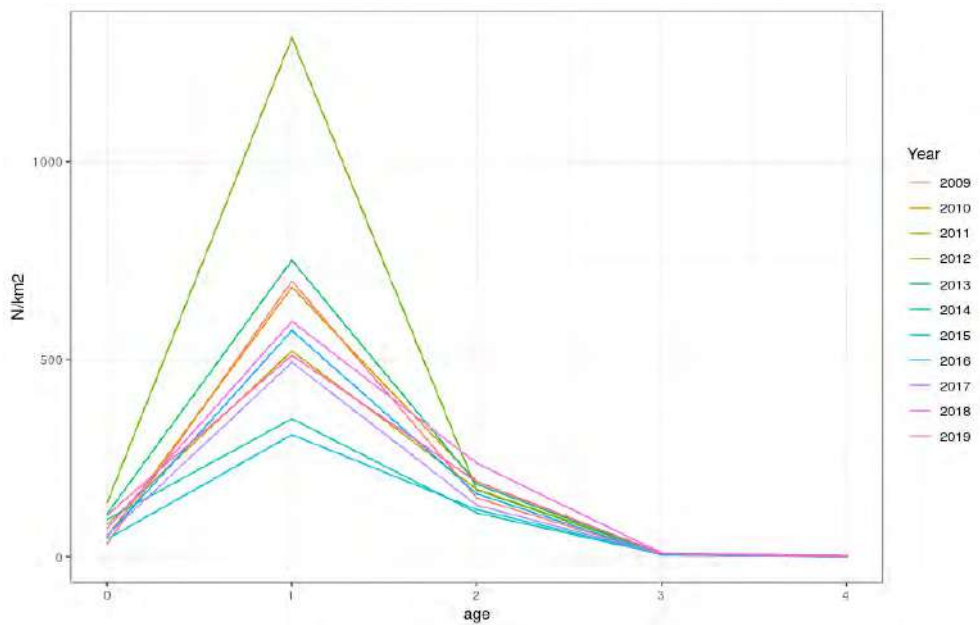
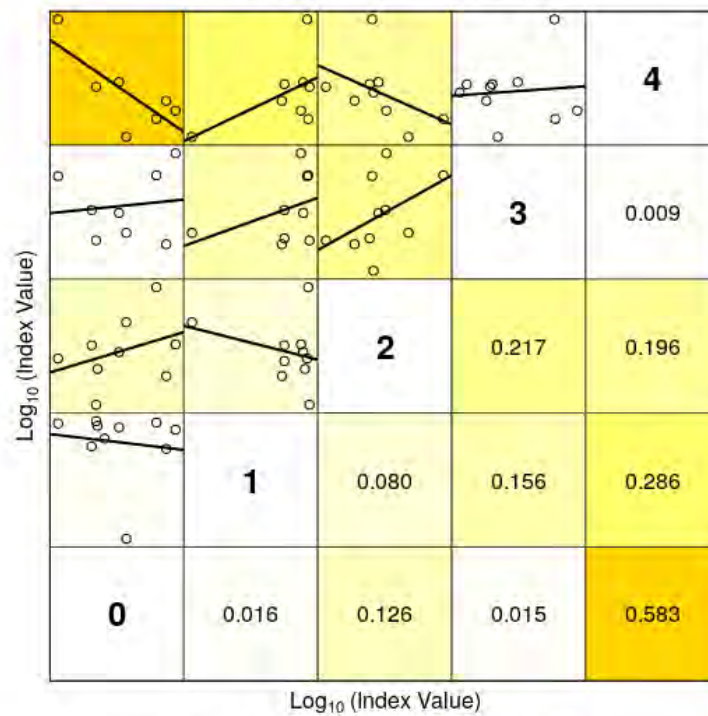


Figure 97: DPS in GSAs 1, 5-16. Catch numbers-at-age in the MEDITS.

Cohort consistency plots for the fishery and survey data are shown in Figures 98 and 99 respectively.

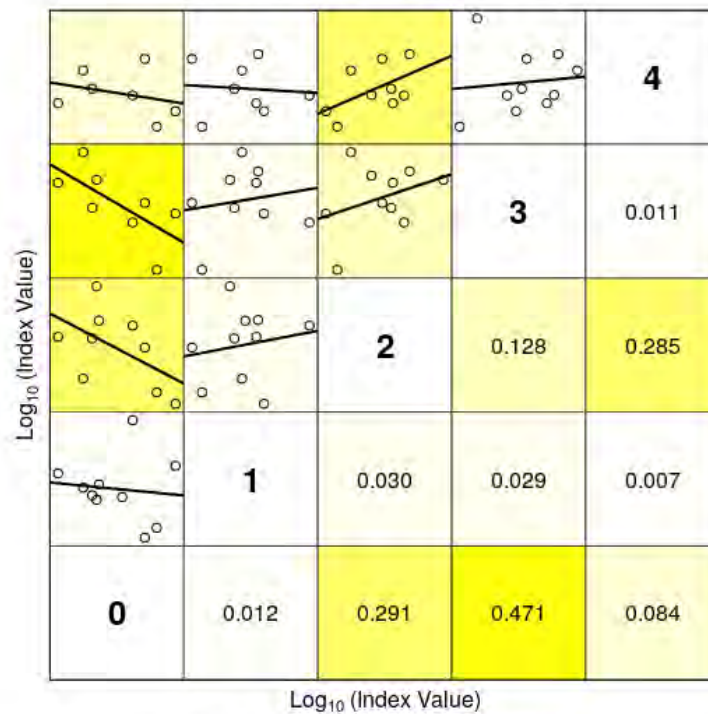
Cohorts consistence of the catch



Lower right panels show the Coefficient of Determination (r^2)

Figure 98: DPS in GSAs 1, 5-16. Cohort consistency in the catch numbers-at-age.

Cohorts consistence in survey



Lower right panels show the Coefficient of Determination (r^2)

Figure 99: DPS in GSAs 1, 5-16. Cohort consistency in the catch numbers-at-age in the MEDITS.

4.6.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment using fisheries and survey data of the 2009-2019 period. A plus group of age 4 was set for the fisheries data, while the number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers at age a x catch weight-at-age a)], to account for the missing catch-at-size data. A true age 4 was used in the survey and F_{bar} range was fixed at 1-3.

The final model settings showing the best diagnostics outputs were:

Fishing mortality sub-model: $fmodel = \sim te(replace(age, age > 3, 3), year, k = c(3, 5)) + s(year, k = 5, by = as.numeric(age == 0)) + s(year, k = 5) + s(replace(age, age > 3, 3), k = 3)$

Catchability sub-model: $qmodel = list(\sim factor(replace(age, age > 3, 3)))$

SR sub-model: $srmodel \sim geomean(CV = 0.2)$

4.6.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in figures 100-105.

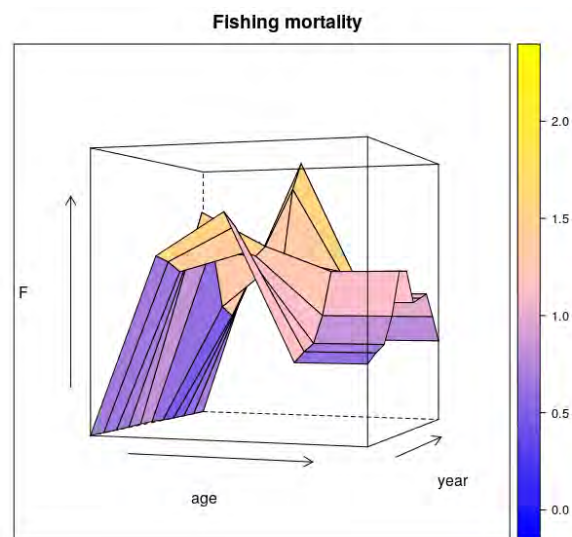


Figure 100: DPS in GSAs 1, 5-16. Fishing mortality by age and year obtained from the a4a model (2009-2019).

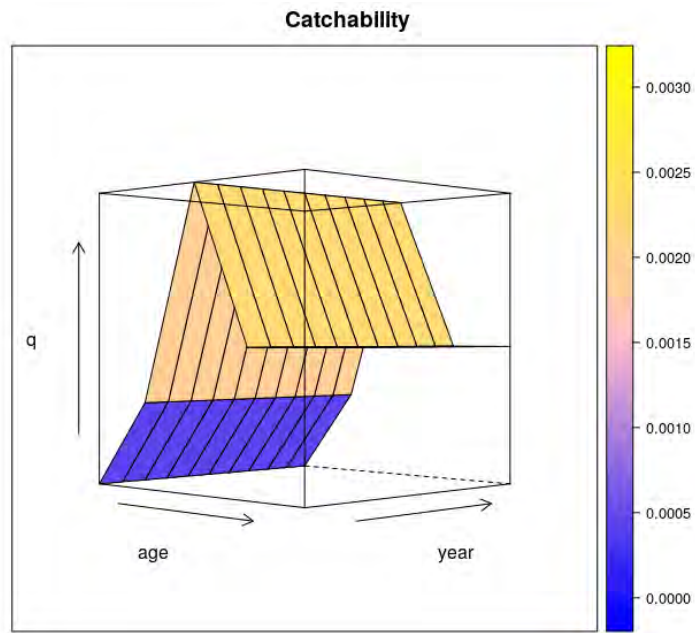


Figure 101: DPS in GSAs 1, 5-16. Catchability of the survey by age and year obtained from the a4a model (2009-2019).

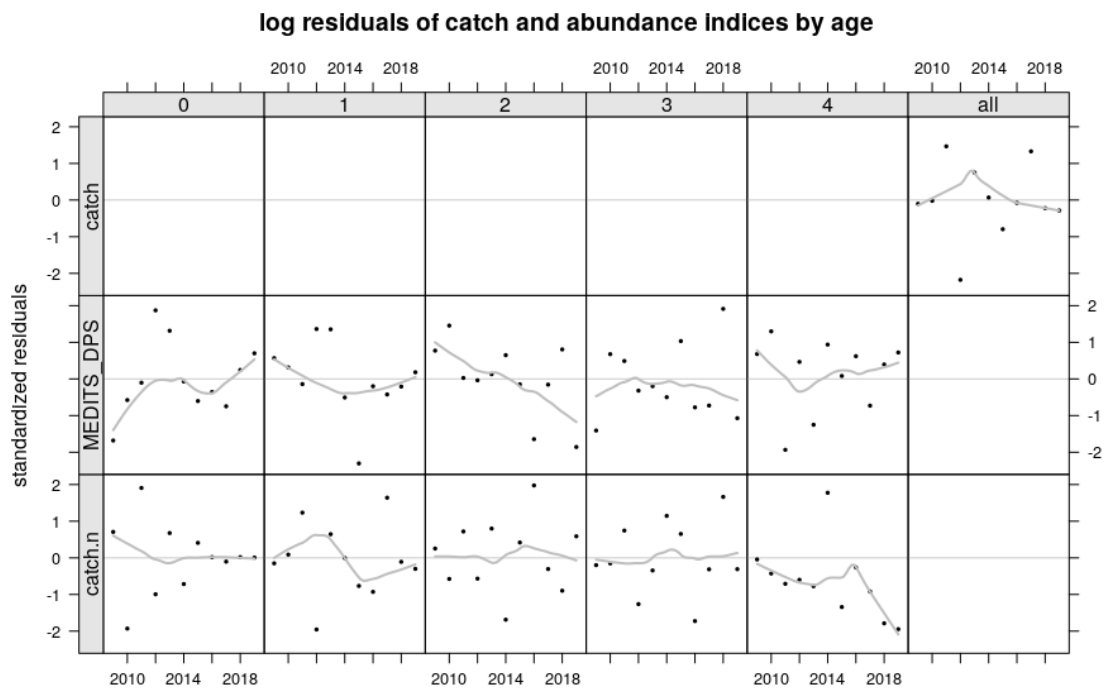


Figure 102: DPS in GSAs 1, 5-16. Log residuals for the catch-at-age fishery and survey data.

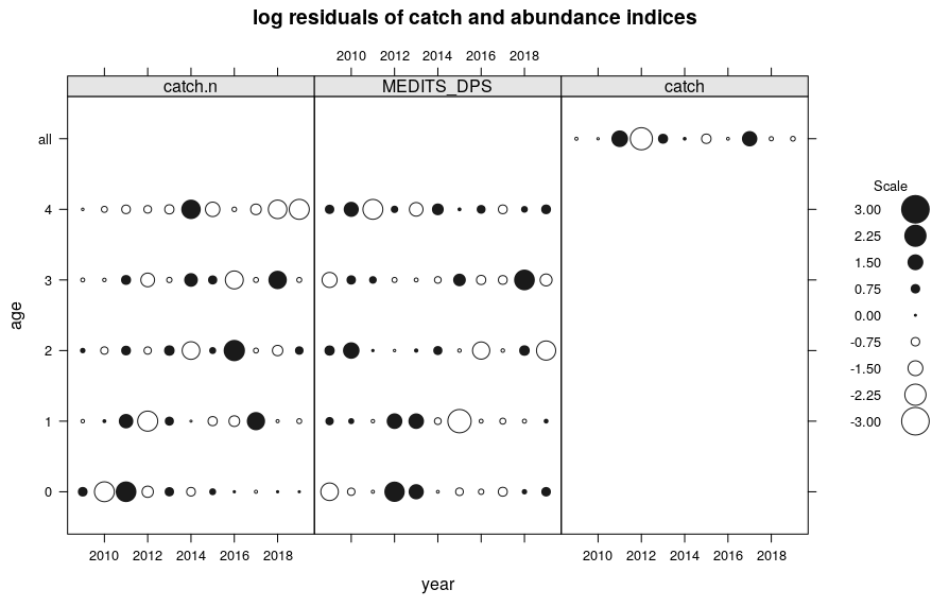


Figure 103: DPS in GSAs 1, 5-16. Bubble plot of the log residuals for the catch-at-age fishery and survey data.

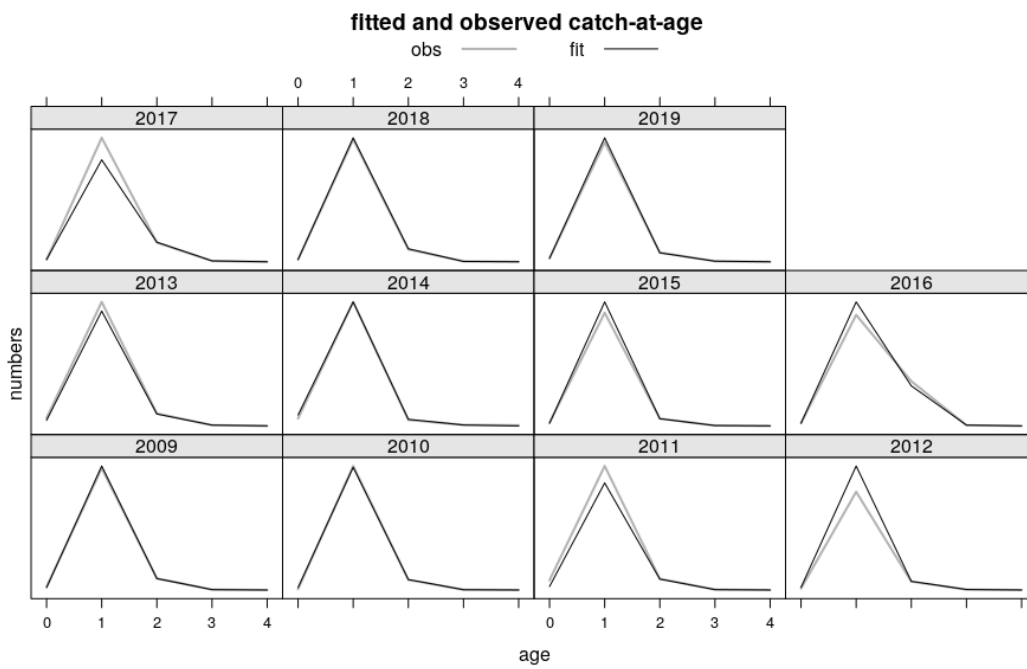


Figure 104: DPS in GSAs 1, 5-16. Fitted vs observed values by age and year for the catches.

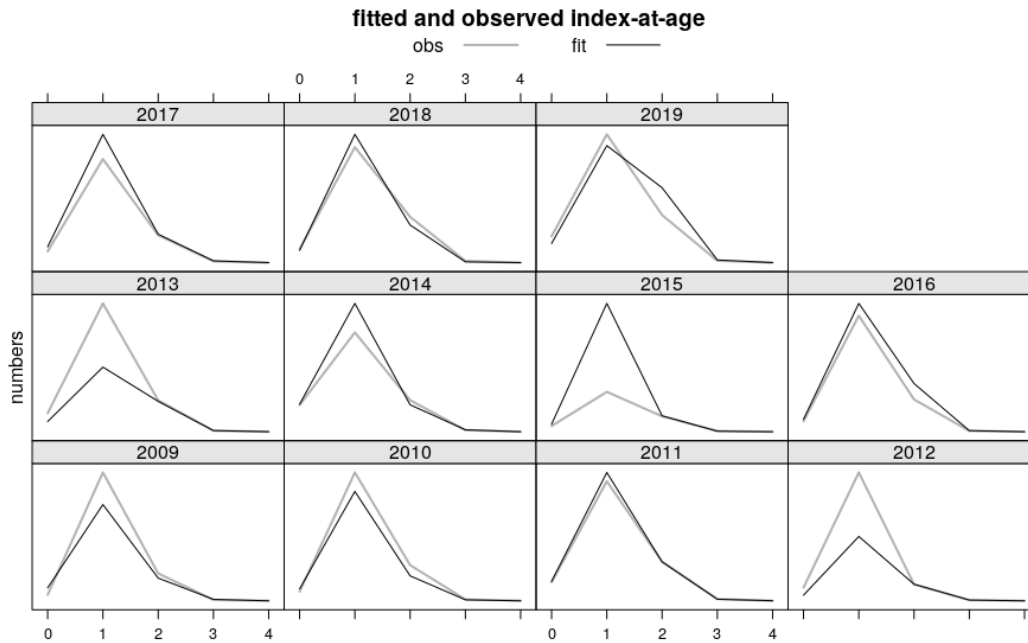


Figure 105: **DPS in GSAs 1, 5-16.** Fitted vs observed values by age and year for the survey.

The retrospective analysis (Figure 106) shows that the assessment model is stable, and the catch estimates obtained by the a4a assessment are fitting well the observed catches. The Mohn's Rho for F, SSB and recruitment are falling within the optimal range -0.2- 0.2 (values are 0.09, 0.11 and 0.017, respectively).

Final assessment outcomes are given in **figures 107-108** and tables 37-39. Value of $F_{0.1}$ calculated by the FLBRP package on the a4a assessment results is equal to 0.6. The Current F value (2019), as calculated by the a4a model, is 1.01 indicating that the stock is under overexploitation ($F/F_{0.1} = 1.68$).

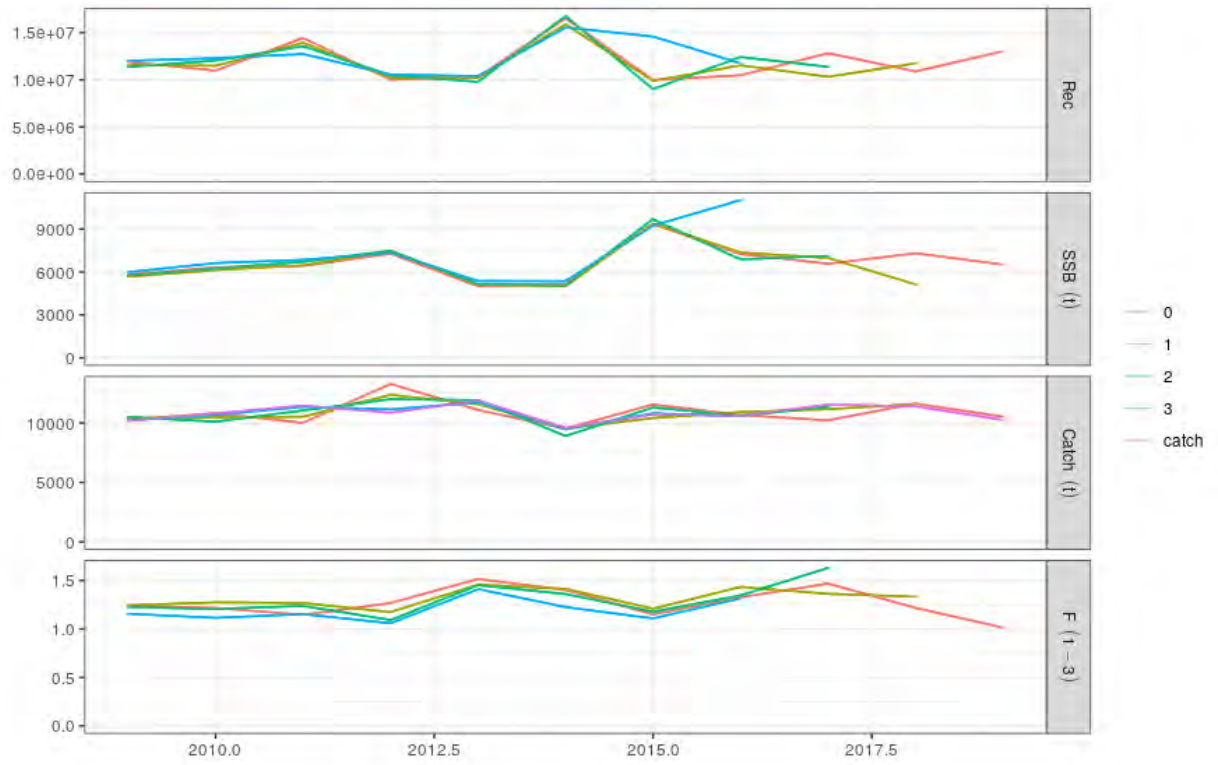


Figure 106: DPS in GSAs 1, 5-16. Retrospective analysis.

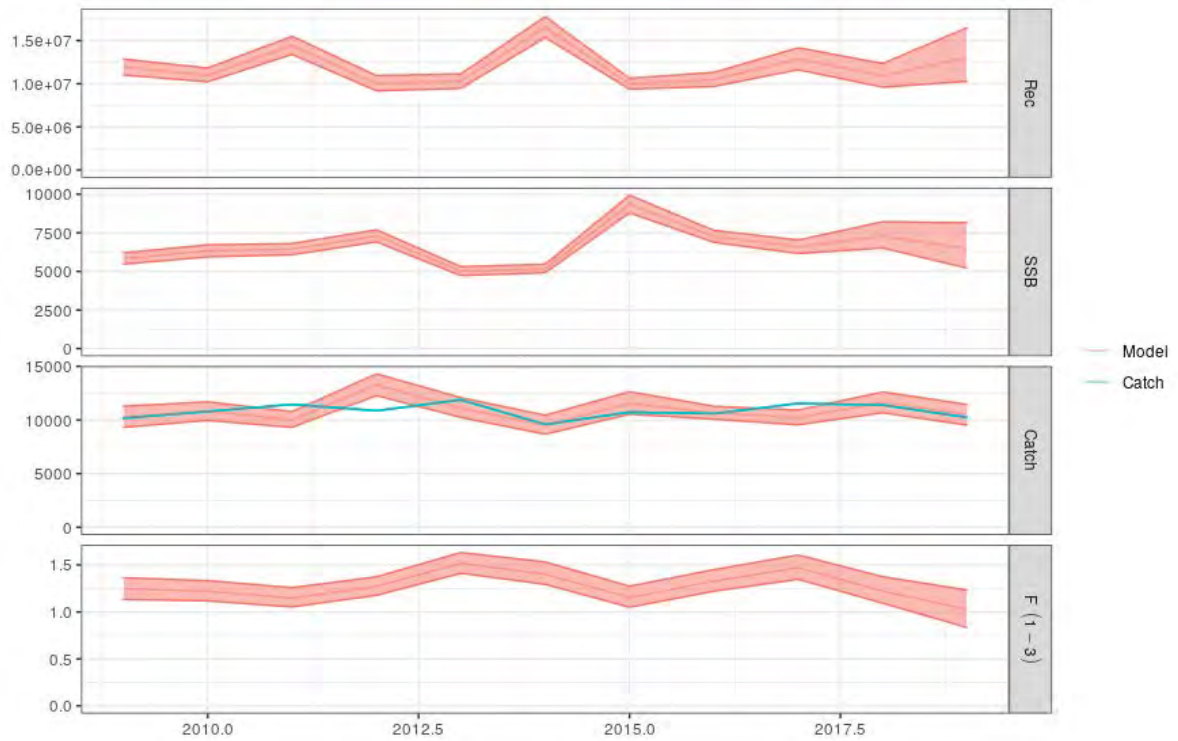


Figure 107: DPS in GSAs 1, 5-16. Outputs of the a4a stock assessment model, with uncertainty; input catch data (light blue line) are plotted against the estimated catches

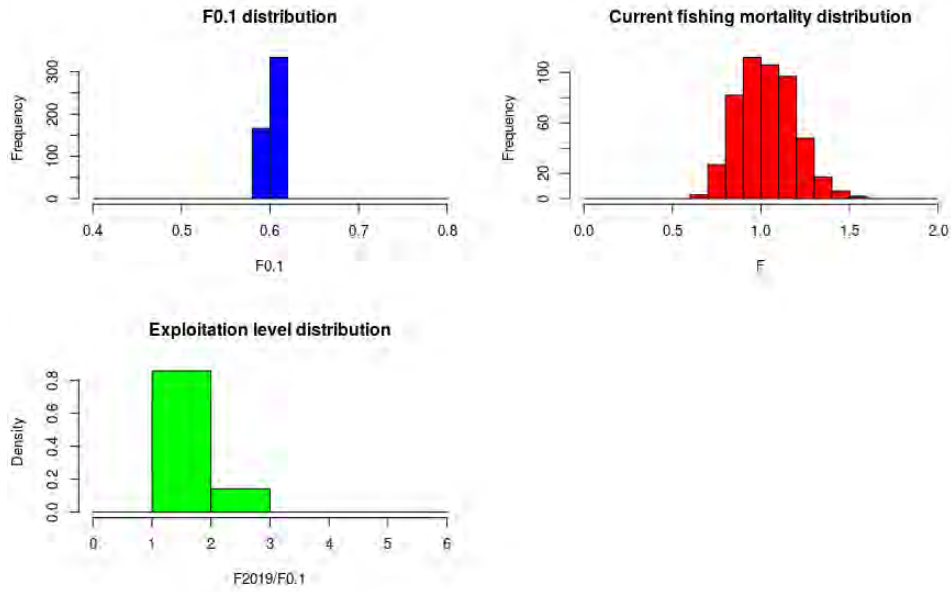


Figure 108: DPS in GSAs 1, 5-16. Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F01 ratio) values.

Table 35: DPS in GSAs 1, 5-16. Stock numbers-at-age

Year	Age 0	Age 1	Age 2	Age 3	Age 4
2009	11925183	2645695	215616	15721	1025
2010	10973265	2870856	219975	13467	3278
2011	14437710	2638725	248585	15166	3054
2012	10008725	3470357	246094	20817	3147
2013	10276314	2401229	270056	21111	3358
2014	16537871	2453020	136653	21300	2421
2015	9983128	3944439	180767	11776	2337
2016	10481773	2397256	489718	15243	1791
2017	12826675	2523971	322987	23201	2088
2018	10896709	3088850	293464	12026	2993
2019	13032163	2621395	293577	23367	2159

Table 36: DPS in GSAs 1, 5-16. Fishing mortality-at-age

Year	Age 0	Age 1	Age 2	Age 3	Age 4
2009	0	1.4	1.72	0.6	0.6
2010	0.01	1.36	1.62	0.67	0.67
2011	0.01	1.28	1.43	0.73	0.73
2012	0.01	1.46	1.41	0.94	0.94
2013	0.01	1.78	1.49	1.28	1.28
2014	0.01	1.52	1.4	1.29	1.29
2015	0.01	1	1.42	1.03	1.03

2016	0	0.91	2	1.07	1.07
2017	0	1.06	2.24	1.1	1.1
2018	0	1.26	1.48	0.91	0.91
2019	0.01	1.58	0.77	0.69	0.69

Table 37: DPS in GSAs 1, 5-16. Summary results of the a4a assessment.

Year	Recruitment (000's)	SSB (t)	Fbar	Catch (t)
2009	11925	5822	1.24	10262
2010	10973	6331	1.22	10810
2011	14438	6436	1.15	10005
2012	10009	7290	1.27	13268
2013	10276	5017	1.52	11084
2014	16538	5173	1.4	9518
2015	9983	9351	1.15	11546
2016	10482	7258	1.33	10667
2017	12827	6583	1.47	10210
2018	10897	7315	1.22	11625
2019	13032	6521	1.01	10517

4.6.4 Comparison with other assessments with a different stock configuration

In the frame of the relevant STECF and GFCM working groups a series of DPS assessments have been performed in the western Mediterranean, either for independent GSAs or jointly for 9-11 and 12-16. Results of the most recent assessments are presented in Table 40. The current results are consistent with the findings of those assessments that indicate overexploitation of DPS stocks in the range 1.05 - 2.28.

Table 38: DPS in GSAs 1, 5-16. Results of DPS assessments in western Mediterranean GSAs.

GSA	Method	Current Levels (2019)	Reference Points	Quantitative Status
Current (1-16)	a4a	$F_c=1.01$	$F_{0.1} = 0.6$	$F/F_{ref} = 1.68$
1	a4a	$F_c = 0.96$	$F_{0.1} = 0.7$	$F/F_{ref} = 1.37$
3	BioDyn and LCA/Yield per Recruit	$F_c = 1$	$F_{0.1} = 0.59$	NA
4	VIT	$F_c = 1.19$	$F_{0.1} = 0.7$	NA
5	XSA	$F_c = 1.85$	$F_{0.1} = 0.81$	$F/F_{ref} = 2.28$
6	XSA	$F_c = 1.27$	$F_{0.1} = 0.79$	$F/F_{ref} = 1.6$
9,10,11	a4a	$F_c = 1.03$	$F_{0.1} = 0.98$	$F/F_{ref} = 1.05$
12,13,14,15,16	XSA	$F_c = 1.23$	$F_{0.1} = 0.84$	$F/F_{ref} = 1.46$

4.7 Deep-water rose shrimp in GSAs 17-20

In the central Mediterranean, deep-water rose shrimp (DPS), *P. longirostris*, is an important commercial demersal species, targeted mainly by bottom trawlers. According to the outcomes of the genetic analysis performed under the MED_UNITS project, a stock configuration including GSAs 17, 18, 19 and 20 was identified and used to perform a statistical catch-at-age stock assessment (Figure 109).

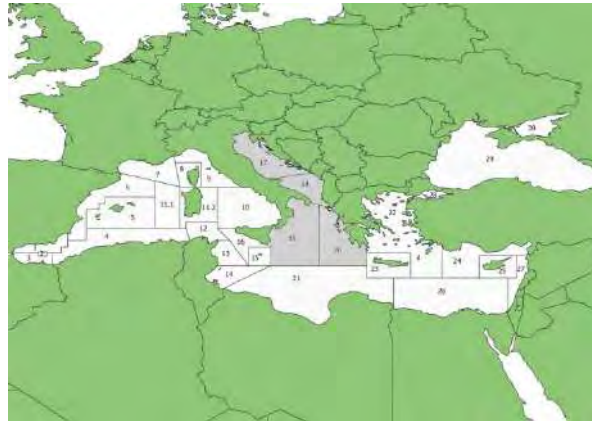


Figure 109: DPS in GSAs 17-20: geographical location of the stock.

4.7.1 Input data

Fishery Dependent data

Commercial data coming from EU DCR/DCF database were used to inform the assessment for the countries Italy, Greece and Croatia (LFDs and biological parameters). While for Albania and Montenegro data from FAO database were used. A deterministic sex combined age slicing was used to obtain age structure of the stock, using growth parameters available under DCF, similar to those of the STECF assessment of DPS 17-19 in 2020 (EWG 20-15; STECF, 2020b). However, in the current assessment an increment of 0.5 was applied to t_0 , taking into account the mid-year spawning of the species. Studies carried out in the Mediterranean indicate a variable reproductive strategy for this species. Some authors found that in the South Ionian the spawning of the deep-water rose shrimp females' is carried out during summer and that is more protracted in Montenegrin waters compared to Ionian waters (Kapiris et al., 2013). From other authors spawning is considered to occur through the year (D' Onghia et al., 1998). For the purposes of this assessment the spawning time was set at the mid-point of the year with 50% F and M occurring before spawning (Table 41). A vector of natural mortality was estimated by the Chen and Watanabe function while the proportion of mature individuals of age 0 was set as 0.4 and for all the other ages as 1 (Table 42).

Fishery Independent data

MEDITS survey in the GSAs 17, 18, 19 and 20 was used to tune the assessment. Density LFDs for the four areas combined were sliced using the same set of growth parameters.

The catch at age structure as well as the abundance at age resulting from the slicing are presented respectively in **figure 110** and **figure 111**

Table 39: DPS in GSAs 17, 18, 19 and 20. VBGF and LW parameters

Linf	k	t0
45.0	0.6	0.3
a		b

0.0024	2.5372
--------	--------

Table 40: DPS in GSAs 17, 18, 19 and 20. Natural mortality and maturity-at-age vectors

Age	Natural Mortality	Proportion of matures
0	1.75	0.4
1	0.938	1
2	0.748	1
3+	0.673	1

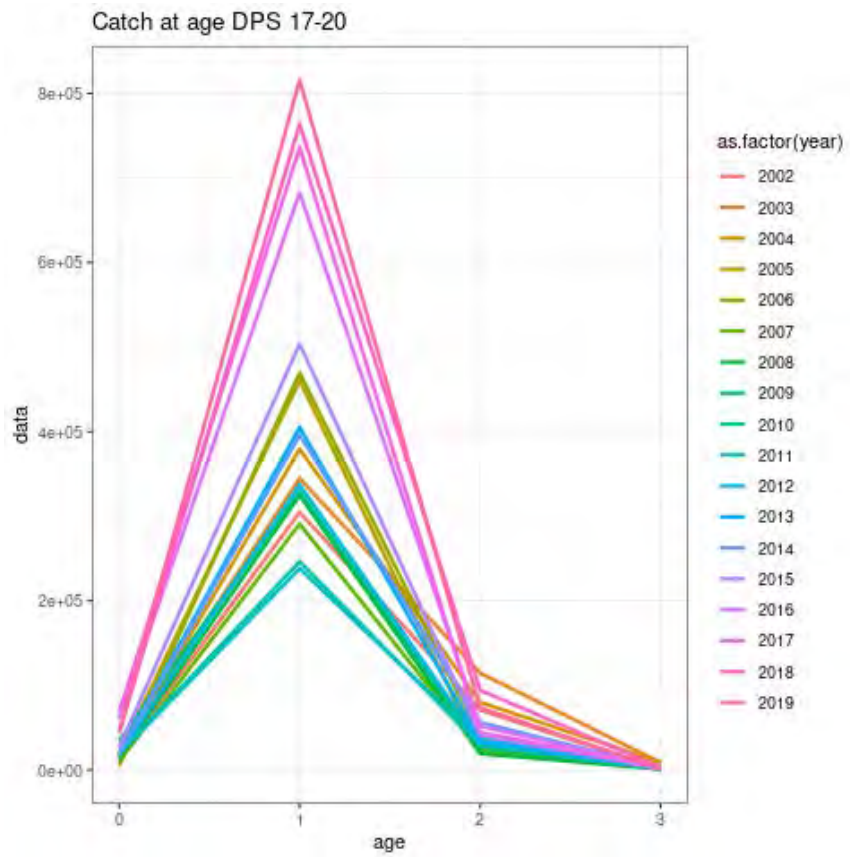


Figure 110: DPS in GSAs 17, 18, 19 and 20. Catch at age

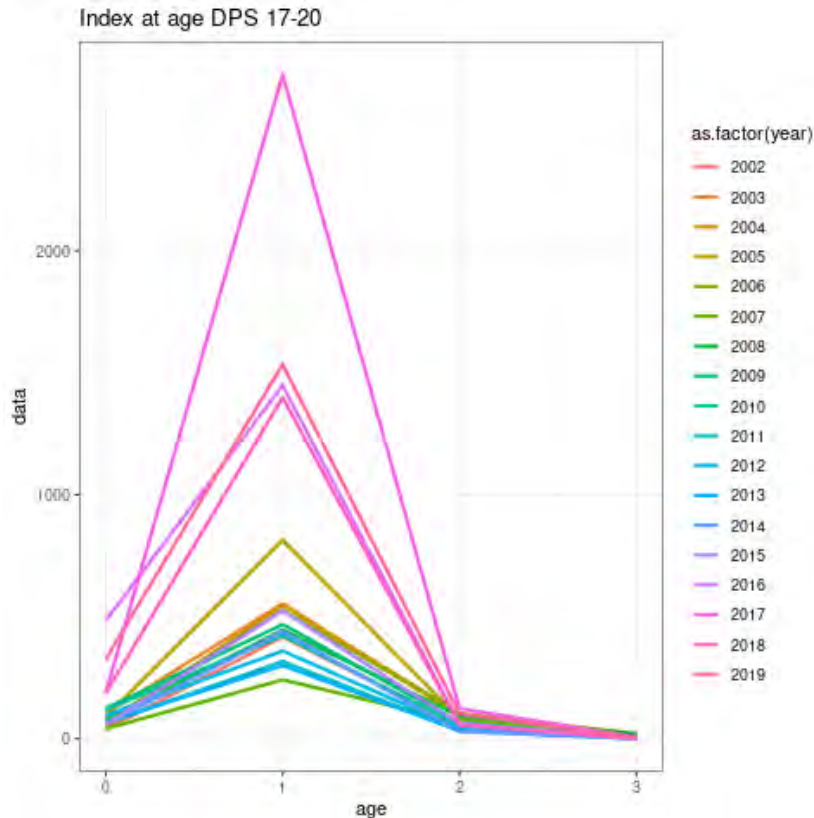


Figure 111: DPS in GSAs 17, 18, 19 and 20. Medits index at age.

4.7.2 Model description

The method for the assessment is a4a, a statistical catch-at-age framework developed by the Joint Research Centre (Jardim et al., 2015). Landings, Discards and Medits index numbers at age were derived from deterministic age slicing using the l2a function of FLR. The age classes range from 0 to 3 in both catch and tuning index; plus group was set at age 3. Given that the catches were composed mainly of individuals of age 1, Fbar ranges from 0 to 2. A SoP (sum of products) correction was applied to catch numbers at age.

Different combination of a4a submodels were tested. The best configuration of submodel according to residuals and retrospective were:

fishing mortality: $\sim \text{factor}(\text{replace}(\text{age}, \text{age} > 2, 2)) + s(\text{year}, k = 6)$

Recruitment: $\sim \text{geomean}(\text{CV} = 0.3)$

Catchability: $\sim \text{factor}(\text{replace}(\text{age}, \text{age} > 2, 2))$

Initial year's abundance: $\sim s(\text{age}, k = 3)$ (Default)

variance: (Default)

catch: $\sim s(\text{age}, k = 3)$

Medits: ~ 1

4.7.3 Outputs from the model

Results are shown in figures 112-116 , namely the estimated recruits, spawning stock biomass, catch and harvest rates for ages 0 – 2. Fishing mortality through all ages and years and catchability of the gear of the MEDITS survey tuning index:

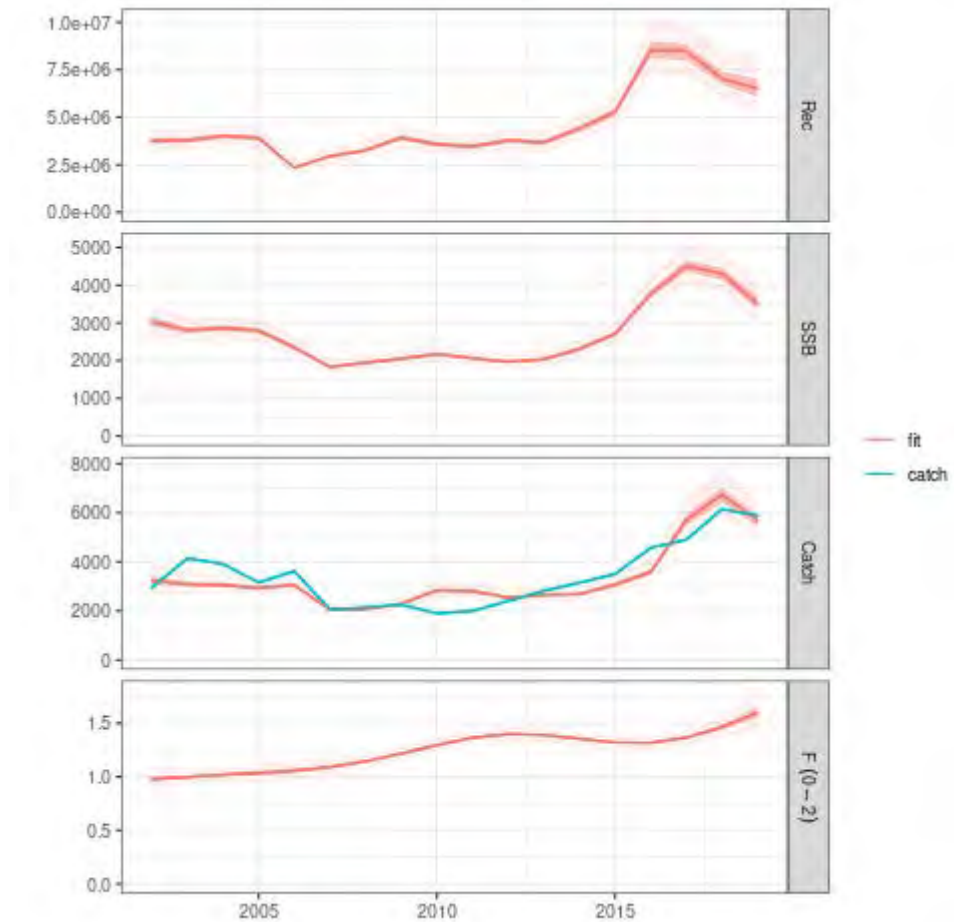


Figure 112: Figure 4 – DPS 17, 18, 19 and 20: Stock summary from the a4a model; recruits, SSB, catch and harvest ($f_{bar}=0-2$).

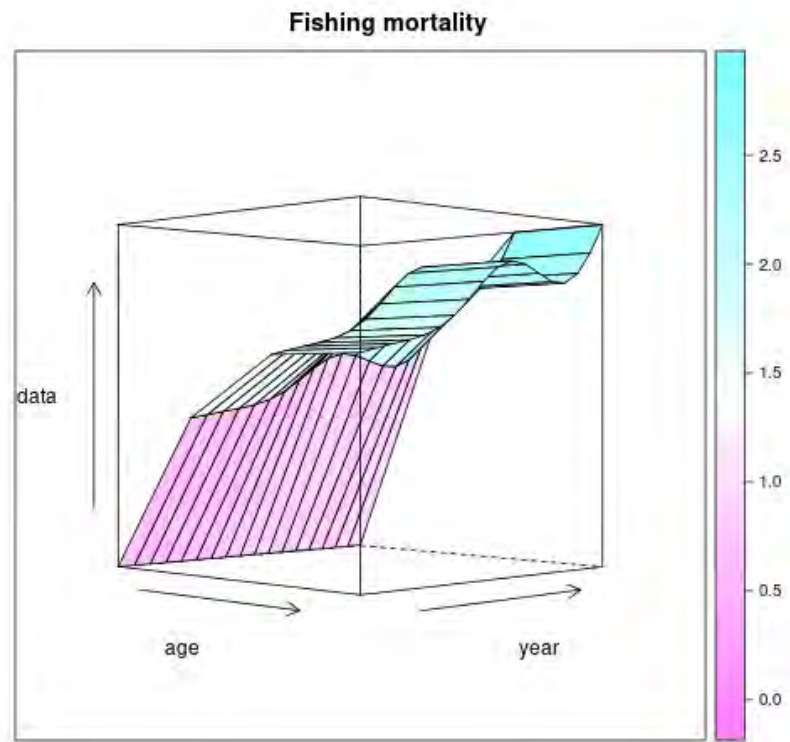


Figure 114: DPS 17, 18, 19 and 20. 3D contour plot of estimated fishing mortality by age and year.

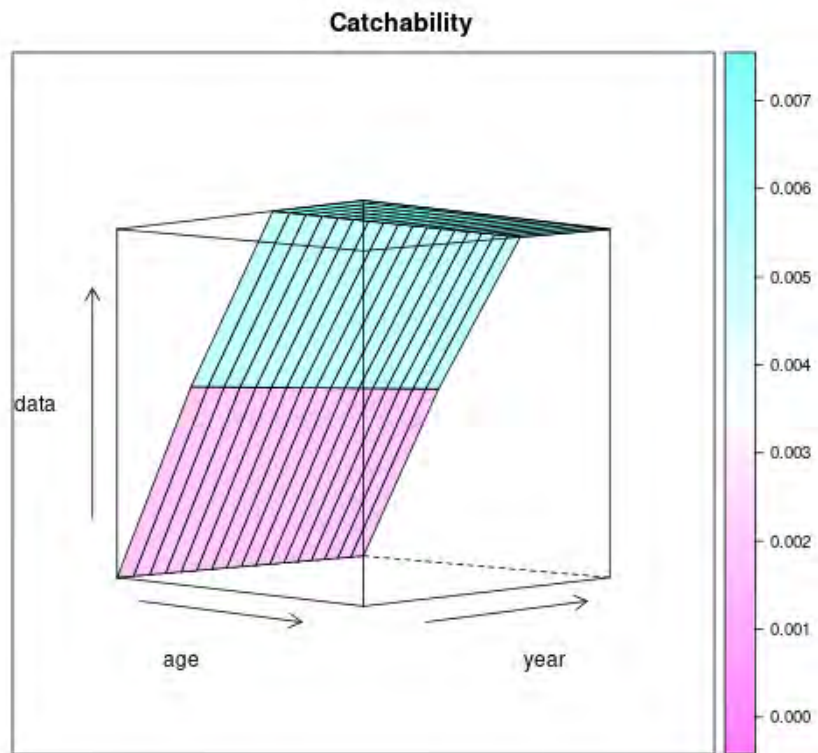
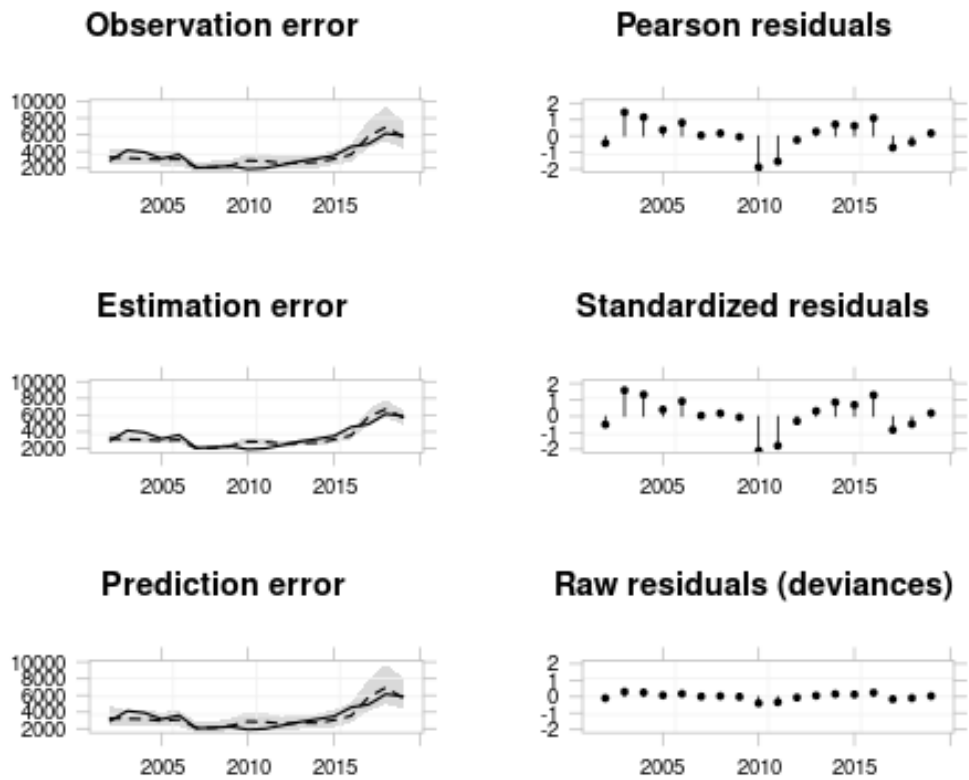


Figure 113: DPS 17, 18, 19 and 20: 3D contour plot of estimated catchability by age and year.

Several diagnostic plots presented below for the goodness of fit of the selected model for the assessment of Deep-water rose shrimp stock. Residuals of the total catch were evenly distributed around zero. Residuals at age in the catch and the survey do not show problematic effects, they are well scattered positive and negative values in the catch and the occasional year effect in the survey

Aggregated catch diagnostics



(shaded area = CI80%, dashed line = median, solid line = observed)

Figure 115: DPS 17, 18, 19 and 20. Aggregated catch diagnostics.

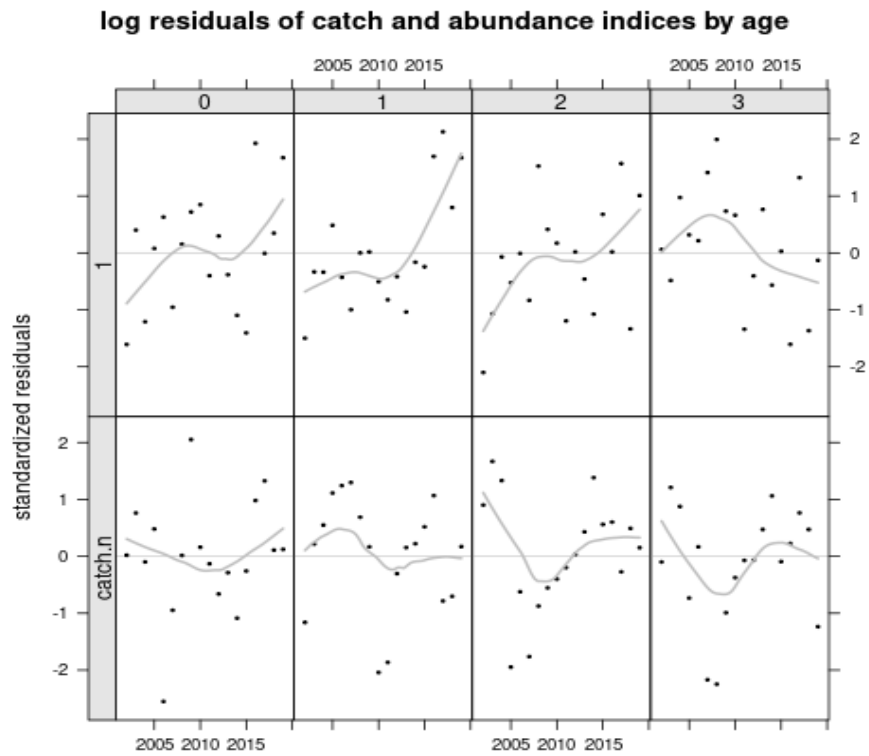


Figure 117: DPS 17, 18, 19 and 20. Standardized log residuals for the fitted model for catch number and index abundances.

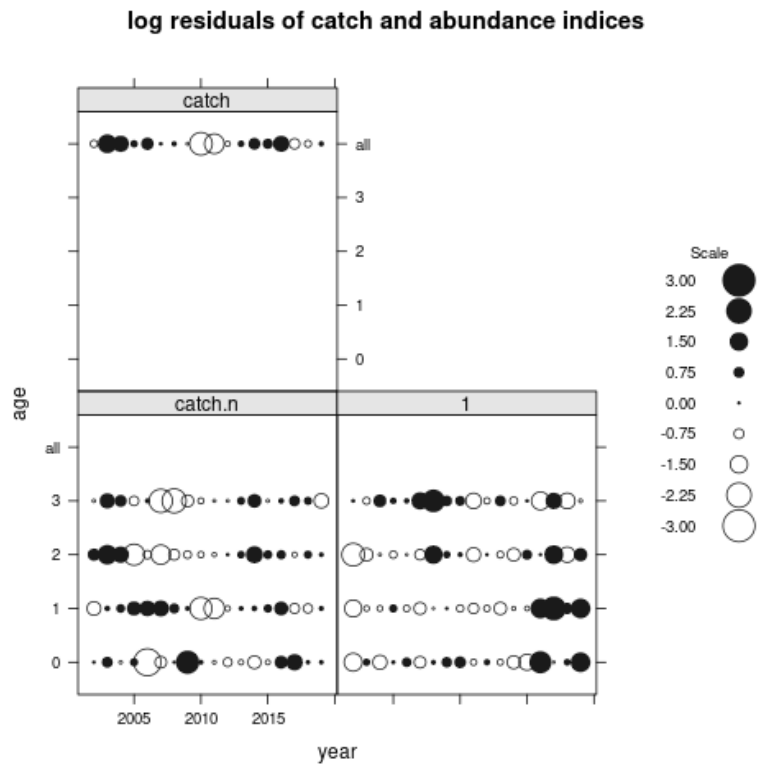


Figure 116: DPS 17, 18, 19 and 20. Standardized log residuals for the fitted model for catch numbers at age, index abundances and total catch presented in a bubble plot.

Fitted versus observed catch at age (Figure 10) show a fairly good fit for the model to the data. Some problems are apparent in the fitted versus index abundance (Figure 11) in the years 2016, 2017 and 2019 in the age 1.

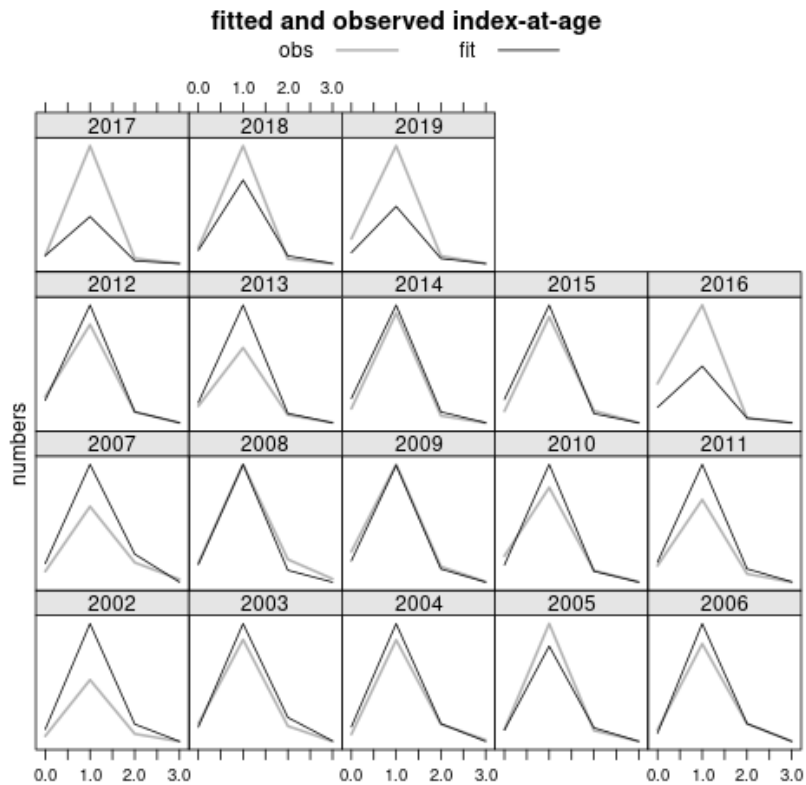


Figure 119: DPS 17, 18, 19 and 20. Estimated versus index at age.

Retrospective analysis shows that the assessment model is stable in terms of catch and SSB while for F is fairly unstable but not with significant patterns (Figure 120). The Mohn's Rho for F, SSB and recruitment fall within the optimal range (-0.2, 0.2); 0.023 for F, 0.27 for SSB and 0.18 for recruitment. The main outputs of the model are summarized in tables 43-45

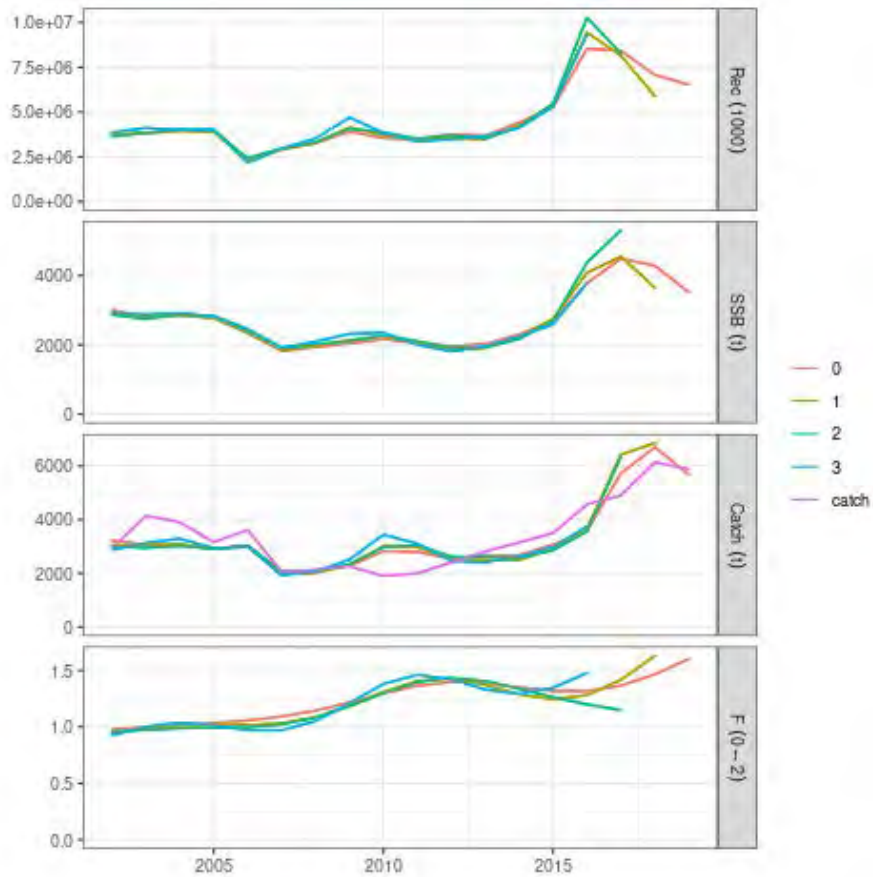


Figure 120: DPS 17, 18, 19 and 20. Retrospective analysis for the a4a model

Table 41: DPS 17, 18, 19 and 20. Stock summary results for a4a model.

Year	Recruitment	SSB	Catch	Fbar	Total Biomass
2002	3753345	3011	3217	0.98	12094
2003	3785214	2791	3074	1.00	11529
2004	3993945	2849	3045	1.02	11980
2005	3908261	2778	2923	1.03	11919
2006	2347841	2346	3045	1.06	9325
2007	2939782	1827	2040	1.09	8187
2008	3249100	1935	2062	1.14	8898
2009	3913725	2040	2283	1.22	9639
2010	3545683	2161	2823	1.30	10169
2011	3468790	2057	2798	1.36	9989
2012	3763175	1961	2541	1.40	9766
2013	3669890	2024	2648	1.39	9981

2014	4411865	2300	2679	1.35	11454
2015	5272645	2705	3068	1.32	13299
2016	8546653	3775	3581	1.32	19291
2017	8424522	4487	5728	1.36	21843
2018	7074668	4285	6705	1.46	21104
2019	6525978	3516	5672	1.60	18465

Table 42: DPS 17, 18, 19 and 20. Stock numbers at age (thousands).

age/year	0	1	2	3
2002	3753345	794512	77735	1563
2003	3785214	646993	91400	6884
2004	3993945	652378	72600	8274
2005	3908261	688239	71511	6608
2006	2347841	673367	73751	6177
2007	2939782	404437	70166	6076
2008	3249100	506251	40361	5459
2009	3913725	559261	47252	2999
2010	3545683	673233	47678	2888
2011	3468790	609493	51900	2527
2012	3763175	595925	43187	2418
2013	3669890	646316	40532	1915
2014	4411865	630342	44465	1810
2015	5272645	758023	45365	2099
2016	8546653	906180	56873	2282
2017	8424522	1468884	68109	2850
2018	7074668	1447284	103940	3145
2019	6525978	1214328	90385	3987

Table 43: DPS 17, 18, 19 and 20: F at age.

age/year	0	1	2	3
2002	0.01	1.22	1.70	1.70
2003	0.01	1.25	1.73	1.73
2004	0.01	1.27	1.76	1.76
2005	0.01	1.30	1.80	1.80
2006	0.01	1.32	1.83	1.83
2007	0.01	1.37	1.89	1.89
2008	0.01	1.43	1.99	1.99
2009	0.01	1.52	2.11	2.11
2010	0.01	1.62	2.25	2.25
2011	0.01	1.71	2.37	2.37
2012	0.01	1.75	2.43	2.43
2013	0.01	1.74	2.41	2.41
2014	0.01	1.69	2.35	2.35
2015	0.01	1.65	2.29	2.29
2016	0.01	1.65	2.29	2.29
2017	0.01	1.71	2.37	2.37
2018	0.01	1.83	2.54	2.54
2019	0.01	2.01	2.78	2.78

The FLBRP package allowed a Yield per Recruit analysis and an estimate of some F-based Reference Points. The fishing mortality rate corresponding to $F_{0.1}$ in the yield per recruit curve is considered here as a proxy of F_{MSY} , the value estimated is 0.44 and the estimated current F is 1.6 which results in a ratio of 3.34 and the stock is considered overexploited (Figure 121).

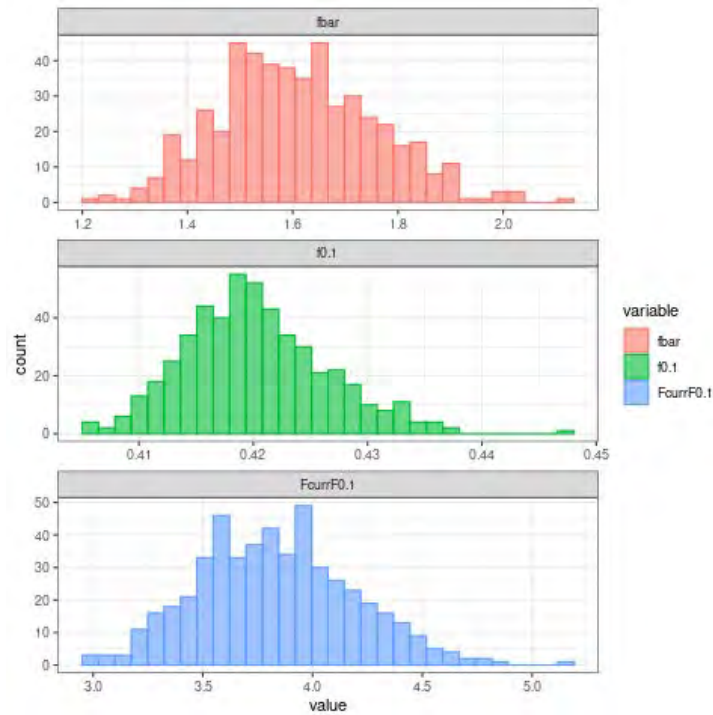


Figure 121: DPS 17, 18, 19 and 20. Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurrent/ F0.1) values.

4.7.4 Comparison with other assessment with different stock configuration

The only comparable assessment for DPS 17 – 20 is the one performed during the STECF 20 – 15 Working Group (STECF, 2020b) which was based in GSAs 17, 18 and 19. The overall perception of the stock is not that different between the two assessments, as for the one of STECF the ratio of $F_{curr}/F_{0.1}$ was estimated to be 2.98.

However, the increment of t_0 changed the overall distribution across ages, so the two stock assessments are not directly comparable. In the present assessment the cohort consistency of both catch and index improved (Figures 122-124). Finally the retrospective pattern in the present assessment appears more stable than the one in the STECF report.

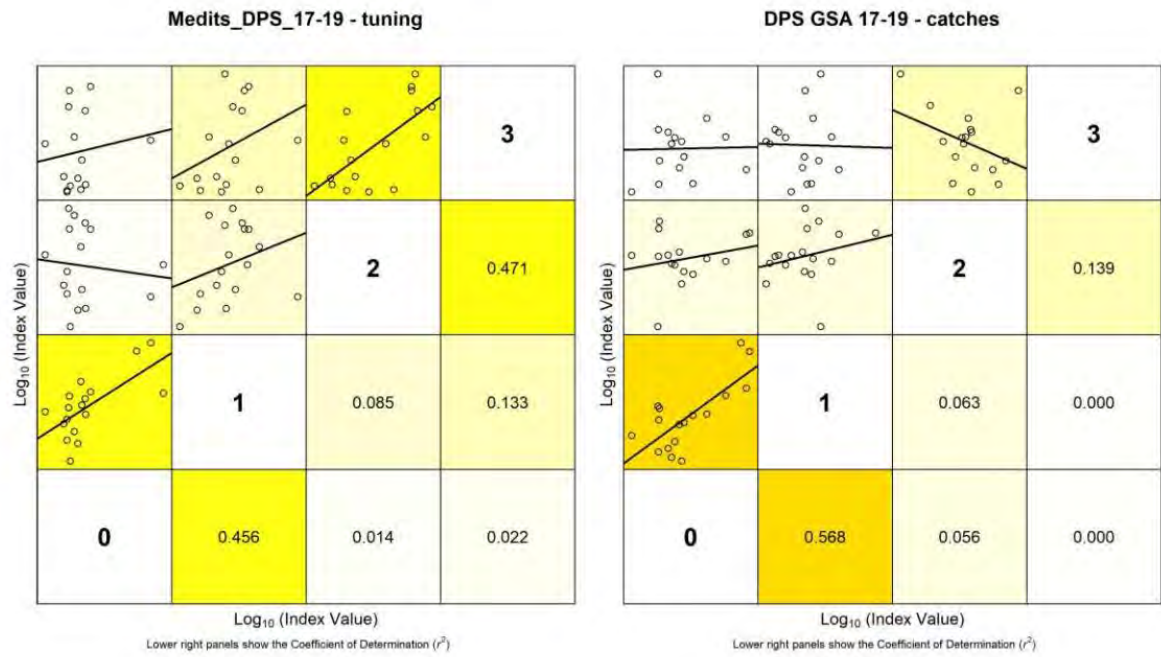


Figure 123: Figure 13 – DPS 17, 18, 19 and 20. Cohort consistency plots of the DPS in GSA 17 – 19 assessment performed in STECF EWG 20-15.

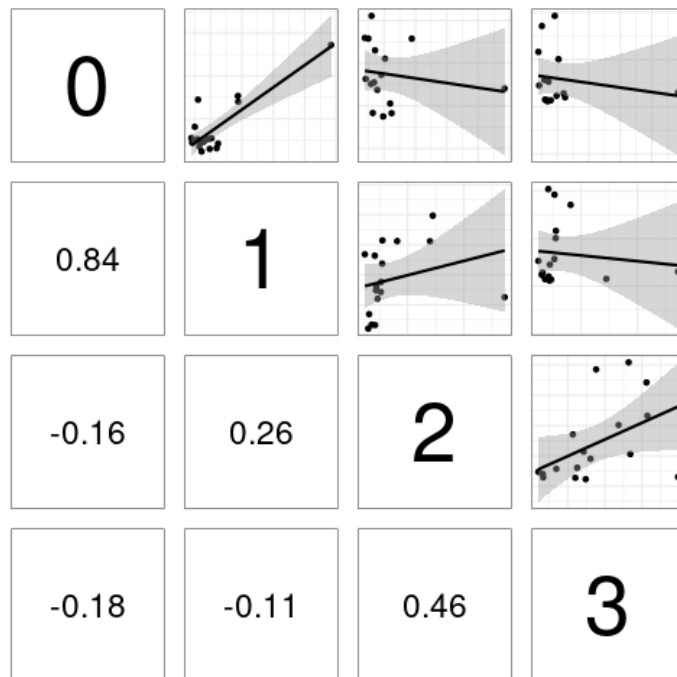


Figure 122: DPS 17, 18, 19 and 20. Cohort consistency plot of the MEDITS index

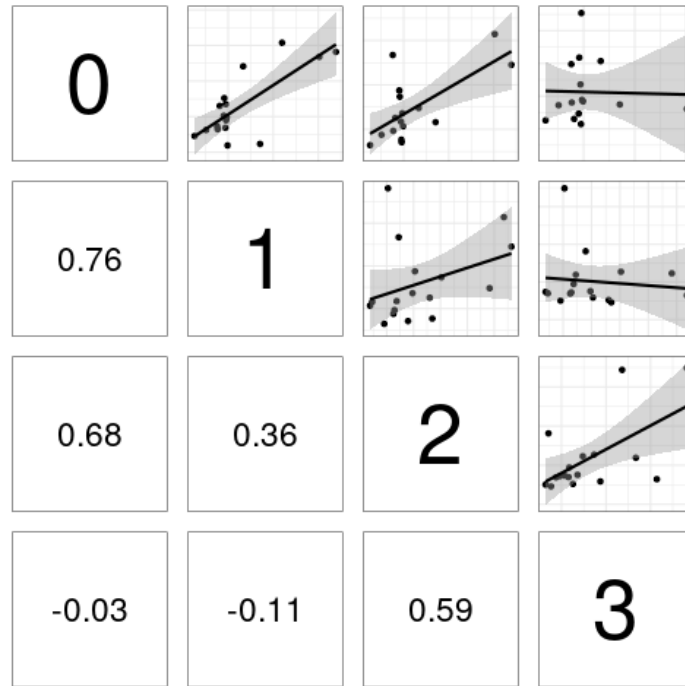


Figure 124: DPS 17, 18, 19 and 20. Cohort consistency plot of the catch at age.

4.8. Norway lobster in GSAs 1, 2, 5-11

In the western Mediterranean, Norway lobster (NEP), *Nephrops norvegicus*, represents one of the most valuable demersal resources for the trawling fleet operating on the muddy bottoms of the upper and middle slope from 300 to 600 m. Norway lobster is particularly important in GSA 6 and GSA 9, where it represents 37.8% and 38.3% respectively of the landing of the species considering the entire historical data series (2002-2019) in the investigated area.

According to the outcomes of the genetic analysis performed under the Med_Units project, a stock configuration including GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11 was identified, and used to perform an assessment (Figure 125).

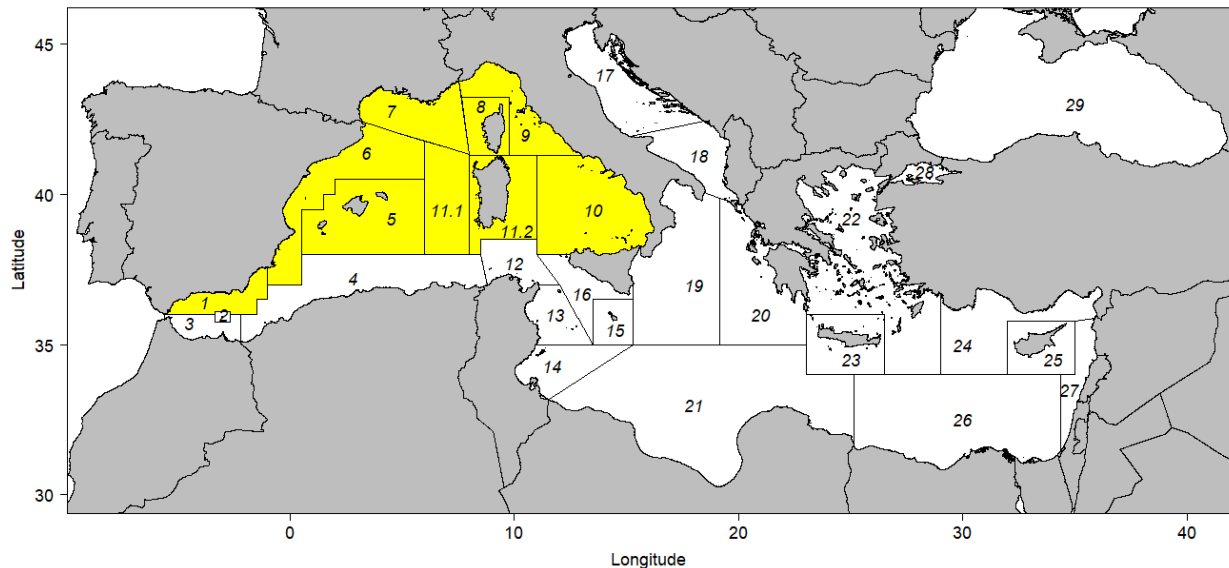


Figure 125: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Geographical location of the stock.

4.8.1 Input data

Fishery-Dependent data

A new stock object was produced and used to assess the stock of NEP in the GSAs of the western Mediterranean (time series 2003-2019). LFDs for the main gear/métier were available for GSAs 1, 5, 6, 7, 9, 10 and 11 and the missing ones were reconstructed from the distributions available for each GSA. LFDs were not available for GSAs 2 and 8. Landings from GSA 8 were added to the stock object, and the catch numbers-at-age was raised to those additional landings. GSA 2 was not included in the assessment as the catches in the area were negligible.

LFDs were sliced (deterministic slicing by sex) to obtain age structures using the growth parameters adopted for the stock assessment in GSA 9 during the STECF EWG 20-09 (STECF, 2020a) and GFCM-WGSAD 2020. In the same way, natural mortality, LW parameters and maturity vector used in the last assessment in GSA 9 were used (Table 46). Natural mortality vector was estimated using the Chen and Watanabe model. Mean weight-at-age were calculated using the LW parameters available under EU DCR/DCF for GSA 9, and applied to the whole time series (Table 47). The same approach was used for the maturity vector.

Fishery-Independent data

LFDs (n/km^2) from the MEDITS surveys in GSAs 1, 5, 7, 8, 9, 10 and 11 were sliced (deterministic age slicing by sex) using the same set of growth parameters used for the commercial data. No data was available for

GSA 2.

Table 44: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. VBGF and LW parameters by sex

Sex	L_{inf}	k	t_0
F	56.0	0.21	0.0
M	72.1	0.17	0.0
	a		b
F	0.00032		3.24848
M	0.00038		3.18164

Table 45: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Natural mortality and maturity-at-age vectors.

Age	Natural mortality	Proportion of matures
1	0.77	0.40
2	0.50	0.75
3	0.39	1.00
4	0.33	1.00
5	0.30	1.00
6	0.27	1.00
7	0.25	1.00
8	0.23	1.00
9+	0.22	1.00

The catch at age structure as well as the abundance at age resulting from the slicing are presented respectively in Figure 126 and Figure 127. Moreover the internal consistencies of such input data is presented respectively in Figure 128 and Figure 129.

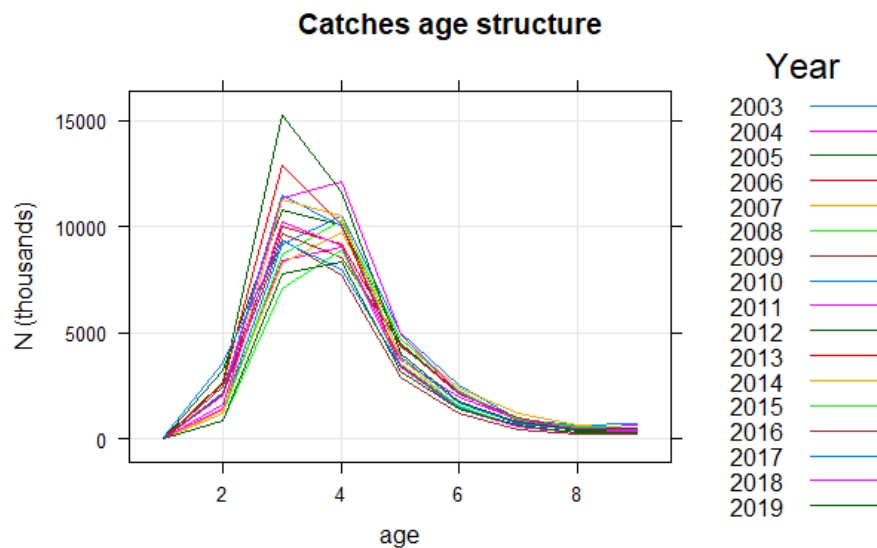


Figure 126: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Catch numbers-at-age.

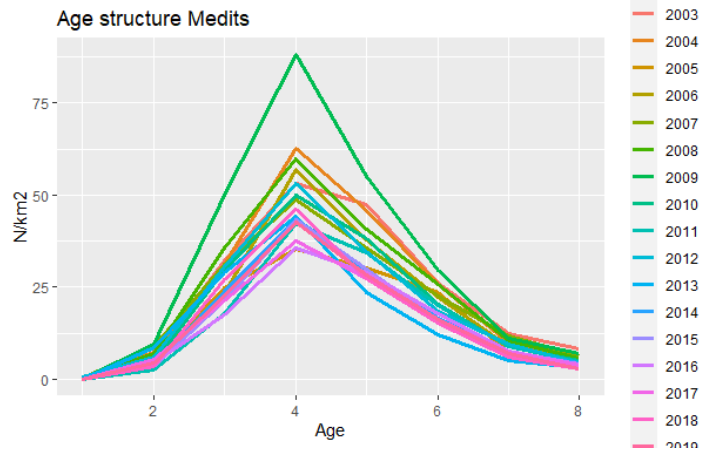


Figure 127: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Catch numbers-at-age in the MEDITS.

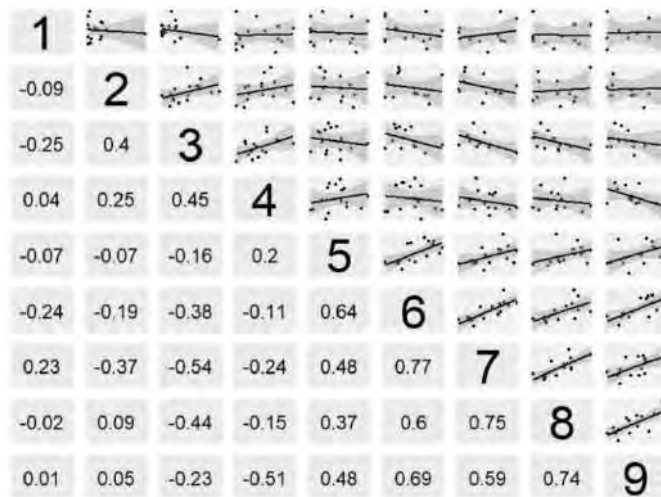


Figure 128: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Cohort consistency in the catch numbers-at-age.

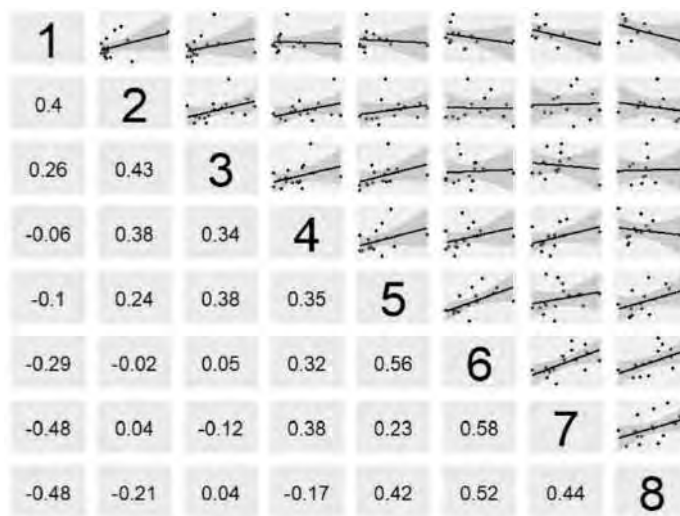


Figure 129: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Cohort consistency in the catch numbers-at-age in the MEDITS.

4.8.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment by means of a4a was carried out using as input data the period 2003-2019 for the catch data and 2003-2019 for the tuning file (MEDITS indices).

A natural mortality vector computed using Chen and Watanabe model was used in the assessment. Length-frequency distributions of commercial catches and surveys were split by sex and then transformed in age classes (plus group was set at age 9) using length-to-age slicing with different growth parameters by sex. A correction of 0.5 was applied to t_0 to align length slicing to assessment year (January to December) to account for spawning at the middle of the year.

The number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers at age $a \times$ catch weight-at-age a)].

In the catches, a plus group at age 9 was set. A true age 8 was used in the survey.

F_{bar} range was fixed at 2-6.

The assessment was performed by sex combined. Given that the landings were composed mainly of individuals between 2 and 6 years, these ages were selected as F_{bar} range.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

```
Fishing mortality sub-model: fmodel = factor(replace(age, age>7,7))+s(year, k=5)
```

```
Catchability sub-model: qmodel = list(~factor(replace(age, age>5,5)))
```

```
SR sub-model: srmod = geomean(CV=0.2)
```

```
Model <- sca(stock = stk, indices = idx, fmodel, qmodel, srmod)
```

The n1model and vmodel used in the final fit are the default ones:

```
n1model <- ~s(age, k = 3)
```

```
vmodel <- list(~s(age, k=3), ~1)
```

4.8.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the following figures. The model estimated 41 parameters out of 289 observations; it is then well below the threshold of 25% ratio between parameters and observations (Figure 130 and Figure 131).

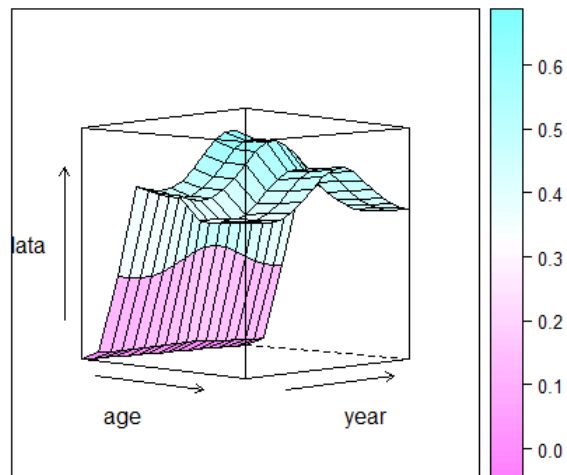


Figure 130: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Fishing mortality by age / year obtained from the a4a model (2003-2019).

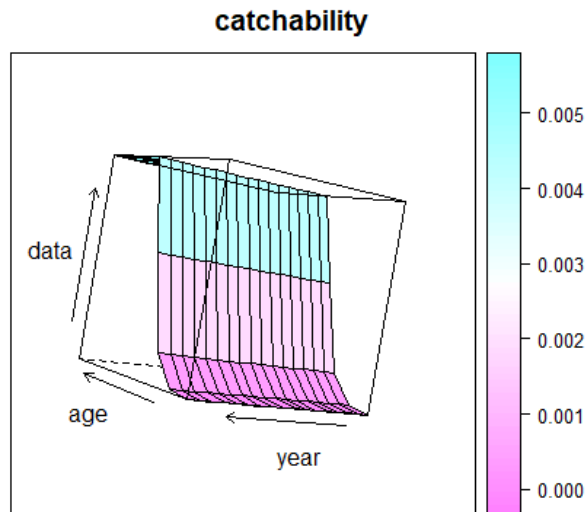


Figure 131: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Catchability of the survey by age and year obtained from the a4a model (2003-2019).

The log residuals for both the catches and the survey do not show any particular trend or issue. The fitting of the survey shows some problems particularly in 2005 and 2009 (Figure 132Figure **135**), probably due to the poor internal consistency of the survey. Despite this, the diagnostics are considered acceptable and the a4a model is acceptable.

log residuals of catch and abundance indices by age

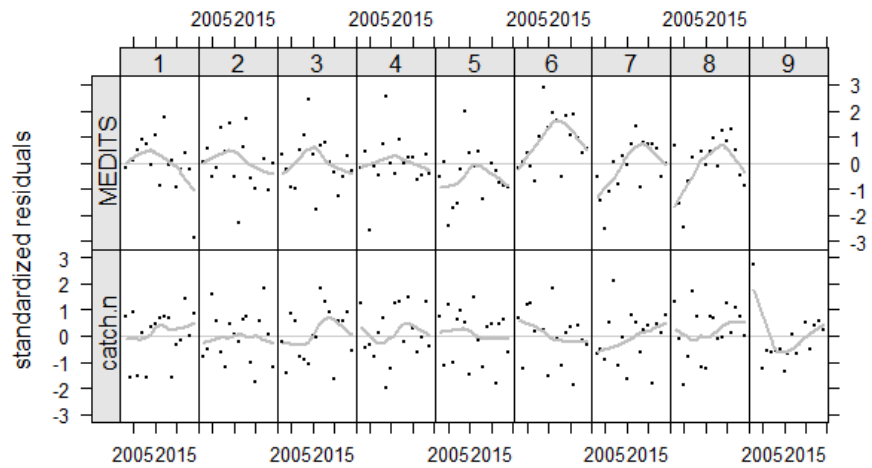


Figure 132: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Log residuals for the catch-at-age data of the fishery and the survey, and the catches.

log residuals of catch and abundance indices

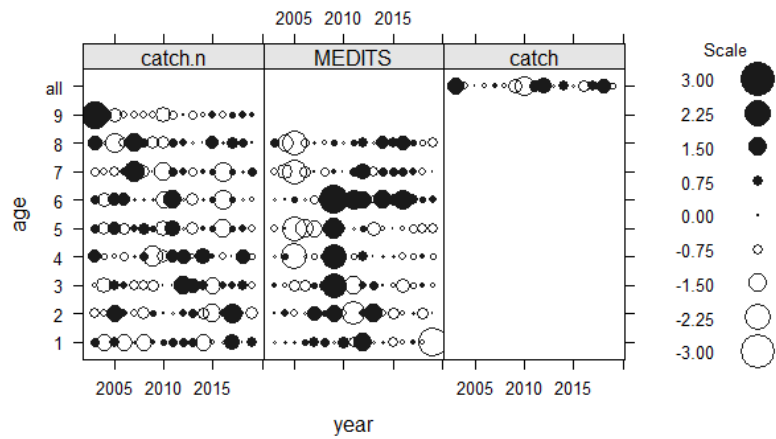


Figure 133: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Bubble plots of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

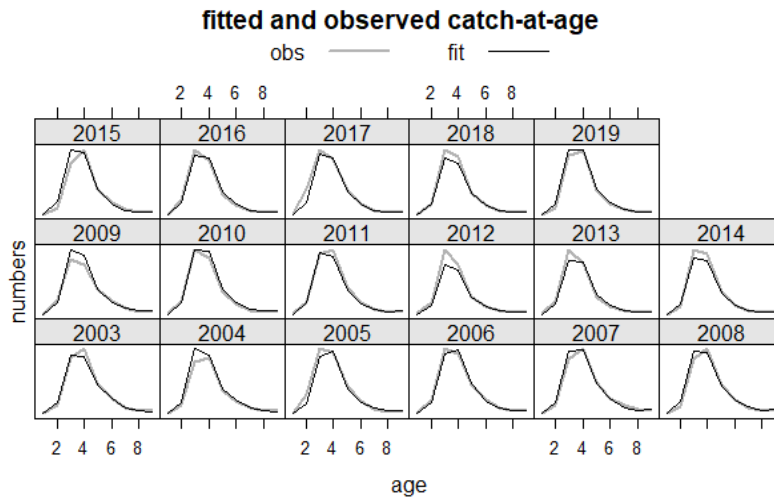


Figure 134: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Fitted vs observed values by age and year for the catches.

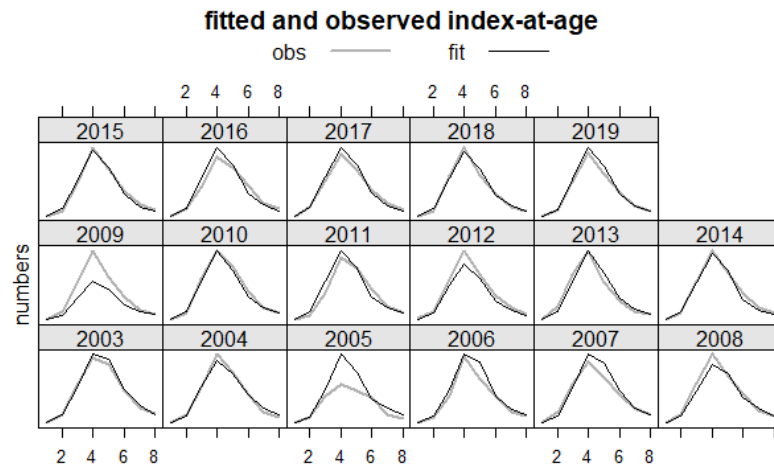


Figure 135: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated, and did not show any relevant issue, as the biomass of the plus group (age 9+) is always below 10% of the total SSB.

The retrospective analysis shows some comb effect in F, SSB and recruitment, indicating that the assessment model is poorly stable (Figure 136). The Mohn's Rho for F, SSB and recruitment are within the optimal range -0.2/0.2 (-0.05 for F, 0.01 for SSB, and 0.02 for recruitment).

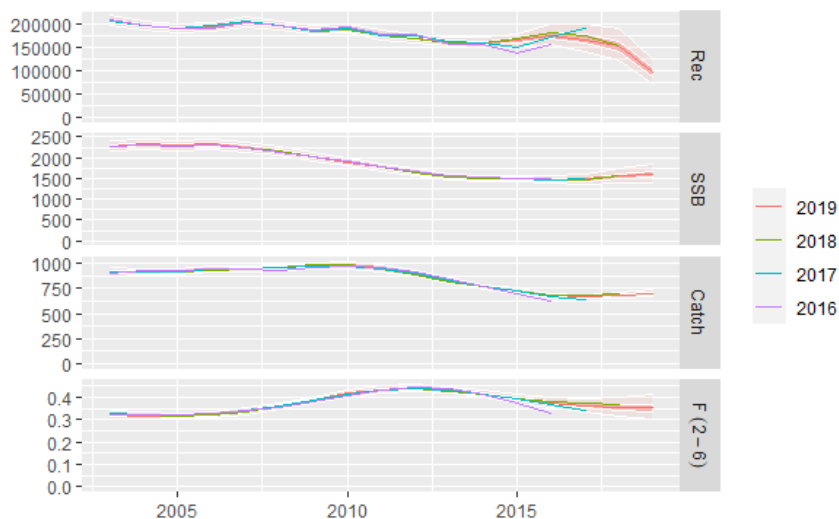


Figure 136: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Retrospective analysis.

Final assessment outcomes are given in Figure 137, and Table 48Table 50. Values of F0.1 calculated by FLBRP package on the a4a assessment results is equal to 0.15 (Figure 138). Current F value (2019), as calculated by model a4a, is 0.35 indicating that the stock is being overexploited.

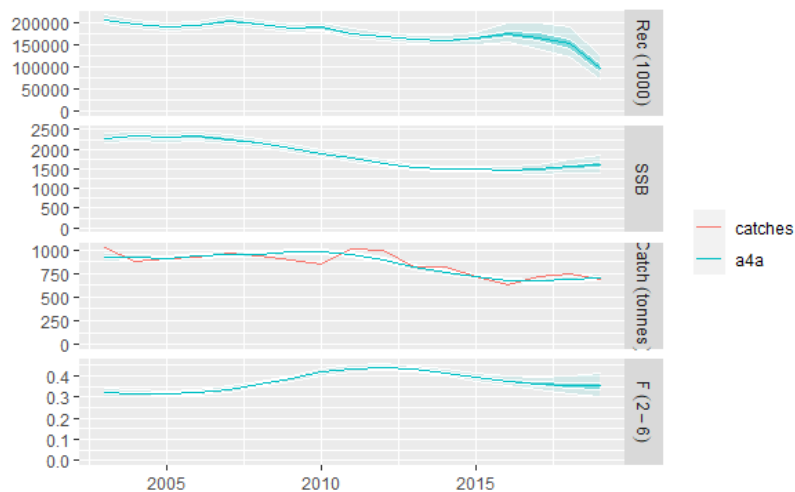


Figure 137: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Outputs of the a4a stock assessment model, with uncertainty; input catch data (red line) are plotted against the estimated catches.

Table 46: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Stock numbers-at-age (thousands).

age/year	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	204907	195158	189060	192001	203021	195087	185386	188199	174303
2	103334	95230	90699	87918	89233	94410	90666	86916	88115
3	55117	61042	56248	53425	52025	52616	55633	53285	51000
4	29110	29556	32892	30549	28746	27607	27495	28414	26735
5	15088	13086	13365	14933	13771	12694	11741	11218	11123

6	7634	7161	6260	6432	7155	6433	5729	5106	4685
7	3771	3750	3564	3112	3174	3462	3011	2569	2207
8	1828	2016	2031	1915	1671	1678	1778	1494	1235
9+	877	1472	1916	2157	2220	2086	1967	1890	1651
age/year	2012	2013	2014	2015	2016	2017	2018	2019	
1	167381	161098	159658	163738	174580	166118	151145	95902	
2	81167	78163	74970	74200	75885	81067	77384	70029	
3	51571	47454	45726	43964	43633	44664	47645	45621	
4	25241	25438	23528	23021	22476	22529	23198	24928	
5	10183	9540	9741	9262	9313	9348	9511	9909	
6	4519	4119	3900	4084	3990	4105	4197	4306	
7	1980	1891	1744	1686	1819	1820	1908	1966	
8	1036	924	892	836	830	915	929	979	
9+	1375	1142	988	920	880	874	929	969	

Table 47: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Fishing mortality-at-age.

age/year	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.025	0.024	0.024	0.024	0.026	0.027	0.030	0.032	0.033
3	0.232	0.228	0.227	0.231	0.242	0.259	0.279	0.299	0.313
4	0.471	0.462	0.460	0.468	0.490	0.525	0.567	0.606	0.634
5	0.449	0.441	0.439	0.447	0.468	0.501	0.540	0.578	0.604
6	0.442	0.434	0.432	0.440	0.460	0.493	0.532	0.569	0.595
7	0.380	0.373	0.371	0.378	0.396	0.423	0.457	0.489	0.511
8	0.380	0.373	0.371	0.378	0.396	0.423	0.457	0.489	0.511
9+	0.380	0.373	0.371	0.378	0.396	0.423	0.457	0.489	0.511
age/year	2012	2013	2014	2015	2016	2017	2018	2019	
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
2	0.033	0.033	0.031	0.030	0.029	0.028	0.027	0.027	
3	0.316	0.310	0.293	0.283	0.270	0.261	0.256	0.253	
4	0.642	0.630	0.604	0.574	0.548	0.530	0.520	0.513	
5	0.612	0.600	0.576	0.548	0.523	0.506	0.495	0.489	
6	0.602	0.591	0.567	0.539	0.515	0.498	0.488	0.482	
7	0.518	0.508	0.487	0.463	0.442	0.428	0.419	0.414	
8	0.518	0.508	0.487	0.463	0.442	0.428	0.419	0.414	
9+	0.518	0.508	0.487	0.463	0.442	0.428	0.419	0.414	

Table 48: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Summary results of the a4a assessment.

	Catch (t)	SSB (t)	Rec ('000)	$F_{\text{bar}(2-6)}$	Total biomass (t)
2003	914.7	2281.5	204907	0.32	3099.1
2004	914.9	2332.6	195158	0.32	3141.7
2005	912.4	2292.0	189060	0.32	3039.5
2006	936.5	2323.9	192001	0.32	3084.7
2007	944.1	2247.6	203021	0.34	3013.2
2008	955.1	2152.9	195087	0.36	2926.9
2009	977.0	2027.9	185386	0.39	2753.3
2010	978.7	1901.8	188199	0.42	2604.4
2011	949.9	1760.2	174303	0.44	2439.3
2012	892.8	1645.6	167380	0.44	2287.1
2013	821.8	1538.2	161098	0.43	2165.8
2014	763.4	1499.8	159658	0.42	2129.6
2015	722.4	1488.3	163738	0.39	2069.4
2016	676.6	1469.4	174580	0.38	2112.3
2017	666.8	1478.6	166118	0.36	2077.9
2018	683.3	1563.8	151145	0.36	2190.2
2019	702.0	1594.9	95902	0.35	2133.5

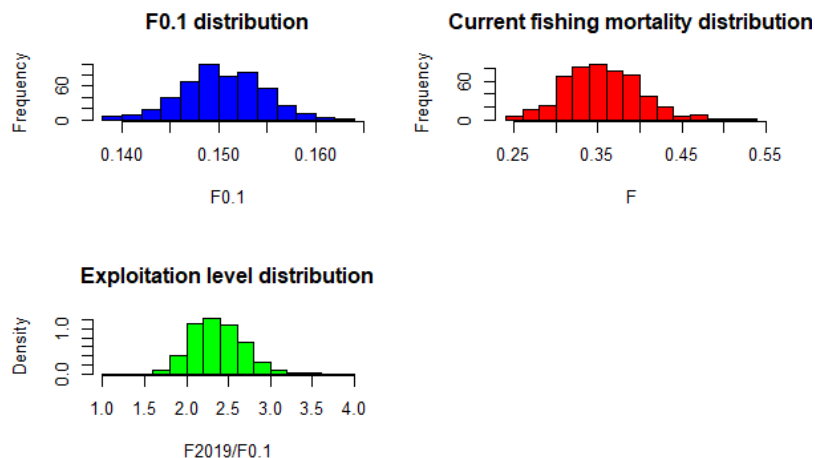


Figure 138: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Histograms of probability for $F_{0.1}$, F_{curr} and level of exploitation ($F_{\text{curr}}/F_{0.1}$ ratio) values.

Besides the model selection procedure, a further check was performed regarding the number of knots (k) of the smoother on year in the fmodel. A test based on AIC, BIC and GCV was performed on k ranging

between 4 and 9. All the model specifications highlight a consistent behaviour in terms of main outcomes (Figure 139). A k value of 5 was retained as the one providing the best retrospective (Figure 140).

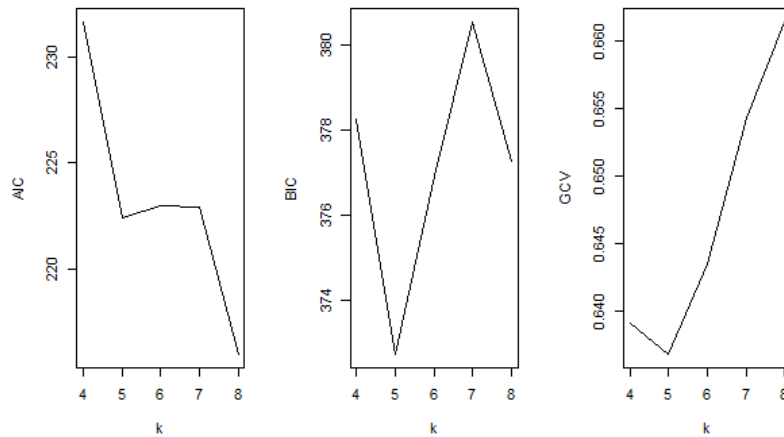


Figure 139: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. AIC, BIC and GCV values estimated on a range of k values of the smoother on year of the fmodel.

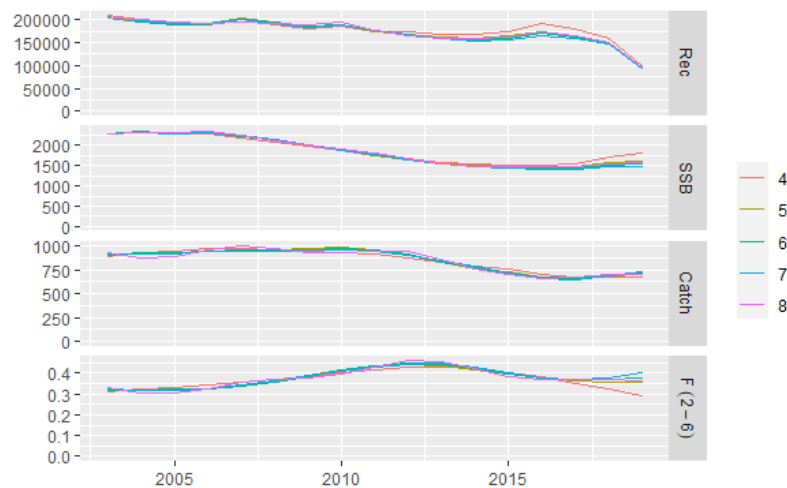


Figure 140: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Outputs of model runs with different k values on the smoother on year in the fmodel.

4.8.4 Comparison with other assessments with a different stock configuration

Considering the output of the genetic, the stock of Norway lobster living in the Mediterranean western basin seems to overlook the conventional GSAs boundaries used in the current assessments, resulting in only one unit. This genetic configuration has been followed and the assessment for the Western Mediterranean stock, including GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11, have been performed. Considering the current assessments developed within the STECF (STECF, 2020a, b) and GFCM (GFCM, 2021) frameworks, the assessment here presented has been compared with those for GSA 6 and GSA 9 in order to verify if there are evident improvements or deterioration of stock assessment quality, and to check the degree of coherence in the stock perception that can be found between the single GSA assessments and the

Med_Units framework.

Model diagnostics

The diagnostics of the models for GSA 6 and GSA 9 performed during the STECF EWG 20-09 (STECF, 2020a) were plotted for comparison with the diagnostic of the model presented in this report (Figure 141). The diagnostics for single GSAs models and for the Med_Units model do not show severe issues. The magnitude and the trend of the residuals are comparable. A slight increase in magnitude is observed in the Med_Units stock assessment, but this is essentially due to the older age classes in the Medits indices.

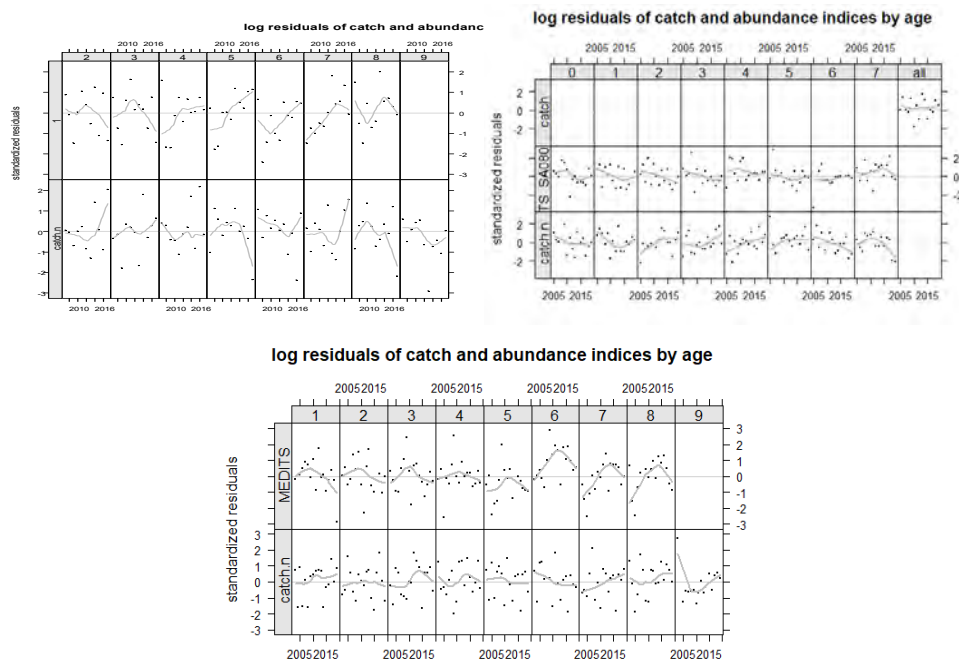


Figure 141: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11: model diagnostic for the stock assessment of STECF EWG 20-09 for GSA 6 (top left) and GSA 9 (top right); and on the MED-UNITS aggregation (bottom).

Retrospective patterns

The retrospective pattern observed for the MED-UNITs stock aggregation had a greater consistency than most of the single GSAs assessment (Figure 142). In particular, the models fitted on GSA 9 is rather unstable, an issue that was not observed in the aggregated assessment.

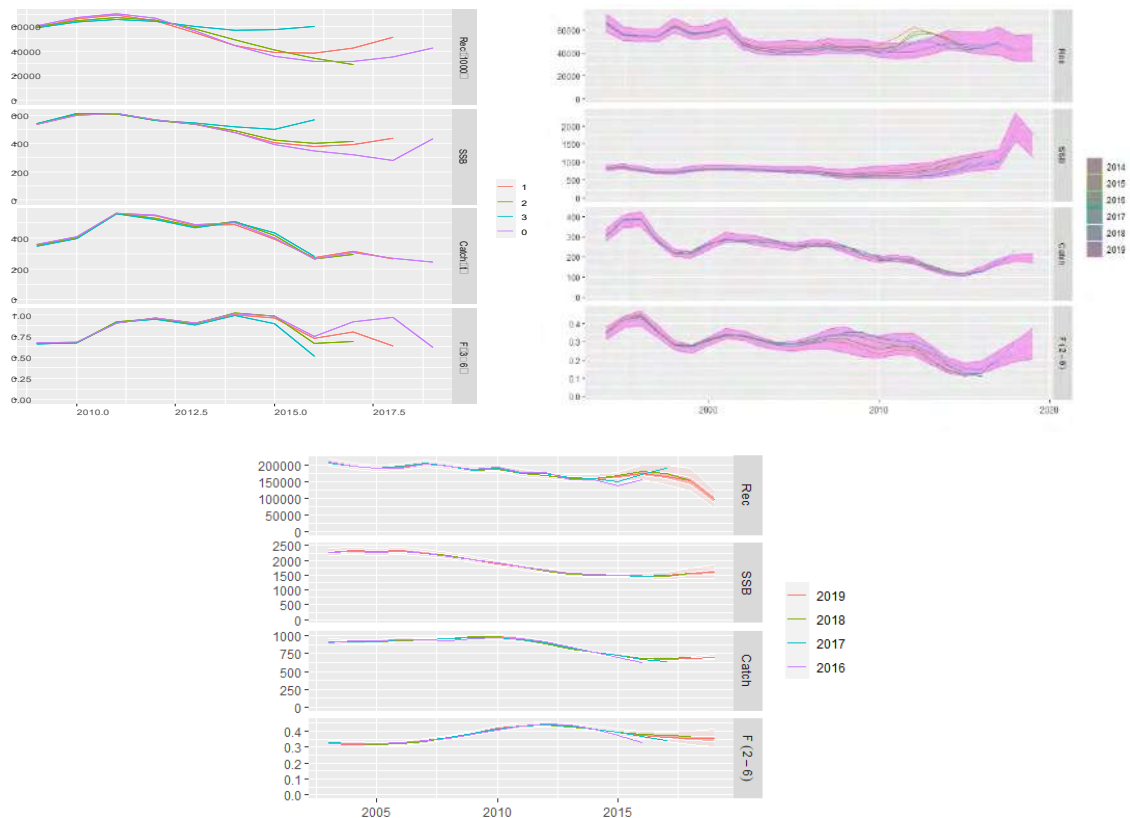


Figure 142: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Retrospective analysis for the stock assessment of STECF EWG 20-09 for GSA 6 (top left) and GSA 9 (top right) and on the Med_Units aggregation (bottom).

Stock trajectories

The trajectories of the single GSA assessments (GSA 6 and GSA 9) have been compared to the Med_Units configuration to check the coherence for SSB and Recruitment (Rec 000) trend estimates, as well as for F_{bar} and F/F_{01} trends (Figure 143). Due to data availability (Norway lobster stock assessment for GSA 6 considers only years from 2009 to 2019), the comparison was restricted to the time-frame 2009-2019. The comparisons for SSB and Recruitment were done by summing the contribution of single GSAs, then data were plotted and Pearson correlation coefficient were computed. The Med_Units configuration estimates higher SSB values, especially in the early years of the time series. Then, the trend tends to converge with very similar values in the last two years. However, it must be taken into account that the Med_Units configuration also includes other GSAs which contribute to about 30% of the total landing in the area. This factor can also have a significant effect on the estimate of the recruitment which turns out to be about double in the Med_Units configuration compared to the sum of the two GSAs. Also in this case a convergence is observed in the last year of the time series.

The F_{bar} and F/F_{01} trends were compared to the results obtained for single GSAs, including an estimation of the average annual value. The two trends appear to be very similar, although a greater inter-annual fluctuation is observed in the average values of the two GSAs.

The F_{bar}/F_{01} results very different between the Med-Units configuration and the single GSAs assessments, in particular when compared with the trend observed for GSA 6.

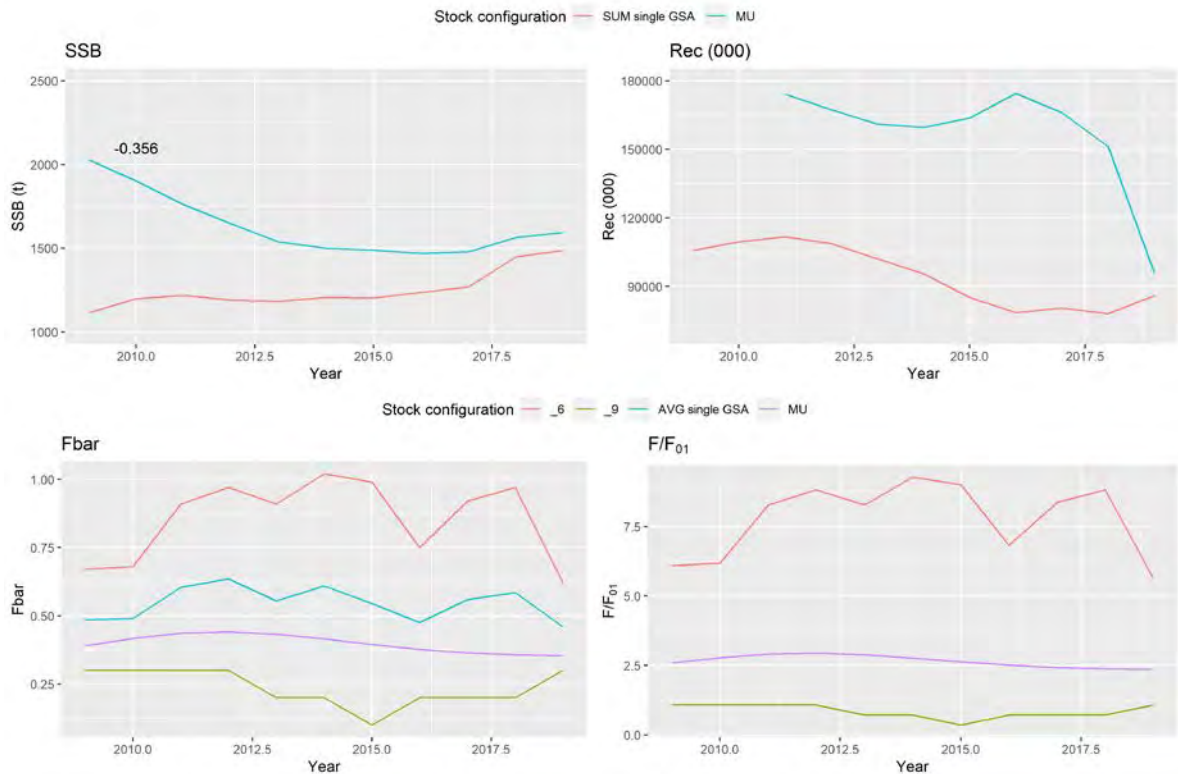


Figure 143: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. Stock trajectories calculated on single GSAs 6 and 9 and on the MED-UNITS aggregation. MU= MED-UNITS aggregation.

F_{bar} (terminal year) and reference points

F_{bar}, F₀₁ and their ratio (F_{bar}/F₀₁) have been displayed for the last available year (2019; Figure 144). The F_{bar} and F₀₁ for the MED-UNITS configuration rank between the values estimated for the other two assessments, whereas both values for the MED-UNITS assessment result more similar to those estimated in the GSA 9 assessment. This result may also have been influenced by the inclusion of other GSAs in the Med_Units configuration.

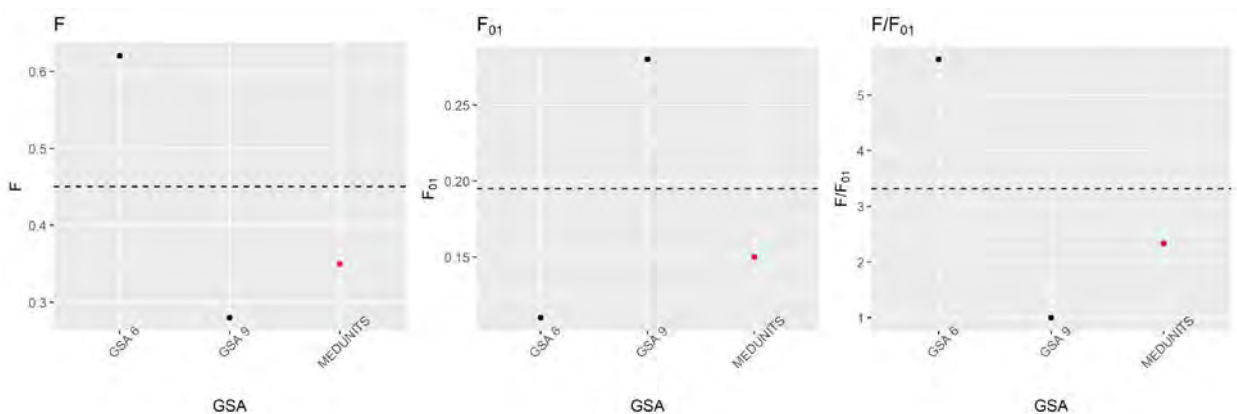


Figure 144: NEP in GSAs 1, 2, 5, 6, 7, 8, 9, 10 and 11. F_{bar}, F₀₁ and their ratio calculated on single GSAs and on the MED-UNITS aggregation.

Model diagnostics did not present large difference in the model fits, whereas an improvement in the retrospective patterns was observed in the MED-UNITS aggregation. The stock trajectories observe important differences between the SSB and recruitment estimated for the Med-Units assessment and

those from the single GSAs assessments. F_{bar} and F_{01} estimated for the Med_Units configuration show intermediate values compared to those observed in the individual assessments for GSA 6 and for GSA 9. Therefore, the result obtained with the Med_Units configuration seems plausible, even if it must always be taken into account that other GSAs have been included in the analysis and that made the results less comparable with the available assessments.

4.9. Norway lobster in GSAs 17-19

In the Adriatic and Ionian Seas, Norway lobster (NEP), *Nephrops norvegicus*, represents one of the most valuable demersal resources for the trawling fleets operating on the muddy bottoms of the upper and middle slope. In GSA 18 the stock is basically distributed on the continental slope, deeper than 200 m depth, both on the eastern (Montenegro, Albania) and western side (Italy, Puglia) of the GSA. The distribution of nursery grounds and spawning areas has been analyzed during the EU project MEDISEH (MAREA tender project). In GSA 17 denser and persistent patches of small specimens occur in the Pomo Pit area. Aggregations of adults were identified in GSA 17 offshore the SW coasts, in the Pomo Pit, and in north and south Croatian waters. In GSA 18 the more persistently abundant adult aggregations occur on the SE and SW edges of the South Adriatic Pit.

According to the outcomes of the genetic analysis performed under the Med_Units project, a stock configuration including GSAs 17, 18 and 19 was identified, and used to perform an assessment (Figure 145).

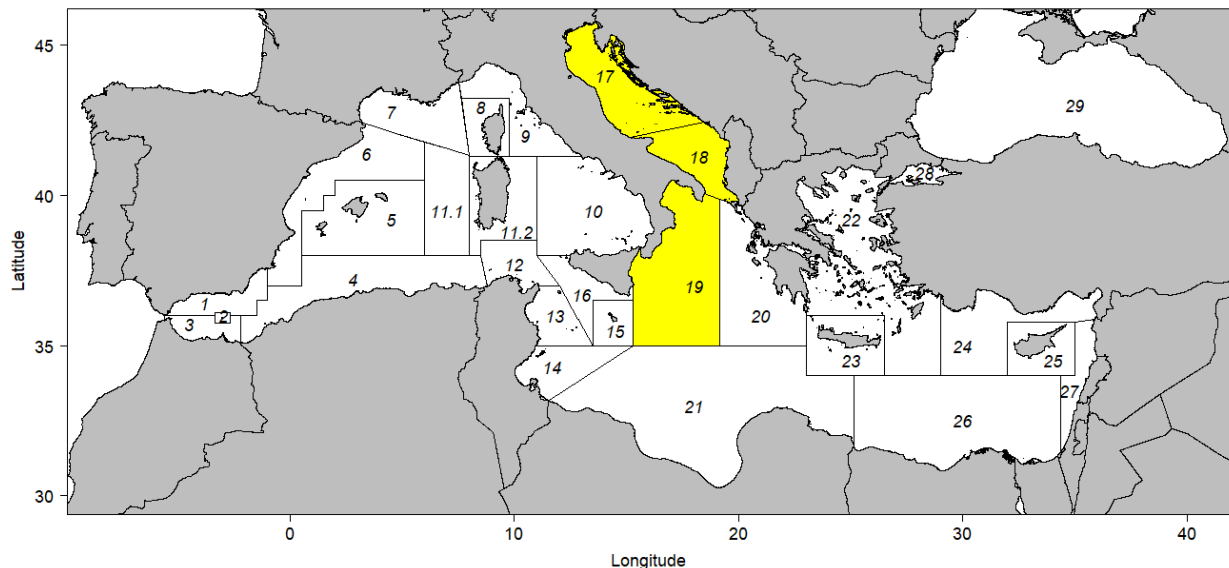


Figure 145: NEP in GSAs 17, 18, 19. Geographical location of the stock.

4.9.1 Input data

Fishery-Dependent data

No data were available for Slovenia because Norway lobster it is not caught in Slovenian fishing grounds. In the following sections Croatian, Italian and Albania data in term of landing s and discards in weight are reported.

Time series of landings starting from 1970 for the Adriatic and Ionian FAO subdivisions were gathered for Italy, Croatia, Montenegro and Albania from the FAO Official Statistics. In recent years, available data from EU DCF were used for Italy and Croatia. These landings were compared with those used for the assessment of NEP in GSAs 17-18 performed by STECF EWG 20-09. As concerns the period covered by DCF, landings from GSA19 were added to the landings used in the assessment of NEP in GSAs 17-18.

A new time series of landings (1970-2019) was then obtained and used to assess the stock of NEP in the GSAs 17, 18 and 19. The contribution in landings of GSA 19 is lower than 10% according to EU DCF data.

Fishery-Independent data

The time series of biomass indices (kg/km^2) from the MEDITS surveys in GSAs 17, 18 and 19 from Italy, Slovenia, Croatia, Montenegro and Albania was used as tuning information in the SPiCT assessment (Figure 146 and Figure 147).

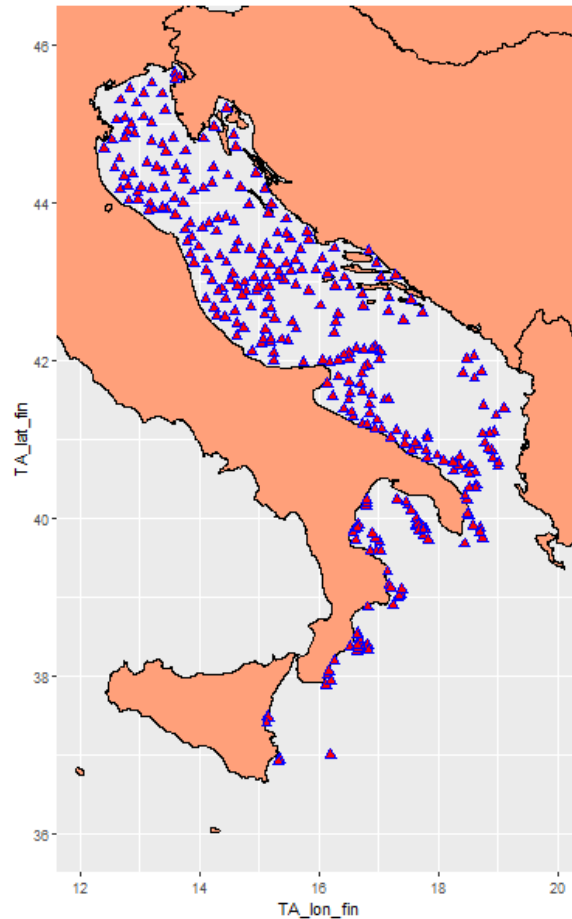


Figure 146: NEP in GSAs 17, 18, 19. Spatial distribution of MEDITS hauls in 2019.

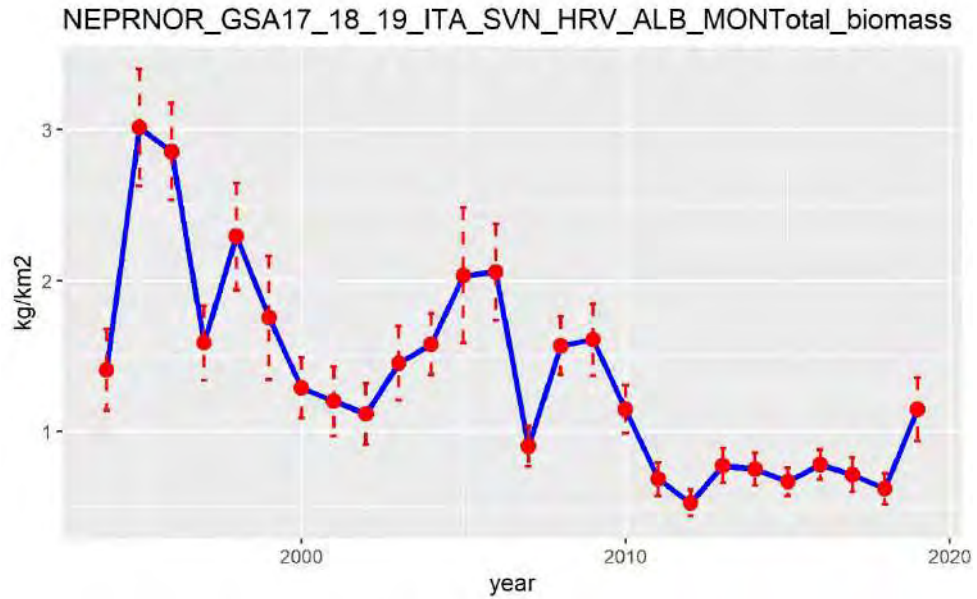


Figure 147: NEP in GSAs 17, 18, 19. Time series of the biomass index (1994-2019).

As additional tuning information, the time series of the surveys from Froglija (1988; time series 1976-1985) and Jukic (1975; time series 1960-1970) were included in the assessment (see STECF, 2020b; Figure 148).

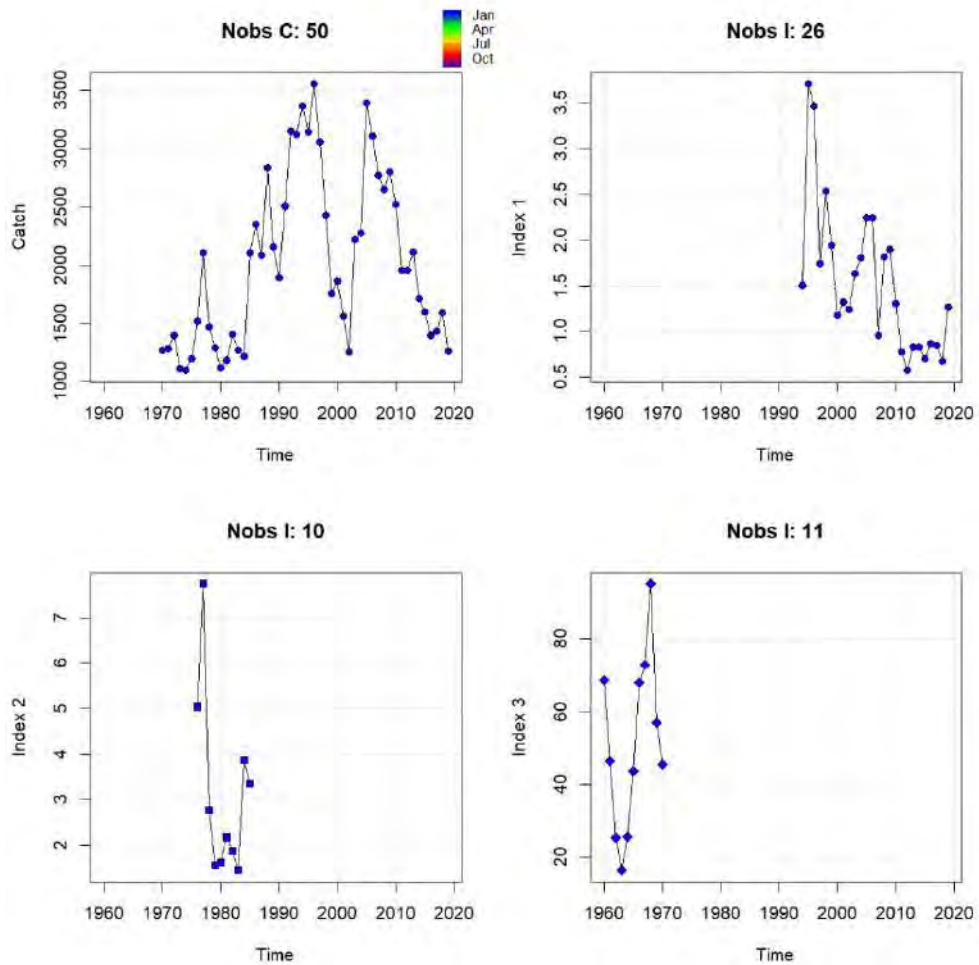


Figure 148: NEP in GSAs 17, 18, 19. Input data: landings from 1970 to 2019 (top-left), MEDITS biomass index (kg/km²) from 1994 to 2019 (top-right); Froglija survey 1976-1985 (bottom-left), and Jukic survey 1960-1970 (bottom-right).

4.9.2 Model description

The Surplus Production in Continuous time (SPiCT) assessment method is described in Pedersen and Berg (2016). The SPiCT assessment method is a state-space version of the Pella-Tomlinson surplus production model (Pella and Tomlinson 1969).

The fitting of the input data (landings vs surveys) are shown in Figure 149 and Figure 150.

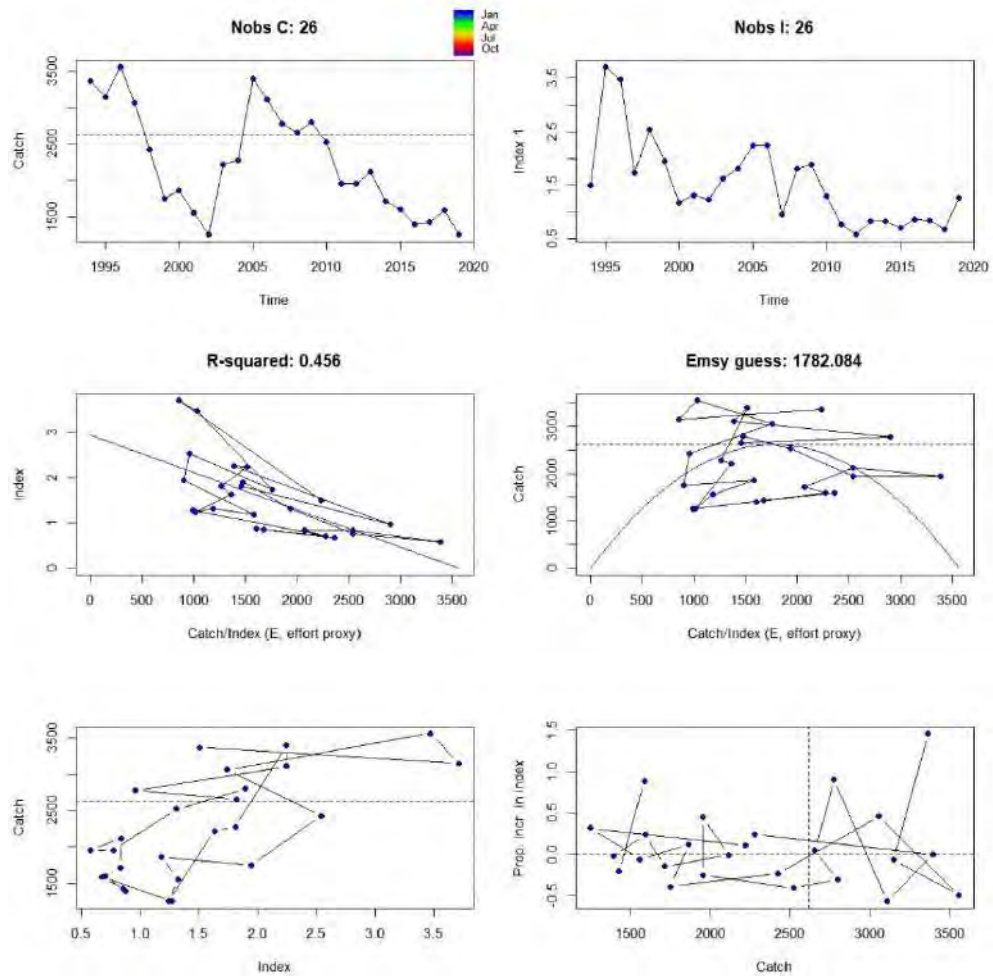


Figure 149: NEP in GSAs 17, 18, 19. Fitting of the input data (landings vs MEDITS biomass index).

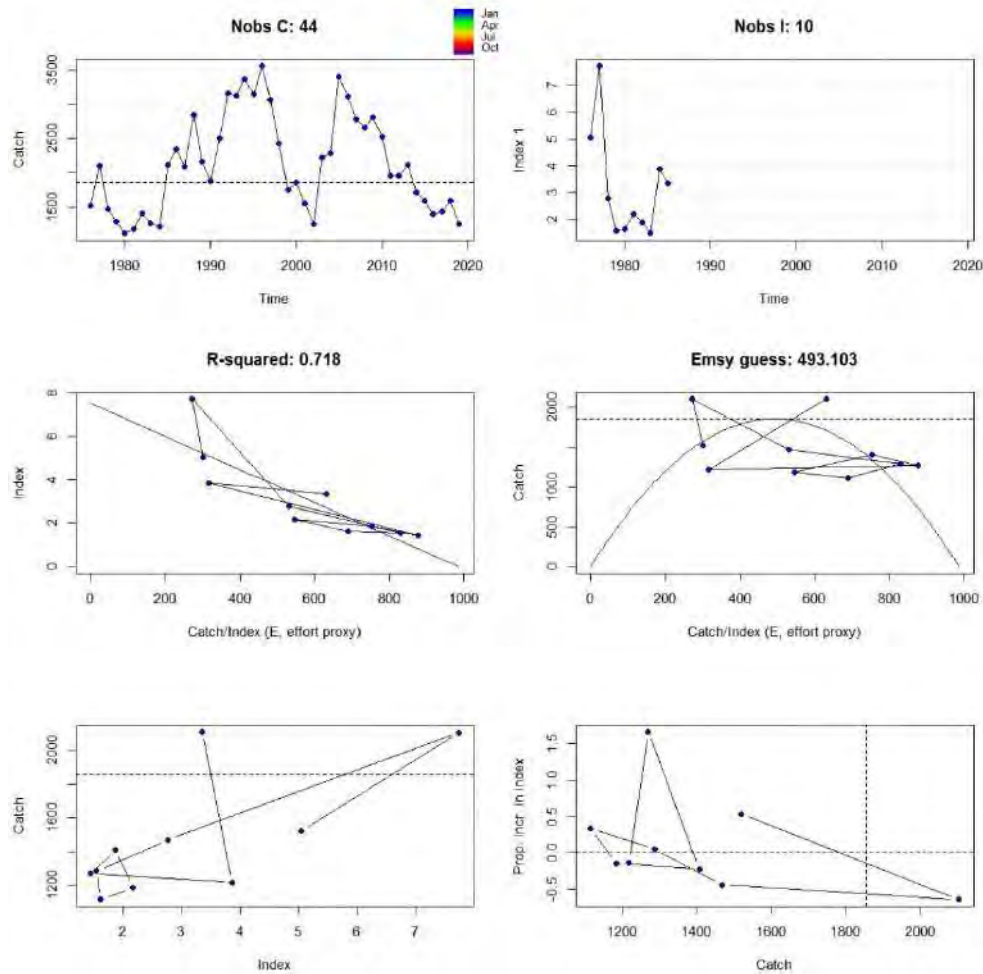


Figure 150: NEP in GSAs 17, 18, 19. Fitting of the input data (landings vs Frogliia survey).

4.9.3 Outputs from the model

SPiCT was run with the default prior settings and no informative priors for initial parameter estimates (Figure 151). The model converged and the diagnostic results (Residuals, Auto correlation and Shapiro p-values) are good for both catches and the 3 tuning indexes (Figure 152).

A retrospective was run with 3 retro years. For production models, the most reliable estimates are in terms of F/F_{MSY} and B/B_{MSY} . The retrospective patterns are consistent across years in terms of B/B_{MSY} with biomass estimated well below B_{MSY} . The consistency of the results indicates the retrospective performance is acceptable (Figure 153).

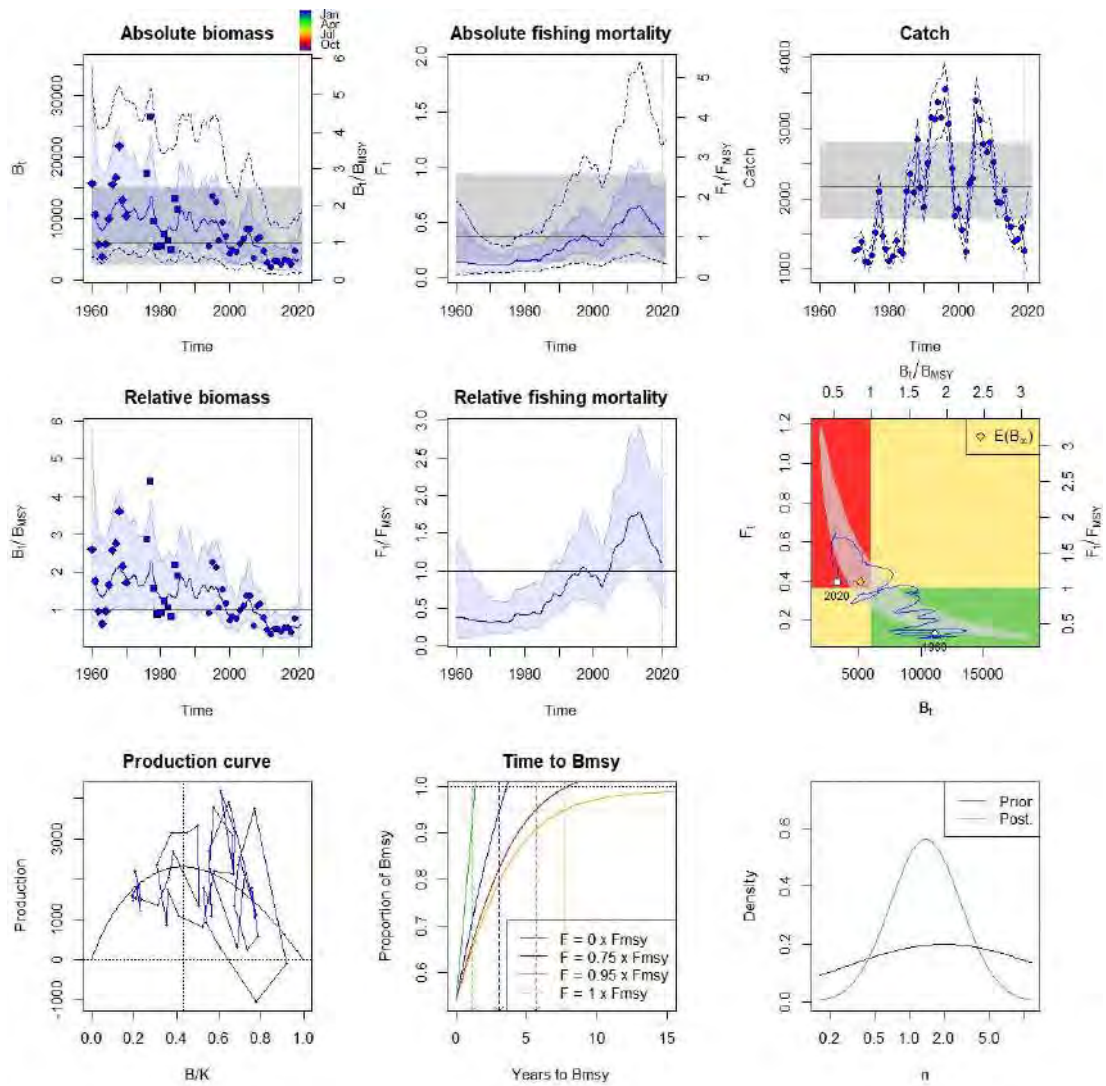


Figure 151: NEP in GSAs 17, 18, 19. Plots of the main results of the model.

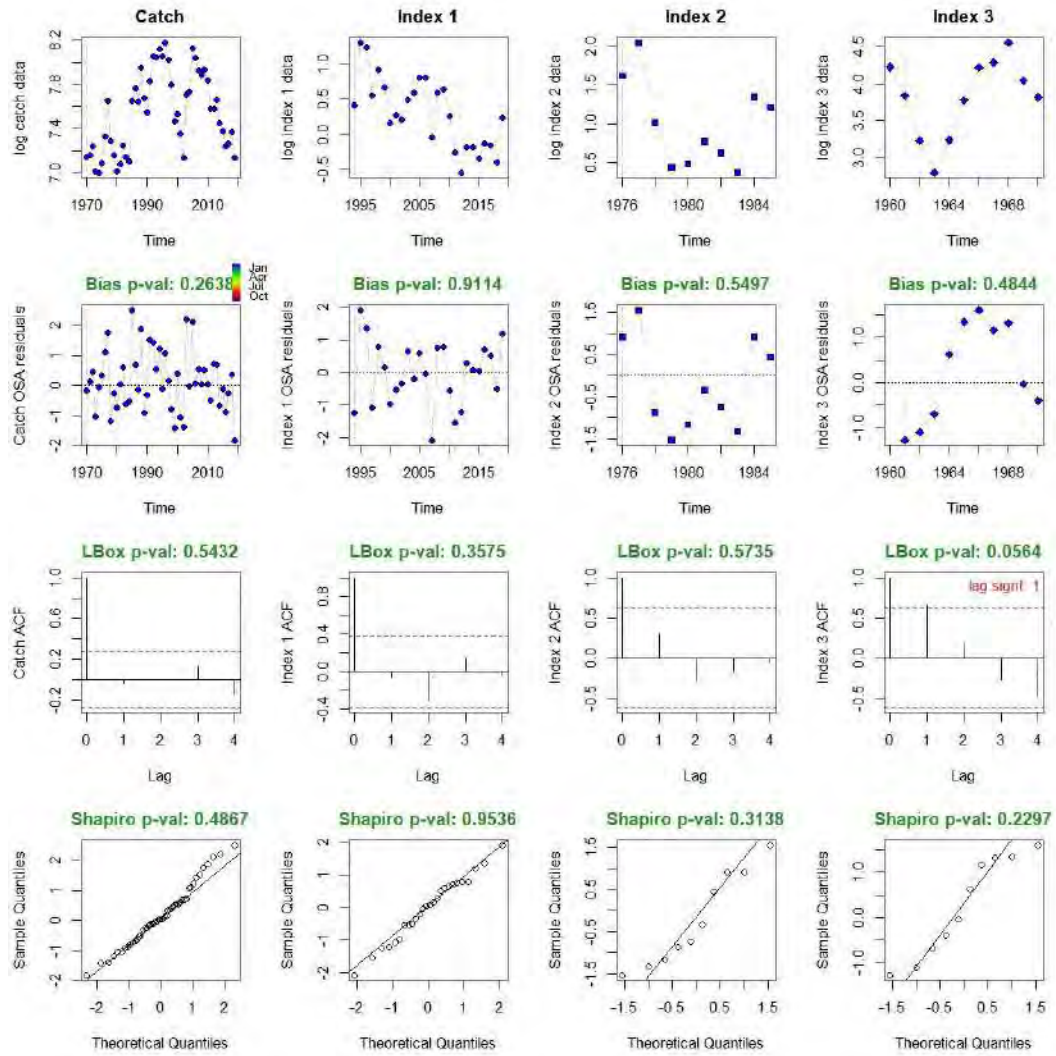


Figure 152: NEP in GSAs 17, 18, 19. Plots of the diagnostics.

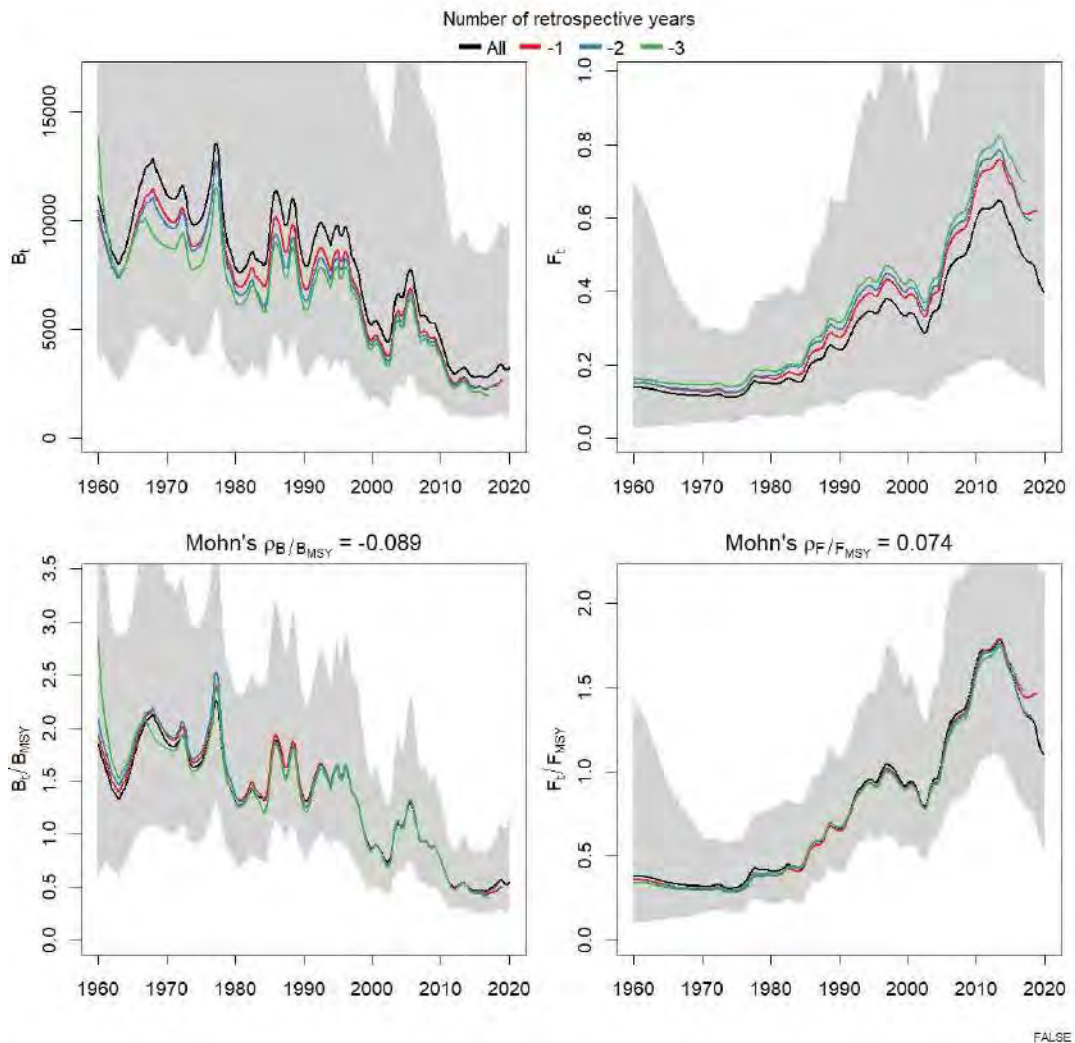


Figure 153: NEP in GSAs 17, 18, 19: Retrospective analysis.

Reference points and outcomes of the assessment are summarized in Table 51.

Table 49: NEP in GSAs 17, 18, 19. Estimated parameters from the SPiCT model

K	r	MSY	B _{MSY}	F _{MSY}	B/B _{MSY}	F/F _{MSY}
14579.6	0.51	2185.7	6024.8	0.36	0.52	1.12

4.9.4 Comparison with other assessments with a different stock configuration

Considering the output of the genetic, the stock of Norway lobster living in the Central Mediterranean and Adriatic basins seems to overlook the conventional GSAs boundaries used in the current assessments, resulting in only one unit. This genetic configuration has been followed and the assessment for the Adriatic and Ionian Seas, including GSAs 17, 18 and 19 was performed. Considering the current assessments developed within the STECF (STECF, 2020) and GFCM (GFCM, 2019) framework, the assessment here presented has been compared with the assessment in GSAs 17-18 (Adriatic Sea) in order to verify if there are evident improvements or deterioration of stock assessment quality, and to check the degree of consistency in the stock perception.

Model diagnostics

The diagnostics of the model for GSAs 17-18 performed during the STECF EWG 20-15 were plotted for comparison (Figure 154).

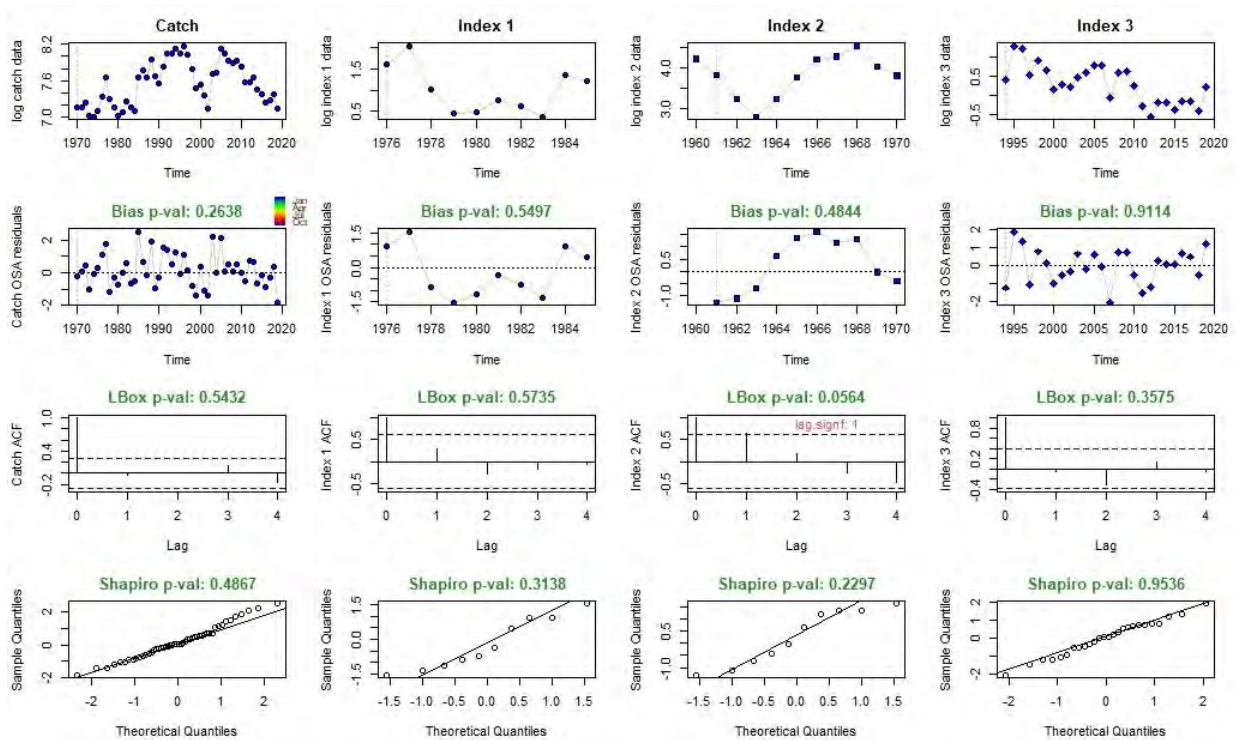


Figure 154: NEP in GSAs 17-18. Model diagnostic for the stock assessment of STECF EWG 20-15.

Retrospective patterns

The retrospective pattern observed for the MED-UNITs stock aggregation is very close to the one of the assessment from GSAs 17-18 (Figure 155).

Stock trajectories

Current estimation of F and B, as well as the reference point of the current assessment are very close to the ones obtained in the assessment of GSAs 17-18 performed at STECF EWG 20-15 (STECF, 2020b).

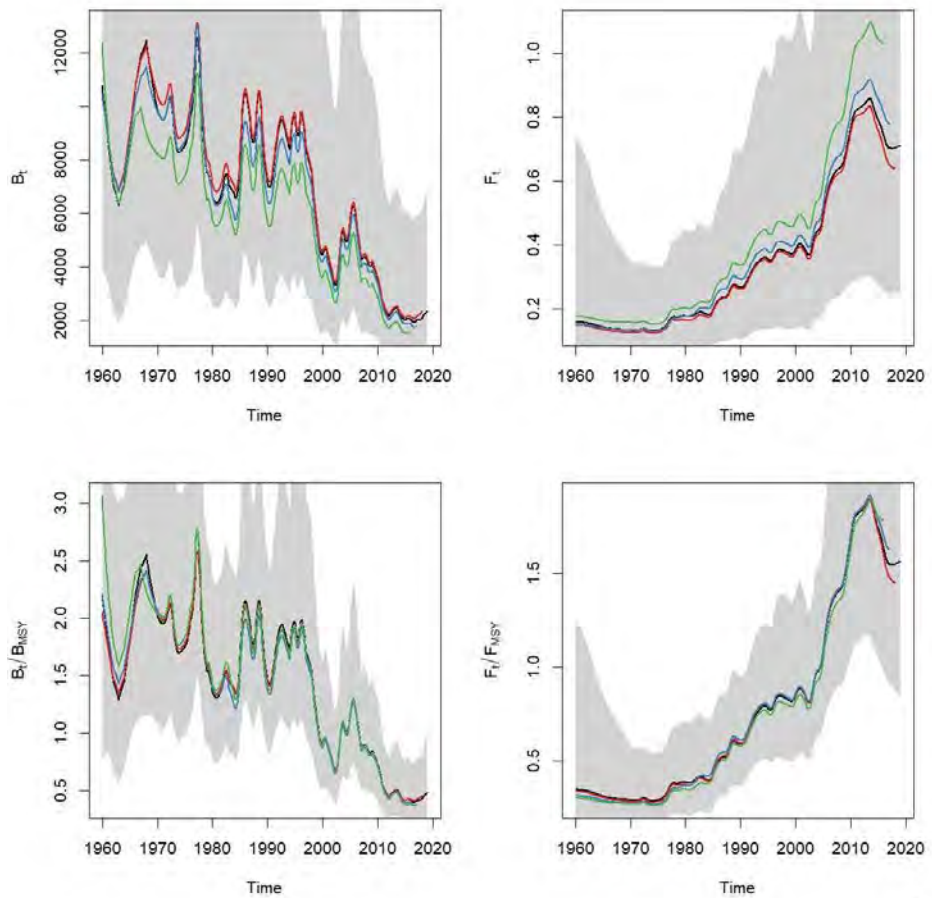


Figure 155: NEP in GSAs 17-18: retrospective for the stock assessment of STECF EWG 20-15.

4.10 Giant red shrimp in GSAs 5, 8-12

In the western Mediterranean, giant red shrimp (ARS), *Aristaeomorpha foliacea*, represents one of the most valuable demersal resources for the trawling fleet operating on the muddy bottoms of the upper and middle slope from 400 to 800 m.

According to the outcomes of the genetic analysis performed under the Med_Units project, a stock configuration including GSAs 5, 8, 9, 10, 11 and 12 was identified, and used to perform an assessment (Figure 156).

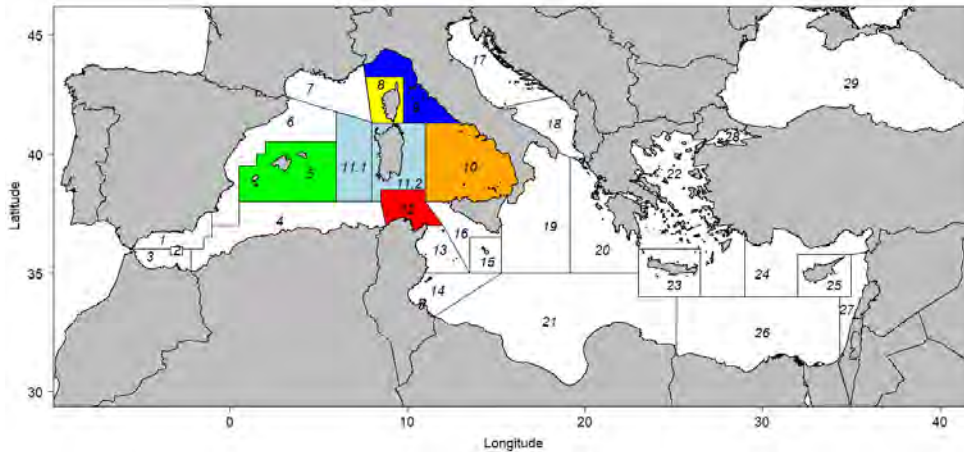


Figure 156: ARS in GSAs 5, 8, 9, 10, 11 and 12. Geographical location of the stock.

4.10.1 Input data

Fishery-Dependent data

The stock object for GSAs 9, 10 and 11 (time series 2005-2019) produced and used to assess the stock of ARS in GSAs 9, 10 and 11 under the STECF EWG 20-09 (STECF, 2020a) and GFCM WGSAD 2020 was used as the basis for this assessment. Although commercial LFDs were available for GSA 5, it was decided not to include them, as they were scattered in time, and covered low amounts of production (lower than 3 t per year). Nonetheless, landings from GSA 5 were added to the stock object, and the catch numbers-at-age was raised to those additional landings. No fishery-dependent data was available for GSA 8 and GSA 12. For GSAs 1, 6 and 7, landings reported under the official EU Data Calls were negligible (e.g., 59 kg in GSA 7 reported only in 2019; zero landing in the other years); therefore, it was decided to not include those GSAs in the assessment.

During the STECF EWG 20-09 and GFCM-WGSAD 2020, LFDs were sliced (deterministic slicing by sex) to obtain age structure using the growth parameters available under EU DCR/DCF (Table 52). Natural mortality vector was estimated using the Chen and Watanabe model (Table 53). Mean weight-at-age were calculated using the LW parameters available under EU DCR/DCF, and applied to the whole time series. The same approach was used for the maturity vector.

Fishery-Independent data

LFDs (n/km^2) from the MEDITS surveys in GSAs 5, 8, 9, 10 and 11 were sliced (deterministic age slicing by sex) using the same set of growth parameters used for the commercial data. No data was available for GSA 12.

Table 50: ARS in GSAs 5, 8, 9, 10, 11 and 12. VBGF and LW parameters by sex

Sex	Linf	k	t0
F	73.0	0.435	-0.10
M	50.0	0.40	-0.10

	a	b
F	0.004	2.52
M	0.003	2.65

Table 51: ARS in GSAs 5, 8, 9, 10, 11 and 12: natural mortality and maturity-at-age vectors.

Age	Natural mortality	Proportion of matures
0	1.89	0.00
1	0.86	0.40
2	0.62	1.00
3	0.53	1.00
4+	0.48	1.00

The catch at age structure as well as the abundance at age resulting from the slicing are presented respectively in Figure 157 and Figure 158. Moreover, the internal consistencies of such input data are presented respectively in Figure 159 and Figure 160

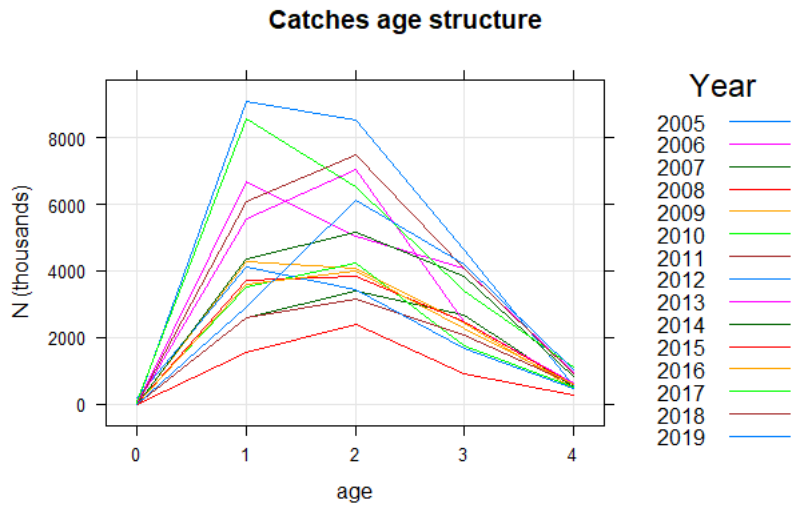


Figure 157: ARS in GSAs 5, 8, 9, 10, 11 and 12. Catch numbers-at-age.

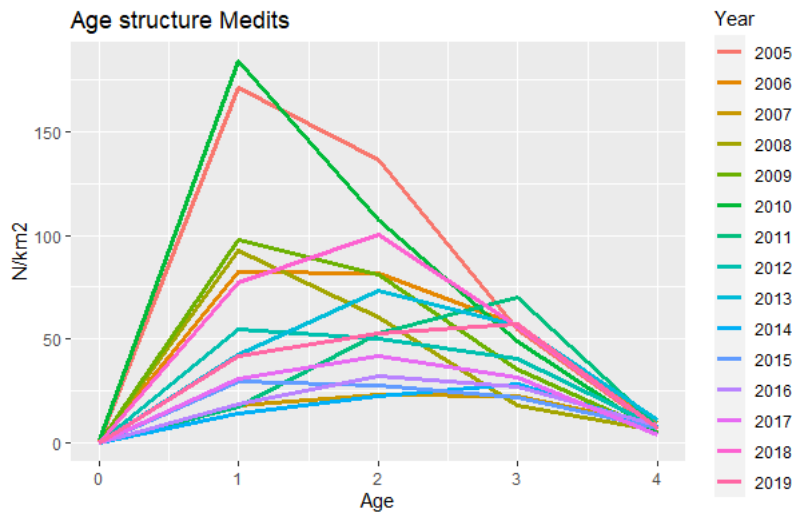


Figure 158: ARS in GSAs 5, 8, 9, 10, 11 and 12. Catch numbers-at-age in the MEDITS.

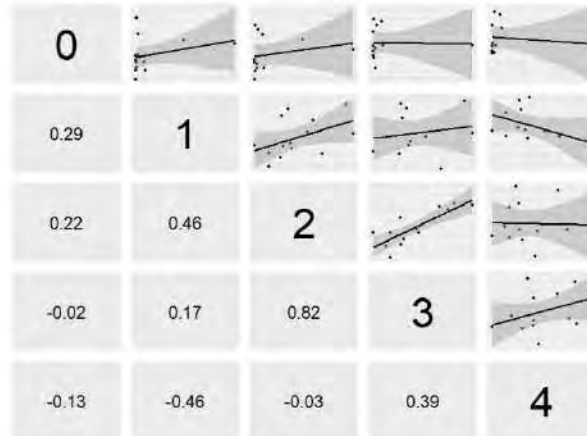


Figure 159: ARS in GSAs 5, 8, 9, 10, 11 and 12. Cohort consistency in the catch numbers-at-age.

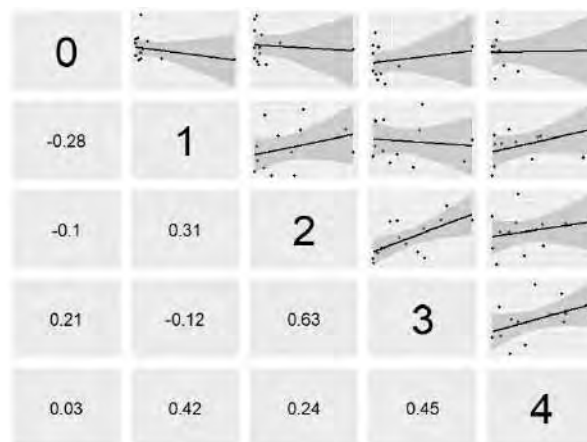


Figure 160: ARS in GSAs 5, 8, 9, 10, 11 and 12. Cohort consistency in the catch numbers-at-age in the MEDITS.

4.10.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment by means of a4a was carried out using as input data the period 2005-2019 for the catch data and 2005-2019 for the tuning file (MEDITS indices).

A natural mortality vector computed using Chen and Watanabe model was used in the assessment. Length-frequency distributions of commercial catches and surveys were split by sex and then transformed in age classes (plus group was set at age 4) using length-to-age slicing with different growth parameters by sex. A correction of 0.5 was applied to t_0 to align length slicing to assessment year (January to December) to account for spawning at the middle of the year.

The number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers-at-age $a \times$ catch weight-at-age a)].

In the catches, a plus group at age 4 was set. A true age 4 was used in the survey.

F_{bar} range was fixed at 1-3.

The assessment was performed by sex combined. Given that the landings were composed mainly of individuals between 1 and 3 years, these ages were selected as F_{bar} range.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $fmodel = \text{factor}(\text{replace}(\text{age}, \text{age} > 3, 3)) + s(\text{year}, k=5)$

Catchability sub-model: $qmodel = \text{list}(\sim \text{factor}(\text{age}))$

SR sub-model: `srmod = geomean(CV=0.2)`

`Model <- sca(stock = stk, indices = idx, fmodel, qmodel, srmod)`

The `n1model` and `vmodel` used in the final fit are the default ones:

`n1model <- ~s(age, k = 3)`

`vmodel <- list(~s(age, k=3), ~1)`

4.10.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the Figure 161 to Figure 166. The model estimated 36 parameters out of 150 observations; it is then well below the threshold of 25% ratio between parameters and observations.

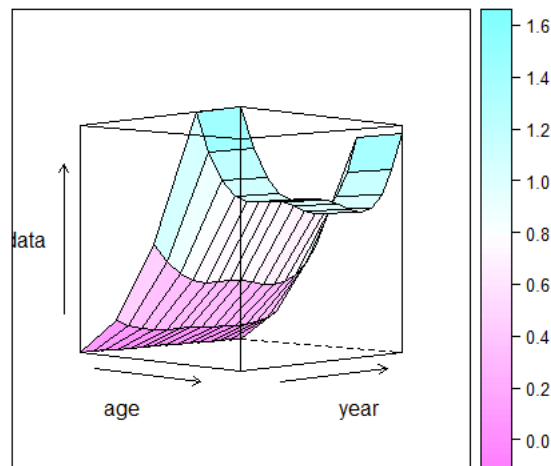


Figure 161: ARS in GSAs 5, 8, 9, 10, 11 and 12. Fishing mortality by age and year obtained from the a4a model (2003-2019).

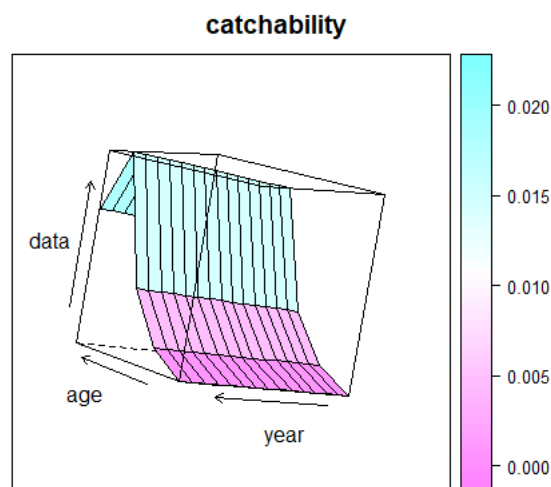


Figure 162: ARS in GSAs 5, 8, 9, 10, 11 and 12. Catchability of the survey by age/year obtained from the a4a model (2003-2019).

The log residuals for both the catches and the survey do not show any particular trend or issue. The fitting of the survey shows some problems (Figure 163), probably due to the poor internal consistency of the survey. Despite this, the diagnostics are considered acceptable and the a4a model is acceptable.

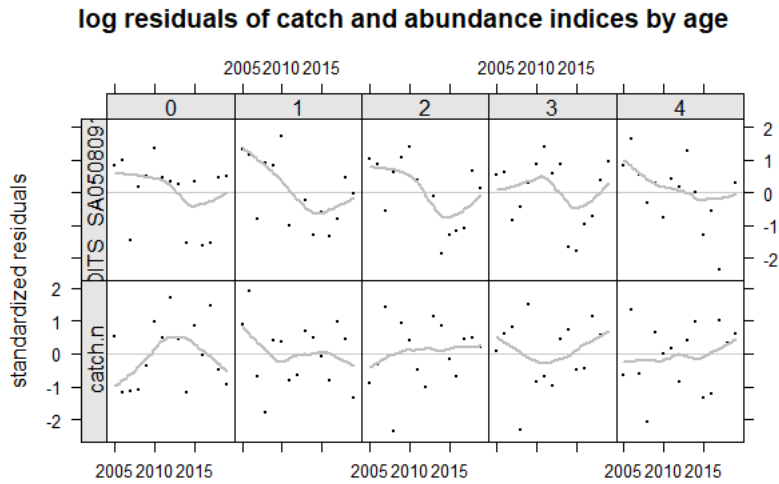


Figure 163: ARS in GSAs 5, 8, 9, 10, 11 and 12. Log residuals for the catch-at-age data of the fishery and the survey, and the catches.

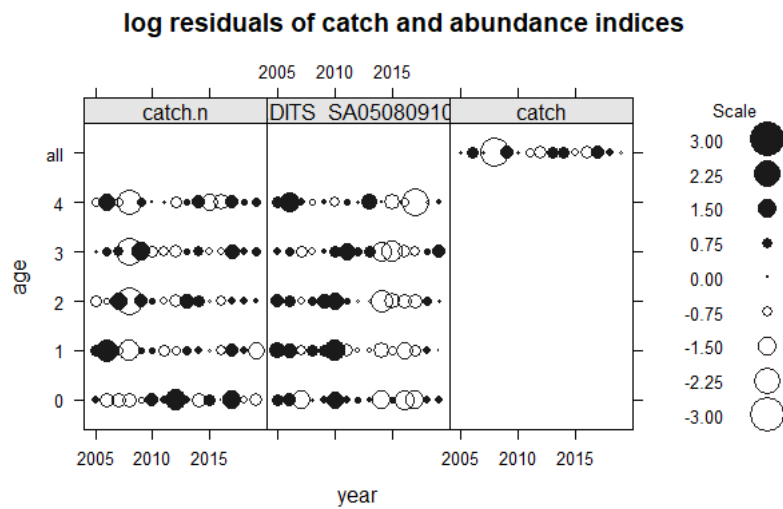


Figure 164: ARS in GSAs 5, 8, 9, 10, 11 and 12: bubble plot of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

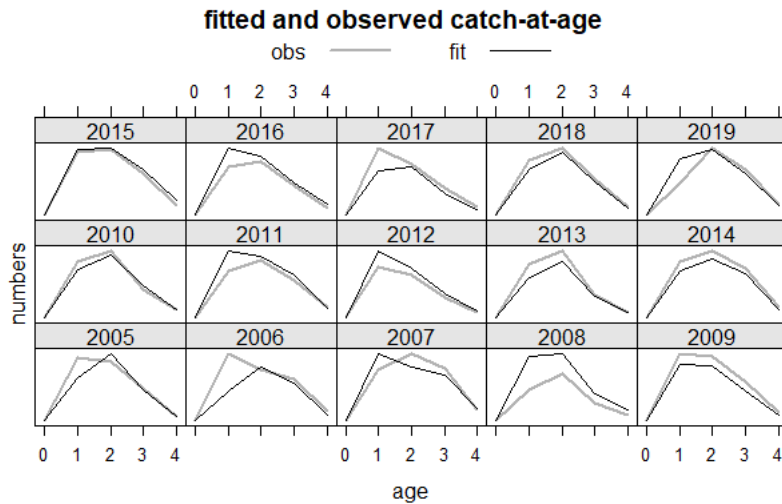


Figure 165: ARS in GSAs 5, 8, 9, 10, 11 and 12. Fitted vs observed values by age and year for the catches.

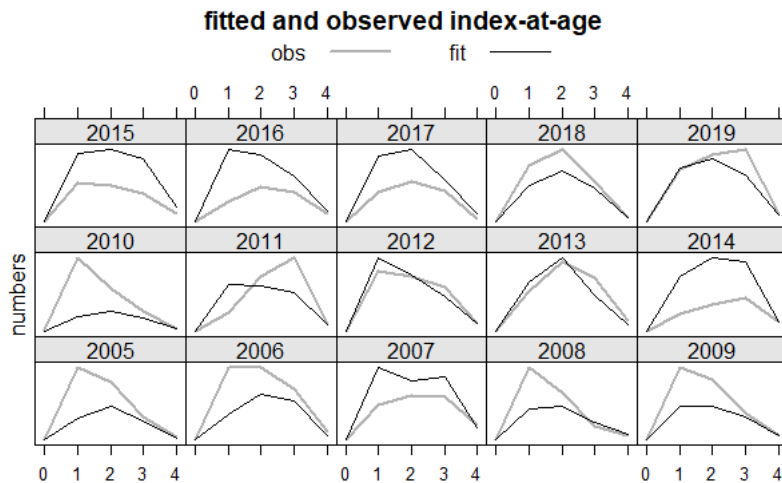


Figure 166: ARS in GSAs 5, 8, 9, 10, 11 and 12. Fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated, and did not show any relevant issue, as the biomass of the plus group (age 4+) is always below 10% of the total SSB.

The retrospective analysis shows some comb effect in F, SSB and recruitment, indicating that the assessment model is poorly stable (Figure 167). The Mohn's Rho for F and SSB are slightly outside the optimal range -0.2/0.2 (0.349 for F, -0.317 for SSB, and -0.156 for recruitment).

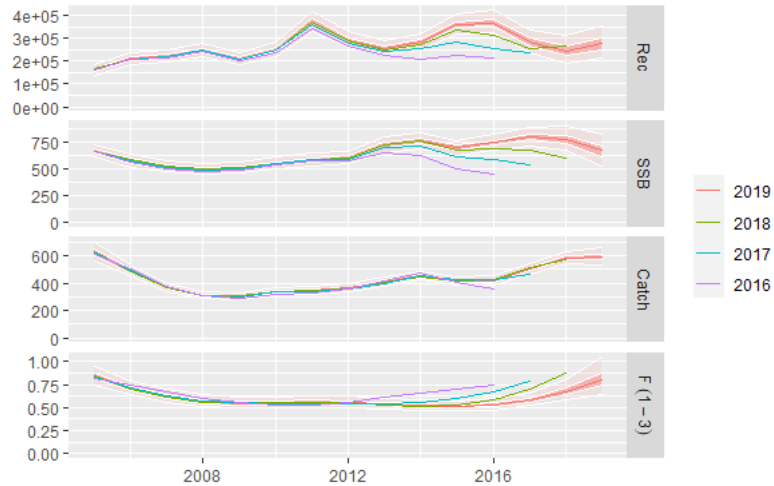


Figure 167: ARS in GSAs 5, 8, 9, 10, 11 and 12. Retrospective analysis.

Final assessment outcomes are given in Figure 168, and Table 54 to Table 56. Values of F0.1 calculated by FLBRP package on the a4a assessment results is equal to 0.46. Current F values (2019), as calculated by model a4a, is 0.81 indicating that the stock is being overexploited (Figure 169).

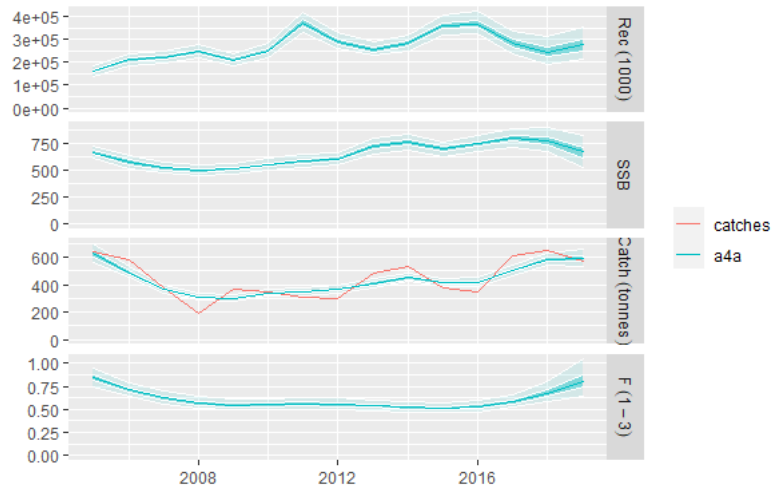


Figure 168: ARS in GSAs 5, 8, 9, 10, 11 and 12. Outputs of the a4a stock assessment model, with uncertainty; input catch data (red line) are plotted against the estimated catches.

Table 52: ARS in GSAs 5, 8, 9, 10, 11 and 12. Stock numbers-at-age (thousands).

age/year	2005	2006	2007	2008	2009	2010	2011	2012
0	160658	211948	219636	247046	207996	250151	371180	291338
1	44078	24271	32019	33180	37321	31422	37791	56074
2	24162	14745	8416	11392	11979	13547	11403	13690
3	7002	6169	4219	2613	3704	3962	4477	3748
4+	1073	1016	1146	1018	761	964	1063	1182
age/year	2013	2014	2015	2016	2017	2018	2019	
0	256152	282887	359542	369021	283363	246518	275002	
1	44013	38697	42736	54316	55748	42808	37241	
2	20324	16014	14152	15667	19825	20055	14999	
3	4507	6774	5425	4831	5273	6373	5930	
4	1058	1222	1813	1675	1463	1374	1326	

Table 53: ARS in GSAs 5, 8, 9, 10, 11 and 12. Fishing mortality-at-age.

age/year	2005	2006	2007	2008	2009	2010	2011	2012
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.235	0.199	0.173	0.159	0.153	0.154	0.155	0.155
2	0.745	0.631	0.55	0.503	0.486	0.487	0.493	0.491
3	1.55	1.313	1.144	1.047	1.011	1.013	1.025	1.021
4+	1.55	1.313	1.144	1.047	1.011	1.013	1.025	1.021
age/year	2013	2014	2015	2016	2017	2018	2019	
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
1	0.151	0.146	0.143	0.148	0.162	0.189	0.226	
2	0.479	0.463	0.455	0.469	0.515	0.598	0.718	
3	0.996	0.962	0.946	0.975	1.071	1.245	1.493	
4	0.996	0.962	0.946	0.975	1.071	1.245	1.493	

Table 54: ARS in GSAs 5, 8, 9, 10, 11 and 12. Summary results of the a4a assessment.

	Catch (t)	SSB (t)	Rec ('000)	F _{bar(1-3)}	Total biomass (t)
2005	620.2	665.3	160658	0.84	1179.6
2006	486.6	578.1	211948	0.71	732.6
2007	366.9	522.9	219636	0.62	823.0
2008	310.1	498.4	247046	0.57	729.8
2009	301.7	514.4	207996	0.55	1107.6
2010	336.8	553.9	250151	0.55	1154.1
2011	347.9	586.3	371180	0.56	1377.9
2012	361.3	601.9	291338	0.56	1352.4
2013	408.7	725.1	256152	0.54	1440.4
2014	451.3	761.6	282887	0.52	1102.4
2015	413.6	698.2	359542	0.51	1564.1
2016	415.2	745.6	369021	0.53	1765.2
2017	504.0	798.4	283363	0.58	1576.7
2018	580.9	781.2	246518	0.68	1327.3
2019	591.6	666.2	275002	0.81	992.5

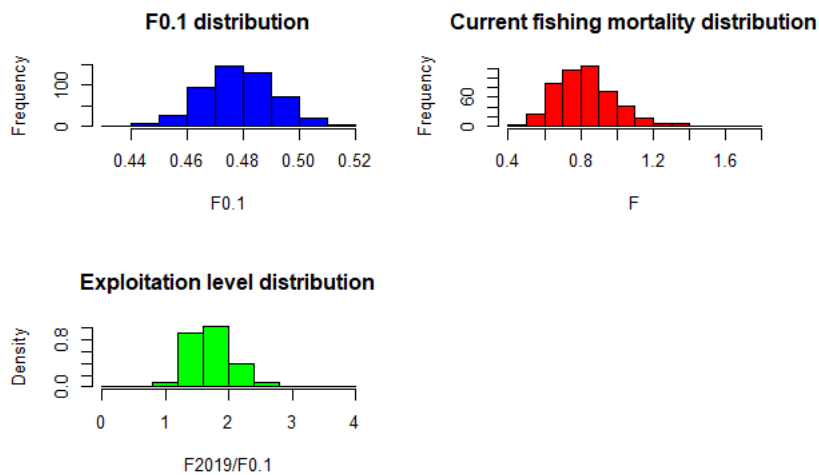


Figure 169: ARS in GSAs 5, 8, 9, 10, 11 and 12. Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F01 ratio) values.

Besides the model selection procedure, a further check was performed regarding the number of knots (k) of the smoother on year in the fmodel. A test based on AIC, BIC and GCV was performed on k ranging between 4 and 9. All the model specifications highlight a consistent behaviour in terms of main outcomes (Figure 170). A k value of 5 was retained as the one providing the best retrospective (Figure 171).

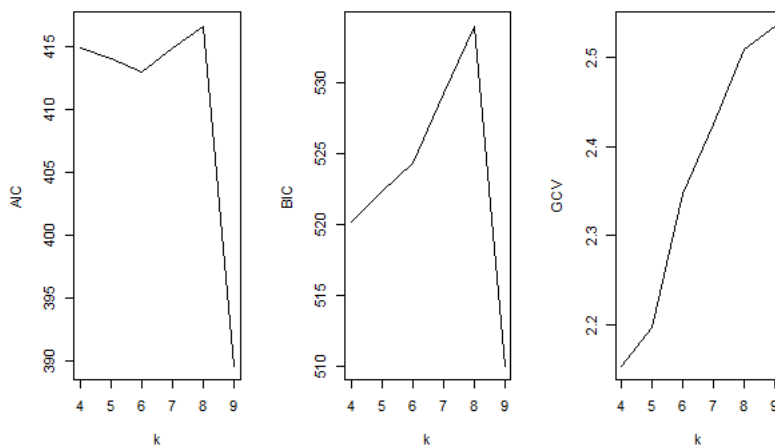


Figure 170: ARS in GSAs 5, 8, 9, 10, 11 and 12: AIC, BIC and GCV values estimated on a range of k values of the smoother on year of the fmodel.

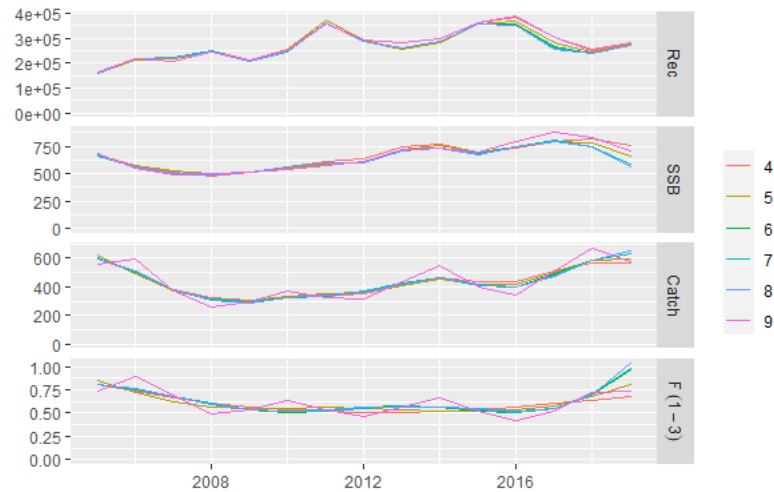


Figure 171: ARS in GSAs 5, 8, 9, 10, 11 and 12. Outputs of model runs with different k values on the smoother on year in the fmodel.

4.10.4 Comparison with other assessments with different stock configuration

Giant red shrimp in the western Mediterranean has been assessed as a single stock unit including GSAs 9, 10 and 11 under GFCM and STECF working groups since 2019. The new stock configuration, which is now including also GSAs 5, 8 and 12, does not show any difference with previous assessments. In fact, the contribution of the additional GSAs in terms of landings and survey data is almost negligible.

4.11 Giant red shrimp in GSAs 15, 16, 18-20

In the central Mediterranean, giant red shrimp (ARS), *Aristaeomorpha foliacea*, represents one of the most valuable demersal resources for the trawling fleet operating on the muddy bottoms of the upper and middle slope from 400 to 800 m.

According to the outcomes of the genetic analysis performed under the Med_Units project, a stock configuration including GSAs 15, 16, 18, 19 and 20 was identified, and used to perform an assessment (Figure 172).

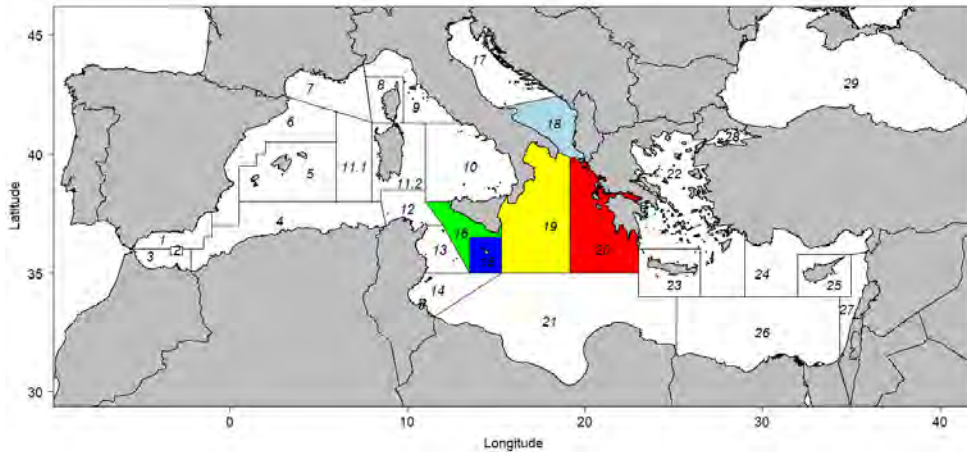


Figure 172 - ARS in GSAs 15, 16, 18, 19 and 20. Geographical location of the stock.

4.11.1 Input data

Fishery-Dependent data

Commercial data (LFDs and biological parameters) coming from EU DCR/DCF database were used to build an FLR stock object with a time series from 2003 to 2019.

LFDs were sliced (deterministic slicing by sex) to obtain age structure using the growth parameters available under EU DCR/DCF (Table 57). Natural mortality vector was estimated using the Chen and Watanabe model. Mean weight-at-age were calculated using the LW parameters available under EU DCR/DCF, and applied to the whole time series (Table 58). The same approach was used for the maturity vector.

Fishery-Independent data: LFDs (n/km^2) from the MEDITS surveys in the five GSAs were sliced (deterministic age slicing by sex) using the same set of growth parameters used for the commercial data.

The catch at age structure as well as the abundance at age resulting from the slicing are presented respectively in Figure 173 and Figure 174. Moreover the internal consistencies of such input data are presented respectively in Figure 175 and Figure 176.

Table 55: ARS in GSAs 15, 16, 18, 19 and 20. VBGF and LW parameters by sex.

Sex	Linf	k	t0
F	74.0	0.44	-0.16
M	53.0	0.36	-0.10
	a		b
F	0.001		2.65
M	0.001		2.75

Table 56: ARS in GSAs 15, 16, 18, 19 and 20. Natural mortality and maturity-at-age vectors.

Age	Natural mortality	Proportion of matures
0	1.89	0.00
1	0.86	0.34
2	0.64	0.98
3	0.51	1.00
4+	0.45	1.00

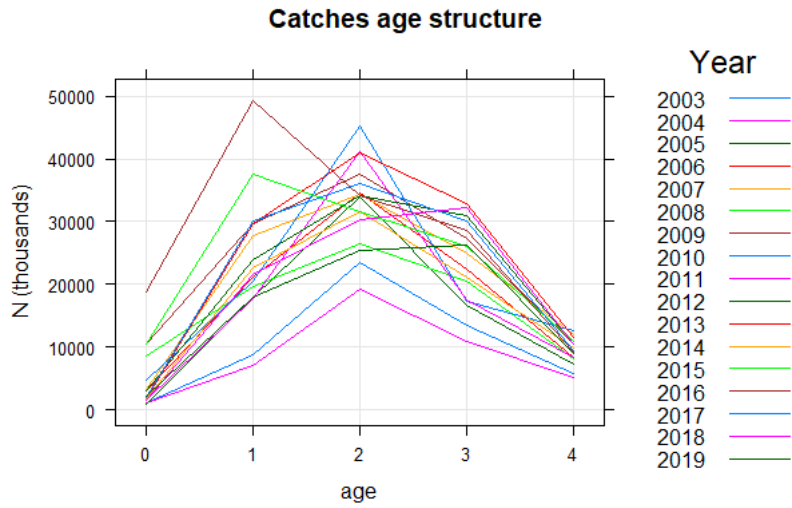


Figure 173: ARS in GSAs 15, 16, 18, 19 and 20. Catch numbers-at-age.

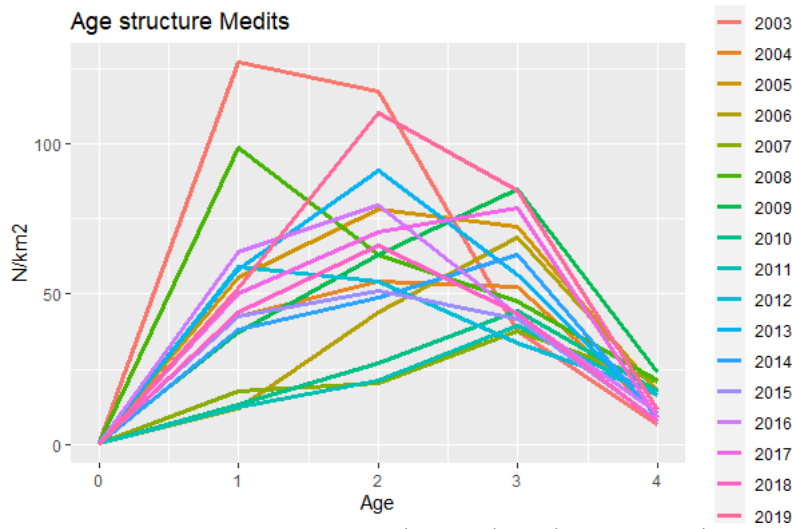


Figure 174: ARS in GSAs 15, 16, 18, 19 and 20. Catch numbers-at-age in the MEDITS.

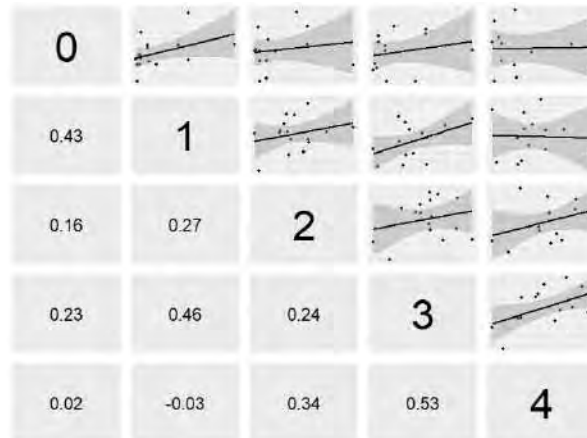


Figure 175: ARS in GSAs 15, 16, 18, 19 and 20. Cohort consistency in the catch numbers-at-age.

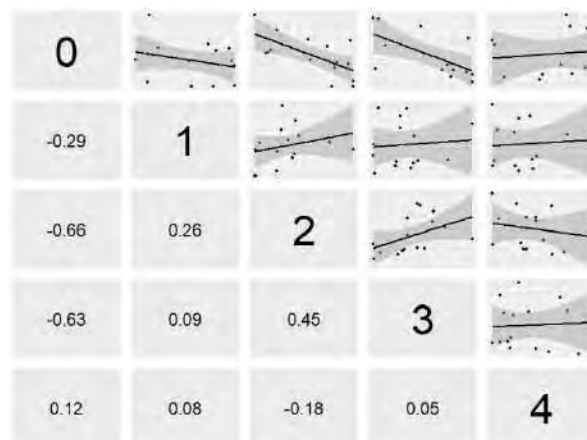


Figure 176: ARS in GSAs 15, 16, 18, 19 and 20. Cohort consistency in the catch numbers-at-age in the MEDITS.

4.11.2 Model description

FLR libraries were employed in order to carry out a Statistical Catch-at-age (a4a) assessment. The assessment by means of a4a was carried out using as input data the period 2003-2019 for the catch data and 2003-2019 for the tuning file (MEDITS indices).

A natural mortality vector computed using Chen and Watanabe model was used in the assessment. Length-frequency distributions of commercial catches and surveys were split by sex and then transformed in age classes (plus group was set at age 4) using length-to-age slicing with different growth parameters by sex. A correction of 0.5 was applied to t_0 to align length slicing to assessment year (January to December) to account for spawning at the middle of the year.

The number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers at age $a \times$ catch weight-at-age a)].

In the catches, a plus group at age 4 was set. A true age 4 was used in the survey.

F_{bar} range was fixed at 1-3.

The assessment was performed by sex combined. Given that the landings were composed mainly of individuals between 1 and 3 years, these ages were selected as F_{bar} range.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $fmodel = factor(replace(age, age>3,3))+s(year, k=6)$

```
Catchability sub-model: qmodel = list(~factor(replace(age, age>3,3)))
```

```
SR sub-model: srmod = geomean(CV=0.3)
```

```
Model <- sca(stock = stk, indices = idx, fmodel, qmodel, srmod)
```

The n1model and vmodel used in the final fit are the default ones:

```
n1model <- ~s(age, k = 3)
```

```
vmodel <- list(~s(age, k=3), ~1)
```

4.11.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the Figure 177 to Figure 182. The model estimated 37 parameters out of 170 observations; it is then well below the threshold of 25% ratio between parameters and observations.

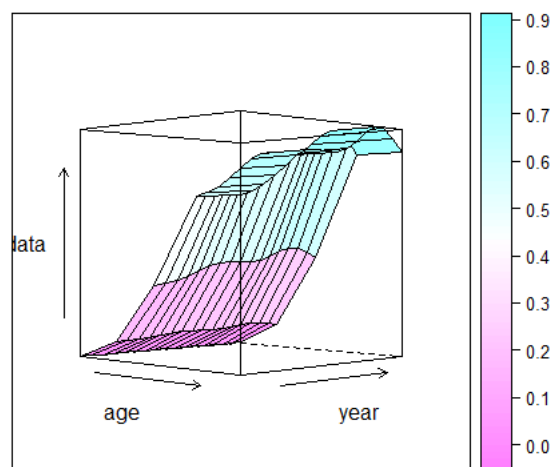


Figure 177: ARS in GSAs 15, 16, 18, 19 and 20. Fishing mortality by age and year obtained from the a4a model (2003-2019).

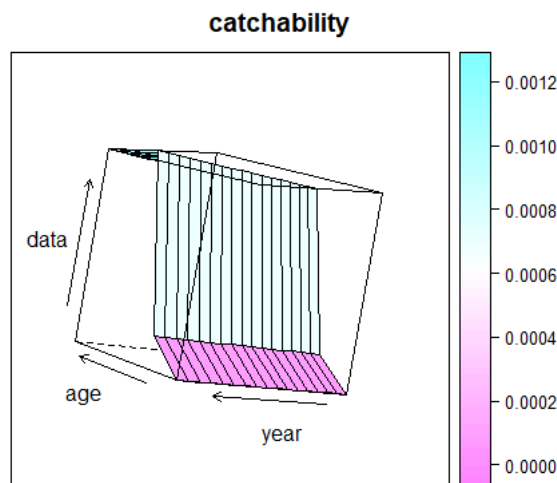


Figure 178: ARS in GSAs 15, 16, 18, 19 and 20. Catchability of the survey by age and year obtained from the a4a model (2003-2019).

The log residuals for both the catches and the survey do not show any particular trend or issue. The fitting of the survey shows some problems (Figure 179), probably due to the poor internal consistency of the survey. Despite this, the diagnostics are considered acceptable and the a4a model is acceptable as a basis

for advice.

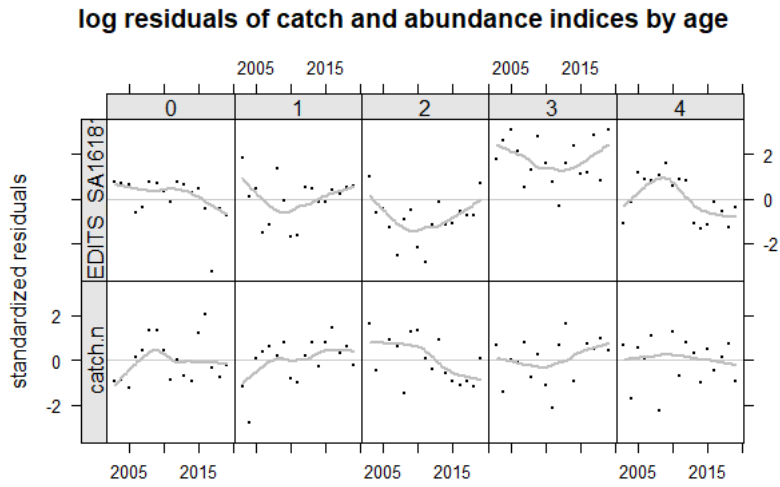


Figure 179: ARS in GSAs 15, 16, 18, 19 and 20. Log residuals for the catch-at-age data of the fishery and the survey, and the catches.

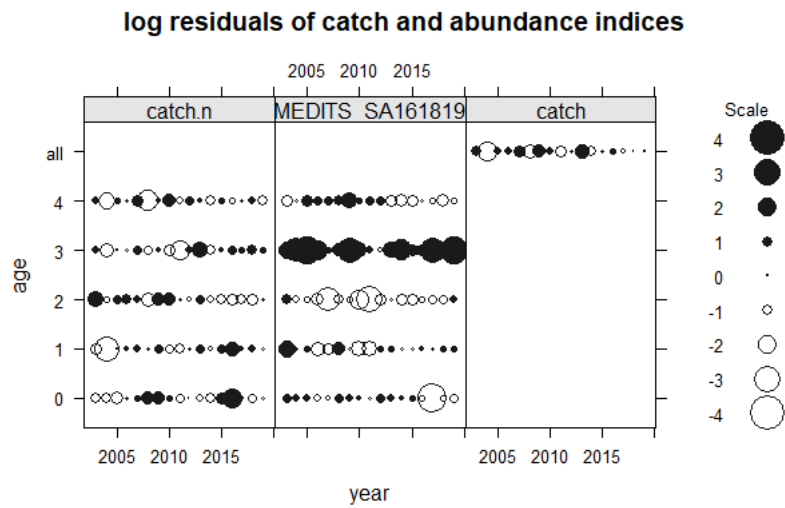


Figure 180: ARS in GSAs 15, 16, 18, 19 and 20. Bubble plots of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

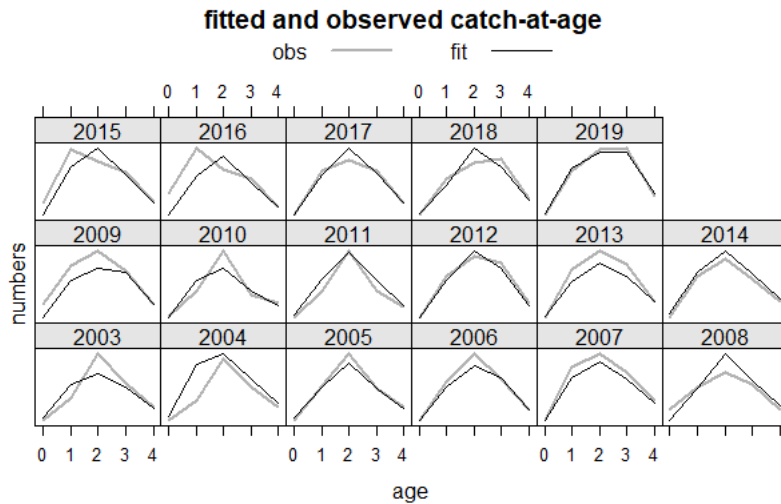


Figure 181: ARS in GSAs 15, 16, 18, 19 and 20. Fitted vs observed values by age and year for the catches.

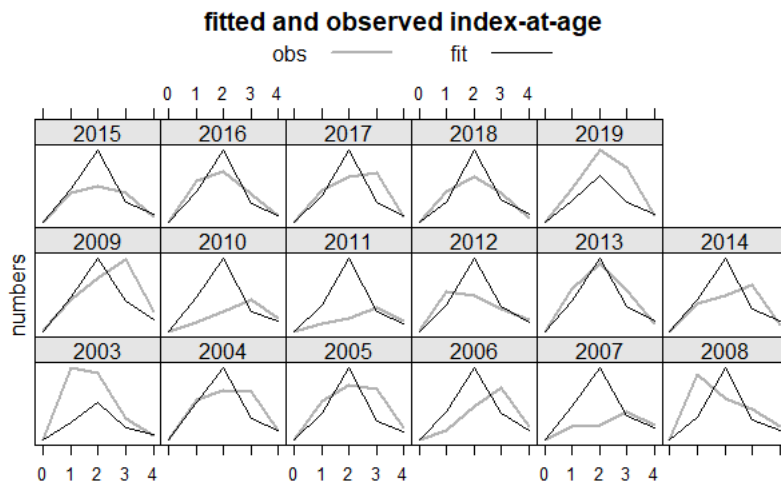


Figure 182: ARS in GSAs 15, 16, 18, 19 and 20: fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated, and did not show any relevant issue, as the biomass of the plus group (age 4+) is always below 10% of the total SSB.

The retrospective analysis shows that the assessment model is moderately stable, and the catch estimates obtained by the a4a assessment are fitting well the observed catches (Figure 183). The Mohn's Rho for F, SSB and recruitment are falling within the optimal range -0.2 - 0.2 (0.035 for F, -0.035 for SSB, and -0.053 for recruitment).

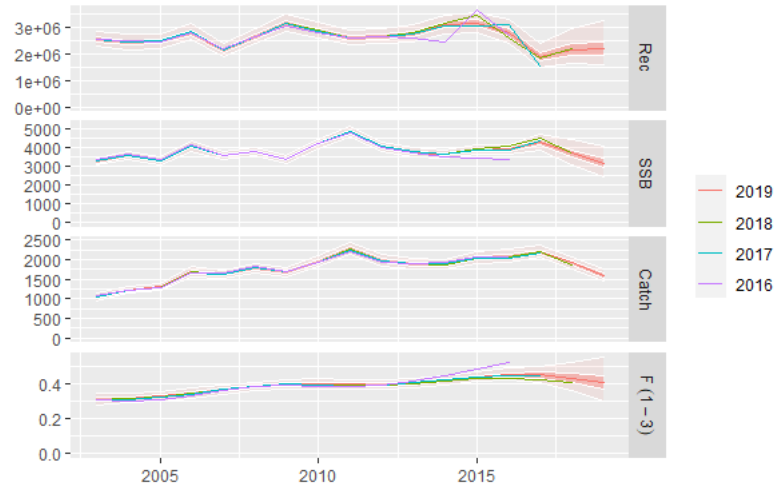


Figure 183: ARS in GSAs 15, 16, 18, 19 and 20. Retrospective patterns

Final assessment outcomes are given in Figure 183, and Table 59 to Table 61. Values of F0.1 calculated by FLBRP package on the a4a assessment results is equal to 0.48. Current F values (2019), as calculated by model a4a, is 0.41 indicating that the stock is being sustainably exploited (Figure 185).

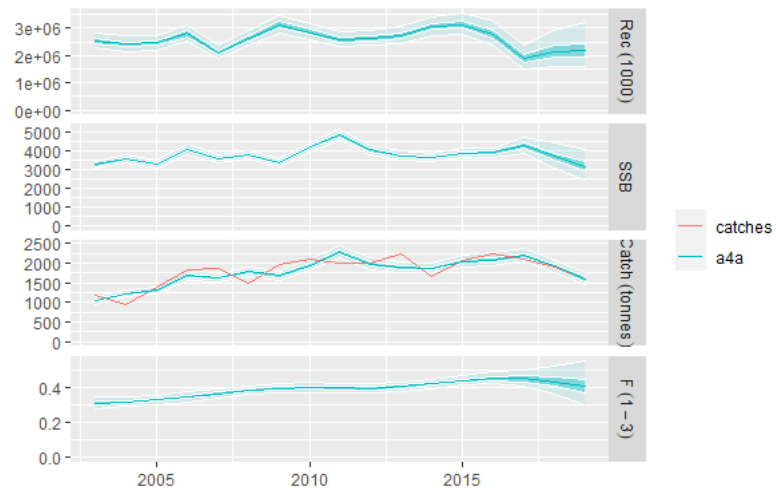


Figure 184: ARS in GSAs 15, 16, 18, 19 and 20. Outputs of the a4a stock assessment model, with uncertainty; input catch data (red line) are plotted against the estimated catches.

Table 57: ARS in GSAs 15, 16, 18, 19 and 20. Stock numbers-at-age (thousands).

age/year	2003	2004	2005	2006	2007	2008	2009	2010	2011
0	2525152	2413987	2464501	2801184	2108368	2635453	3132897	2852827	2579266
1	284560	378716	362002	369439	419852	333797	394824	469430	427611
2	92289	112962	149023	143064	143463	163920	128093	151811	180385
3	34424	37163	45011	59563	55453	54997	62266	47894	56339
4+	14767	16838	18200	20767	25403	24762	23596	24882	20903
age/year	2012	2013	2014	2015	2016	2017	2018	2019	
0	2614063	2734023	3068154	3138012	2771317	1885879	2154178	2201424	
1	386412	392008	409781	459819	470160	415131	282115	322696	
2	163597	148987	150559	156606	172914	175861	158062	107968	
3	66436	60699	54714	54669	55385	60617	62007	56895	
4	21964	25479	24592	21918	20320	19648	20879	22386	

Table 58: ARS in GSAs 15, 16, 18, 19 and 20. Fishing mortality-at-age.

age/year	2003	2004	2005	2006	2007	2008	2009	2010	2011
0	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003
1	0.070	0.071	0.074	0.078	0.082	0.086	0.089	0.090	0.089
2	0.277	0.285	0.295	0.310	0.328	0.344	0.355	0.357	0.355
3	0.582	0.598	0.620	0.651	0.688	0.723	0.745	0.750	0.746
4+	0.582	0.598	0.620	0.651	0.688	0.723	0.745	0.750	0.746
age/year	2012	2013	2014	2015	2016	2017	2018	2019	
0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	
1	0.089	0.091	0.095	0.099	0.102	0.101	0.097	0.092	
2	0.355	0.362	0.377	0.394	0.406	0.404	0.388	0.366	
3	0.746	0.761	0.792	0.828	0.852	0.848	0.815	0.768	
4	0.746	0.761	0.792	0.828	0.852	0.848	0.815	0.768	

Table 59: ARS in GSAs 15, 16, 18, 19 and 20. Summary results of the a4a assessment.

	Catch (t)	SSB (t)	Rec ('000)	$F_{\text{bar}(1-3)}$	Total biomass (t)
2003	1034.9	3261.8	2525152	0.31	9713.2
2004	1211.0	3574.7	2413987	0.32	10857.2
2005	1293.6	3301.1	2464501	0.33	10032.7
2006	1674.6	4074.6	2801184	0.35	11901.5
2007	1618.6	3591.7	2108368	0.37	8968.1
2008	1773.7	3795.9	2635453	0.38	9926.7
2009	1671.1	3364.3	3132897	0.40	10621.7
2010	1941.5	4216.7	2852827	0.40	12704.1
2011	2278.2	4858.5	2579266	0.40	12592.1
2012	1965.0	4062.6	2614063	0.40	11082.5
2013	1872.5	3752.9	2734023	0.40	12707.9
2014	1866.2	3648.1	3068154	0.42	13502.0
2015	2036.7	3881.0	3138012	0.44	13419.8
2016	2077.0	3906.3	2771317	0.45	12062.8
2017	2190.2	4281.6	1885879	0.45	9988.0
2018	1909.7	3713.8	2154178	0.43	9836.0
2019	1593.0	3176.8	2201424	0.41	9335.4

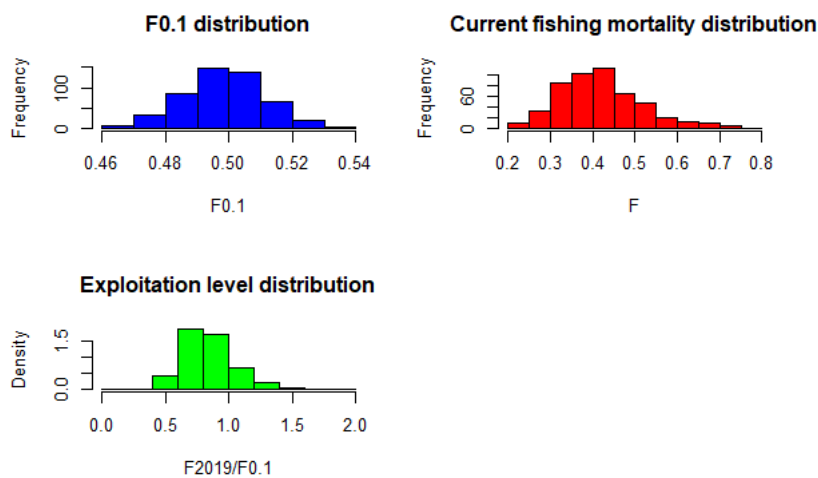


Figure 185: ARS in GSAs 15, 16, 18, 19 and 20. Histograms of probability for F0.1, Fcurr and level of exploitation (Fcurr/F01 ratio) values.

Besides the model selection procedure, a further check was performed regarding the number of knots (k) of the smoother on year in the fmodel. A test based on AIC, BIC and GCV was performed on k ranging between 4 and 9 (Figure 186). This analysis confirmed that the k value performing the best is 6 (as a rule of thumb, the GCV is the most important parameter), as specified in the final model selected. Nonetheless, all the model specifications highlight a consistent behaviour in terms of main outcomes (Figure 187).

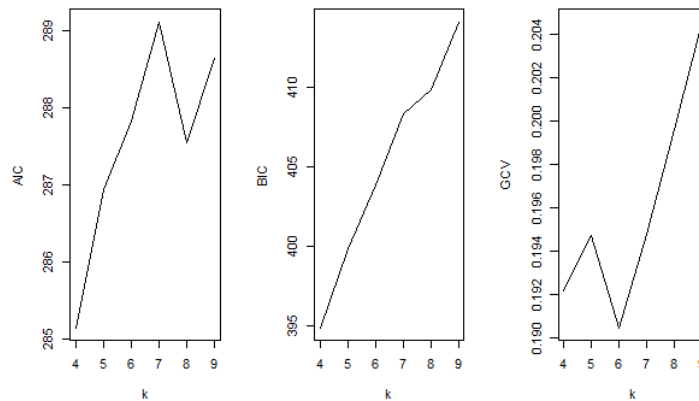


Figure 186: ARS in GSAs 15, 16, 18, 19 and 20. AIC, BIC and GCV values estimated on a range of k values of the smoother on year of the fmodel.

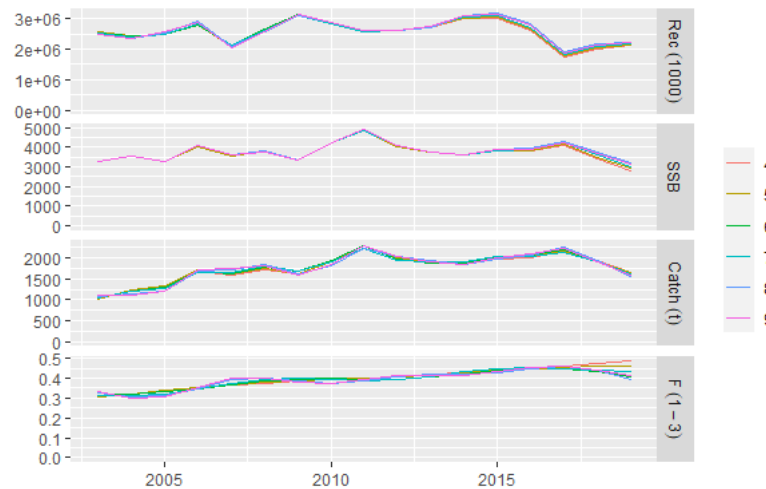


Figure 187: ARS in GSAs 15, 16, 18, 19 and 20: Outputs of model runs with different k values on the smoother on year in the fmodel.

4.11.4 Comparison with other assessments with different stock configuration

A recent assessment covering GSAs 18 and 19 was presented at the GFCM WGSAD 2020. This assessment was run with a4a, with very similar sub-model settings. The inclusion of GSA 16 represents the major source of change compared to the previous assessment. In fact, GSA 16 produces the bulk of the catches (almost 1000 t per year), while GSAs 18 and 19 account for around 500 t per year. The contribution from GSA 15 and GSA 20 is almost negligible.

In terms of cohort consistency, the new stock configuration has a slightly better consistency between age 0 and age 1 in the catches, while the consistency of older age classes remains almost stable. In contrast,

there is a worsening in the cohort consistency between age 0 and 1 in the survey, while the consistency of older age classes improved.

In terms of diagnostics, there are no major differences between the run covering GSAs 18 and 19, and the new stock configuration. The latter shows a slightly better retrospective pattern. However, this could be due to the sub-model settings. Although the two assessments are very similar, the new assessment has a lower k value (6 vs 9) in the smoother of the year in the F sub-model. This might have improved the stability of the retrospective.

In terms of main outcomes, the new stock configuration shows a sustainable exploitation status, while assessing GSAs 18-19 alone showed an overexploitation status. This difference could be due to the major contribution provided by GSA 16.

4.12 Blue and red shrimp in GSAs 1, 5, 6, 9-12

The stock assessment of the blue and red shrimp in the Western Mediterranean Sea is here presented. Considering the WP1 and WP2 outputs, the GSAs 1, 5, 6 and from 9 to 11 were lumped together (Figure 188). No data were available for GSA 12.

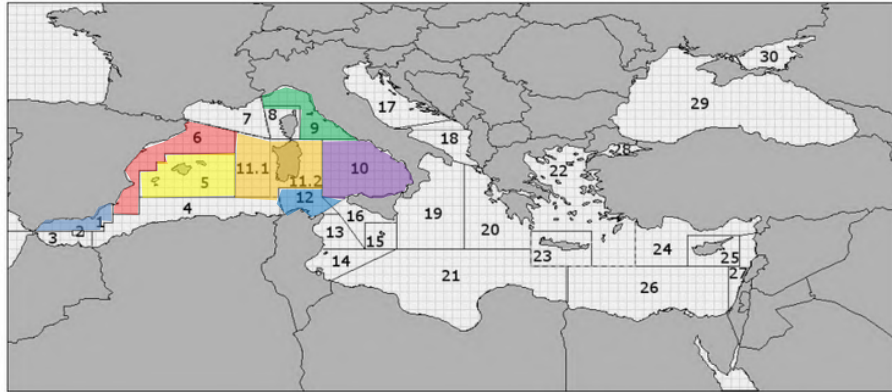


Figure 188: ARA in GSAs 1, 5, 6, 9, 10 and 11 (colored GSAs): geographical location of the stock

The blue and red shrimp is distributed throughout the Mediterranean and the Eastern Atlantic, from Portugal to Cabo Verde Islands (Arrobas and Ribeiro-Cascalho 1987). In the Mediterranean, the species shows a longitudinal gradient, being more abundant in the north-western areas than in the central and eastern ones (D'Onghia et al. 1998, Cau et al. 2002, Cardinale et al. 2017, Guijarro et al. 2019). It is an euribatic species, with a wide bathymetric distribution between 80 down to 3300 m depth (D'Onghia et al. 1998, Sardà et al. 2003, Politou et al. 2004). At depths shallower than 1000 m, the population is mostly composed of adults, with females dominant and showing clear spatiotemporal variations in its population dynamics. Below 1000 m depth, the population is more stable, mostly composed of juveniles and dominated by males (Sardà et al., 1994, 2003).

In the Western Mediterranean, the blue and red shrimp is the most valuable deep water resource and the target of a well-developed bottom trawl fishery (Sardà et al. 2003). The species shows a large geographical and interannual variability along the Mediterranean (Guijarro et al. 2019). The inter-annual variability has been related to large-scale climatic indices, such as the North Atlantic Oscillation (NAO, Carbonell et al. 1999, Maynou 2008, D'Onghia et al. 2012) or the Mediterranean Oscillation (MO; Massutí et al. 2008) indices, although this variability can also show different patterns even from nearby ports (Maynou 2008, Hidalgo et al. 2015). The way these indices affect variability may be related to changes in the oceanographic features that may also be linked with more adequate conditions for the species, including temperature and salinity and an enhancement of food supply. Spatial distribution has also been related to cascading of dense shelf waters along the slope (Company et al. 2008), geomorphology (presence of canyons and seamounts; Tudela et al. 2003, Sabatini et al. 2007, 2011), type of bottom (Cartes et al. 2008) and oceanographic features (Guijarro et al. 2008).

The reproductive period for this species is during late spring and summer, but with small geographical differences even at a small scale (Sardà and Demestre, 1987; Carbonell et al., 1999; García-Rodríguez and Esteban, 1999). The period of recruitment seems to show some geographical differences but occurs between December to June, both in the exploited fishing grounds and in the deeper unexploited waters (Sardà and Cartes, 1997; Sardà and Company, 2012; Guijarro et al., 2008).

4.12.1 Input data

Fishery dependent data

Commercial data (catch data, LFDs and biological parameters) for Spain and Italy were collected from the

EU DCR/DCF database, considering only data from OTB (Table 62). No data were available for GSA 12. Considering the data available, this assessment was carried out using years from 2005 to 2019.

Table 62. ARA in GSAs 1, 5, 6 9-12: Catch data divided by country, GSA and year.

Year	ESP - GSA 1	ESP - GSA 5	ESP - GSA 6	ITA – GSA 9	ITA – GSA 10	ITA – GSA 11
2002	156.96	141.45	197.95			
2003	335.74	122.01	316.99	76.95	18.52	
2004	225.20	193.58	448.11	82.41	120.19	
2005	232.10	191.48	294.33	154.92	63.93	97.72
2006	288.82	213.89	396.22	92.70	51.69	171.73
2007	178.43	239.12	527.41	47.37	39.49	56.54
2008	133.48	232.85	736.62	63.46	22.97	74.64
2009	144.59	126.16	515.05	123.50	27.41	65.25
2010	152.09	153.24	508.82	186.40	20.07	53.33
2011	131.42	111.24	663.37	174.69	48.50	59.41
2012	148.57	201.14	703.45	192.62	31.47	57.27
2013	124.96	188.60	678.88	170.44	34.28	40.52
2014	184.03	141.28	545.65	83.56	8.72	46.42
2015	170.23	160.15	689.39	90.68	66.91	57.42
2016	138.22	138.10	570.28	66.60	95.41	89.40
2017	99.19	171.35	522.69	62.39	75.96	110.02
2018	123.21	249.68	606.38	77.23	135.00	284.49
2019	132.09	205.90	546.72	101.03	141.47	107.00

LFDs from the commercial data GSAs 1, 5, 6, 9, 10 and 11 were sliced by sex using the same set of growth parameters used for the assessment of GSAs 9-10-11 carried out in the STECF-20-09 which come from Orsi-Relini, 1998 (Table 63). Plus group was set at age 6. Maturity ogive and M vector were also those used for GSAs 9-10-11 in STECF-20-09 (Table 64).

Table 63: ARA in GSAs 1, 5, 6, 9, 10 and 11: VBGF and LW parameters.

Sex	Linf	k	t0
Females	76.9	0.21	-0.02
Males	46	0.21	-0.02
	a		b
Females	0.0042		2.3237
Males	0.0028		2.4652

Table 64: ARA in GSAs 1, 5, 6, 9, 10 and 11:: natural mortality and maturity-at-age vectors.

Ages	Natural mortality	Proportion of matures (combined)
0	2.023	0
1	0.768	0.204
2	0.511	0.786
3	0.402	0.983
4	0.342	0.999
5	0.301	1
6+	0.281	1

LFDs (n/km^2) from the MEDITS surveys in GSAs 1, 5, 6, 9, 10 and 11 were sliced using the same set of growth parameters used for the commercial data (Table 63).

The catch at age from the commercial fleet and the survey are presented respectively in Figure 189 and Figure 190. Moreover the internal consistencies of such input data are presented respectively in Figure 191 and Figure 192.

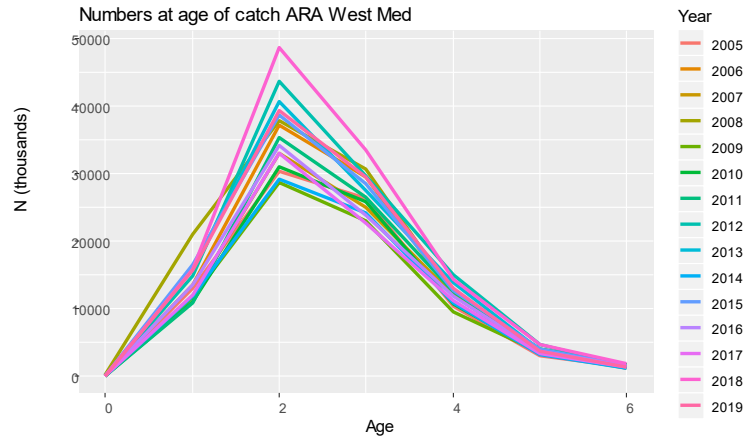


Figure 189: ARA in GSAs 1, 5, 6, 9, 10 and 11: catch numbers-at-age for the commercial fleet.

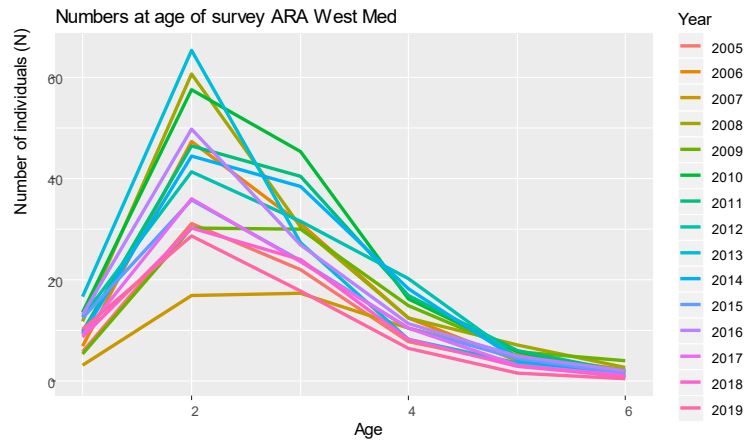


Figure 190: ARA in GSAs 1, 5, 6, 9, 10 and 11: catch numbers-at-age for the MEDITS survey.

Cohorts consistence in the catch

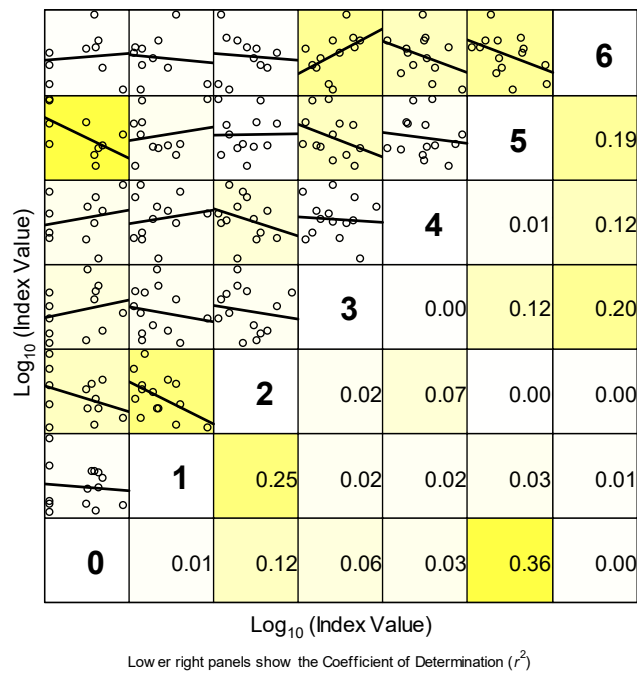


Figure 191: ARA in GSAs 1, 5, 6, 9, 10 and 11: cohort consistency numbers-at-age in the commercial catch.

Cohorts consistence in the MEDITS :

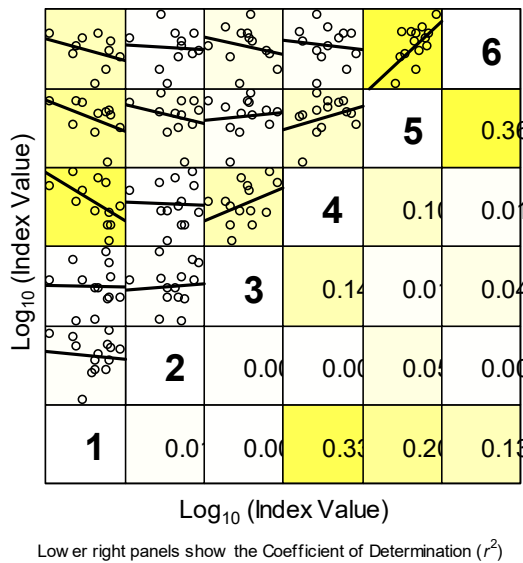


Figure 192: ARA in GSAs 1, 5, 6, 9, 10 and 11: cohort consistency in the catch numbers-at-age in the MEDITS.

4.12.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment using a4a was carried out using as input data the period 2005-2019 for both the catch data the tuning file (MEDITS indices).

The number of individuals by age was SOP corrected [SOP = Landings / $\sum a$ (total catch numbers at age a x catch weight-at-age a)].

F_{bar} range was fixed at 1-4, since most individuals belong to these age groups. Survey indices (N/km²) from MEDITS survey were used. Age 0 was removed from the analysis.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $f_{\text{model}} = \sim\text{factor}(\text{age}) + s(\text{year}, k = 7)$

Catchability sub-model: $q_{\text{model}} = \text{list}(\sim\text{factor}(\text{replace}(\text{age}, \text{age} > 4, 4)))$

Stock-recruitment sub.model: $sr_{\text{model}} = \sim\text{factor}(\text{year})$

4.12.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the following figures. Fishing mortality by age and year showed an increasing trend for the last years (Figure 193). Survey catchability showed a stable pattern along the years (Figure 194). The residuals of catch and abundance indices are shown in Figures 195 and 196 (bubble plot). The comparison between fitted and observed values are found in Figures 197 and 198.

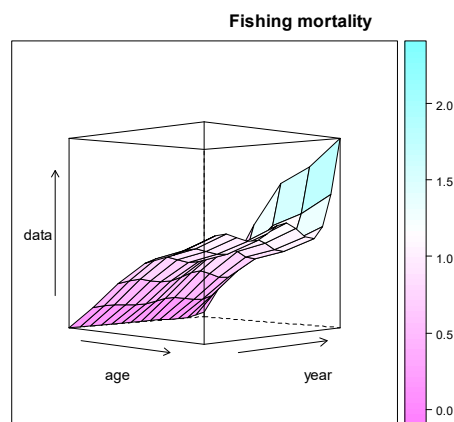


Figure 193: ARA in GSAs 1, 5, 6, 9, 10 and 11: fishing mortality by age and year obtained from the a4a model (2005-2019).

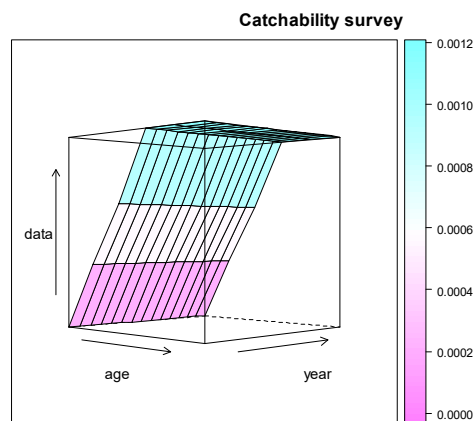


Figure 194: ARA in GSAs 1, 5, 6, 9, 10 and 11: catchability of the survey by age and year obtained from the a4a model (2005-2019).

log residuals of catch and abundance

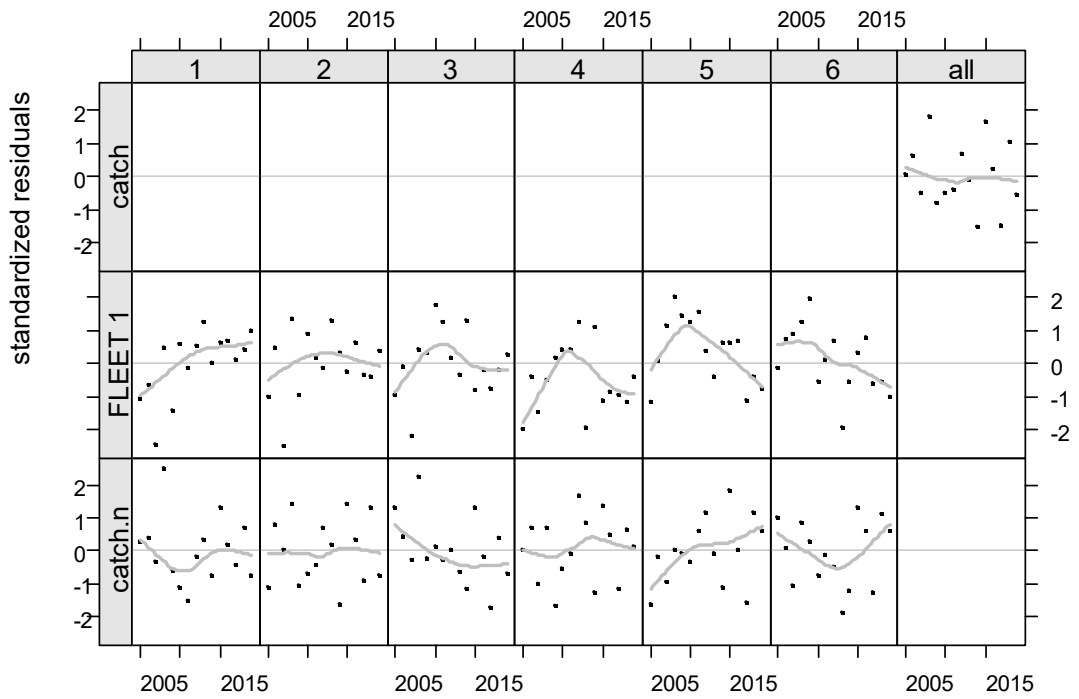


Figure 195: ARA in GSAs 1, 5, 6, 9, 10 and 11: log residuals for the catch-at-age data of the fishery and the survey.

log residuals of catch and abundance

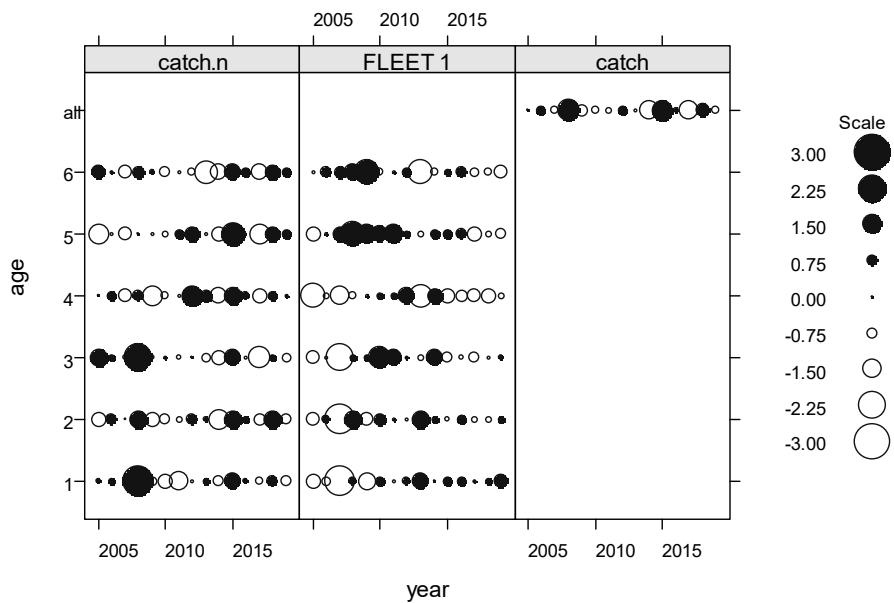


Figure 196: ARA in GSAs 1, 5, 6, 9, 10 and 11: bubble plot of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

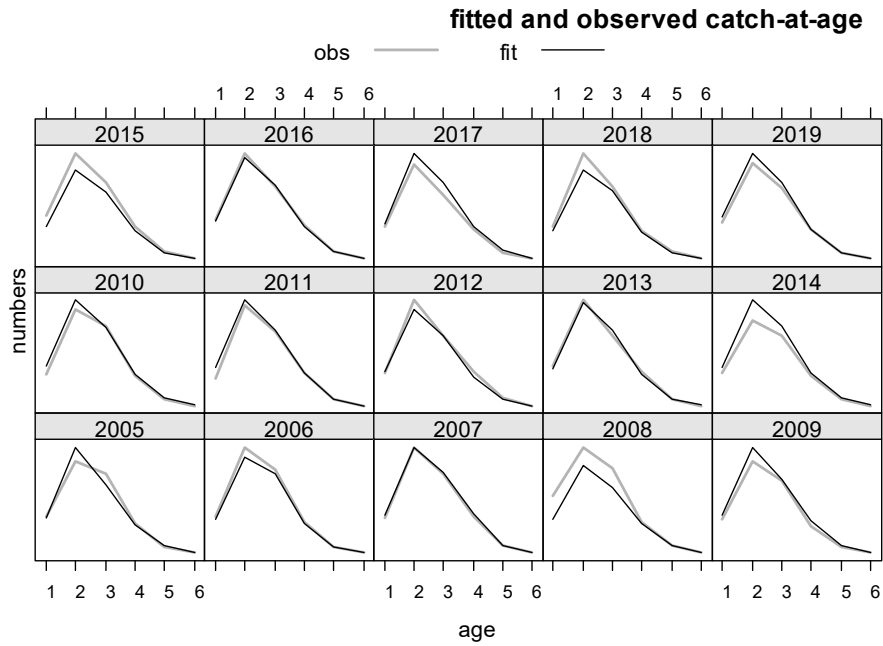


Figure 197: ARA in GSAs 1, 5, 6, 9, 10 and 11: fitted vs observed values by age and year for the catches.

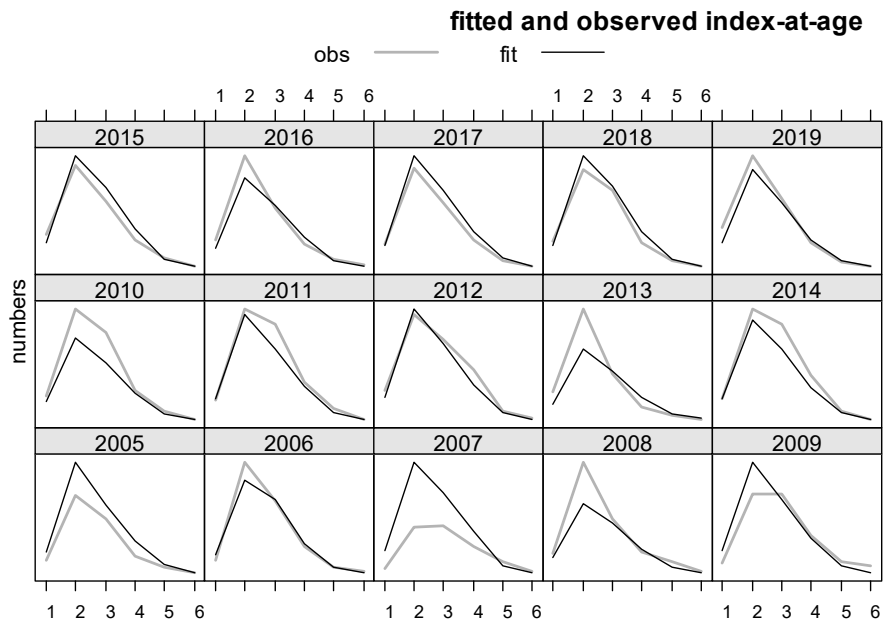


Figure 198: ARA in GSAs 1, 5, 6, 9, 10 and 11: fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated and showed that biomass of the plus group (age 6+) is between 2-4% of the total SSB (average 2.8%), with the lowest values for the time series in 2019 (2.4%) (Figure 12).

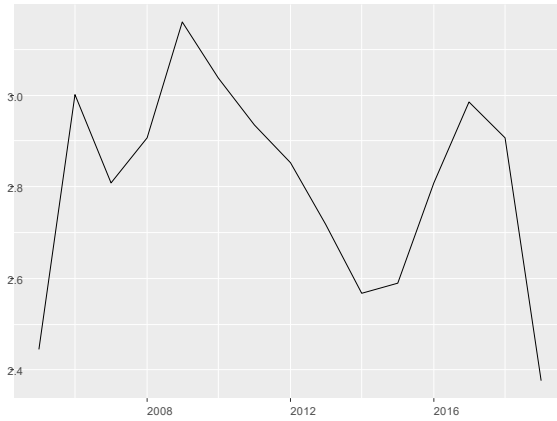


Figure 199: ARA in GSAs 1, 5, 6, 9, 10 and 11: cryptic biomass.

Figure 200 shows the final result of this assessment. The retrospective analysis shows a rather stable model (Figure 201). Figure 202 shows the difference between the estimated and the observed catches, with similar trends, except for some specific years. Stock numbers-at-age, fishing mortality-at-age and summary results of the model are presented respectively in Tables 65, 66 and 67 respectively.

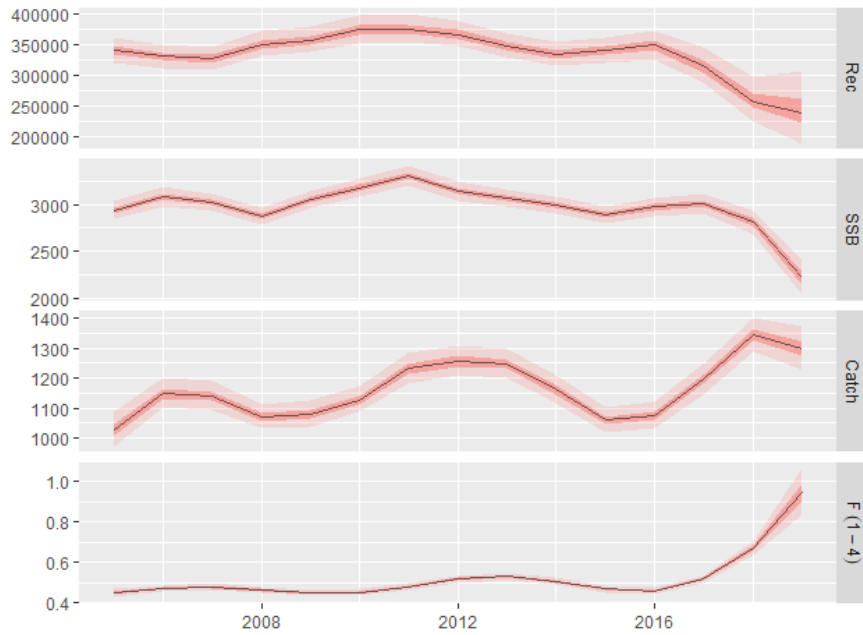


Figure 200: ARA in GSAs 1, 5, 6, 9, 10 and 11: output of the a4a model.

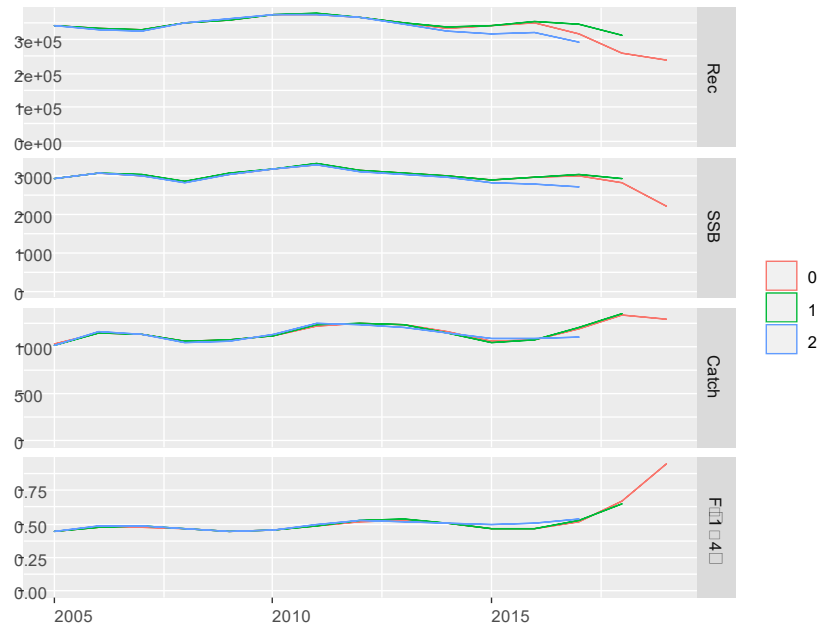


Figure 201: ARA in GSAs 1, 5, 6, 9, 10 and 11: retrospective analysis.

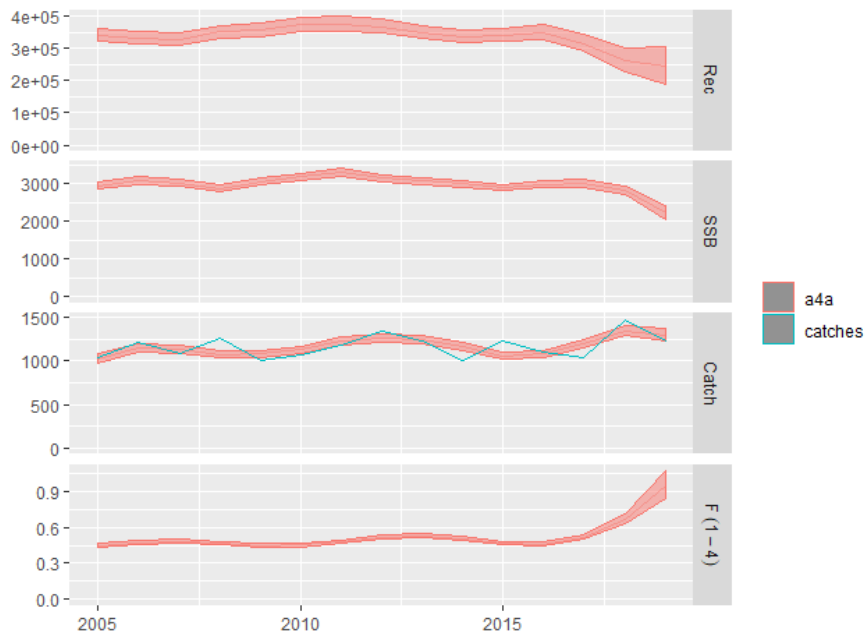


Figure 202: ARA in GSAs 1, 5, 6, 9, 10 and 11: outputs of the a4a stock assessment model, with uncertainty; input catch data (blue line) are plotted against the estimated catches.

Table 65: ARA in GSAs 1, 5, 6, 9, 10 and 11: Stock numbers-at-age (thousands).

Age/Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	340771.6	331643.3	327481.6	350678.5	358363.2	374516	375876.5	367626.5	348056.3
2	163455.2	149916.7	145486.1	143544.6	153973.8	157660.1	164709.1	164747.2	160426.1
3	60817.1	71631.4	64602.5	62396.3	62182.9	67487.6	68961.1	70613.4	68823.5
4	21266.3	22396.4	25548.8	22834.6	22478.5	22906.8	24763.6	24357.4	23742.4
5	6813.1	6568.6	6615.9	7452.6	6840.1	6946.2	7039.9	7216.2	6626.5

6+	2051.3	2535.1	2458.8	2418.1	2720.8	2717.9	2731.1	2597.6	2415.5
Age/Year	2014	2015	2016	2017	2018	2019			
1	334760.4	340675.7	349873.5	315788.7	259506.6	241104.2			
2	151647.9	146283.2	149533.3	153723.4	137820	111220.3			
3	66399.5	63863.7	63249	65034.2	64260	51753.2			
4	22735.9	22670.1	22924.1	22957.9	21893.6	17642.7			
5	6302.2	6319	6756.6	6939.4	6256.6	4489.4			
6+	2162.7	2138.1	2315.4	2527.7	2334	1525.9			

Table 66: ARA in GSAs 1, 5, 6, 9, 10 and 11: Fishing mortality-at-age.

Age/Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	0.053	0.056	0.057	0.055	0.053	0.053	0.057	0.061	0.063
2	0.314	0.331	0.336	0.326	0.314	0.316	0.336	0.362	0.371
3	0.597	0.629	0.638	0.619	0.597	0.601	0.639	0.688	0.706
4	0.833	0.877	0.890	0.863	0.832	0.838	0.891	0.960	0.984
5	0.923	0.972	0.986	0.957	0.922	0.928	0.987	1.064	1.091
6+	1.070	1.127	1.143	1.109	1.069	1.076	1.144	1.233	1.264
Age/Year	2014	2015	2016	2017	2018	2019			
1	0.060	0.055	0.054	0.061	0.079	0.112			
2	0.354	0.327	0.322	0.361	0.468	0.661			
3	0.673	0.623	0.611	0.687	0.891	1.258			
4	0.938	0.869	0.853	0.958	1.242	1.754			
5	1.040	0.962	0.945	1.062	1.377	1.944			
6+	1.205	1.115	1.096	1.230	1.596	2.253			

Table 67: ARA in GSAs 1, 5, 6, 9, 10 and 11: summary results of the a4a assessment. Age 1 individuals have been considered as recruitment.

Year	Catch (t)	SSB (t)	Rec (000)	F_{bar} (1-4)	Total biomass (t)
2005	1029.5	2936.5	340771.6	0.449	4469.4
2006	1148.0	3084.0	331643.3	0.473	4446.8
2007	1139.3	3021.6	327481.6	0.480	4490.8
2008	1070.3	2869.9	350678.5	0.466	4123.9
2009	1080.9	3057.4	358363.2	0.449	4653.4
2010	1127.4	3177.7	374516.0	0.452	4839.9
2011	1231.2	3303.2	375876.5	0.481	4984.3
2012	1254.9	3141.5	367626.5	0.518	4611.9
2013	1246.7	3064.9	348056.3	0.531	4343.9
2014	1161.8	2990.8	334760.4	0.506	4482.6
2015	1059.8	2890.0	340675.7	0.468	4392.1
2016	1076.1	2968.4	349873.5	0.460	4509.8
2017	1194.5	3006.1	315788.7	0.517	4433.4
2018	1342.8	2810.6	259506.6	0.670	4021.3
2019	1297.7	2214.0	241104.2	0.946	3292.7

4.12.4 Comparison with other assessments with a different stock configuration

The current assessment has been compared with previous assessments in the area carried out within the STECF (STECF, 2020) and GFCM (GFCM, 2021). In the first case, the GSAs were grouped as follows: GSA 1, GSA 5, GSA 6-7 and GSAs 9-10-11. In the second case, the GSAs assessed were GSA 1, GSA 2, GSA 5, GSA 6 and GSAs 9-10-11. For GSA 1, the time period covered was the same in both case (2002-2019); for GSA 5, STECF provided advice based on the survey index and GFCM based on a4a (1992-2019); for GSA 6-7 from STECF, the time period was 2002-2019; for GSA 6 from GFCM the time period was 2000-2019 and for GSA 9-10-11 the assessment endorsed by GFCM was the one carried out within STECF (2006-2019).

Model diagnostics

The diagnostics of the models performed during the STECF EWG 20-09 (STECF; 2020) were plotted for comparison with the diagnostic of the model described above (Figure 203).

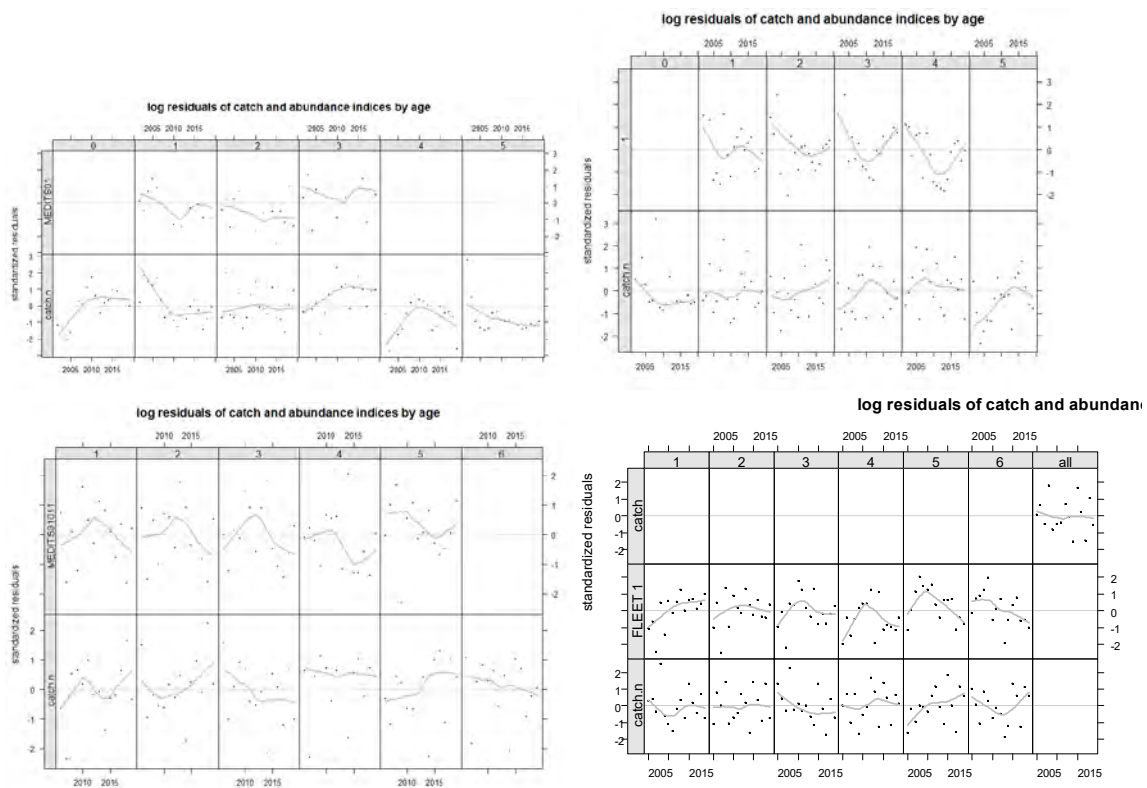


Figure 203: ARA in the western Mediterranean: model diagnostic for the stock assessment of STECF EWG 20-09 for GSA 1 (up left), GSA 6-7 (up right), GSA 9-10-11 (bottom left) and the current model (bottom right).

Retrospective patterns

Figure 204 shows the retrospective patterns for the different GSA aggregations.

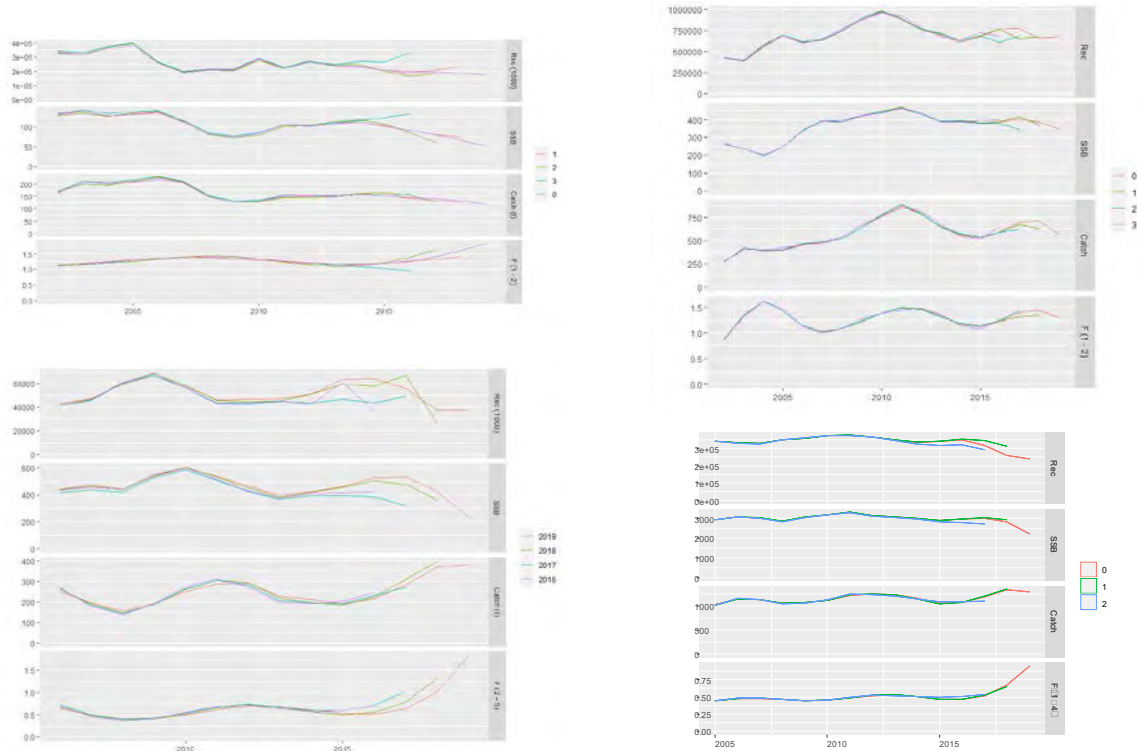


Figure 204: ARA in the western Mediterranean: retrospective patterns of STECF EWG 20-09 GSA 1 (up left), GSA 6-7 (up right), GSA 9-10-11 (bottom left) and the current model (bottom right).

Stock trajectories

The comparison with previous models was restricted to the time-frame 2006-2019 due to data availability (Figure 205). The SUM of SSB by STECF approach represented approximately 30% on average of the SSB obtained from the MED_UNITS approach, although the trends were consistent, with a Pearson correlation coefficient of 0.91. These differences may be related to different biological parameters (maturity ogives and growth parameters) used in the different approaches. Similarly, recruitment estimates by STECF approach represent approximately 50% of the MED_UNITS approach, with a Pearson correlation coefficient of 0.771. It should be noted that recruitment was considered as age 1 for some of the assessments and age 0 for others. Both SSB and recruitment, in both approaches, showed a decreasing trend for the last years. The Fbar and F/F01 trends were also compared, including an estimation of the average annual value. F showed an increasing trend in the MED_UNITS approach, also detected for GSAs 1, 9-10-11, but not for GSAs 6-7. The ration F/F01 showed values higher than 5 for the MED_UNITS approach but also for GSA 1 and GSAs 9-10-11.

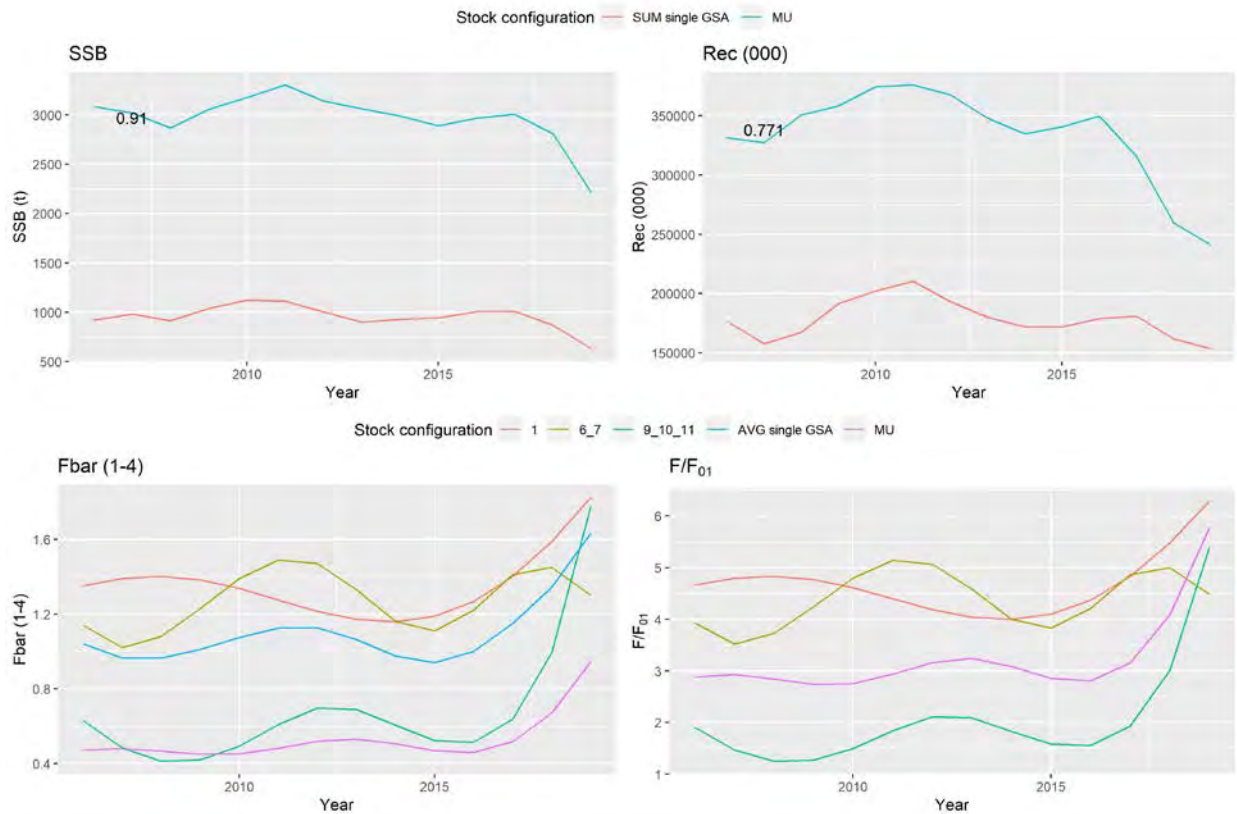


Figure 205: ARA in the western Mediterranean: results of the stock assessment of STECF EWG 20-09 for GSA 1, GSA 6-7 and GSA 9-10-11 and the current model.

F_{bar} (terminal year) and reference points

Figure 206 shows a comparison between last available year (2019) for Fbar, F01 and their ratio. Both Fbar and F01 estimated within the MED_UNITS framework was lower than any of the other approaches. Thus, the ratio F/F01 was close to the mean estimated from the different approaches.

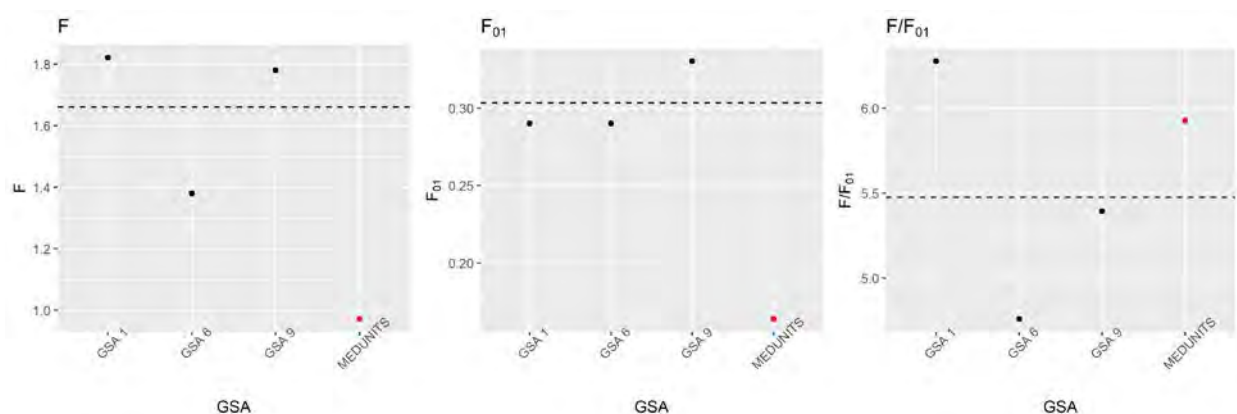


Figure 206: ARA in the western Mediterranean: Fbar, F01 and their ratio from the stock assessment of STECF EWG 20-09 for GSA 1, GSA 6-7 (labeled as GSA 6) and GSA 9-10-11 (labeled as GSA 9) and the MED_UNITS model.

4.13 Blue red shrimp in GSAs 15, 16, 18-20

The stock assessment of the blue and red shrimp in the Central Mediterranean Sea is here presented, based on the WP1 and WP2 outputs (Figure 207).

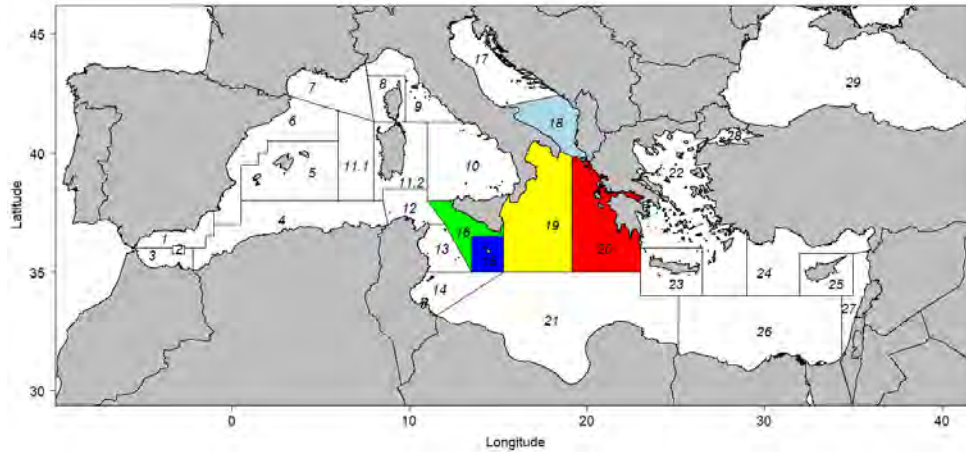


Figure 207: ARA in Central Mediterranean

The blue and red shrimp is distributed throughout the Mediterranean and the Eastern Atlantic, from Portugal to Cabo Verde Islands (Arrobas and Ribeiro-Cascalho 1987). In the Mediterranean, the species shows a longitudinal gradient, being more abundant in the north-western areas than in the central and eastern ones (D'Onghia et al. 1998, Cau et al. 2002, Cardinale et al. 2017, Guijarro et al. 2019). It is an euribatic species, with a wide bathymetric distribution between 80 down to 3300 m depth (D'Onghia et al. 1998, Sardà et al. 2003, Politou et al. 2004). At depths shallower than 1000 m, the population is mostly composed of adults, with females dominant and showing clear spatiotemporal variations in its population dynamics. Below 1000 m depth, the population is more stable, mostly composed of juveniles and dominated by males (Sardà et al., 1994, 2003) and thus, the deep waters seem to act as a refuge area mostly for the recruitment of this species and, to a lesser extent, for spawners (D'Onghia et al., 2009).

In the central Mediterranean, its exploitation is overlapped with the giant red shrimp *Aristaeomorpha foliacea* and this is the species mainly exploited in southern Sardinia (Ragonese 1995). Since the exploited population mostly consists of reproductive females, conditions of recruitment overfishing might occur (D'Onghia et al., 2009). In this area, the exploitation is mainly carried out by Italian fleets, while the north-western Greek coast is considered an unexploited area (D'Onghia et al., 2005). In the Ionian Sea, the recruitment was observed on a wide depth range with an increasing frequency of occurrence at greatest depths of small individuals and males and seems to occur as a discrete phenomenon (D'Onghia et al., 2009). Although very small individuals are found between October to November (Mura et al., 1997), in the Greek Ionian Sea the young of the year appeared to be fully recruited only in the late winter/early spring (Papaconstantinou and Kapiris, 2001).

4.13.1 Input data

Fishery dependent data

Commercial data (catch data, LFDs and biological parameters) for GSAs 16, 18 and 19 were collected from the EU DCR/DCF database, considering only data from OTB. No data were available for GSA 15 and, for GSA 20, only catches from the last two years were available and they represent less than 0.05%, so they were not considered. Considering the data available, this assessment was carried out using years from 2006 to 2019 (Table 68).

Table 68: ARA in GSAs 16, 18 and 19: Catch data divided by country, GSA and year.

Year	ITA - GSA 16	ITA - GSA 18	ITA - GSA 19
2006	162.790	21.327	437.565
2007	164.410	14.168	359.648
2008	135.090	4.628	201.853
2009	47.510	14.073	225.077
2010	54.231	21.594	206.525
2011	59.800	24.837	159.986
2012	92.020	4.325	263.387
2013	71.190	4.414	242.598
2014	116.601	2.697	299.460
2015	385.011	10.470	78.971
2016	402.291	16.757	103.020
2017	797.058	36.313	27.628
2018	277.915	67.936	335.692
2019	247.511	51.947	406.020

LFDs from the commercial data GSAs 16, 18 and 19 were sliced by sex using the set of growth parameters available from the DCF in GSA 19, with both sexes combined (Table 69). Plus group was set at age 6. Maturity ogive used comes from the DCF in GSA 19 and M vector was computed from Chen and Watanabe (Table 70).

Table 69: ARA in GSAs 16, 18 and 19: VBGF and LW parameters.

Sex	Linf	k	t0
Females+Males	66	0.243	-0.2
	a		b
Females+Males	0.032	2.4017	

Table 70: ARA in GSAs 16, 18 and 19: natural mortality and maturity-at-age vectors.

Ages	Natural mortality	Proportion of matures (combined)
0	0	1.554
1	0.428	0.718
2	0.984	0.505
3	1	0.410
4	1	0.357
5	1	0.324
6+	1	0.301

Fishery independent data

LFDs (n/km^2) from the MEDITS surveys in GSAs 16, 18 and 19 were sliced using the same set of growth parameters used for the commercial data (Table 69).

The catch at age from the commercial fleet and the survey are presented respectively in Figure 208 and Figure 209. Moreover the internal consistencies of such input data are presented respectively in Figure 210 and Figure 211.

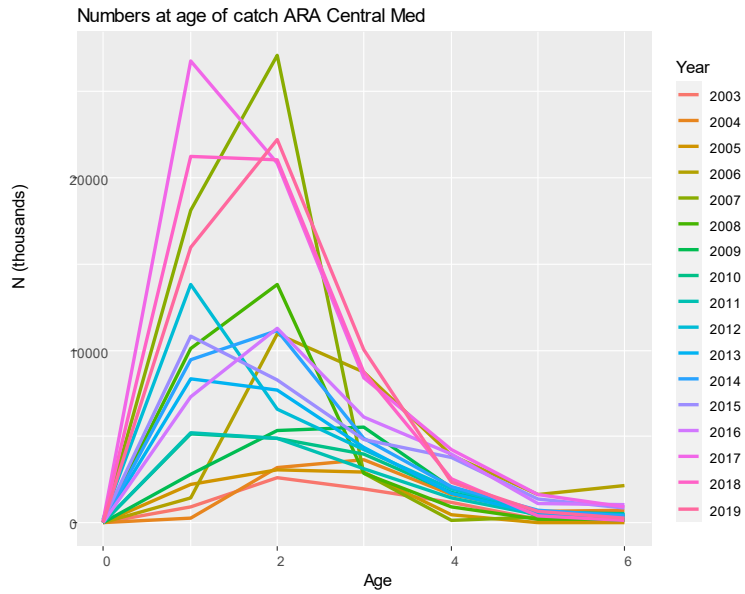


Figure 208: ARA in GSAs 16, 18 and 19: catch numbers-at-age for the commercial fleet.

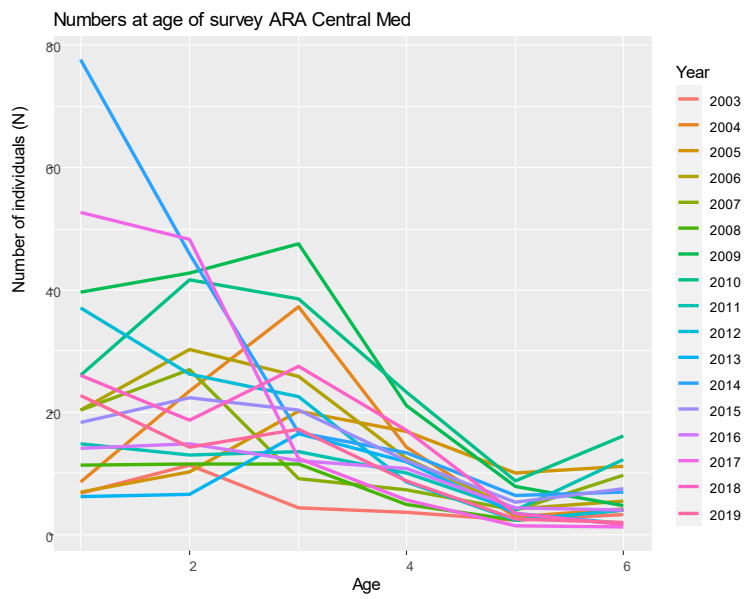


Figure 209: ARA in GSAs 16, 18 and 19: catch numbers-at-age for the MEDITS survey.

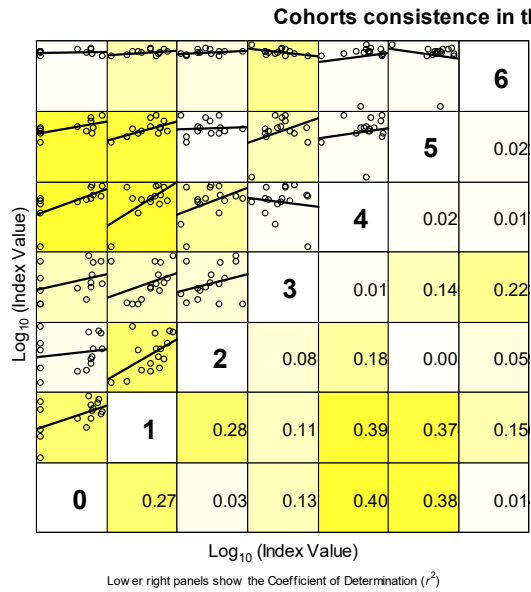


Figure 210: ARA in GSAs 16, 18 and 19: cohort consistency numbers-at-age in the commercial catch.

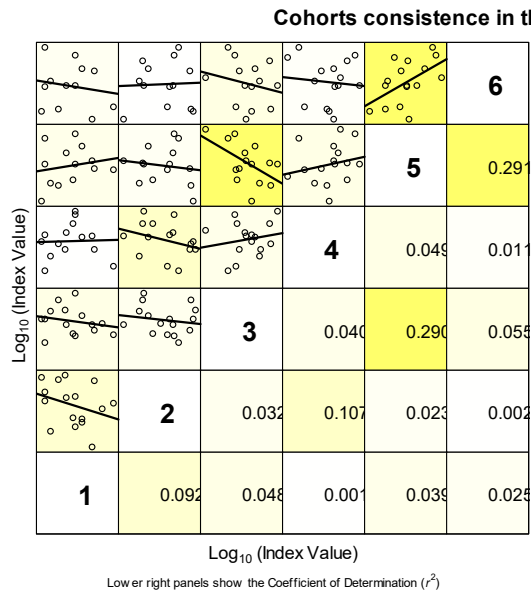


Figure 211: ARA in GSAs 16, 18 and 19: cohort consistency in the catch numbers-at-age in the MEDITS.

4.13.2 Model description

FLR libraries were employed to carry out a Statistical Catch-at-age (a4a) assessment. The assessment using a4a was carried out using as input data the period 2003-2019 for both the catch data the tuning file (MEDITS indices).

The number of individuals by age was SOP corrected [$SOP = Landings / \sum a$ (total catch numbers at age $a \times$ catch weight-at-age a)].

F_{bar} range was fixed at 1-4, since most individuals belong to these age groups. Survey indices (N/km^2) from MEDITS survey were used. Age 0 was removed from the analysis.

The model settings that minimized the residuals and showed the best diagnostics outputs were used for the final assessment, and are the following:

Fishing mortality sub-model: $f_{\text{model}} = \sim s(\text{year}, k=7) + \text{factor}(\text{replace}(\text{age}, \text{age}>2, 2))$

Catchability sub-model: $q_{\text{model}} = \text{list}(\sim \text{factor}(\text{replace}(\text{age}, \text{age}>3, 3)))$

Stock-recruitment sub.model: $s_{\text{model}} = \sim \text{factor}(\text{year})$

4.13.3 Outputs from the model

The diagnostics and outputs of the assessment run are presented in the following figures. Fishing mortality by age and year showed an increasing trend for the last years (Figure 212). Survey catchability showed a stable pattern along the years (Figure 213). The residuals of catch and abundance indices are shown in Figures 214 and 215 (bubble plot). The comparison between fitted and observed values are found in Figures 216 and 217.

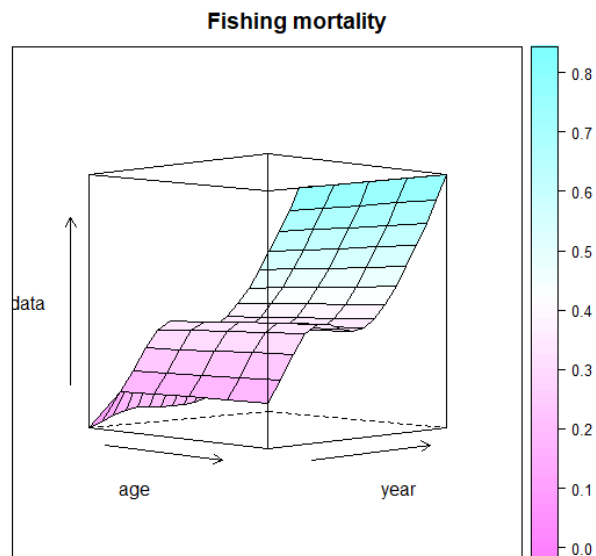


Figure 212: ARA in GSAs 16, 18 and 19: fishing mortality by age and year obtained from the a4a model.

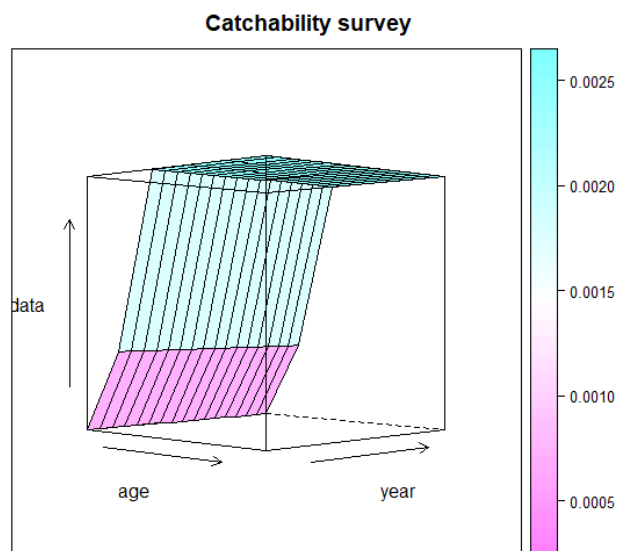


Figure 213: ARA in GSAs 16, 18 and 19: catchability of the survey by age and year obtained from the a4a model.

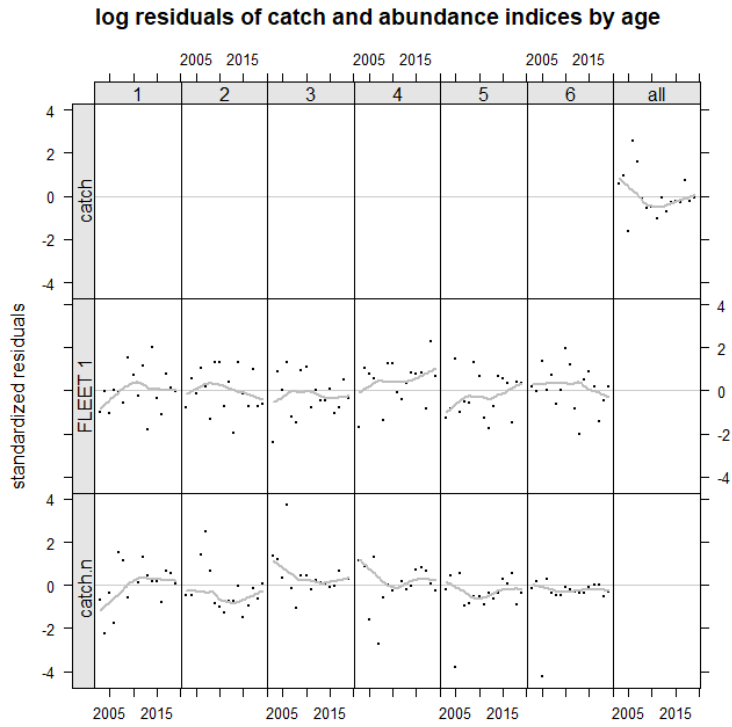


Figure 214: ARA in GSAs 16, 18 and 19: log residuals for the catch-at-age data of the fishery and the survey.

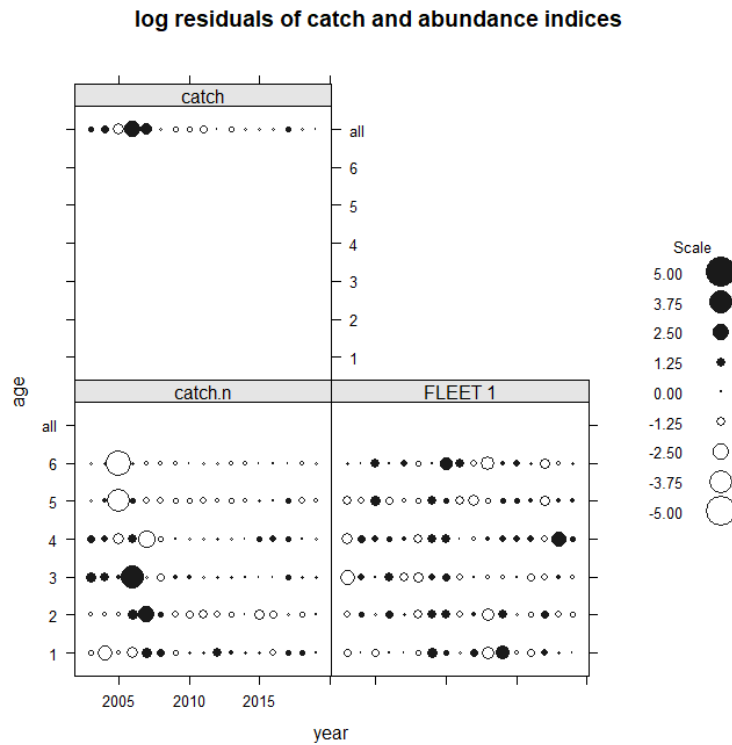


Figure 215: ARA in GSAs 16, 18 and 19: bubble plot of the log residuals for the catch-at-age data of the fishery and the survey, and the catches.

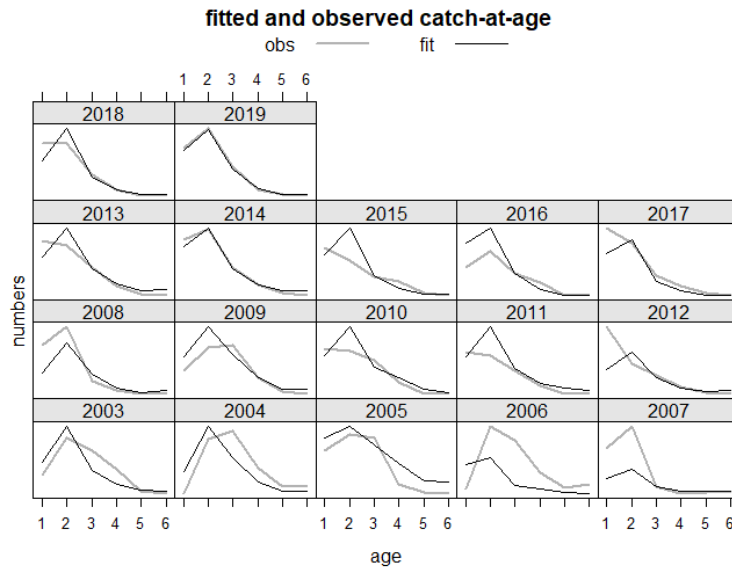


Figure 216: ARA in GSAs 16, 18 and 19: fitted vs observed values by age and year for the catches.

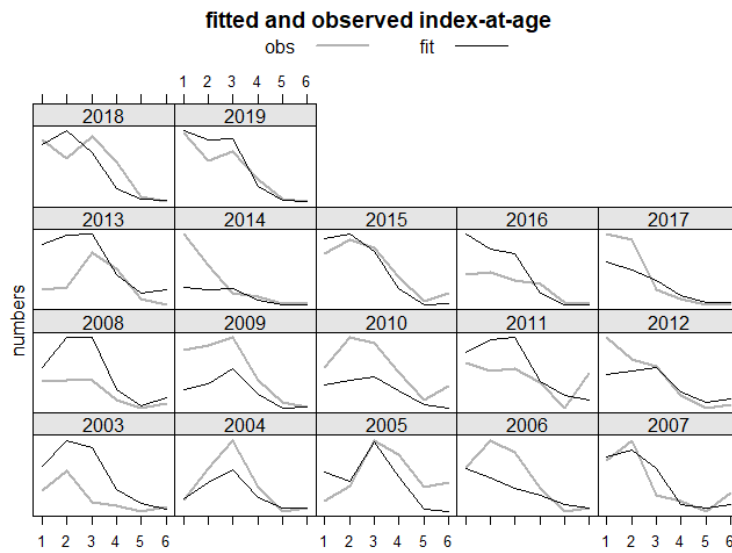


Figure 217: ARA in GSAs 16, 18 and 19: fitted vs observed values by age and year for the survey.

The effect of cryptic biomass was investigated and showed that biomass of the plus group (age 6+) is between 4-14% of the total SSB (average 9.5%), with the lowest values for the time series in 2018-2019 (Figure 218).

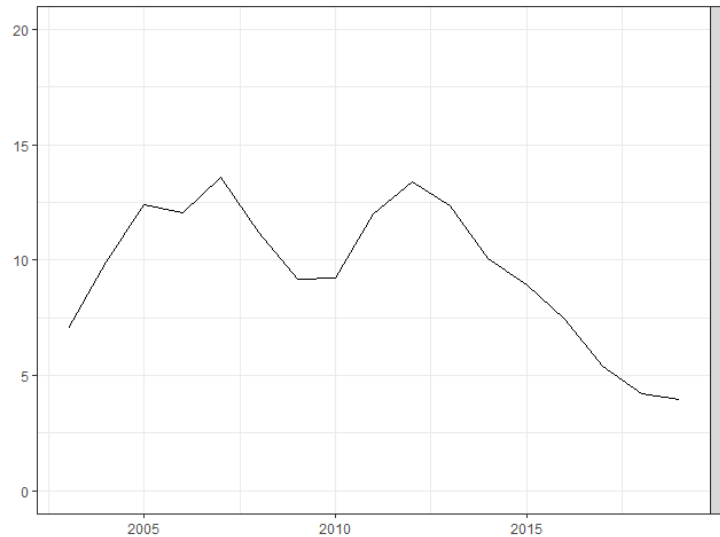


Figure 218: ARA in GSAs 16, 18 and 19: cryptic biomass.

Figure 219 shows the final result of this assessment. The retrospective analysis shows certain instability in the model (Figure 220). Figure 221 shows the difference between the estimated and the observed catches, with similar trends, except for some specific years. Stock numbers-at-age, fishing mortality-at-age and summary results of the model are presented respectively in Tables 71, 72 and 73 respectively.

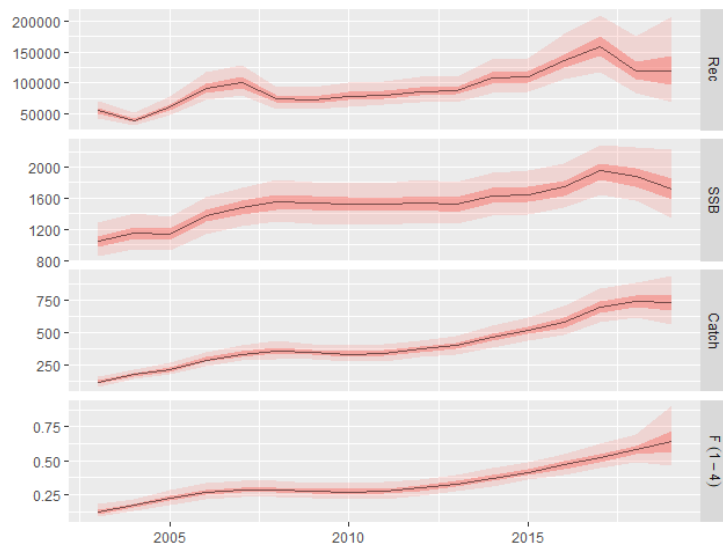


Figure 219: ARA in GSAs 16, 18 and 19: output of the a4a model.

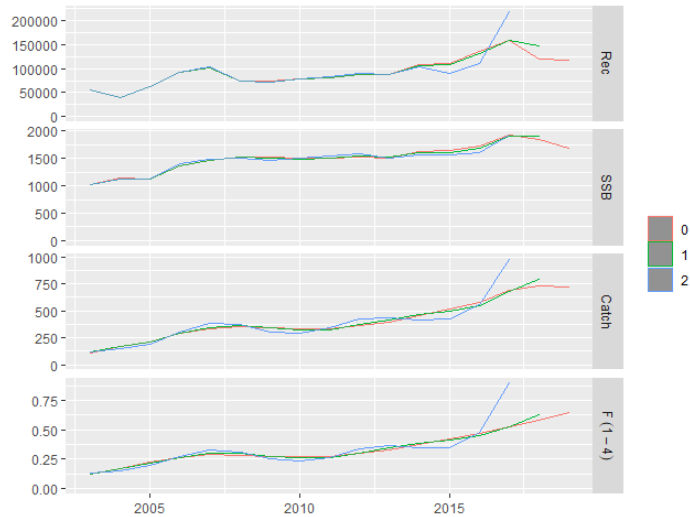


Figure 220: ARA in GSAs 16, 18 and 19: retrospective analysis.

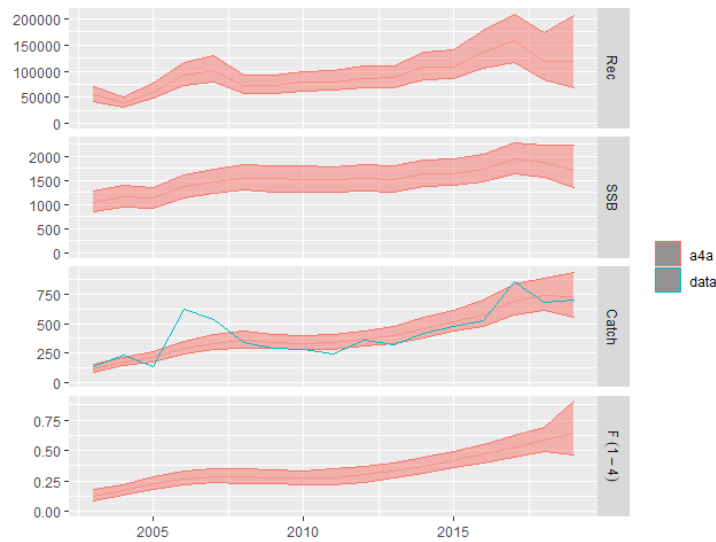


Figure 221: ARA in GSAs 16, 18 and 19: outputs of the a4a stock assessment model, with uncertainty; input catch data (blue line) are plotted against the estimated catches.

Table 71: ARA in GSAs 16, 18 and 19: Stock numbers-at-age (thousands).

Age/Year	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	55436.6	39773.9	61349	92228.2	101126.6	73803.6	73134.8	78573.3	80922.3
2	27556.4	26019.2	18402.5	27930.4	41437.2	45155.1	32970.4	32774.3	35272.2
3	9710.2	14252.2	12704.6	8421.1	12119.8	17539.5	19149.9	14159.7	14173.4
4	4133.4	5522.6	7652.5	6393.1	4018.3	5641.3	8179.7	9043.8	6733.7
5	2296.8	2478.8	3126.7	4060.4	3216.7	1972.2	2774.1	4073.2	4534.9
6+	1541.7	2401.4	2888.4	3335.5	3886.5	3649	2900.1	2955.1	3678.1
Age/Year	2012	2013	2014	2015	2016	2017	2018	2019	
1	86551.1	87526.2	107918.4	109951.6	136235.5	158463.2	120377.9	118237.6	
2	36255.8	38536.9	38582.4	46969.9	47173.7	57561.1	65866.6	49157.6	
3	15134.4	15171.8	15489.8	14734.7	16933.9	15991.1	18271.1	19471.1	
4	6687.5	6964.4	6706	6505.1	5841.7	6312.4	5581.8	5939.5	
5	3350.1	3244.9	3245.8	2969.5	2719.4	2296.1	2323.3	1913.3	

6+ 4267.2 3869.9 3470.5 3110.9 2658.4 2209.8 1733.6 1451.5

Table 72: ARA in GSAs 16, 18 and 19: Fishing mortality-at-age.

Age/Year	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	0.038	0.053	0.069	0.082	0.088	0.088	0.085	0.083	0.085
2	0.154	0.212	0.277	0.330	0.355	0.353	0.340	0.333	0.341
3	0.154	0.212	0.277	0.330	0.355	0.353	0.340	0.333	0.341
4	0.154	0.212	0.277	0.330	0.355	0.353	0.340	0.333	0.341
5	0.154	0.212	0.277	0.330	0.355	0.353	0.340	0.333	0.341
6+	0.154	0.212	0.277	0.330	0.355	0.353	0.340	0.333	0.341

Age/Year	2012	2013	2014	2015	2016	2017	2018	2019
1	0.091	0.101	0.114	0.128	0.144	0.160	0.178	0.197
2	0.366	0.406	0.458	0.515	0.577	0.643	0.714	0.792
3	0.366	0.406	0.458	0.515	0.577	0.643	0.714	0.792
4	0.366	0.406	0.458	0.515	0.577	0.643	0.714	0.792
5	0.366	0.406	0.458	0.515	0.577	0.643	0.714	0.792
6+	0.366	0.406	0.458	0.515	0.577	0.643	0.714	0.792

Table 73: ARA in GSAs 16, 18 and 19: summary results of the a4a assessment. Age 1 individuals have been considered as recruitment.

Year	Catch (t)	SSB (t)	Rec (000)	F _{bar} (1-4)	Total biomass (t)
2003	112.0	1026.9	55437.0	0.125	1255.1
2004	170.1	1135.5	39774.0	0.172	1323.0
2005	212.1	1119.0	61349.0	0.225	1333.3
2006	286.3	1355.4	92228.0	0.268	1783.2
2007	331.1	1460.3	101127.0	0.288	1873.2
2008	354.9	1530.7	73804.0	0.287	1834.9
2009	342.4	1519.3	73135.0	0.276	1819.0
2010	331.7	1503.9	78573.0	0.271	1825.3
2011	337.0	1499.9	80922.0	0.277	1830.7
2012	366.7	1526.6	86551.0	0.297	1831.2
2013	394.2	1506.2	87526.0	0.330	1814.0
2014	456.6	1619.8	107918.0	0.372	2060.0
2015	512.9	1637.4	109951.6	0.418	2024.5
2016	572.4	1718.4	136235.5	0.468	2273.0
2017	690.0	1918.8	158463.2	0.522	2565.2
2018	732.4	1845.6	120377.9	0.580	2340.3
2019	716.8	1682.3	118237.6	0.643	2165.9

F_{bar}(1-4) for the last year was 0.643 and the reference point F_{0.1} was 0.276; thus, the ratio F_{bar}/F_{0.1} was 2.33.

4.13.4 Comparison with other assessments with a different stock configuration

This species has not been previously assessed in the GSAs included in this approach and thus it was not possible to perform a comparison.

5. Discussion and conclusion

Under Task 3.3, 13 novel stock configuration assessments were carried out and compared, when possible, with the assessments routinely carried out at the level of single GSA or combination of GSAs. The results of the analyses, presented in this deliverable, Deliverable 3.3 “Report on the stock assessments with the new stock configurations for the 6 target species of the study”, show that in most of the cases the new stock assessments do not present particular improvements of diagnostics; this can be due to several reasons, apart from the new stock configurations (increased data heterogeneity when the number of aggregated GSAs is increased, model settings, etc...). It must be acknowledged that the trials attempted under Task 3.3 of the MED_UNITS project represent a first approach to the assessment of the new stock configurations, and further investigation shall be implemented before scientific advice can be provided in a reliable and robust way.

European hake was assessed considering the entire western Mediterranean (GSAs 1-12) and the part of the eastern Mediterranean (GSAs 22-27). The assessments performed within the MED_UNITS project does not seem to improve the existing assessments. In particular, for the western Mediterranean, although genetic and otolith results indicate a panmictic population, the fishing pattern and the environmental conditions over the investigated area are very heterogeneous, resulting in a wide range of growth parameters and exploitation patterns. Therefore, the MED_UNITS configuration can potentially hinder the possibility to rely on a unique view for growth parameters and F trends. This evidence caused a deterioration of the population structure as estimated on the aggregated area, and an exploitation pattern that was just an average of many different facets. The eastern Mediterranean configuration seems improving the existing assessment. However, this new aggregation cannot resolve the uncertainties in the catch data, already mentioned in the stock assessment of hake in GSA 22 performed by STECF EWG 20-15 (STECF, 2020b). In addition, to cover the whole stock distribution as identified by the MED_UNITS results, the present model includes catch data from Syria and Israel that can increase the uncertainty, whereas no catch data are available for Turkey. The MED-UNITS assessment agrees with the stock status described by the SPiCT assessment developed in GSA 22 only, thus resulting in conflict with the results depicted by the a4a model developed in GSA 22 (STECF, 2020b). Considering these aspects, further investigations are needed for giving a clear picture of the status of this stock.

Red mullet was assessed considering three distinct stocks distributed in the western Mediterranean (GSAs 1, 5-16), the Adriatic and Ionian region (GSAs 17-20) and the eastern Mediterranean (GSAs 22-25). In the western area, the model performed on such a large aggregation presented more disadvantages than advantages. Even considering that genetic and otolith results indicate a panmictic population, the fishing pattern and the environmental conditions over the investigated area are very heterogeneous, resulting in a wide range of growth parameters, though with variability probably due to the otolith reading and growth interpretation at such spatial scale (Carbonara et al., 2019), and exploitation patterns, thus hindering the possibility to rely on a unique view for growth parameters and F trends. This evidence caused a deterioration of the population structure as estimated on the aggregated area, and an exploitation pattern that was just an average of many different facets. In Adriatic and Ionian region, the comparison of cohort consistencies indicated a higher stability in the MEDITS index when aggregating GSAs. By contrast, aggregated catch cohort consistency shows a deterioration in respect to the stock assessment carried out in GSAs 17-18 and separately in GSA 19. Several reasons could be behind these controversial results, as the increased heterogeneity in the catch data when aggregating more GSAs, differences in the exploitation patterns more pronounced when aggregate more GSAs, difference in the stock assessment models used (e.g., for 17_18 a4a, for 19 XSA) and stock assessment settings (0 age in tuning used or not). Regarding model diagnostics, no large difference was observed in the model fits, while an improvement in the retrospective patterns was observed in the MED_UNITS stock configuration. The genetic and otolith results do not indicate well-defined and confined sub-populations for this species. However, comparing the results of different assessments is complex and could lead to a high uncertainty represented by

different fishing patterns, environmental conditions, and growth parameters used. The analyses carried out for red mullet in the eastern sector of the Mediterranean showed large variations in data availability among GSAs and the lack of information about potential differences in the exploitation pattern among fleets. Therefore, it seems preferable to avoid aggregation of such a large area and rely on separate assessments for providing advice. Only for GSAs 22 and 23, which share similar data gaps and are exploited by the same fleets, performing a joint assessment seems suitable, given the findings of the current study.

Deep-water rose shrimp was assessed in the western Mediterranean (GSAs 1, 5-16) and the Adriatic and Ionian region (GSAs 17-20). In the first case, considering variations in data availability among GSAs, it seems preferable to avoid aggregation of such a large area and rely on smaller GSA aggregations for providing advice, similarly to recent STECF and GFCM approaches. In the Adriatic and Ionian region, the overall perception of the stock is not that different between the assessment of STECF 20-15 and the MED_UNITS one. However, the increment of t_0 in the STECF 20-15 changed the overall distribution across ages, so the two stock assessments are not directly comparable. In the MED_UNITS assessment the cohort consistency of both catch and index improved and the retrospective pattern appears more stable than the one in the STECF report.

Norway lobster was assessed considering the western Mediterranean (GSAs 1, 2, 5-11) and the Adriatic and Ionian region (GSAs 17-19). In the first case, the new assessment seems to improve the existing assessments in terms of some diagnostics. The use of separate VBGF and L-W relationship parameters (borrowed from GSA9, and previously used for the input data of the assessment of NEP in GSA9) allowed obtaining a robust population structure, consistent with what was observed in the single GSAs (GSA 6 and GSA 9) assessments. The exploitation pattern obtained for the combination of GSAs was close to the average of the two available assessments for GSA 6 and GSA 9. The Adriatic and Ionian region assessment of Norway lobster is consistent with the assessment performed in GSAs 17-18. GSAs 17-18, in fact, mostly drive the assessment, as the contribution of GSA 19 in terms of landings is relatively low (less than 10% as from DCF data). However, in the case of less vagile species such as Norway lobster, the new stock configurations should be considered with caution because can generate the first type of error defined by Punt (2003) (e.g., different stocks are assessed as a single unit), and further investigation may be needed. The results of the recent study by Melaku Canu et al. (2020) on Norway lobster connectivity in the Adriatic Sea suggest the presence of at least three different subpopulations within the basin, which would need to be independently assessed and managed taking into account their different life history traits (see Angelini et al., 2020).

Giant red shrimp was assessed considering the western Mediterranean (GSAs 5, 8-12) and the central Mediterranean and Ionian region (GSAs 15, 16, 18-20). In the first case, the new stock configuration does not show any difference with previous assessments (GSAs 9-11). In fact, the contribution of the additional GSAs in terms of landings and survey data is almost negligible. Differently, in the second scenario, the inclusion of GSA 16 data, characterized by higher landings, represented the main change compared to previous assessments including GSAs 18 and 19 only. This has affected significantly the overall results of the assessment, which provided a stock status of sustainable exploitation, despite not improving the cohort consistency, and diagnostics performance of the model run.

Blue and red shrimp was assessed considering part of the western Mediterranean (GSAs 1, 5, 6, 9-12) and part of the central and Ionian region (GSAs 15, 16, 18-20). In the first case, absolute values for SSB and recruitment showed differences between both approaches, giving larger values in the case of the MED_UNITS approach. We should take into account that, in the STECF approach, different growth parameters and maturity ogives are considered; additionally, recruitment in some cases is considered as age 0 and in others as age 1, depending on the growth parameters used. In any case, the trends were consistent in both approaches, with a certain decreasing trend for the last year. Regarding F , an increasing trend was detected in all cases, both in the STECF and in the MED_UNITS approach, except for GSA 6-7. In any case, the ratio between F and $F_{0.1}$ is quite consistent with values around 5-6. For the central and Ionian region, it was not possible to perform a comparison as there are not previous assessment on this GSAs. In this case, the fit of the model was not very robust, especially for the survey, although the residuals were

good.

In the case of red shrimps (*A. foliacea* and *A. antennatus*), the use of the outputs available from Tasks 3.1 and 3.2 in terms of fishing grounds and spatial overlap with the resources can improve the quality of the assessments in the future. Actually, in most of the cases, the vessels targeting red shrimps exploit fishing grounds in GSAs and areas far from the declared landing points (e.g., a considerable portion of Italian landings of *A. foliacea* reported in GSA 16 are probably coming from the eastern Mediterranean, see Armelloni et al., 2021). Therefore, a revision of the input data in relation to the new stock configurations envisaged by the genetic analyses could provide a more accurate picture of the stock status in future assessments.

6. Deviation from the Workplan

The present deliverable has been delayed due to the late availability of WP1 and WP2 outcomes and COVID-19 impacts.

Moreover, due to COVID-19 restrictions was not possible to organize a workshop in presence as foreseen in the Inception Report.

7. Remedial actions

Most of the meetings and the activities have been carried out remotely soon after the availability of WP1 and WP2 outputs by target species.

8. References

- Abella A., Serena F., Ria M., 2005. Distributional response to variations in abundance over spatial and temporal scales for juveniles of European hake (*Merluccius merluccius*) in the Western Mediterranean Sea. *Fish. Res.* 71, 295–310.
- Abella A., Fiorentino F., Mannini A., Orsi Relini L., 2008. Exploring relationships between recruitment of European hake (*Merluccius merluccius* L. 1758) and environmental factors in the Ligurian Sea and Strait of Sicily (Central Mediterranean). *Journal of Marine Systems* 71 (2008) 279–293. doi: 10.1016/j.jmarsys.2007.05.010
- Angelini, S., Martinelli, M., Santojanni, S., Colella, S., 2020. Biological evidence of the presence of different subpopulations of Norway lobster (*Nephrops norvegicus*) in the Adriatic Sea (Central Mediterranean Sea), *Fisheries Research*, 221. <https://doi.org/10.1016/j.fishres.2019.105365>
- Armelloni, E.A., Tassetti, A.N., Ferrà, C., Galdelli, A., Scanu, M., Mancini, A., Fabi, G., Scarcella, G. 2021. AIS data, a mine of information on trawling fleet mobility in the Mediterranean Sea. *Marine Policy*, 129. <https://doi.org/10.1016/j.marpol.2021.104571>.
- Bartolino V., Colloca F., Sartor P., Ardizzone G., 2008a. Modelling recruitment dynamics of hake, *Merluccius merluccius*, in the central Mediterranean in relation to key environmental variables. *Fish. Res.* 92(2-3): 277-288. <https://doi.org/10.1016/j.fishres.2008.01.007>
- Bartolino V., Ottavi A., Colloca F., Ardizzone G., Stefánsson G., 2008 b. Bathymetric preferences of juvenile European hake (*Merluccius merluccius*). *ICES Journal of Marine Science*, 65: 963–969. <https://doi.org/10.1093/icesjms/fsn079>
- Butterworth, D. S., and Geromont, H. F. 2001. Simulation testing as an approach to evaluate the reliability of assessment methods: An example involving initial consideration of the one/two stock hypotheses for north Atlantic bluefin tuna. *ICCAT Collective Volume of Scientific Papers*. 52: 1115:1129.
- Begg, G. A., Waldman, J. R. 1999. A holistic approach to fish stock identification. *Fisheries Research*. 43: 35-44.
- Biagi F., Sartor P., Ardizzone G. D., Belcari P., Belluscio A., Serena F., 2002. Analysis of demersal assemblages off the Tuscany and Latium coasts (north-western Mediterranean). *Sci.Mar.*, 66 (suppl.2): 233-242
- Brodziak, J., J. Ianelli, K. Lorenzen, and R.D. Methot Jr. (eds). 2011. Estimating natural mortality in stock assessment applications. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/SPO-119, 38 p.
- Brodziak, J., T. Gedamke, C. Porch, J. Walter, D. Courtney, J. O'Malley, and Benjamin Richards. 2012. A workshop on methods to estimate total and natural mortality rates using mean length observations and life history parameters. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-32, 26 p. + Appendix.
- Cadrin, S. X., Friedland, K. D., and Waldman, J. R. 2005. Stock identification methods: An overview. In *Stock identification methods: Applications in fishery science*. Edited by S.X. Cadrin, K.D. Friedland and J.R. Waldman. Elsevier Academic Press, pp. 356.
- Carbonara P., Porcu C., Donnalioia M., Pesci P., Sion L., Spedicato M. T., Zupa W., Vitale F., Follesa M. C., 2019. The spawning strategy of European hake (*Merluccius merluccius*, L. 1758) across the Western and Central Mediterranean Sea. *Fish. Res.* 105333. <https://doi.org/10.1016/j.fishres.2019.105333>
- Carbonara P., Zupa W., Anastasopoulou A., Bellodi A., Bitetto I., Charilaou C., Chatzisprou C., Elleboode R., Esteban A., a Follesa M.C., Isajlovic I., Jadaud A., García-Ruiz C., Giannakaki A., Guijarro B., Kiparissis S.E., Ligas A., Mahé K., Massaro A., Medvesek D., Mytilineou C., Ordines F., Pesci P., Porcu C., Peristeraki P., Thasitis I., Torres P., Spedicato M.T., Tursi A., Sion L. 2019. Explorative analysis on red mullet (*Mullus*

- barbatus) ageing data variability in the Mediterranean. *Sci. Mar.* 83S1: 271-279. <https://doi.org/10.3989/scimar.04999.19A>
- Carbonara, P., Intini, S., Kolutari, J., Joksimović, A., Milone, N., Lembo, G., Casciaro, L., et al. 2018. A holistic approach to the age validation of *Mullus barbatus* L., 1758 in the Southern Adriatic Sea (Central Mediterranean). *Scientific Reports*, 8: 1–19.
- Carbonell, A., Carbonell, M., Demestre, M., Grau, A., Monserrat, S., 1999. The red shrimp *Aristeus antennatus* (Risso, 1816) fishery and biology in the Balearic Islands, Western Mediterranean. *Fish. Res.* 44: 1–13.
- Cardinale, M., Osio, G.C., Scarcella, G., 2017. Mediterranean Sea: A failure of the European Fisheries Management System. *Front. Mar. Sci.* 4: 72.
- Cau, A., Carbonell, A., Follesa, M.C., Mannini, A., Norrito, G., Orsi-Relini, L., Politou, C.Y., Ragonese, S., Rinelli, P., 2002. MEDITS-based information on the deep-water red shrimps *Aristaeomorpha foliacea* and *Aristeus antennatus* (Crustacea: Decapoda: Aristeidae). *Sci. Mar.* 66: 103–124.
- Colloca F., Bartolino V., Lasinio G.J., Maiorano L., Sartor P., Ardizzone G., 2009. Identifying fish nurseries using density and persistence measures. *Mar. Ecol. Prog. Ser.* 381: 287-296. <https://doi.org/10.3354/meps07942>
- Colloca F., Cardinale M., Belluscio A., Ardizzone G., 2003. Pattern of distribution and diversity of demersal assemblages in the central Mediterranean Sea. *Estuarine, Coastal and Shelf Science* Volume 56, Issue 3-4, Pages 469 – 480
- Company, J.B., Puig, P., Sardà, F., Palanques, A., Latasa, M., Scharek, R., 2008. Climate influence on deep sea populations. *PLoS One* 3: e1431.
- Cope, J. M., and Punt, A.E. 2009. Drawing the lines: Resolving fishery management units with simple fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences.* 66: 1256-12.
- D’Onghia, G., Capezzuto, F., Mytilineou, Ch., Maiorano, P., Kapiris, K., Carlucci, R., Sion, L., Tursi, A., 2005. Comparison of the population structure and dynamics of *Aristeus antennatus* (Risso, 1816) between exploited and unexploited areas in the Mediterranean Sea. *Fish. Res.* 76: 22-38.
- D’Onghia, G., Giove, A., Maiorano, P., Carlucci, R., Minerva, M., Capezzuto, F., Sion, L., Tursi, A., 2012. Exploring Relationships between Demersal Resources and Environmental Factors in the Ionian Sea (Central Mediterranean). *J. Mar. Biol.* 27946: 12.
- D’Onghia, G., Maiorano, P., Capezzuto, F., Carlucci, R., Battista, D., Giove, A., Sion, L., Tursi, A., 2009. Further evidences of deep-sea recruitment of *Aristeus antennatus* (Crustacea: Decapoda) and its role in the population renewal on the exploited bottoms of the Mediterranean. *Fish. Res.* 95, 236-245.
- D’Onghia G., Matarrese A., Maiorano P., Perri F., 1998. Valutazione di *Parapenaeus longirostris* (Lucas, 1846) (Crustacea, Decapoda) nel Mar Ionio. *Biologia Marina Mediterranea*, 5: 273-283.
- Froese R., Pauly D., 2019. FishBase, version 05/2019. World Wide Web electronic publication. www.fishbase.org (last accessed 7 November 2019)
- Frogia, C.; 1988. An estimate of growth and mortality parameters for Norway lobster (*Nephrops norvegicus*) in the central Adriatic Sea [1988] FAO, Rome (Italy). General Fisheries Council for the Mediterranean eng; Technical Consultation on Stock Assessment in the Adriatic and Ionian Seas eng 1-5 Jun 1987 5 Bari (Italy); Gramitto, M.E. (Consiglio Nazionale delle Ricerche, Ancona (Italy). *Ist. Ricerche Pesca Marittima*)
- García-Rodríguez, M., Esteban, A., 1999. On the biology and fishery of *Aristeus antennatus* (Risso, 1816), (Decapoda, Dendrobranchiata) in the Ibiza Channel (Balearic Islands, Spain). *Sci. Mar.* 63 (1), 27–37.
- GFCM, 2019. General Fisheries Commission for the Mediterranean (GFCM), 2019. Scientific Advisory Committee for fisheries (SAC) - Working Group on Stock Assessment of Demersal Species (WGSAD).

Benchmark session for the assessment of European hake in GSAs 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19,20, 22,23 and 26FAO headquarters, Rome, Italy, 2–7 December 2019. Report. <http://www.fao.org/gfcm/technical-meetings/en/>

GFCM, 2021. General Fisheries Commission for the Mediterranean (GFCM), 2021. Scientific Advisory Committee on Fisheries (SAC) - Working Group on Stock Assessment of Demersal species (WGSAD). Online, 18–23 January 2021(western Mediterranean). Report. <http://www.fao.org/gfcm/technical-meetings/en/>

Guijarro, B., Massutí, E., Moranta, J., Díaz, P., 2008. Population dynamics of the red shrimp *Aristeus antennatus* in the Balearic Islands (western Mediterranean): Short spatio-temporal differences and influence of environmental factors. *J. Mar. Syst.* 71: 385–402.

Guijarro, B., Bitetto, I., D’Onghia, G., Follesa, M.C., Kaporis, K., Mannini, A., Marković, O., Micallef, R., Ragonese, S., Skarvelis, K., Cau, A., 2019. Spatial and temporal patterns in the Mediterranean populations of *Aristaeomorpha foliacea* and *Aristeus antennatus* (Crustacea: Decapoda: Aristeidae) based on the MEDITS surveys. *Sci. Mar.*, 83S1: 57-70

Hammer, C. Zimmermann, C. 2005. The role of stock identification in formulating fishery management Advice. In *Stock identification methods: Applications in fishery science*. Edited by S.X. Cadrin, K.D. Friedland and J.R. Waldman. Elsevier Academic Press, pp. 631-658.

Hidalgo, M., Rueda, L., Molinero, J.C., Guijarro, B., Massutí, E., 2015. Spatial and temporal variation of seasonal synchrony in the deep-sea shrimp *Aristeus antennatus* in the Western Mediterranean. *J. Mar. Syst.* 148: 131–141.

Hilborn, R., Quinn, T. P., Schindler, D. E., Rogers, D. E. 2003. Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences.* 100: 6564-6568.

Hurtado-Ferro F., Szuwalski C. S., Valero J. L., Anderson S. C., Cunningham C. J., Johnson K. F., Licandeo R., McGilliard C. R., Monnahan C. C., Muradian M. L., Ono K., Vert-Pre K. A., Whitten A. R., Punt A. E., 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES Journal of Marine Science*, Volume 72, Issue 1, January 2015, Pages 99–110, <https://doi.org/10.1093/icesjms/fsu198>

Jardim E, Millar CP, Mosqueira I, Scott F, Osio GC, Ferretti M, et al. (2015) What if stock assessment is as simple as a linear model? The a4a initiative. *ICES Journal of Marine Science* 72: 232–236. [10.1093/icesjms/fsu050](https://doi.org/10.1093/icesjms/fsu050)

Kaporis K., Kasalica O., Klaoudatos D., Djurovic M., 2013. Contribution to the biology of *Parapenaeus longirostris* (Lucas, 1846) in the South Ionian and South Adriatic Sea. *Turkish Journal of Fisheries and Aquatic Sciences*, 13: 647-665.

Kerr, L. A., Cadrin, S. X., Secor, D. H. 2010. The role of spatial dynamics in the stability, resilience, and productivity of an estuarine fish population. *Ecological Applications.* 20: 497-507.

Kerr, L.A., Hintzen, N.T., Cadrin, S.X., Worsøe Clausen, L., Dickey-Collas, M., Goethel, D.R., Hatfield, E.M.C., Kritzer, J.P., & Nash, R.D.M. 2016. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. *ICES Journal of Marine Science*.

Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L. W., Dickey-Collas, M., Goethel, D. R., Hatfield, E. M. C., et al. 2017. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. <https://academic.oup.com/icesjms/article-pdf/doi/10.1093/icesjms/fsw188/19178517/fsw188.pdf> (Accessed 13 September 2019).

Kristensen K., Nielsen A., Berg C.W., Skaug H. J., Bell B., 2015. TMB: automatic differentiation and Laplace approximation. *J. Stat. Softw.* 70, 1–21. <https://doi.org/10.18637/jss.v070.i05>

Massutí, E., Monserrat, S., Oliver, P., Moranta, J., López-Jurado, J.L., Marcos, M., Hidalgo, M., Guijarro, B.,

- Carbonell, A., Pereda, P., 2008. The influence of oceanographic scenarios on the population dynamics of demersal resources in the western Mediterranean: Hypothesis for hake and red shrimp off Balearic Islands. *J. Mar. Syst.* 71: 421–438.
- Maynou, F., 2008. Environmental causes of the fluctuations of red shrimp (*Aristeus antennatus*) landings in the Catalan Sea. *J. Mar. Syst.* 71: 294–302.
- Melaku Canu, D., Laurent, C., Morello, E.B., Querin, S., Scarcella, G., Vrgoc, N., Froggia, C., Angelini, S. and Solidoro, C. (2021), *Nephrops norvegicus* in the Adriatic Sea: Connectivity modeling, essential fish habitats, and management area network. *Fish Oceanogr*, 30: 349-365. <https://doi.org/10.1111/fog.12522>
- Mildenberger T. K., Kokkalis A., Berg C. W., 2019. Guidelines for the stochastic production model in continuous time (SPiCT).
- Mura, I., Cau, A., 1989. Sul dimorfismo sessuale e sex-ratio in *Aristeus antennatus* (Risso, 1816), *Ebalia* XV(2), 811±814.
- Mureno M., Cau A., Colloca F., Sartor P., Fiorentino F., Garofalo G., Piccinetti C., Manfredi C., D’Onghia G., Carlucci R., Donnalioia L., Lembo P., 2010. Mapping the potential locations of European hake nurseries in Italian waters (049-068). International Fishery GIS Society, 2010.
- Papaconstantinou, C., Kapiris, K., 2001. Distribution and population structure of the red shrimp (*Aristeus antennatus*) on an unexploited fishing ground in the Greek Ionian Sea. *Aquat. Living Resourc.* 14, 303-312.
- Pedersen M. W., Berg C. W., 2017. A stochastic surplus production model in continuous time. *Fish Fish.* 18, 226–243. <https://doi.org/10.1111/faf.12174>
- Pella J. J., Tomlinson P. K., 1969. A generalized stock production model. *Bulletin of the Inter-American Tropical Tuna Commission* 13, 421–458.
- Punt, A. E. 2003. The performance of a size-structured stock assessment method in the face of spatial heterogeneity in growth. *Fisheries Research.* 65: 391-409.
- Ragonese, S., 1995. Geographical distribution of *Aristaeomorpha foliacea* (Crustacea-Aristeidae) in the Sicilian Channel (Mediterranean Sea). *ICES J. Mar. Sci. Symp.*, 199: 183-188.
- Recasens L., Chiericoni V., Belcari P., 2008. Spawning pattern and batch fecundity of the European hake (*Merluccius merluccius* (Linnaeus, 1758)) in the western Mediterranean. *Sci. Mar*, 72(4):721-32.
- Sabatini, A., Follesa, M.C., Locci, I., Pendugiu, A.A., Pesci, P., Cau, A., 2007. Assemblages in a submarine canyon: Influence of depth and time. *Hydrobiologia* 580: 265–271.
- Sabatini, A., Follesa, M.C., Locci, I., Matta, G., Palmas, F., Pendugiu, A.A., Pesci, P., Cau, A., 2011. Demersal assemblages in two trawl fishing lanes located on the Baronie Seamount (Central Western Mediterranean). *J. Mar. Biol. Ass. U. K.* 91: 65-75.
- SAMED, 2002. SAMED EU PROJECT nº 99-047, Final Report (2002)
- Sardà, F., Cartes, J.E., Norbis, W., 1994. Spatio-temporal structure of the deep-water shrimp *Aristeus antennatus* (Decapoda: Aristeidae) population in the western Mediterranean. *Fish. Bull.* 92, 599–607.
- Sardà, F., Cartes, J.E., 1997. Morphological features and ecological aspects of early juvenile specimens of the aristeid shrimp *Aristeus antennatus* (Risso, 1816). *Mar. Freshw. Res.* 48, 73–77.
- Sardà, F., Demestre, M., 1987. Estudio biológico de la gamba *Aristeus antennatus* (Risso, 1816) en el Mar Catalán (NE de España). *Inv. Pesq.* 51 (1), 213–232.
- Sardà F., Company, J.B., 2012. The deep-sea recruitment of *Aristeus antennatus* (Risso, 1816)(Crustacea: Decapoda) in the Mediterranean Sea. *J. Mar. Sys.* 105-108, 145-151.
- Sardà F., Company, J.B., Maynou, F., 2003. Deep-sea shrimp *Aristeus antennatus* Risso 1816 in the Catalan Sea: a review and perspectives. *J. Northwest Atl. Fish. Sci.* 31: 127–136.

Secor, D. H. 1999. Specifying divergent migrations in the concept of stock: The contingent hypothesis. *Fisheries Research*. 43: 13-34.

STECF. 2020a. Scientific, Technical and Economic Committee for Fisheries (STECF)- Stock Assessments: demersal stocks in the western Mediterranean Sea (STECF-20-09). EUR 28359 EN, Publications Office of the European Union, Luxembourg, 2020. https://stecf.jrc.ec.europa.eu/documents/43805/1446742/STECF+16-14+-+Methods+for+MED+stock+assessments_JRC102680.pdf.

STECF. 2020b. Scientific, Technical and Economic Committee for Fisheries (STECF) Stock Assessments in the Mediterranean Sea – Adriatic, Ionian and Aegean Seas (STECF-20-15). EUR 28359 EN, Publications Office of the European Union, Luxembourg, 2020. https://stecf.jrc.ec.europa.eu/documents/43805/1446742/STECF+16-14+-+Methods+for+MED+stock+assessments_JRC102680.pdf.

Stephenson, R. L. 1999. Stock complexity in fisheries management: A perspective of emerging issues related to population sub-units. *Fisheries Research*. 43: 247-249.

Tudela, S., Maynou, F., Demestre, M., 2003. Influence of submarine canyons on the distribution of the deep-water shrimp, *Aristeus antennatus* (Risso, 1816) in the NW Mediterranean. *Crustaceana* 76: 217–225.

Ying, Y., Chen, Y., Lin, L., Gao, T. 2011. Risks of ignoring fish population spatial structure in fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*. 68: 2101-2120.



Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201
– SC03- Project MED_UNITS

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP4 - SYNTHESIS AND PROPOSALS

Task 4.2 – Integrating results by different WPs and proposals of new management units

Deliverable 4.2 – Report with the maps of population units and discontinuities

Responsible: Vinko Bandelj (OGS)

Compiled by: Vinko Bandelj (OGS), Fabrizio Gianni (OGS)

24. 8. 2021

Version: Final

Index

Index	2
List of acronyms	5
Executive summary	6
Introduction	8
Objectives	10
Workplan	11
Materials and methods	12
Environmental data set	12
Individual dataset	12
GSA dataset	14
Analysis of species	16
Data	16
Fuzzy clustering	16
Cluster interpretation	17
Spatial variables	17
Gradient analysis and variation partitioning	18
<i>Merluccius merluccius</i>	20
Data	20
Results and discussion	23
Fuzzy clustering	23
Spatial analysis	36
Environmental analysis	50
Variation partitioning and best model selections	58
Conclusions	67
<i>Aristaeomorpha foliacea</i>	70
Data	70
Results and discussion	70
Fuzzy clustering	70
Spatial analysis	73
Environmental analysis	75
Variation partitioning and best model selection	75

Conclusions	77
<i>Aristeus antennatus</i>	78
Data	78
Results and discussion	78
Fuzzy clustering	78
Spatial analysis	80
Environmental analysis	81
Variation partitioning and best model selection	82
Conclusions	82
<i>Nephrops norvegicus</i>	83
Data	83
Results and discussion	83
Fuzzy clustering	83
Spatial analysis	88
Environmental analysis	97
Variation partitioning and best model selection	100
Conclusions	106
<i>Parapenaeus longirostris</i>	108
Data	108
Results and discussion	108
Fuzzy clustering	108
Spatial analysis	111
Environmental analysis	117
Variation partitioning and best model selection	119
Conclusions	124
<i>Mullus barbatus</i>	126
Data	126
Results and discussion	128
Fuzzy clustering	128
Spatial analysis	142
Environmental analysis	158
Variation partitioning and best model selection	166
Conclusions	172
General conclusions	174

List of acronyms

CF - Coefficient of Fuzziness

CMEMS – Copernicus Marine Environment Monitoring Service

DA – Discriminant Analysis

FKM – Fuzzy k-means

GIS – Geographic Information System

GSA – Geographic Subarea

NC – Number of Clusters

MEM – Moran Eigenvector Maps

MPC – Modified Partition Coefficient

PC – Partition Coefficient

PCA – Principal Component Analysis

PE – Partition Entropy

RDA – Redundancy Analysis

SIL – Silhouette Method

SIL.F – Fuzzy Silhouette Method

WP – Workpackage

XB – Xie and Beni Index

Executive summary

The main objectives of Task 4.2 were to identify possible management units for the species under study by combining the data coming from other WPs (genetic data from WP1 and otolith data from WP2) with environmental data collected in this Task and explicit spatial predictors. In particular, the Workplan envisaged the use of fuzzy clustering methods (Bezdek, 1981), explicit spatial predictors, and direct gradient analysis methods such as the Redundancy analysis (RDA) (van den Wollenberg, 1977).

As fuzzy clustering method the fuzzy k-means (FKM), a general case of the well-known partitive algorithm k-means (Bezdek, 1984), was chosen since it provides the needed flexibility: it does identify discontinuities in the data (i.e., clusters), but it also assesses the level of fuzziness in the association of data to clusters. In this way, FKM represents an intermediate solution between clustering methods and gradient analysis methods. As explicit spatial method the Moran Eigenvector Maps (MEM) (Borcard & Legendre, 2002; Dray et al., 2006) were selected since they can model all-scales spatial patterns by computing independent continuous variables from the coordinates of the sampling locations. Both, FKM membership grades, as variables to be explained, and MEMs, as predictors, can be used in RDA along with environmental variables.

The environmental data were collected from the products available on the Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>), in particular from the physical and biogeochemical reanalysis. The data were extracted at the bottom and for the euphotic zone (0-200 m) in the depth range of each species, and several statistics were computed over different time intervals (5 years, 3 years, 1 year). For the needs of Task 4.2 only the median values of the extracted environmental variables were used.

In Task 4.2, with FKM we first searched for possible clusters existing in genetic data (fish and shrimp species), otolith shape and microchemistry data (fish species only), and combinations of them. Internal cluster validation indexes were applied in order to assess if the clusters found were well separated and internally homogeneous. Spatial and environmental variables were used as external validations for the fuzzy clusters obtained: the more variance the spatial and environmental variables were able to explain, the more those clusters do show a spatially structured and environmentally based distribution. Spatial analysis was performed by using depth, geographic coordinates of the sampling locations, and the MEMs computed from the geographic coordinates of the sampling locations. Environmental analysis was performed by using variables over the three-time intervals and the two-depth of extractions. For the clusterizations that were better explained by both, the spatial and the environmental predictors, a variation partitioning procedure (Borcard & Legendre, 1992) and a forward selection procedure (Blanchet et al., 2008) were applied in order to disentangle the mutual relationships between predictors and to select the best final models.

Maps of the distribution of the fuzzy clusters and of the habitat characteristics of each species were produced as GIS layers and used for figure production and for other activities of Task 4.2 (see D.4.3).

For the fish species, where more different datasets were available on which to base the sub-population identification, the results showed that the combination of different data, e.g., the genetic and the otolith shape data, or the otolith shape and the otolith microchemistry data, or all three together, is a sound strategy as it addresses the differences emerging between different stocks due to genetic isolation or habitat characteristics. For both fish species the combined datasets showed strengthened relationships with the spatial and environmental

variables, as if similar patterns in both datasets would more clearly emerge, thus increasing the possibility to detect and interpret the observed differences between sub-populations. The results did show that different numbers of potential management units (stocks) can be identified in the Mediterranean for the different species under study. In particular, for *Merluccius merluccius* three stocks were identified; for *Nephrops norvegicus* four stocks; and for *Parapenaeus longirostris* two stocks. No stock was clearly identified for either *Aristeomorpha foliacea*, *Aristeus antennatus* and *Mullus barbatus*. The three stocks of *M. merluccius* were identified on the combination of the genetic and otolith shape data, and were distributed one in the Eastern Mediterranean, one in the Adriatic-Ionian-Tyrrhenian basins, and one in the Western Mediterranean. For *N. norvegicus* one stock was characterizing the Adriatic Sea and the sampled subareas of the Ionian and Aegean Sea, while the other three were distributed in the Western Mediterranean. For *P. longirostris*, the discontinuity between the eastern and the western stocks was positioned in the Ionian Sea south of Italy. Obviously, these results depend critically on the sampling design, which was based on the GSA subareas. Even if in some GSA subareas there were more than one sampling locations, a much denser sampling design would be needed to correctly and consistently reconstruct possible sub-populations at a scale smaller than that of the GSA subareas.

These same considerations hold true also for the interpretation of the results for *A. foliacea*, *A. antennatus* and *M. barbatus*. For the two shrimps no clusterization did show a recognizable spatial distribution, and the relation with the spatial and environmental variables was low, non-significant and did not show any pattern related to the number of clusters. The only possible conclusion is that for these two species, based on the data produced in this study and with the methods applied here, only one population could be identified in the sampled areas of the Mediterranean. For *M. barbatus* the conclusion is much the same, even though there were some rather inconclusive evidences for the existence of 2 to 3 sub-population in the Mediterranean. Other evidences pointed to the possibility of a divergence of *M. barbatus* populations at scales smaller than those sampled here, but also these were inconclusive since limited by the scale of the sampling design. A more dense sampling design, perhaps restricted to smaller areas, might help in confirming or rejecting this hypothesis.

Introduction

The quest of identifying fish management units needs different information. First, sub-populations of a population must be identified based on either genetic data, or other data for which divergent characteristics in the sub-populations can be observed and described (e.g., for fish otolith shape and otolith microchemistry data). A management unit should also be spatially defined, i.e., if a genetic, morphological or chemical composition divergence is observed in a population, it cannot be considered a possible management unit unless it has a clearly defined spatial distribution. Obviously, genetic differences cannot appear if the sub-populations are not at least to a certain point segregated, either in space, or time, or both. The analysis of possible temporal segregations of sub-populations was not an objective of this study, hence we focused on the spatial patterns present in the genetic, otolith shape and otolith microchemistry data. Furthermore, the environmental characteristics of a specific area may influence the morphology and the chemical composition of the tissues of the individuals.

Thus, in Task 4.2, we first searched for possible clusters existing in genetic data (fish and shrimp species), otolith shape and microchemistry data (fish species only), and combinations of them. The clustering method chosen was the fuzzy k-means (FKM), a general case of the well-known partitive algorithm k-means (Bezdek, 1984), since it provides the needed flexibility: it does identify discontinuities in the data (i.e., clusters), but it also assesses the level of fuzziness in the association of data to clusters. In this way, FKM represents an intermediate solution between clustering methods and gradient analysis methods. Furthermore, FKM membership grades are continuous variables and can be used in Redundancy analysis (RDA) (van den Wollenberg, 1977) to find relationships with possible explanatory variables. We applied several Internal validation methods, which are often applied in order to assess the most appropriate number of clusters, as well as the quality of clustering results (i.e., internal homogeneity and external separation) in terms of the variables used to derive the clusters.

Nevertheless, in this study, we also used spatial and environmental variables as external validations for the obtained fuzzy clusters. We considered that the more variance the spatial and environmental variables were able to explain, the more those clusters did show a spatially structured and environmentally based distribution, thus helping in delineating possible fish management units.

Spatial analysis was performed by using depth, geographic coordinates of the sampling locations, and the Moran Eigenvector Maps (MEMs) (Borcard & Legendre, 2002; Dray et al., 2006) computed from the geographic coordinates of the sampling locations. MEMs are a type of explicit spatial method the Moran Eigenvector Maps (MEM) that can model all-scales spatial patterns (Borcard et al., 2004) by computing independent continuous variables from the coordinates of the sampling locations. Such spatial components can be further used as explanatory variables in an RDA.

The environmental data were collected from the products available on the Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>), in particular from the physical and biogeochemical reanalysis products. The data were extracted at the bottom and for the euphotic zone (0-200 m) in the depth range of each species, and several statistics were computed over different time intervals (5 years, 3 years, 1 year). For the needs of Task 4.2 only the median values of the extracted environmental variables were used. Environmental analysis was performed by using the extracted median variables over the three-time intervals and the two-depth of extractions to explain the variability in the fuzzy clusters obtained from the genetic, the otolith shape, the otolith microchemistry data, or from combinations of them (just for fish species). The best models of spatial and environmental predictors explaining fuzzy cluster memberships were used in a variation partitioning procedure (Borcard et al., 1992) in order to disentangle the mutual

relationships between predictors, and in a forward selection procedure (Blanchet et al., 2008) to select the best final models.

With the combinations of the methods above applied to the data produced in the project we were able to: 1) assess the existence of possible different sub-populations based on genetic, otolith shape and otolith microchemistry data, and inspect their spatial distribution; 2) to quantitatively relate the sub-populations to spatial and environmental variables, determining the best predictors and their relationships; 3) and, finally, to provide information as to the existence and spatial distribution of possible management units for the species under study. In this way all the objectives of Task 4.2 were appropriately addressed.

Objectives

The main objectives of Task 4.2 were to identify possible management units for the species under study by combining the data coming from other WPs (genetic data from WP1 and otolith data from WP2) with environmental data collected in this Task, and with explicit spatial predictors.

Workplan

The main aim of the Task 4.2 is to deliver maps of the delineated population units by species on the basis of biological information (genetic and otolith results) from WPs 1-2. In addition, this task will also combine the results from biological units (WPs 1-2) to those of management units (WP3).

Results of biological analysis (Genetic data produced in WP1, Otolith Shape and Otolith microchemistry in WP2) will be analysed with multivariate analysis techniques also by including environmental variables in order to obtain a division in separate, yet possibly overlapping fish population units. Fuzzy clustering methods (Bezdek, 1981) will be used since they provide a separation of data based on their similarity, yet with also a different membership

linking each sample to each identified cluster/population unit. The main characteristics of each population unit will be described.

The results of fuzzy clustering will be integrated with a node-based seascape analysis through a direct gradient analysis (RDA - Redundancy analysis), in which environmental variables will be used as explanatory factors of the observed spatial distribution of different population units. For this purpose, environmental variables will be collected from available in situ-measurements, relevant satellite observations and modelling products as those provided by Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>).

Spatially explicit methods will be applied to disentangle the singular contribution of environmental and spatial variables, highlighting discontinuities/edge effects.

Appropriate visualization methods are needed to deliver complex multivariate data to a wide community of non-scientific users. Thus, throughout this Task innovative and tailor-made visualization solutions will be pursued in order to facilitate the communication of the relevant results to the stakeholders. Visualization will be based on open source GIS (Geographic Information System) software and will integrate both, the results of clustering of genetic data and the information on the distribution of significant explanatory variables.

Materials and methods

Environmental data set

Individual dataset

Environmental data provided by OGS in MED-Units Task 4.2 were elaborated based on data extracted from the models produced by the E.U. Copernicus Marine Service and available on the Copernicus website (<https://marine.copernicus.eu/>).

Three variables (Table 1) were extracted from the Mediterranean Sea Physics Reanalysis (product identifier MEDSEA_MULTIYEAR_PHY_006_004) from 1987 to 2019 on a grid with $1/24^\circ \times 1/24^\circ$ horizontal resolution and 125 vertical levels of thickness increasing with depth (Escudier et al., 2020). The model assimilates satellite sea surface temperature and sea level and in-situ temperature-salinity profiles.

Table 1: Names and measure units for physical variables extracted from the Mediterranean Sea Physics Reanalysis.

SHORT NAME	VARIABLE DESCRIPTION	MEASURE UNIT
sal	Salinity	psu
T	Temperature	°C
vel	Velocity module	m s ⁻¹

Ten variables (Table 2) were extracted from the Mediterranean Sea Biogeochemistry Reanalysis (product identifier MEDSEA_MULTIYEAR_BGC_006_008) from 1999 to 2019 on a grid with $1/24^\circ \times 1/24^\circ$ horizontal resolution and 125 vertical levels of thickness increasing with depth (Teruzzi et al., 2021). The model assimilates satellite chlorophyll and is driven by physical forcing fields produced as output by the Mediterranean physical model.

Table 2: Names and measure units for biogeochemical variables extracted from the Mediterranean Sea Biogeochemistry Reanalysis.

SHORT NAME	VARIABLE DESCRIPTION	MEASURE UNIT
ALK	Alkalinity	$\mu\text{mol kg}^{-1}$
chl _f	Concentration of chlorophyll (as carbon) in sea water	mg(C) m^{-3}
DIC	Dissolved inorganic carbon	$\mu\text{mol kg}^{-1}$
NH ₄	Mole concentration of ammonium in sea water	mmol m^{-3}
NO ₃	Mole concentration of nitrate in sea water	mmol m^{-3}
O _{2o}	Mole concentration of dissolved molecular oxygen in sea water	mmol m^{-3}
pH	pH	
phy	Concentration of phytoplankton biomass (as carbon) in sea water	mg(C) m^{-3}
PO ₄	Mole concentration of phosphate in sea water	mmol m^{-3}

ppn	Net primary production of biomass per day	mg m ⁻³ day ⁻¹
-----	---	--------------------------------------

Data were extracted as median values over three temporal windows: 5 years (2014-2018), 3 years (2016-2018), and 1 year (2018).

Values were provided both at the bottom and in the euphotic zone. In particular, the bottom values are medians over the last two levels above the bottom according to the internal bathymetry of the model, in order to avoid possible bottom effects. The euphotic zone values are medians over the water column extending from surface to 200 m of depth or to the bottom if the bottom is less than 200 m deep. The reason to provide also the values of the euphotic zone is that secondary production at the bottom depends critically on the primary production in the euphotic zone. In the biogeochemistry model the processes at the bottom might be less well resolved than the processes related to primary production in the euphotic zone. The advice is to use either the biogeochemistry variables of the euphotic zone or the biogeochemistry variables of the bottom, whichever set is better related to the variables to explain, not to mix the two datasets.

The environmental variables were extracted only in areas where bottom depth falls in the depth range of each species. A priori we defined the ranges for each species based on MED-Units Deliverable 3.2 and on expert knowledge from the project. Since some depth values of the sampling locations reported in the MED-Units dataset exceeded these depth ranges, we stretched them to the max/min values of samples for each species. The model bathymetry and the bathymetry reported in the samples' dataset did not always match, due to model resolution and simplification of geometry, and to uncertainties in position and bathymetry recordings. Thus, the definitive extraction range was defined based on the model depths at the coordinates of the sampling points. In Table 3, we report the originally defined ranges, the reported ranges of the samples in the MED-Units samples dataset, and the final extraction range referring to the model bathymetry for each species.

Table 3: Depth ranges for each species.

	SHORT NAME	DEPTH RANGE	SAMPLES DEPTH RANGE	MODEL EXTRACTION RANGE
<i>Merluccius merluccius</i>	HKE	30 - 500 m	15 – 642 m	14 – 507 m
<i>Mullus barbatus</i>	MUT	20 - 300 m	12 – 550 m	11 – 326 m
<i>Nephrops norvegicus</i>	NEP	15 - 800 m	30 – 688 m	22 - 890 m
<i>Parapenaeus longirostris</i>	DPS	20 - 600 m	23 – 709 m	13 – 610 m
<i>Aristaeomorpha foliacea</i>	ARS	400 - 800 m	450 – 744 m	339 – 804 m
<i>Aristeus antennatus</i>	ARA	400 - 800 m	165 - 1037 m	310 – 988 m

No extraction was performed for Atlantic samples: while there are in Copernicus models covering that area, they are different from the models used for the Mediterranean Sea and the extraction of variables from different models could have introduced a bias in the analysis.

The samples in the individual dataset for which coordinates were missing or incongruent, or for which the coordinates were already those of the GSA or Subarea centroid, were associated to the GSA or Subarea centroid coordinates: the number of such cases was limited for all species. The variables calculated for these points were medians over the entire bottom area of the GSA or Subarea included into the extraction range of each species (i.e., not the values associated to the centroid coordinates). All these cases are identified as “GSA Centroid” in the field “Comments” and the GSA or Subarea coordinates are provided in the fields “new_lat” and “new_long”.

Due to inaccuracy in position reporting or to the simplification of the coastline in the Copernicus models, the coordinates of some samples fell on land cells of the models. We moved them to the nearest sea cell and these cases are identified as “coord on land, moved to sea” in the field “Comments” and the extraction coordinates are provided in the fields “new_lat” and “new_long”.

For a few sampling locations (1 for MUT; 1 for DPS; 2 for ARS; 1 for ARA) there was a considerable difference between the reported depth and the model depth. In these cases, the coordinates were moved to the nearest point inside a model cell included into the extraction range of each species. These cases are identified either as “sample too deep: moved to the closest cell within the species' depth range” or as “sample too shallow: moved to the closest cell within the species' depth range” in the field “Comments” and the extraction coordinates are provided in the fields “new_lat” and “new_long”.

For one sampling location of ARA (samples 5c-Aa-01 to 5c-Aa-50), there was no reported depth, while the model depth (2562 m) was outside the range of the species. We calculated the environmental variables as medians over the entire bottom area of the GSA included into the extraction range of ARA. These samples are identified as “sample too deep: considered as a GSA centroid” in the field “Comments” and the GSA or Subarea centroid coordinates are provided in the fields “new_lat” and “new_long”.

The sampling coordinates for NEP samples 20a-Nn-01 to 20a-Nn-48 were too deep in the Gulf of Corinth, an area not covered by the mesh grid of the Copernicus models. Since there were no sea cells nearby in the model where to move the sampling point, we did not provide environmental variables for this point. The records are identified as “coord on land, impossible to move the point to sea: environmental data not extracted” in the field “Comments”.

GSA dataset

The GSA dataset has the same variables, the same time intervals and species-specific depth ranges as the individual dataset.

For each GSA or Subarea, the value of the environmental variables was calculated as the median over the bottom areas of the GSA or Subarea inside the extraction range of each species. Computations of GSA or Subarea centroid coordinates was performed in QGIS in WGS84 coordinate reference.

Analysis of species

Data

Data for the four shrimp species were made up by the most relevant Principal Components (PCA) extracted from the genetic data generated in WP1 (D.1.5.1, D.1.5.2). For the two fish species, also the most relevant PCA extracted from otolith shape data were provided (D.2.4) [, as well as the concentrations of chemical elements found in the otoliths (D.2.5). Thus, for shrimps only one dataset was used in the analysis, while for fish three datasets and their combinations were used in the analysis.

Fuzzy clustering

Fuzzy k-means (FKM) (Bezdek, 1981) was used as the main clustering technique. FKM is a generalization of the hard k-means partitive clustering algorithm (Bezdek, 1984) and provides a membership grade linking each observation to each cluster. Thus, observations are not assigned exclusively to one cluster, as in the so-called hard clustering techniques, but with different memberships to all clusters. In this way a more realistic representation is obtained, without the imposition of arbitrary discontinuities among groups.

Two parameters must be specified by the user: the coefficient of fuzziness (CF) and the number of clusters (NC). The CF controls the dispersion of the membership grades among the clusters: a CF=1 produces a hard clustering result, where each observation is assigned to only one cluster, thus equaling the result of a standard hard k-means algorithm. CF bigger than 1 produce increasingly dispersed membership grades, up to the case where each membership grade is equal to $1/NC$ (i.e., each observation is assigned with an equal probability to each cluster). The CF value at which this happens must be determined empirically for each dataset trying different CF values. Usually, intermediate CF values between 1 and the CF value for which membership grades are equal to $1/NC$ are preferred.

The number of clusters also has to be determined empirically by the user for each dataset, but in this case several internal indexes for clustering quality exist and can be applied in order to help in the choice of the best NC. We explored the results of fuzzy clustering for different values of NC., taking into account also the number of observations to be clustered, since the lower the number of observations, the lower the NC should be. For each clustering result we computed 6 different internal quality indexes: the partition coefficient (PC) (Bezdek, 1974), the partition entropy (PE) (Bezdek, 1981), the modified partition coefficient (MPC) (Davè, 1996), the silhouette (SIL) ([Kaufman & Rousseeuw, 1990](#)), the fuzzy silhouette (SIL.F) (Campello & Hruschka, 2006), and the Xie and Beni index (XB) (Xie & Beni, 1991). While the method for the calculation of each index is different, they all rely on the assumption that a good clustering means internally homogenous clusters that are well separated from other clusters. In Table 4 is indicated for each index the the value reached for an optimal number of clusters.

Table 4: Internal cluster validity indexes applied in this study and the values for an optimal number of clusters.

NAME	ACRONYM	BEST VALUE
PARTITION COEFFICIENT	PC	MAX
PARTITION ENTROPY	PE	MIN

MODIFIED PARTITION COEFFICIENT	MPC	MAX
SILHOUETTE	SIL	MAX
FUZZY SILHOUETTE	SIL.F	MAX
XIE AND BENI	XB	MIN

All FKM computations were performed with 100 random initializations and a maximum number of 10^6 iterations. For FKM and cluster validity indexes computations the fclust package for R (Ferraro et al., 2019) was used.

Cluster interpretation

Since each individual was assigned with the fuzzy k-means algorithm with different memberships to all clusters, we calculated the mean membership grade for each cluster over all the individuals belonging to each GSA Subarea and to each GFCM subregions. Thus, each GSA subarea and GFCM subregion was assigned with a different membership grade, similarly to the fuzzy assignments of each individual, to the clusters. In this way it was possible to assess the spatial characterization of the clusters, as well as the strength of the association of each subarea/subregion to the clusters.

Spatial variables

The spatial variables used in the present study were depth, the (metric) longitude and latitude, and the Moran Eigenvector Maps (MEMs) (Dray et al., 2006). Depth for each individual sample was obtained from the MED_UNITS dataset. In case of missing information for depth, we used the depth of the corresponding coordinate point in the COPERNICUS models. In cases where the coordinates of an individual were those of the centroid of the GSA subarea in which was sampled, we calculated depth associated to the centroid as the median over the extraction range of the species in that GSA subarea.

Before the analysis, the coordinates of the individuals sampling locations were transformed from geographical (WGS84) to projected metric coordinates in the UTM system. As the UTM system expresses coordinates relative to one of the zones in which the Earth is divided, we transformed the geographic coordinates into UTM zone 33N coordinates, since this zone lies in the middle of the 7 UTM zones covering the Mediterranean Sea. All longitude coordinates are thus referred to the central meridian of the UTM 33N zone. The transformation from geographical coordinates to metric was needed in order to correctly calculate distances between sampling points, as requested by spatially explicit methods (e.g., MEMs).

The MEMs (Dray et al., 2006) are a method to explicitly model the spatial signal by decomposing it into a set of orthogonal, i.e., mutually independent, spatial components (Borcard & Legendre, 2002). The components thus represent the whole spatial signal, from the smallest scale (local autocorrelation) to the largest scale (patterns related to climatic zones). Single or groups of MEMs can be later used as explanatory variables along or in contrast to other explanatory variables, such as, e.g., environmental variables. In this way also the spatial structuring of environmental variables, and its scale, can be assessed. Thus, MEMs can serve both, to describe or to control for spatial patterns (Bauman et al., 2018a).

The MEMs are obtained by diagonalization of a doubly centered spatial weighting matrix (Bauman et al., 2018b). The spatial matrix is obtained by the product of a connectivity matrix and a weighting matrix. Different choices exist to select the connectivity and the weighting matrix (Bauman et al., 2018b). The connectivity matrix can be built based on known constraints deriving from the spatial features of the area under study (e.g., landmass preventing direct connections between two aquatic ecosystems), or based on distances. Bauman et al., (2018b), showed that in case of irregular sampling designs the most appropriate way of building the connectivity matrices is by using graph-based connection schemes, such as the Delaunay triangulation, the Gabriel's graph, the relative neighborhood graph, or the minimum spanning tree. As for the weighting matrix it can also be built based on known characteristics of the processes contributing to the spatial shaping (for instance, connections in an area of strong currents might be stronger than in an area of weak currents), but usually specific weighting functions are used, such as the linear function, the concave-down function, or the concave-up function Bauman et al., 2018a; Bauman et al., 2018b].

While MEMs can represent every spatial signal, also large linear gradients, these latter can be more efficiently modelled using longitude and latitude (Borcard et al., 2011)]. Thus, the first step before deriving the MEMs consisted in detrending the datasets, i.e., removing the linear gradient as modelled by longitude and latitude. On the residuals of the spatial model of longitude and latitude the MEMs were then derived. Because of this procedure the MEMs that we derived represented additional independent spatial components, not correlated to longitude and latitude.

We determined the best combination of the connectivity matrix and weighting matrix by exploring for each different dataset and FKM results 84 different spatial weighting matrices obtained by combining four graph-based methods for connectivity definition (Delaunay triangulation, Gabriel's graph, relative neighborhood graph, minimum spanning tree) and three weighting functions (linear function, concave-down function, concave-up function): for the concave-down and the concave-up function 10 values for the coefficient were evaluated (from 0.1 to 1 for the concave-up; from 1 to 10 for the concave-down).

The 84 matrices obtained for each dataset and FKM result from each spatial weighting matrix were subjected to forward selection with a double-stopping criterion to select a sub-list of significant MEMs. The best sub-list of MEMs in terms of adjusted R^2 (Ezekiel, 1930) was then extracted as the relevant spatial components for each dataset and used in further analyses.

For all computations involving MEMs the R packages *adespatial* (Dray et al., 2021) and *vegan* (Oksanen et al., 2020) were used.

Gradient analysis and variation partitioning

Redundancy analysis (RDA) is a wide used statistical method for direct gradient analysis (van den Wollenberg, 1977). A matrix of variables is explained by another matrix of variables, called explanatory variables. In this study RDA was used to assess the explanatory power on each dataset and on each fuzzy clustering results of spatial variables (longitude and latitude, significant MEMs), and of environmental variables. Variation partitioning (Borcard et al.,

1992) was applied in order to understand the shared portion of variance between each set of explanatory variables. Forward selection procedures were applied on the models to obtain parsimonious model of a subset of significant variables. A double-stopping criterion was used for variables selection (Blanchet et al., 2008)

For all computations involving RDA, forward selection and variation partitioning the R packages *adespatial* (Dray et al., 2021) and *vegan* (Oksanen et al., 2020) were used.

Merluccius merluccius

Data

Data for the hake were made up by 20 Principal Components (PCA) (Table 5) extracted from the genetic data(D.1.5.1); 6 PCA (Table 5) extracted from otolith shape data (D.2.4); and concentrations of 7 chemical elements (Table 5) found in the otoliths (D.2.5).

The data from the Atlantic were excluded from the datasets before analysis since no environmental variables were retrieved in this area (see Material and Methods in Environmental data set).

All variables showed different means and standard deviations, both inside each dataset and across different datasets. To prevent variables with highest means and standard deviations to unduly influence the analyses, all variables were standardized before the analysis by subtracting the mean and dividing by the standard deviation.

The number of individuals with complete records for each of the three datasets was different, as it was the number of sampling location, i.e., the spatial coverage of each dataset (Fig. 1). Genetic data and otolith shape data had the highest number of individuals with complete records and the widest spatial coverage. In contrast, the otolith microchemistry data had the lowest number of individuals analyzed and presented the lowest spatial coverage (Fig. 1).

Since the ultimate scope of the Task 4.2 was the integration of the results of different work packages, we also considered the following combinations of the datasets produced respectively by the genetic, otolith shape and otolith microchemistry working groups: a combination of all three datasets; a combination of the genetic and otolith shape variables; and a combination of the otolith shape and otolith microchemistry data. The combination of all three datasets produced the highest number of variables to include in the analysis, but its spatial coverage was limited by that of otolith microchemistry data. The rationale for the combination of the genetic and otolith shape was that both datasets have a wide spatial coverage and the highest numbers of individuals analyzed. It was already shown that the two datasets can corroborate each other (ICES, 2020). The combined, genetic-otolith shape dataset also shows quite a high spatial coverage and number of individuals. The rationale for the combination of the otolith shape and otolith microchemistry data was that both refer to characteristics of the otoliths. Variations in otolith shape of fish from different geographic areas are at least partly expressed during its life history, thus representing a phenotypic measure of stock identification. Otolith chemistry has been used successfully to address issues related to stock identification and fish movements, as the otolith microchemistry is influenced by the physical and chemical properties of the environment. Thus, their combination might give us more insight into possible diversification among populations based on the otoliths, while again its spatial coverage was limited by that of otolith microchemistry data. In Table 5, there also the acronyms by which each dataset or combination of datasets will be referred to from now on.

Table 5: Datasets used in the analysis for Merluccius. Data taken from individuals of the Atlantic Ocean were discarded before the analysis.

DATASET	ACRONYM	N VARIABLES	TYPE OF VARIABLE	COMPLETE RECORDS	SAMPLING LOCATIONS
GENETIC	GEN	20	Principal components	1624	127
OTOLITH SHAPE	OTHO	6	Principal components	1640	126
MICROCHEMISTRY	MICRO	7	Concentrations	218	21
GENETIC + OTOLITH SHAPE	GEN.OTHO	26	Principal components	1368	114
OTOLITH SHAPE + MICROCHEMISTRY	OTHO.MICRO	13	Principal components + Concentrations	193	20
ALL (GENETIC + OTOLITH SHAPE + OTOLITH MICROCHEMISTRY)	ALL	33	Principal components + Concentrations	167	19

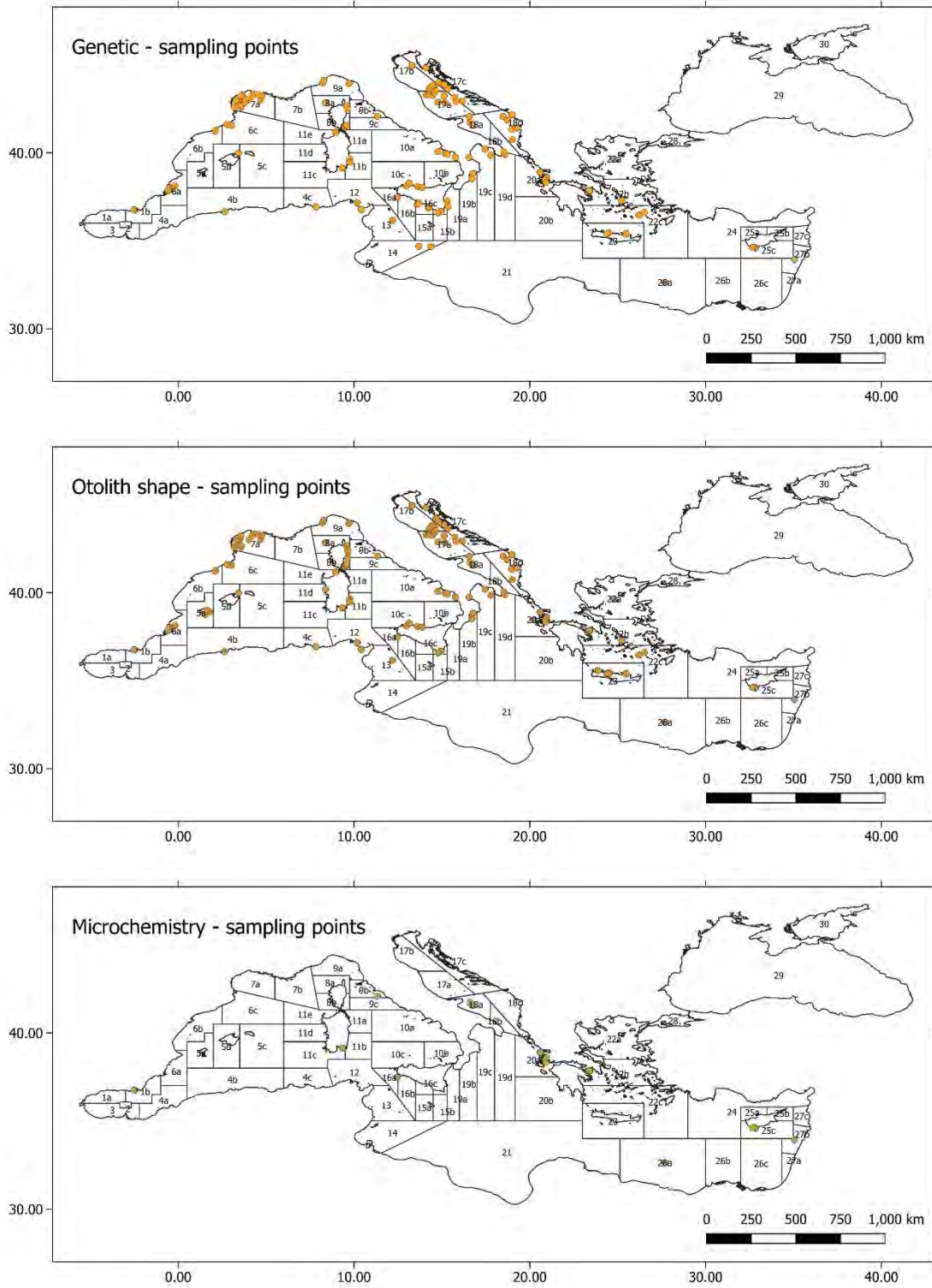


Figure 1: Comparison between the sampling locations for the genetic dataset (upper panel), the otolith shape dataset (middle panel), and the otolith microchemistry dataset (lower panel) of *Merluccius merluccius*.

Results and discussion

Fuzzy clustering

The full results of the internal cluster validity indexes for hake are not presented, but in Table 6 these results are summarized reporting how many times each number of clusters was chosen as the optimal clusterization or the second optimal clusterization by the six internal cluster validity indexes applied.

Table 6: The number of times each NC for each dataset of Merluccius merluccius was selected by the six internal cluster validity indexes applied as the optimal number of clusters or the second optimal number of clusters.

SELECTED	GEN		OTHO		MICRO		GEN.OTHO		OTHO.MICRO		ALL	
	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd
NC = 2	2		3	1	4	1	4		4		4	
NC = 3		2		2		4		3		2		2
NC = 4	1											
NC = 5	1	3		2				2	2	2		2
NC = 6	2	1	3	1	2	1	2	1		2	2	2

The applied internal cluster validity indexes gave quite different results. It seems obvious that the differences in how they are computed influence the results: for instance, the XB consistently choose the highest or the second highest NC as the optimal choice, while PC and PE consistently indicated the lowest NC (i.e., NC = 2) as the best choice. Since it was not the goal of this study to critically compare the performances of different cluster validity indexes, we considered the majority vote criterion, i.e., the NC that most cluster validity indexes choose as best. Furthermore, in several cases the difference in the value of the index between each NC was very low, possibly indicating a lesser amount of diversification in these datasets.

The GEN dataset showed the less clear results, with 2 and 6 clusters both chosen by two indexes, and 4 and 5 clusters chosen once each. For all other datasets, NC = 2 and NC = 6 seem clearly the best choices, but for the OTHO.MICRO dataset, for which NC = 2 was the best option for 4 validity indexes, while NC = 5 was the best option for 2 validity indexes. These results indicate that the clusterization on different datasets might give different results, i.e., that there is no clear internal structure common to all datasets, due to different processes that influence their variability. Nevertheless, NC = 2 seems a fairly good “least common denominator” that it is worth of considering for all datasets. NC = 3, while never chosen as the optimal choice, is the most chosen second best choice. Given that we expect a low number of clusters, NC = 3 is also worth investigating as to the spatial characterization of the ensuing clusters. Finally, NC = 6 for most datasets might be also worth considering. In any case, the existence of a cluster structure in the examined datasets does not guarantee that this structure has a spatial or environmental rationale, nor that this is the same across different datasets for the same NC.

For the GEN dataset the membership grades of individuals grouped by GFCM subregion at NC = 2 showed a separation between West-Central Mediterranean and East-Adriatic Mediterranean (Table 7). Nevertheless, this picture blurred somewhat looking at the results per GSA subarea (Table 8). In fact, three subareas of the West Mediterranean (11b, 11c, 11e)

joined the cluster with most East and Adriatic subarea, as did subareas 19b, 19c, 19d, encompassing much of the Ionian. Individuals of subarea 19a showed the same average membership for both clusters. At NC = 3 the division between the West-Central and the East-Adriatic Mediterranean remained largely unchanged (but for 9b joining the East-Adriatic cluster), while a new cluster grouped 3 subareas of the Ionian (19a, 19b, 19c), three coastal subareas of the Eastern Mediterranean (22a, 26a, 27b) and one Adriatic (17b). At NC = 6 the Eastern Mediterranean subareas and the Adriatic subareas together with most of the Ionian subareas (19b, 19c, 19d) were clearly separated from the rest and made up respectively cluster 6 and cluster 2 (this latter joined also by western subarea 11c). Cluster 1 was mainly found in the West-Central basin, but only subarea 9b showed a preference for it. Cluster 4 was the main West-Central cluster, with some subareas preferring cluster 3 (mostly those along the African coast, 4b, 4c, 12, 14, plus 11e and 10a). Cluster 5 was much dispersed across the basin, with only subareas 19a and 20 in the Central Mediterranean showing a high average membership grade for it.

Table 7: Spatial interpretation of fuzzy clusters results for the GEN dataset of *Merluccius merluccius* considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

GEN	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
West	0.444	0.556	0.378	0.295	0.326	0.151	0.138	0.196	0.234	0.163	0.118
Center	0.469	0.531	0.368	0.312	0.320	0.128	0.179	0.203	0.216	0.171	0.103
Adriatic	0.567	0.433	0.296	0.335	0.369	0.106	0.331	0.163	0.133	0.170	0.098
East	0.600	0.400	0.263	0.360	0.376	0.076	0.124	0.125	0.096	0.163	0.416

Table 8: Spatial interpretation of fuzzy clusters results for the GEN dataset of *Merluccius merluccius* considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

GEN		2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
1b	West	0.443	0.557	0.390	0.309	0.302	0.196	0.168	0.122	0.208	0.168	0.139
4b	West	0.414	0.586	0.376	0.322	0.302	0.187	0.128	0.243	0.192	0.187	0.063
4c	West	0.393	0.607	0.443	0.303	0.254	0.124	0.157	0.228	0.205	0.183	0.103
5b	West	0.397	0.603	0.412	0.298	0.289	0.114	0.098	0.198	0.288	0.197	0.105
6a	West	0.400	0.600	0.438	0.241	0.321	0.144	0.137	0.175	0.293	0.140	0.111
6b	West	0.415	0.585	0.398	0.276	0.326	0.187	0.155	0.209	0.227	0.149	0.073
6c	West	0.423	0.577	0.382	0.273	0.344	0.218	0.086	0.137	0.338	0.137	0.083
7a	West	0.434	0.566	0.397	0.276	0.327	0.071	0.112	0.198	0.332	0.159	0.128
8a	West	0.468	0.532	0.379	0.278	0.343	0.140	0.129	0.187	0.260	0.158	0.125
8b	West	0.383	0.617	0.533	0.154	0.313	0.116	0.201	0.125	0.416	0.047	0.094
9a	West	0.402	0.598	0.415	0.290	0.295	0.118	0.104	0.226	0.275	0.138	0.139
9b	West	0.479	0.521	0.324	0.301	0.374	0.214	0.080	0.176	0.191	0.181	0.158
10a	West	0.416	0.584	0.393	0.335	0.272	0.134	0.111	0.234	0.163	0.171	0.187
10c	West	0.451	0.549	0.381	0.333	0.286	0.082	0.128	0.222	0.228	0.204	0.137
11b	West	0.507	0.493	0.341	0.257	0.402	0.201	0.173	0.181	0.202	0.123	0.120
11c	West	0.528	0.472	0.325	0.301	0.374	0.184	0.215	0.157	0.195	0.136	0.113
11e	West	0.503	0.497	0.297	0.336	0.368	0.087	0.199	0.233	0.173	0.203	0.105

12	Center	0.407	0.593	0.428	0.267	0.305	0.137	0.124	0.244	0.238	0.141	0.115
13	Center	0.494	0.506	0.340	0.319	0.341	0.057	0.226	0.169	0.271	0.200	0.077
14	Center	0.392	0.608	0.448	0.290	0.262	0.191	0.116	0.244	0.218	0.116	0.114
16b	Center	0.473	0.527	0.345	0.310	0.345	0.192	0.141	0.183	0.211	0.164	0.109
16c	Center	0.433	0.567	0.397	0.299	0.304	0.074	0.122	0.248	0.257	0.177	0.122
16d	Center	0.413	0.587	0.399	0.222	0.378	0.073	0.114	0.191	0.411	0.106	0.105
17a	Adriatic	0.576	0.424	0.312	0.330	0.358	0.102	0.370	0.159	0.122	0.152	0.095
17b	Adriatic	0.557	0.443	0.289	0.389	0.322	0.068	0.323	0.157	0.146	0.221	0.085
17c	Adriatic	0.578	0.422	0.287	0.314	0.399	0.045	0.425	0.146	0.111	0.170	0.103
18a	Adriatic	0.603	0.397	0.265	0.304	0.431	0.180	0.332	0.151	0.123	0.117	0.096
18c	Adriatic	0.518	0.482	0.316	0.335	0.349	0.101	0.229	0.197	0.158	0.203	0.112
19a	Center	0.500	0.500	0.321	0.380	0.298	0.129	0.115	0.185	0.206	0.240	0.124
19b	Center	0.540	0.460	0.320	0.357	0.323	0.126	0.336	0.138	0.168	0.156	0.075
19c	Center	0.581	0.419	0.299	0.364	0.337	0.088	0.269	0.254	0.107	0.190	0.092
19d	Center	0.630	0.370	0.211	0.253	0.536	0.060	0.321	0.171	0.230	0.100	0.118
20a	Center	0.441	0.559	0.389	0.332	0.279	0.147	0.192	0.180	0.166	0.228	0.087
22a	East	0.619	0.381	0.247	0.399	0.354	0.082	0.169	0.161	0.091	0.191	0.306
22b	East	0.550	0.450	0.306	0.327	0.367	0.124	0.187	0.199	0.101	0.167	0.221
22c	East	0.615	0.385	0.210	0.379	0.412	0.075	0.106	0.106	0.150	0.221	0.343
23	East	0.672	0.328	0.232	0.328	0.440	0.034	0.173	0.085	0.093	0.159	0.456
25c	East	0.625	0.375	0.229	0.363	0.409	0.089	0.055	0.102	0.106	0.124	0.524
26a	East	0.531	0.469	0.348	0.361	0.291	0.060	0.062	0.080	0.071	0.131	0.595
27b	East	0.559	0.441	0.294	0.359	0.347	0.062	0.081	0.127	0.052	0.136	0.542

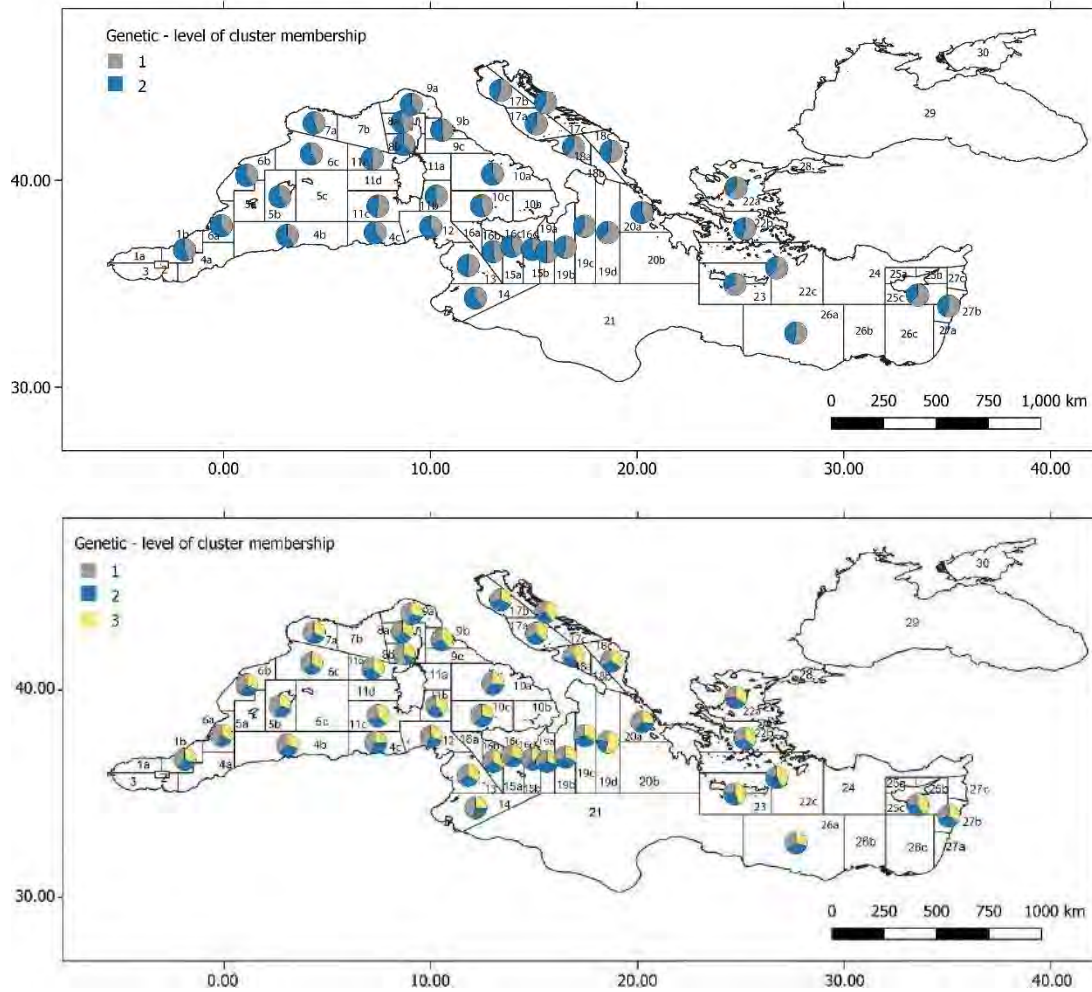


Figure 2: Pie charts with average membership grades per GSA subarea for the GEN dataset of *Merluccius merluccius* at NC = 2 (upper panel) and NC = 3 (lower panel).

The OTHO dataset on the contrary showed a first division between the East-Central Mediterranean and the Adriatic-West Mediterranean (Table 9). Nevertheless, looking at the single subarea average memberships (Table 10), the situation was less clear and no clear spatial pattern in terms of GSA subareas in the membership grades seemed to emerge. At NC = 3 and NC = 6 the situation was similar and clearly the GFCM subregions did not seem to match the pattern emerged from the OTHO dataset. This suggest that other factors might influence the emergence of the observed pattern in OTHO data, such as possibly bottom depth or environmental characteristics of the sampled areas.

Table 9: Spatial interpretation of fuzzy clusters results for the OTHO dataset of *Merluccius merluccius* considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

OTHO	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
West	0.575	0.425	0.258	0.406	0.336	0.192	0.129	0.096	0.147	0.204	0.232
Center	0.469	0.531	0.371	0.297	0.332	0.182	0.182	0.156	0.178	0.145	0.156
Adriatic	0.572	0.428	0.248	0.385	0.367	0.222	0.122	0.128	0.120	0.177	0.231
East	0.491	0.509	0.290	0.316	0.395	0.219	0.105	0.206	0.158	0.137	0.175

Table 10: Spatial interpretation of fuzzy clusters results for the OTHO dataset of *Merluccius merluccius* considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

OTHO		2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
1b	West	0.377	0.623	0.500	0.258	0.242	0.110	0.240	0.185	0.233	0.110	0.122
4b	West	0.577	0.423	0.183	0.333	0.484	0.361	0.123	0.091	0.098	0.141	0.187
4c	West	0.399	0.601	0.380	0.258	0.362	0.204	0.228	0.179	0.161	0.071	0.157
5a	West	0.402	0.598	0.329	0.239	0.432	0.226	0.284	0.119	0.121	0.177	0.073
5b	West	0.462	0.538	0.263	0.263	0.475	0.294	0.123	0.169	0.174	0.091	0.150
6a	West	0.294	0.706	0.541	0.170	0.289	0.151	0.422	0.115	0.170	0.060	0.082
6b	West	0.573	0.427	0.267	0.421	0.312	0.197	0.108	0.076	0.150	0.206	0.263
6c	West	0.622	0.378	0.225	0.417	0.358	0.168	0.109	0.076	0.173	0.287	0.187
7a	West	0.456	0.544	0.447	0.335	0.219	0.108	0.196	0.120	0.228	0.231	0.118
8a	West	0.485	0.515	0.318	0.299	0.382	0.239	0.226	0.120	0.111	0.111	0.193
8b	West	0.427	0.573	0.460	0.302	0.238	0.114	0.194	0.070	0.299	0.069	0.255
9a	West	0.717	0.283	0.165	0.564	0.271	0.089	0.048	0.077	0.069	0.563	0.154
9b	West	0.652	0.348	0.187	0.477	0.335	0.204	0.079	0.070	0.120	0.194	0.332
10a	West	0.735	0.265	0.126	0.560	0.314	0.185	0.066	0.047	0.088	0.277	0.336
10c	West	0.664	0.336	0.147	0.431	0.422	0.286	0.048	0.067	0.125	0.206	0.268
11b	West	0.726	0.274	0.134	0.552	0.314	0.175	0.051	0.052	0.087	0.342	0.293
11c	West	0.741	0.259	0.134	0.592	0.273	0.147	0.047	0.037	0.112	0.243	0.414
11d	West	0.445	0.555	0.407	0.310	0.283	0.148	0.172	0.095	0.280	0.102	0.203
11e	West	0.579	0.421	0.276	0.440	0.284	0.164	0.109	0.118	0.182	0.122	0.305
12	Center	0.356	0.644	0.516	0.193	0.291	0.181	0.338	0.177	0.123	0.075	0.106
13	Center	0.341	0.659	0.520	0.209	0.271	0.138	0.241	0.226	0.195	0.116	0.084
16b	Center	0.602	0.398	0.169	0.373	0.457	0.265	0.042	0.066	0.215	0.153	0.259
16d	Center	0.667	0.333	0.144	0.445	0.411	0.292	0.054	0.090	0.087	0.138	0.339
17a	Adriatic	0.620	0.380	0.219	0.434	0.347	0.205	0.110	0.087	0.111	0.251	0.236
17b	Adriatic	0.499	0.501	0.301	0.307	0.392	0.247	0.169	0.120	0.148	0.139	0.178
17c	Adriatic	0.486	0.514	0.311	0.353	0.336	0.180	0.157	0.256	0.090	0.091	0.227
18a	Adriatic	0.531	0.469	0.305	0.338	0.357	0.205	0.162	0.101	0.141	0.198	0.194
18c	Adriatic	0.635	0.365	0.170	0.428	0.402	0.265	0.054	0.154	0.111	0.121	0.296
19b	Center	0.382	0.618	0.495	0.273	0.232	0.094	0.175	0.207	0.290	0.097	0.137
19c	Center	0.518	0.482	0.312	0.342	0.345	0.170	0.121	0.098	0.245	0.206	0.159
19d	Center	0.646	0.354	0.194	0.444	0.362	0.145	0.077	0.041	0.249	0.282	0.206
20a	Center	0.520	0.480	0.300	0.341	0.359	0.184	0.132	0.197	0.135	0.216	0.136
22a	East	0.478	0.522	0.316	0.320	0.364	0.229	0.120	0.139	0.168	0.110	0.234
22b	East	0.645	0.355	0.172	0.485	0.344	0.190	0.045	0.113	0.182	0.164	0.307
22c	East	0.609	0.391	0.199	0.406	0.395	0.234	0.062	0.141	0.113	0.235	0.215
23	East	0.551	0.449	0.263	0.371	0.365	0.184	0.098	0.149	0.189	0.189	0.191
25b	East	0.404	0.596	0.249	0.191	0.560	0.340	0.077	0.281	0.118	0.094	0.090
26a	East	0.373	0.627	0.377	0.216	0.407	0.221	0.193	0.244	0.153	0.070	0.119
27b	East	0.400	0.600	0.444	0.250	0.306	0.120	0.139	0.374	0.178	0.098	0.091

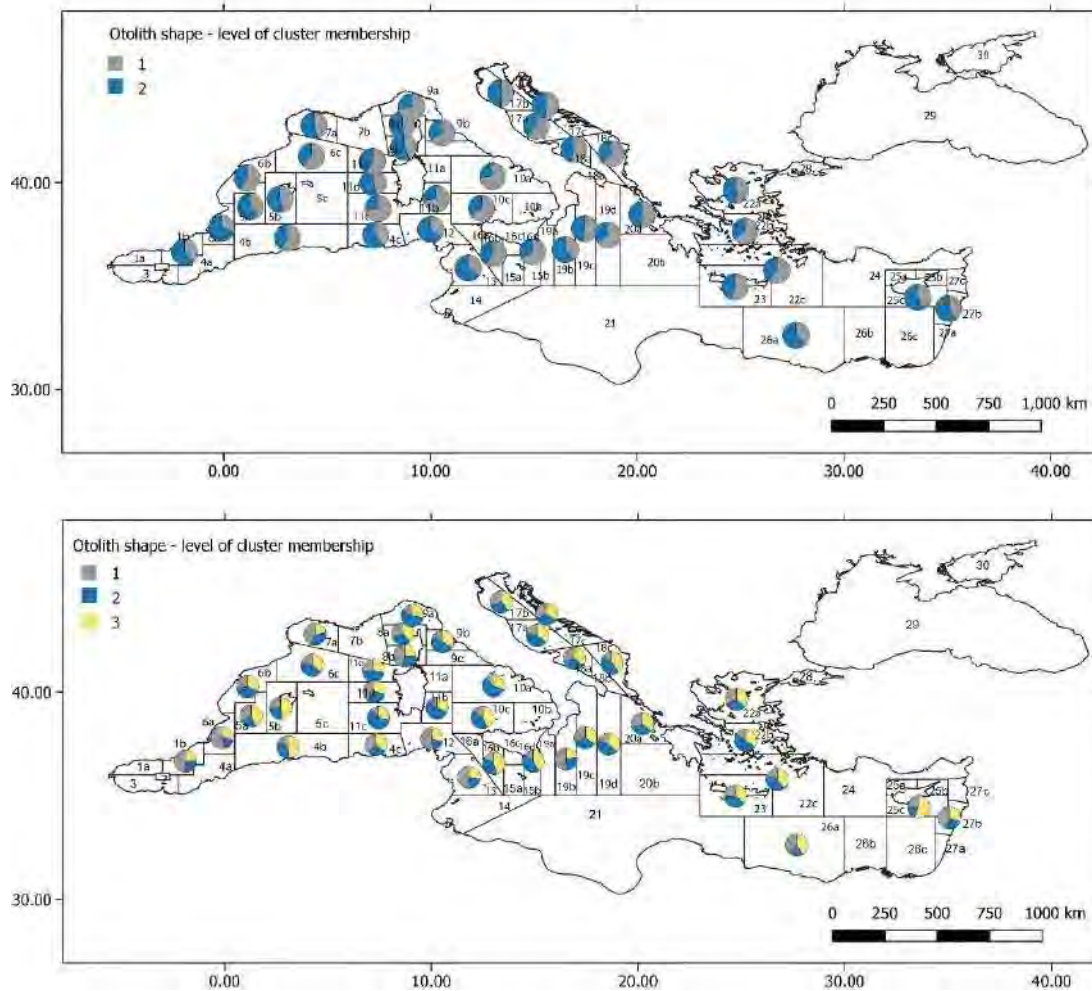


Figure 3: Pie charts with average membership grades per GSA subarea for the OTHO dataset of *Merluccius merluccius* at NC = 2 (upper panel) and NC = 3 (lower panel).

The MICRO dataset is the one with the lowest number of sampling locations, 21, covering 10 GSA subareas: 3 in the Western Mediterranean, 1 in the Adriatic Sea, 2 in Central Mediterranean and 4 in the Eastern basin. At NC = 2 the division is between the West-Central Mediterranean and the East-Adriatic basin, similarly to what happened for the GEN dataset (Table 11). At the level of subareas (Table 12) in the first group are assigned not only all western subareas, but also two eastern (25c, 26a). The two central subareas were assigned one to each cluster (16b along with western subareas to cluster 1, 20a along with the Adriatic subarea and the eastern subareas 22b and 27 b to cluster 2). At NC = 3 subareas 1b, 25c and 26a made a cluster by their own, while cluster 2 remained a western cluster and cluster 3 an Adriatic and eastern cluster. The different behavior of subareas 16b and 20a might support the idea that there is a difference between the western and eastern part of the Mediterranean in the hake population, while samples from subareas 1b, 25c and 26a should be investigated as to their assignment to the same group. Given the low number of sampled subareas, there is not much sense in discussing partitions with higher numbers of NC.

Table 11: Spatial interpretation of fuzzy clusters results for the MICRO dataset of *Merluccius merluccius* considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

MICRO	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5
West	0.731	0.269	0.411	0.440	0.150	0.139	0.216	0.368	0.065	0.055	0.157
Center	0.552	0.448	0.202	0.452	0.347	0.070	0.181	0.136	0.290	0.041	0.283
Adriatic	0.388	0.612	0.151	0.338	0.511	0.182	0.057	0.219	0.321	0.026	0.195
East	0.433	0.567	0.531	0.123	0.345	0.418	0.133	0.236	0.144	0.048	0.021

Table 12: Spatial interpretation of fuzzy clusters results for the MICRO dataset of *Merluccius merluccius* considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

MICRO	AREA	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
		GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5
1b	West	0.679	0.321	0.463	0.367	0.170	0.144	0.172	0.340	0.059	0.148	0.137
9b	West	0.807	0.193	0.427	0.488	0.085	0.138	0.230	0.438	0.036	0.008	0.150
11c	West	0.706	0.294	0.342	0.465	0.194	0.134	0.246	0.328	0.101	0.008	0.183
16b	Center	0.785	0.215	0.188	0.727	0.086	0.033	0.182	0.216	0.038	0.032	0.500
18a	Adriatic	0.388	0.612	0.151	0.338	0.511	0.182	0.057	0.219	0.321	0.026	0.195
20a	Center	0.298	0.702	0.217	0.152	0.631	0.110	0.180	0.049	0.565	0.050	0.046
22b	East	0.263	0.737	0.441	0.090	0.469	0.537	0.045	0.143	0.177	0.087	0.012
25c	East	0.584	0.416	0.597	0.185	0.218	0.309	0.254	0.261	0.082	0.069	0.025
26a	East	0.667	0.333	0.715	0.156	0.129	0.323	0.195	0.394	0.046	0.012	0.030
27b	East	0.298	0.702	0.415	0.077	0.507	0.446	0.082	0.188	0.253	0.008	0.023

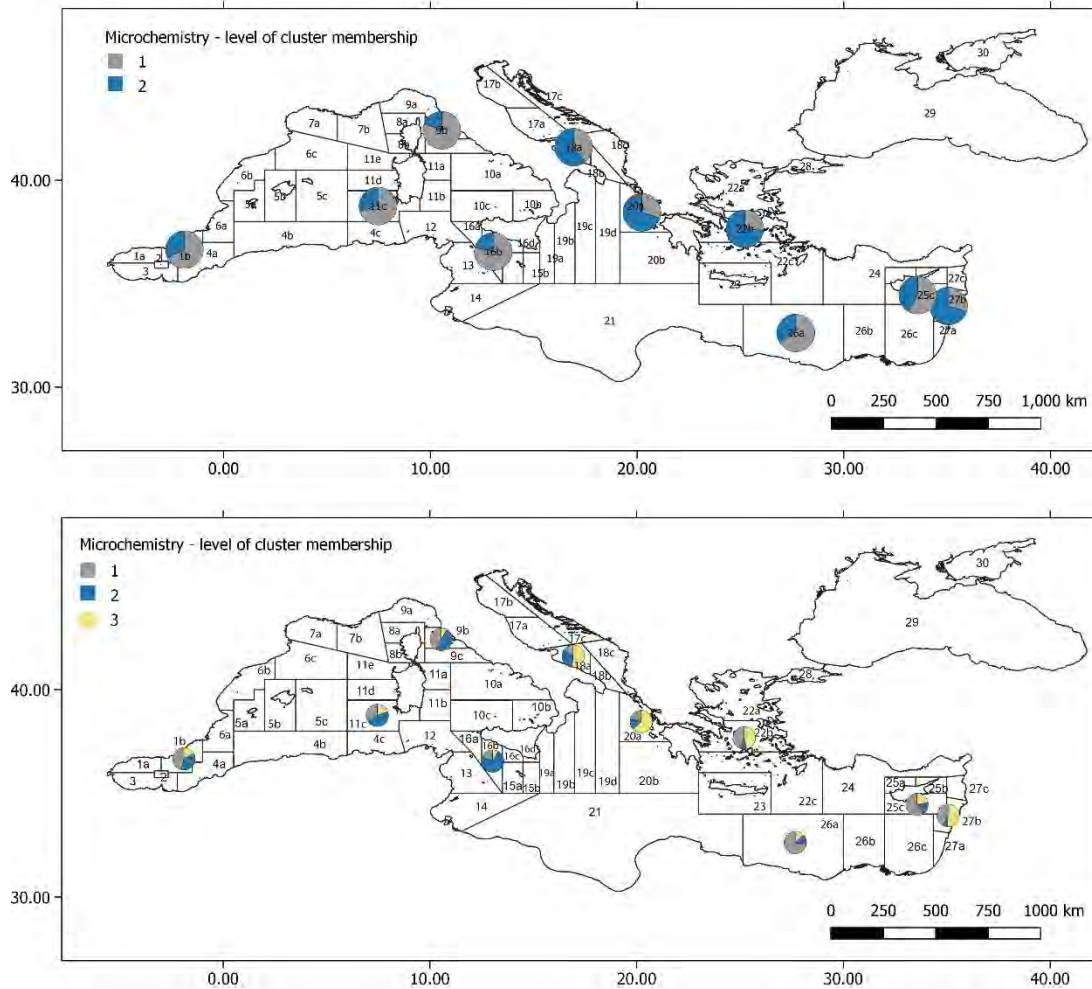


Figure 4: Pie charts with average membership grades per GSA subarea for the MICRO dataset of *Merluccius merluccius* at NC = 2 (upper panel) and NC = 3 (lower panel).

The GEN.OTHO dataset offered the highest number of variables considered together at the highest possible spatial coverage. At the level of GFCM subregions (Table 13), at NC = 2 the West Mediterranean and the Adriatic Sea were grouped together in one cluster, while the Eastern Mediterranean was assigned to the other cluster. The subregion of the Central Mediterranean was assigned equally to the two clusters. At the level of subareas (Table 14), the majority of Western Mediterranean subareas, all Adriatic subareas and the majority of Central Mediterranean subareas were assigned to cluster 1. High memberships for cluster 2 had all but one Eastern Mediterranean subareas (namely 22b), and some subareas along Spanish (1b, 6a, 5b), French (7a, 8a) and African coasts (4c, 12, 13). At NC = 3 the Eastern Mediterranean subregion characterized cluster 2, the Central Mediterranean subregion characterized cluster 3, and the Western and Adriatic characterized cluster 1. As for subareas, cluster 1 still represented most Western, Adriatic and Central subareas, but cluster 2 only grouped Eastern Mediterranean subareas, while all other subareas from Western and Central Mediterranean that at NC = 2 were grouped together with the Eastern, now split into cluster 3. Thus, it seems that NC = 3 might have a better spatial rationale, and that the main difference in the GEN.OTHO data is between the East Mediterranean on one side, and all the

rest subareas on the other. At a finer level, there seem to exist a differentiation in this latter group of subareas that needs to be investigated further. At NC = 6 the Eastern Mediterranean subareas were still clearly separated from the rest (FKM6). Also, the “coastal” Western and Central subareas group was still recognizable in FKM1 (indicating that this group is internally homogeneous), while the Adriatic subareas were split between FKM4 and FKM5, and the other Western subareas among FKM2, FKM3, FKM4 and FKM5.

Table 13: Spatial interpretation of fuzzy clusters results for the combined GEN.OTHO dataset of Merluccius merluccius considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

GEN.OTHO	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
West	0.54	0.46	0.37	0.27	0.36	0.18	0.17	0.20	0.19	0.18	0.08
Center	0.50	0.50	0.33	0.30	0.37	0.22	0.15	0.17	0.18	0.18	0.10
Adriatic	0.55	0.45	0.37	0.28	0.34	0.14	0.15	0.18	0.21	0.23	0.09
East	0.41	0.59	0.27	0.49	0.25	0.10	0.10	0.14	0.14	0.14	0.37

Table 14: Spatial interpretation of fuzzy clusters results for the combined GEN.OTHO dataset of Merluccius merluccius considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

GEN.OTHO	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS						
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
1b	West	0.429	0.571	0.260	0.331	0.409	0.337	0.172	0.125	0.121	0.135	0.111
4b	West	0.521	0.479	0.345	0.284	0.371	0.162	0.236	0.206	0.193	0.123	0.080
4c	West	0.464	0.536	0.291	0.325	0.384	0.258	0.145	0.171	0.167	0.146	0.113
5b	West	0.471	0.529	0.317	0.339	0.343	0.198	0.134	0.176	0.194	0.183	0.115
6a	West	0.423	0.577	0.241	0.332	0.428	0.416	0.116	0.114	0.125	0.100	0.130
6b	West	0.592	0.408	0.404	0.222	0.374	0.177	0.211	0.187	0.171	0.212	0.043
6c	West	0.572	0.428	0.395	0.249	0.356	0.146	0.207	0.247	0.169	0.166	0.065
7a	West	0.470	0.530	0.309	0.311	0.381	0.335	0.092	0.170	0.142	0.159	0.102
8a	West	0.494	0.506	0.322	0.310	0.368	0.227	0.161	0.223	0.142	0.145	0.101
8b	West	0.524	0.476	0.296	0.233	0.471	0.288	0.136	0.273	0.116	0.132	0.054
9a	West	0.590	0.410	0.413	0.243	0.344	0.123	0.144	0.294	0.169	0.185	0.086
9b	West	0.579	0.421	0.418	0.261	0.321	0.131	0.214	0.184	0.220	0.176	0.075
10a	West	0.534	0.466	0.373	0.302	0.325	0.115	0.175	0.209	0.185	0.218	0.098
10c	West	0.568	0.432	0.398	0.262	0.340	0.130	0.150	0.226	0.215	0.189	0.089
11b	West	0.591	0.409	0.418	0.241	0.341	0.115	0.215	0.172	0.231	0.210	0.057
11c	West	0.615	0.385	0.433	0.216	0.351	0.111	0.234	0.193	0.213	0.200	0.048
11e	West	0.602	0.398	0.431	0.236	0.333	0.139	0.103	0.165	0.267	0.250	0.076
12	Center	0.437	0.563	0.268	0.344	0.388	0.323	0.137	0.136	0.142	0.127	0.135
13	Center	0.455	0.545	0.276	0.308	0.416	0.318	0.097	0.148	0.163	0.153	0.121
16b	Center	0.553	0.447	0.379	0.270	0.351	0.121	0.217	0.189	0.223	0.180	0.070
16d	Center	0.572	0.428	0.394	0.264	0.343	0.115	0.131	0.175	0.259	0.242	0.079
17a	Adriatic	0.558	0.442	0.377	0.267	0.356	0.134	0.150	0.193	0.192	0.249	0.082
17b	Adriatic	0.537	0.463	0.346	0.276	0.377	0.183	0.151	0.195	0.177	0.199	0.095
17c	Adriatic	0.506	0.494	0.344	0.330	0.326	0.162	0.073	0.171	0.224	0.233	0.137
18a	Adriatic	0.553	0.447	0.382	0.274	0.344	0.142	0.200	0.164	0.207	0.204	0.083

18c	Adriatic	0.560	0.440	0.395	0.286	0.319	0.116	0.130	0.188	0.234	0.228	0.103
19b	Center	0.482	0.518	0.303	0.304	0.393	0.270	0.143	0.139	0.158	0.178	0.112
19c	Center	0.522	0.478	0.354	0.299	0.348	0.159	0.141	0.190	0.173	0.235	0.101
19d	Center	0.625	0.375	0.431	0.204	0.365	0.107	0.150	0.225	0.225	0.248	0.045
20a	Center	0.584	0.416	0.414	0.251	0.336	0.122	0.142	0.245	0.188	0.219	0.083
22a	East	0.430	0.570	0.282	0.442	0.276	0.129	0.120	0.151	0.160	0.160	0.279
22b	East	0.542	0.458	0.391	0.336	0.273	0.076	0.137	0.192	0.191	0.236	0.168
22c	East	0.496	0.504	0.352	0.399	0.249	0.081	0.115	0.182	0.212	0.192	0.217
23	East	0.384	0.616	0.242	0.517	0.242	0.114	0.069	0.115	0.135	0.157	0.409
25c	East	0.339	0.661	0.200	0.589	0.211	0.073	0.086	0.147	0.083	0.091	0.519
26a	East	0.322	0.678	0.185	0.591	0.224	0.115	0.080	0.094	0.089	0.070	0.552
27b	East	0.334	0.666	0.195	0.555	0.249	0.132	0.076	0.086	0.078	0.084	0.544

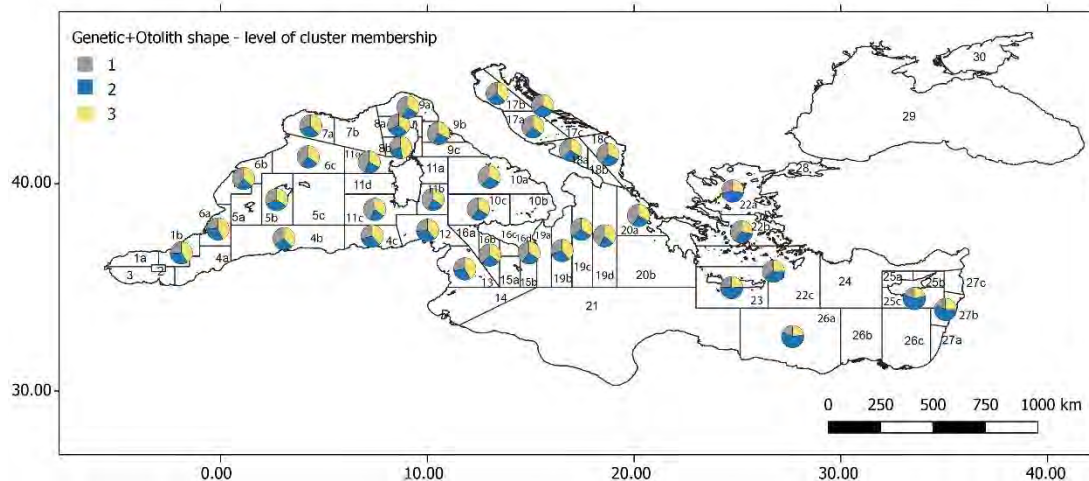
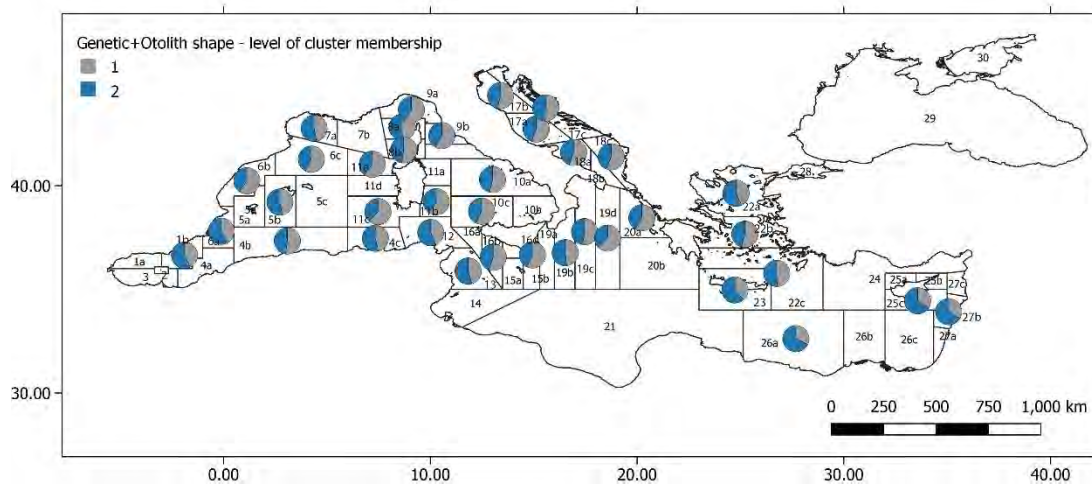


Figure 5: Pie charts with average membership grades per GSA subarea for the GEN.OTHO dataset of *Merluccius merluccius* at NC = 2 (upper panel) and NC = 3 (lower panel).

The OTHO.MICRO dataset is interesting since it groups together all variables related to different characteristics of the otoliths, but with reduced spatial coverage due to the reduced spatial coverage of the MICRO dataset. Nevertheless, it seems that the MICRO variables are driving the differentiation in the dataset, since Western and Central Mediterranean group together, and Adriatic Sea and Eastern Mediterranean together, as happened also in the clusterization based only on MICRO variables (Table 15). At NC = 3 the

Adriatic split from the Eastern Mediterranean subregions, again as for the MICRO dataset. At the level of subareas (Table 16), all Western and the westernmost Central subarea were assigned with high average membership grades to FKM2. Subarea 26a was also assigned to this cluster, but the difference in membership assignment to the two clusters is very small. The single Adriatic subarea (18a), the easternmost subarea of Central Mediterranean, and the remaining Eastern basin subareas (22b, 25c, 27b) were assigned to FKM1. At NC = 3 western subarea 1b joined two eastern subareas (25c, 26a) in FKM3, while the other two eastern subareas (22b, 27b) joined 18a and 20a in FKM2. The other two western subareas and the Central Mediterranean 16b subarea showed a preference for FKM1. No clear spatial pattern seemed to emerge at this NC, as well as at NC = 6.

Table 15: Spatial interpretation of fuzzy clusters results for the combined OTHO.MICRO dataset of *Merluccius merluccius* considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

OTHO.MICRO	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
West	0.345	0.655	0.201	0.457	0.342	0.242	0.069	0.209	0.051	0.192	0.237
Center	0.386	0.614	0.303	0.490	0.207	0.180	0.242	0.326	0.030	0.111	0.111
Adriatic	0.507	0.493	0.401	0.356	0.243	0.206	0.169	0.181	0.087	0.220	0.138
East	0.634	0.366	0.336	0.175	0.489	0.140	0.098	0.049	0.294	0.187	0.232

Table 16: Spatial interpretation of fuzzy clusters results for the combined OTHO.MICRO dataset of *Merluccius merluccius* considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

OTHO.MICRO	AREA	GFCM	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
			FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
1b	West		0.476	0.524	0.278	0.324	0.398	0.167	0.074	0.157	0.095	0.275	0.231
9b	West		0.281	0.719	0.137	0.523	0.340	0.278	0.029	0.219	0.036	0.185	0.253
11c	West		0.293	0.707	0.199	0.506	0.295	0.272	0.103	0.245	0.029	0.126	0.225
16b	Center		0.209	0.791	0.136	0.698	0.165	0.226	0.049	0.513	0.018	0.099	0.095
18a	Adriatic		0.507	0.493	0.401	0.356	0.243	0.206	0.169	0.181	0.087	0.220	0.138
20a	Center		0.623	0.377	0.525	0.213	0.262	0.118	0.500	0.077	0.047	0.126	0.132
22b	East		0.700	0.300	0.417	0.170	0.413	0.110	0.148	0.063	0.178	0.260	0.242
25c	East		0.552	0.448	0.248	0.220	0.532	0.228	0.073	0.053	0.252	0.187	0.207
26a	East		0.497	0.503	0.144	0.181	0.675	0.192	0.017	0.037	0.232	0.155	0.368
27b	East		0.738	0.262	0.469	0.133	0.398	0.053	0.127	0.038	0.499	0.134	0.149

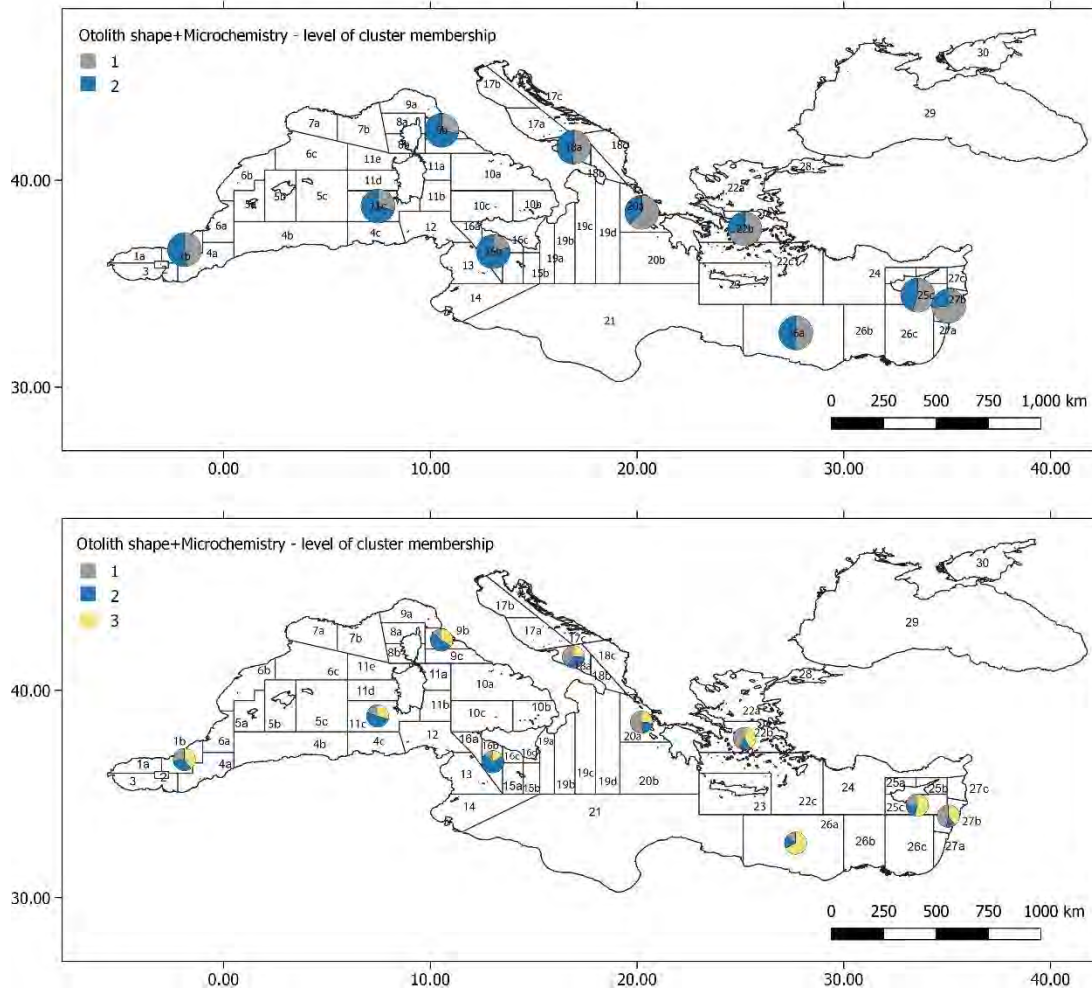


Figure 6: Pie charts with average membership grades per GSA subarea for the OTHO.MICRO dataset of *Merluccius merluccius* at NC = 2 (upper panel) and NC = 3 (lower panel).

The ALL dataset allowed to consider all variables, but with a reduced spatial coverage again limited by the reduced spatial coverage of the MICRO dataset. A first division among GFCM subregions split the Eastern Mediterranean from the other subregions (Table 17). At NC = 3 The Adriatic subregion further split from the Western and Central Mediterranean subregions. Looking at subareas (Table 18), at NC = 2 there is a neat separation between all western subareas, the Adriatic one and the westernmost Central Mediterranean subarea 16b on one hand, and the easternmost Central Mediterranean subarea and all eastern basin subareas on the other. At NC = 3 the Adriatic 18a and the Central 20a subarea joined together into cluster FKM1, while the other two clusters remained strongly characterized in term of longitude: FKM3 as the “western” cluster, FKM2 as the “eastern” cluster. The results for the whole dataset thus seem to represent a very good spatial pattern. First, there is a clear difference between the eastern and western part of the Mediterranean, with a discontinuity positioned south of Sicily that separates the two groups. At a finer level, it seems that also the central, or better northern-central part of the Mediterranean (i.e., the Adriatic and the Ionian Sea) might represent a different population of the hake. It is interesting also to note, that at NC = 3 as well as at NC = 6, many subareas showed very high

average memberships for specific clusters, indicating a rather robust result. By contrast, the membership grades of 1b subarea individuals were almost equally distributed among the three clusters at NC = 3. This result might indicate that in 1b a greater variability is observed, perhaps due to mixing with Atlantic populations of hake.

Table 17: Spatial interpretation of fuzzy clusters results for the combination of all three datasets of *Merluccius merluccius* considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

ALL	2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
West	0.344	0.656	0.271	0.196	0.533	0.187	0.025	0.289	0.217	0.031	0.251
Center	0.326	0.674	0.463	0.056	0.481	0.246	0.251	0.251	0.095	0.005	0.151
Adriatic	0.337	0.663	0.488	0.070	0.442	0.427	0.050	0.138	0.077	0.015	0.293
East	0.827	0.173	0.206	0.678	0.116	0.022	0.052	0.057	0.304	0.417	0.148

Table 18: Spatial interpretation of fuzzy clusters results for the combination of all three datasets of *Merluccius merluccius* considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

ALL		2 CLUSTERS		3 CLUSTERS			6 CLUSTERS					
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6
1b	West	0.466	0.534	0.340	0.311	0.350	0.151	0.012	0.127	0.261	0.079	0.369
9b	West	0.389	0.611	0.244	0.239	0.516	0.266	0.005	0.186	0.250	0.020	0.271
11c	West	0.208	0.792	0.244	0.067	0.690	0.138	0.053	0.511	0.152	0.005	0.141
16b	Center	0.109	0.891	0.313	0.033	0.653	0.380	0.019	0.298	0.080	0.004	0.220
18a	Adriatic	0.337	0.663	0.488	0.070	0.442	0.427	0.050	0.138	0.077	0.015	0.293
20a	Center	0.660	0.340	0.692	0.091	0.217	0.041	0.608	0.179	0.118	0.008	0.045
22b	East	0.795	0.205	0.400	0.444	0.156	0.043	0.147	0.085	0.293	0.226	0.207
25c	East	0.752	0.248	0.177	0.643	0.181	0.012	0.004	0.095	0.343	0.338	0.207
26a	East	0.820	0.180	0.036	0.870	0.094	0.019	0.001	0.029	0.534	0.356	0.061
27b	East	0.952	0.048	0.117	0.867	0.016	0.007	0.014	0.004	0.069	0.820	0.087

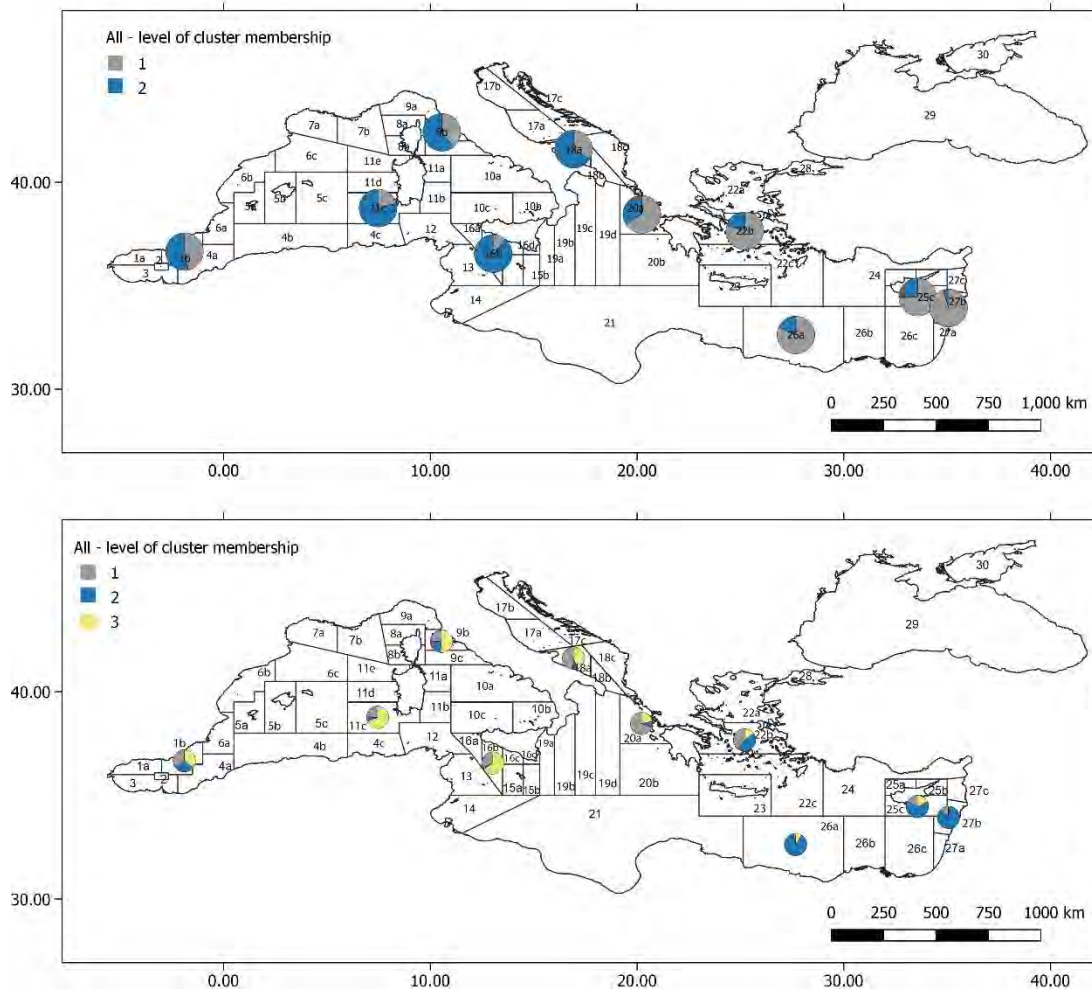


Figure 7: Pie charts with average membership grades per GSA subarea for the ALL dataset of *Merluccius merluccius* at NC = 2 (upper panel) and NC = 3 (lower panel).

Spatial analysis

In Table 19 the results of the RDA built to explain the three hake datasets and their combinations with spatial variables (i.e., depth, longitude and latitude, significant MEMs) are presented. Adjusted R² are reported not only for each cluster for different NC, but also for the entire dataset for comparison. In general, the variance of the original datasets, i.e., of all sampled and analyzed individuals together, is much greater than that of the FKM clusters and the spatial explanatory variables can explain only a minor fraction of it compared to that of the clusters. All models were built on original, non-detrended data, which means that the explanatory power of depth, geographic coordinates and MEMs might be correlated one with another. Thus, it is not correct to assume that the sum of the adjusted R² of the three (groups of) variables is equal to the total explanatory power of the spatial predictors: given the shared effects among the different spatial components, the combined adjusted R² will be lower than the sum of the adjusted R² of each single spatial component. Furthermore, all three spatial components might share some fraction of their explanatory power also with the set of environmental variables, which are also usually spatially structured. These latter aspects will be addressed in detail in the variation partitioning procedures described below.

Table 19: Spatial analysis results for *M. merluccius*. For each dataset the explanatory power of each spatial variable or group of variables is reported as adjusted R². All reported models with depth and LongLat were tested for significance with an ANOVA and “ns” indicates those that were not significant. The significance of MEMs was tested with a forward selection procedure and double stopping criterion. Only significant MEMs are reported. For MEMs also the number of significant MEMs, their names and the winner spatial weighting matrix are reported.

GEN	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	0.03	0.05	25	MEM4, MEM12, MEM16, MEM29, MEM18, MEM35, MEM20, MEM2, MEM19, MEM36, MEM50, MEM51, MEM8, MEM3, MEM58, MEM17, MEM9, MEM45, MEM23, MEM1, MEM21, MEM52, MEM7, MEM6, MEM14	MST_Up_0.1
2 CLUSTERS	ns	0.09	0.08	12	MEM18, MEM5, MEM41, MEM49, MEM17, MEM39, MEM21, MEM27, MEM57, MEM18, MEM5, MEM41, MEM49, MEM17, MEM39, MEM21, MEM27, MEM57	Relative_Down_3
3 CLUSTERS	ns	0.02	0.02	8	MEM18, MEM27, MEM34, MEM49, MEM41, MEM57, MEM47, MEM5	Relative_Down_5
4 CLUSTERS	ns	0.01	0.01	7	MEM36, MEM38, MEM51, MEM16, MEM43, MEM48, MEM46	Delaunay_Up_0.6
5 CLUSTERS	0.00	0.01	0.03	14	MEM3, MEM18, MEM4, MEM16, MEM47, MEM44, MEM35, MEM9, MEM58, MEM8, MEM12, MEM45, MEM2, MEM28	MST_Up_0.1
6 CLUSTERS	ns	0.06	0.10	23	MEM8, MEM18, MEM1, MEM20, MEM30, MEM10, MEM4, MEM5, MEM2, MEM39, MEM7, MEM15, MEM23, MEM37, MEM13, MEM3, MEM59, MEM25, MEM17, MEM14, MEM52, MEM38, MEM11	Gabriel_Down_9
otho	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	0.03	0.14	38	MEM38, MEM50, MEM36, MEM58, MEM18, MEM10, MEM9, MEM7, MEM33, MEM17, MEM56, MEM5, MEM22, MEM49, MEM42, MEM11, MEM13, MEM1, MEM54, MEM6, MEM35, MEM37, MEM14, MEM60, MEM39, MEM51, MEM3, MEM19, MEM26, MEM4, MEM53, MEM34, MEM52, MEM28, MEM32, MEM24, MEM15, MEM20	Relative_Up_0.1

2 CLUSTERS	0.00	0.05	0.28	23	MEM38, MEM45, MEM36, MEM49, MEM6, MEM52, MEM26, MEM3, MEM4, MEM9, MEM57, MEM35, MEM7, MEM51, MEM59, MEM47, MEM10, MEM53, MEM21, MEM34, MEM44, MEM29, MEM1	MST_Up_0.4
3 CLUSTERS	0.00	0.03	0.17	29	MEM36, MEM27, MEM51, MEM52, MEM4, MEM25, MEM6, MEM39, MEM49, MEM53, MEM9, MEM8, MEM60, MEM3, MEM1, MEM44, MEM48, MEM28, MEM42, MEM29, MEM32, MEM45, MEM13, MEM56, MEM16, MEM20, MEM50, MEM35, MEM5	MST_Up_0.6
4 CLUSTERS	0.00	0.03	0.16	36	MEM37, MEM6, MEM49, MEM52, MEM28, MEM12, MEM3, MEM51, MEM48, MEM40, MEM46, MEM18, MEM34, MEM58, MEM20, MEM30, MEM44, MEM33, MEM5, MEM7, MEM25, MEM50, MEM14, MEM11, MEM41, MEM39, MEM53, MEM23, MEM15, MEM56, MEM9, MEM60, MEM10, MEM26, MEM55, MEM36	Relative_Up_0.2
5 CLUSTERS	0.00	0.02	0.13	35	MEM39, MEM51, MEM58, MEM43, MEM32, MEM27, MEM44, MEM8, MEM52, MEM5, MEM25, MEM6, MEM37, MEM23, MEM49, MEM9, MEM7, MEM13, MEM59, MEM28, MEM40, MEM35, MEM48, MEM60, MEM3, MEM4, MEM18, MEM19, MEM17, MEM50, MEM53, MEM45, MEM24, MEM21, MEM26	MST_Up_1
6 CLUSTERS	0.00	0.02	0.12	36	MEM58, MEM51, MEM37, MEM43, MEM27, MEM32, MEM38, MEM44, MEM25, MEM4, MEM6, MEM48, MEM8, MEM60, MEM52, MEM23, MEM53, MEM9, MEM47, MEM20, MEM28, MEM5, MEM39, MEM59, MEM13, MEM18, MEM2, MEM7, MEM17, MEM24, MEM49, MEM26, MEM3, MEM57, MEM15, MEM35	MST_Up_0.9
micro	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX

DATA	0.02	0.05	0.17	9	MEM10, MEM6, MEM7, MEM4, MEM3, MEM1, MEM8, MEM5, MEM2	Relative_Up_0.2
2 CLUSTERS	0.12	0.11	0.28	6	MEM7, MEM2, MEM9, MEM4, MEM1, MEM6	Relative_Linear
3 CLUSTERS	0.05	0.12	0.27	8	MEM10, MEM3, MEM5, MEM2, MEM1, MEM7, MEM9, MEM8	Relative_Up_0.1
4 CLUSTERS	0.03	0.10	0.22	9	MEM10, MEM6, MEM7, MEM5, MEM2, MEM1, MEM8, MEM9, MEM4	MST_Up_0.1
5 CLUSTERS	0.03	0.07	0.17	8	MEM10, MEM6, MEM7, MEM2, MEM5, MEM3, MEM1, MEM9	Relative_Up_0.1
6 CLUSTERS	0.03	0.06	0.15	6	MEM3, MEM7, MEM4, MEM9, MEM1, MEM8	Relative_Down_4
gen.otho	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	0.03	0.08	38	MEM2, MEM35, MEM23, MEM27, MEM10, MEM22, MEM11, MEM1, MEM5, MEM3, MEM29, MEM21, MEM6, MEM4, MEM48, MEM49, MEM12, MEM25, MEM17, MEM28, MEM47, MEM9, MEM7, MEM41, MEM15, MEM32, MEM40, MEM30, MEM16, MEM55, MEM8, MEM39, MEM46, MEM14, MEM18, MEM20, MEM13, MEM45	Relative_Down_9
2 CLUSTERS	ns	0.15	0.29	20	MEM39, MEM15, MEM42, MEM11, MEM3, MEM25, MEM19, MEM1, MEM2, MEM9, MEM16, MEM36, MEM14, MEM26, MEM22, MEM7, MEM8, MEM10, MEM30, MEM20	Delaunay_Down_3
3 CLUSTERS	ns	0.20	0.30	25	MEM30, MEM13, MEM10, MEM7, MEM2, MEM51, MEM14, MEM8, MEM55, MEM22, MEM46, MEM52, MEM17, MEM15, MEM23, MEM9, MEM38, MEM47, MEM6, MEM12, MEM42, MEM39, MEM50, MEM3, MEM5	Relative_Linear
4 CLUSTERS	0.00	0.20	0.31	27	MEM30, MEM13, MEM10, MEM7, MEM2, MEM51, MEM14, MEM23, MEM55, MEM8, MEM22, MEM17, MEM15, MEM52, MEM46, MEM34, MEM6, MEM9, MEM38, MEM12, MEM5, MEM3, MEM39, MEM32, MEM4, MEM42, MEM21	Relative_Linear

5 CLUSTERS	0.00	0.18	0.28	26	MEM30, MEM2, MEM7, MEM10, MEM51, MEM23, MEM14, MEM55, MEM13, MEM15, MEM8, MEM34, MEM17, MEM22, MEM52, MEM46, MEM5, MEM6, MEM9, MEM12, MEM3, MEM39, MEM16, MEM4, MEM48, MEM47	Relative_Linear
6 CLUSTERS	0.00	0.12	0.18	24	MEM30, MEM2, MEM7, MEM10, MEM51, MEM14, MEM23, MEM55, MEM41, MEM8, MEM15, MEM13, MEM17, MEM47, MEM5, MEM39, MEM3, MEM46, MEM34, MEM48, MEM20, MEM4, MEM35, MEM12	Relative_Linear
otho.micro	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.02	0.05	0.16	8	MEM8, MEM5, MEM2, MEM3, MEM9, MEM7, MEM1, MEM6	Relative_Linear
2 CLUSTERS	0.15	0.16	0.35	4	MEM3, MEM4, MEM1, MEM7	Delaunay_Up_0.2
3 CLUSTERS	0.07	0.13	0.36	7	MEM3, MEM8, MEM4, MEM9, MEM2, MEM5, MEM1	Relative_Down_4
4 CLUSTERS	0.06	0.11	0.27	7	MEM6, MEM8, MEM3, MEM2, MEM5, MEM1, MEM9	Relative_Linear
5 CLUSTERS	0.04	0.10	0.25	6	MEM8, MEM3, MEM1, MEM4, MEM7, MEM2	Relative_Down_3
6 CLUSTERS	0.03	0.08	0.24	8	MEM9, MEM8, MEM6, MEM5, MEM1, MEM7, MEM3, MEM4	Relative_Up_0.5
all	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.01	0.04	0.08	6	MEM8, MEM7, MEM4, MEM9, MEM5, MEM1	Relative_Up_1
2 CLUSTERS	0.08	0.31	0.51	5	MEM6, MEM9, MEM1, MEM7, MEM2	Relative_Up_1
3 CLUSTERS	0.06	0.24	0.36	5	MEM7, MEM1, MEM3, MEM5, MEM2	Gabriel_Linear
4 CLUSTERS	0.05	0.22	0.33	7	MEM8, MEM5, MEM4, MEM7, MEM3, MEM1, MEM9	Relative_Up_0.3
5 CLUSTERS	0.03	0.13	0.26	8	MEM8, MEM5, MEM4, MEM7, MEM6, MEM1, MEM9, MEM3	Relative_Up_0.3
6 CLUSTERS	0.02	0.11	0.23	8	MEM8, MEM5, MEM4, MEM7, MEM3, MEM6, MEM1, MEM9	Relative_Up_0.2

Depth did not have a significant effect in explaining the variance of the datasets, but for those with the lesser number of individuals and sampling locations, i.e., the MICRO, the OTHO.MICRO and the ALL datasets. In particular, neither GEN data, nor OTHO data seem to be organized according to depth. The highest explanatory power of depth was observed for NC = 2 in the MICRO (0.12), OTHO.MICRO (0.15) and ALL (0.08) datasets. In all three datasets the explanatory power of depth decreased with increasing NC. Possibly the main reason for such low explanatory power of depth in particular for the datasets with a higher number of individuals and sampling locations, is the association of multiple individuals with their variability to the same depth coordinates (thus depth cannot act as an adequate explanatory variable); other possible reasons include low accuracy in reported depth data; and cases for which depth and coordinates data were not reported and were thus substituted with either the GSA subarea median depth associated to the subarea centroid or with the depth extracted from Copernicus model.

Longitude and latitude coordinates were used to model the large-scale linear gradient in the Mediterranean. Again, the results were different for the different datasets. For the GEN dataset at NC = 2 and NC = 3 the variance explained by this gradient is significant. In the OTHO dataset the variances explained did not exceed 0.05 for NC = 2. Thus, it seems that the GEN dataset is more structured according to the horizontal gradient modelled by longitude and latitude than the OTHO dataset, while the opposite is true regarding the depth gradient. Again, by combining OTHO and GEN data we got a much higher variance explained (0.15-0.20) for NC = 2 to 4. Up to 11-12% of variance of the MICRO dataset at NC = 2 and 3 was also explained by the metric coordinates, and even more (16-13% of the variance of the combined OTHO.MICRO dataset). For the combination of all variables, the variance explained by the large-scale linear gradient was 0.31 and 0.24 at NC = 2 and NC = 3 respectively.

The MEMs are able to model the whole spectrum of scales in spatial structuring, from local scale spatial autocorrelation to large climatic zone spatial organization. Nevertheless, since we derived them on the detrended data, the MEMs used in this study could be only related to spatial features at a scale smaller than the scale of the whole sampling design of the project. Generally, MEMs identified by low numbers do model large-scale spatial features, while MEMs identified by large numbers do model small-scale spatial features (Borcard et al., 2011). From Table 19 it is immediately clear that the more complex sampling designs of datasets GEN and OTHO and their combination needed more spatial components to be adequately described (total MEMs for these two datasets were around 60). Among the significant MEMs extracted for these datasets there are some large-scale components, but prevailed medium to small-scale components. In contrast, the MICRO dataset and its combine datasets were described by a lesser number of spatial components (total MEMs for this dataset were around 9), since the sampling design in this case is simpler, with less sampling locations well apart one from the other. Thus, the scale range of the MEMs for these datasets was reduced.

The spatial structuring of the GEN dataset was not very strong. In fact, the significant MEMs explanatory power followed the one of the geographic coordinates, being low except for NC = 2 and NC = 6. Much higher was the explanatory power of the MEMs for the other two

datasets, OTHO and MICRO. It seems thus that the information provided by otolith shape and microchemistry is more related to spatial features. The explanatory power of MEMs for these two datasets was higher for NC = 2 and NC = 3, decreasing for increasing NC and with minimum usually for the highest NC. The combined datasets always outperformed the single datasets as to the adjusted R2 of the models built with all significant MEMs. In fact, this speaks in favor of the combination of different datasets in order to analyze the spatial structure of the Mediterranean hake population. In the ALL dataset the explanatory power of the selected MEMs was 51% (!) at NC = 2. This latter was by far the highest explanatory power of any MEM based spatial model, and together with the 31% explained by the geographic coordinates and the 8% explained by depth granted a very high explanation of the spatial structuring of this combined dataset.

Some examples of significant MEMs selected for the different datasets, as well as the different connectivity schemes explored in this study, are presented in Figg. 8-13.

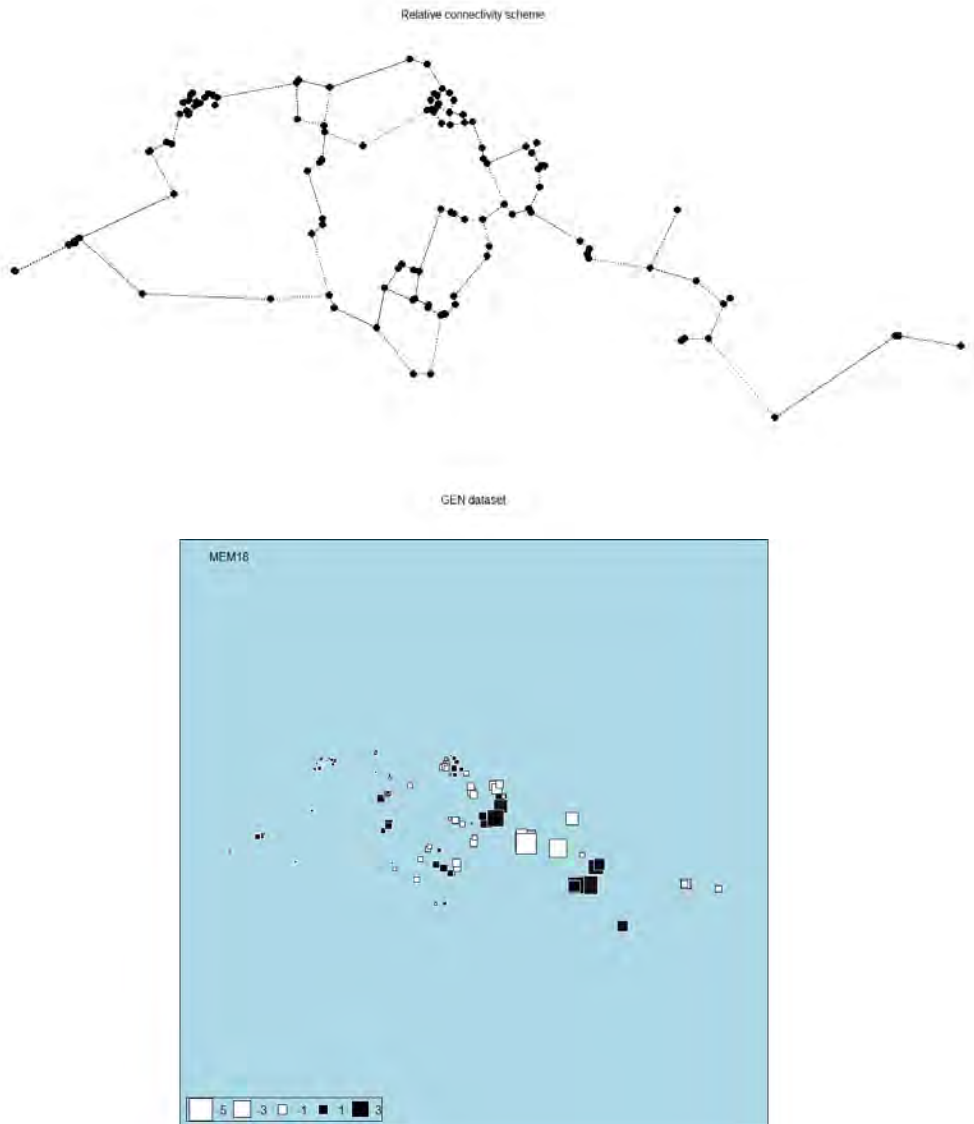


Figure 8: Relative connectivity scheme for the sampling location of the GEN dataset (above) and the significant medium scale MEM18 derived on it. Black squares denote positive values of the MEM, white squares denote negative values of the MEM. Size of squares is proportional to the absolute value at each sampling location.

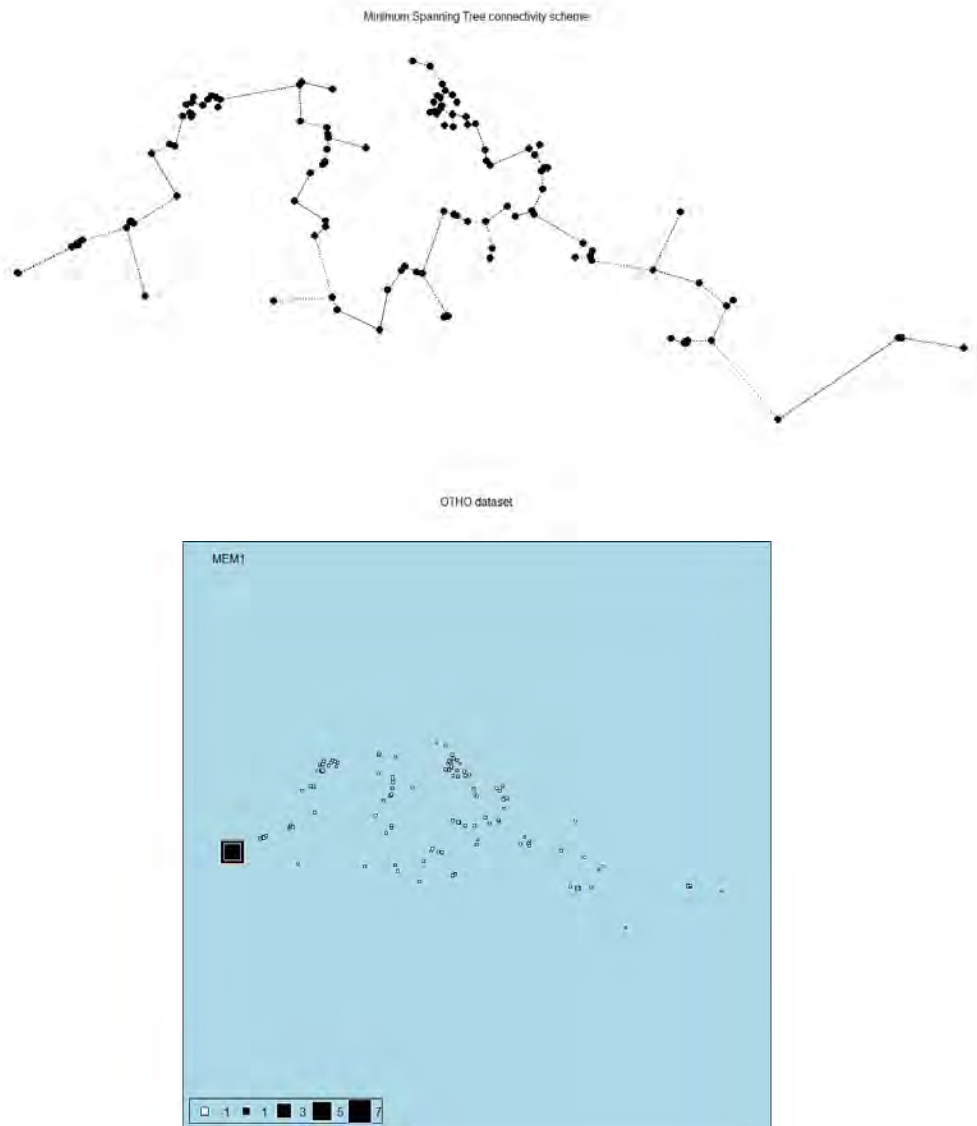


Figure 9: Minimum Spanning Tree connectivity scheme for the sampling location of the OTHO dataset (above) and the significant large scale MEM1 derived on it. Black squares denote positive values of the MEM, white squares denote negative values of the MEM. Size of squares is proportional to the absolute value at each sampling location.

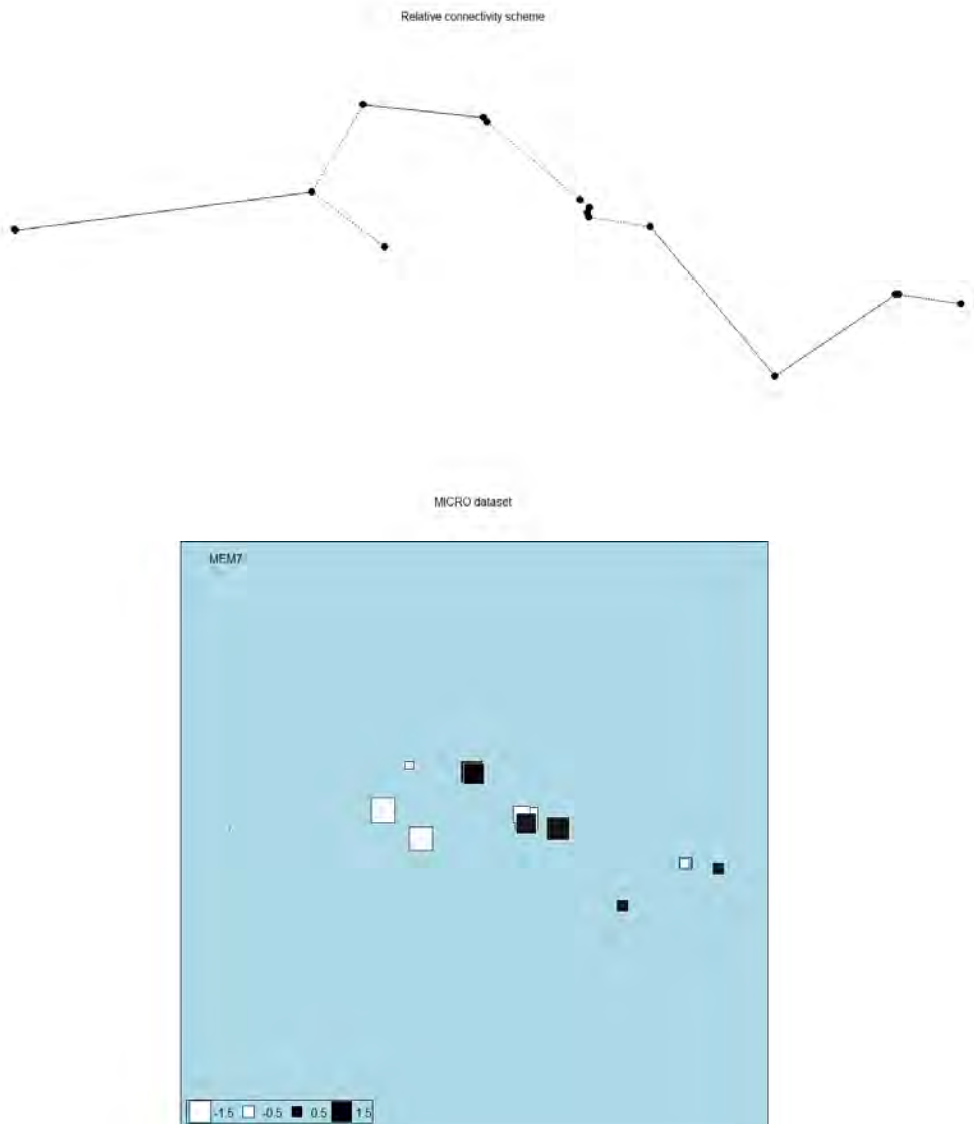


Figure 10: Relative connectivity scheme for the sampling location of the MICRO dataset (above) and the significant small scale MEM7 derived on it. Black squares denote positive values of the MEM, white squares denote negative values of the MEM. Size of squares is proportional to the absolute value at each sampling location.

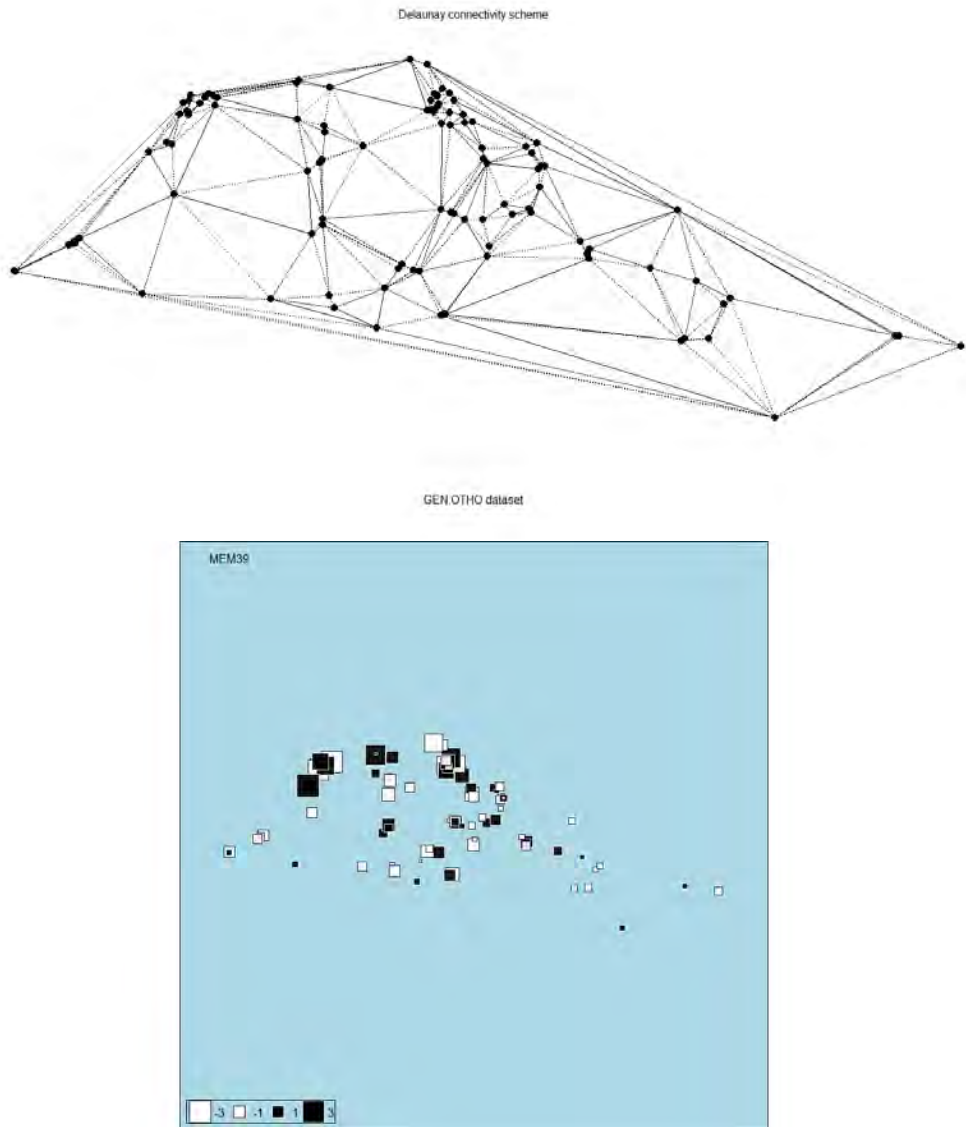


Figure 11: Delaunay's connectivity scheme for the sampling location of the GEN.OTHO dataset (above) and the significant small scale MEM39 derived on it. Black squares denote positive values of the MEM, white squares denote negative values of the MEM. Size of squares is proportional to the absolute value at each sampling location.

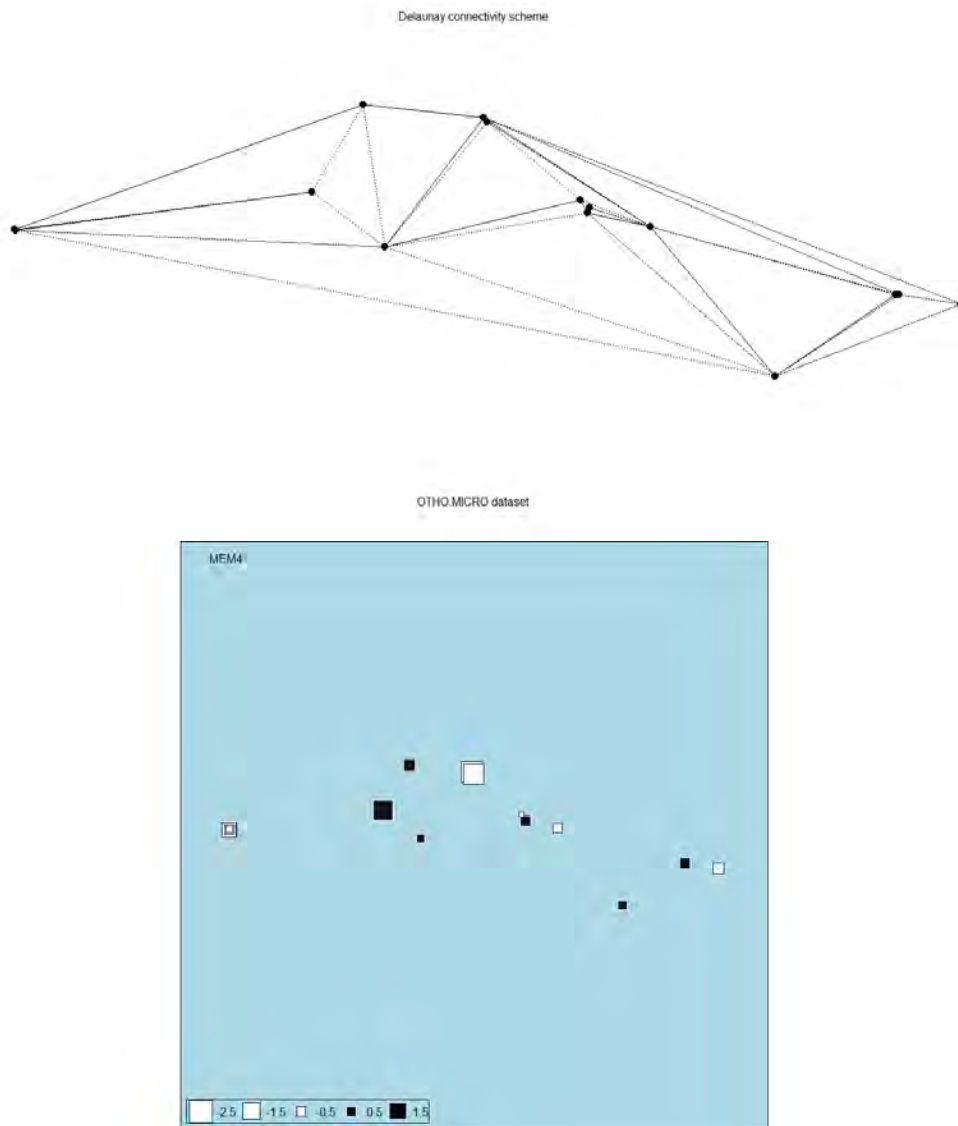


Figure 12: Delaunay's connectivity scheme for the sampling location of the OTHO.MICRO dataset (above) and the significant medium scale MEM4 derived on it. Black squares denote positive values of the MEM, white squares denote negative values of the MEM. Size of squares is proportional to the absolute value at each sampling location.

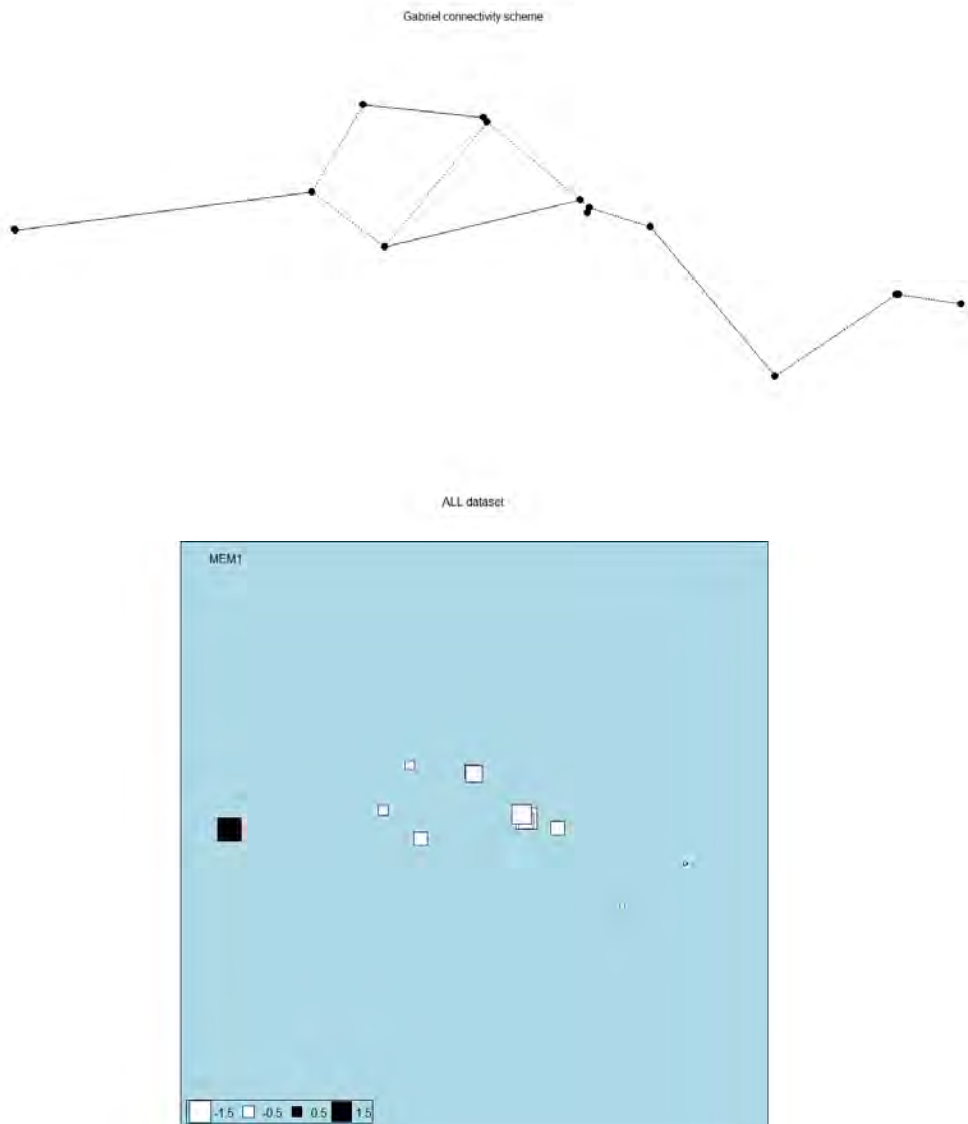


Figure 13: Gabriel's connectivity scheme for the sampling location of the ALL dataset (above) and the significant large scale MEM1 derived on it. Black squares denote positive values of the MEM, white squares denote negative values of the MEM. Size of squares is proportional to the absolute value at each sampling location.

With the significant spatial components, we built a model for GEN.OTHO and ALL at $N = 2$ and $NC = 3$. The model for the GEN.OTHO dataset at $NC = 2$ explained an adjusted R^2 of 0.31 and clearly separated between the Eastern Mediterranean (high longitude, low latitude) and the rest of the basin. At $NC = 3$ along the first RDA this same gradient is visible (Fig. 14). The second RDA separates mostly between cluster 1 at higher latitudes and longitudes (central and northern Mediterranean, including the Adriatic), and cluster 3 found at lower latitudes (along the Spanish and African coast) and longitudes (in the Western Mediterranean). Nevertheless, the variance explained by this second gradient is much lower (5%).

In the spatial model for ALL dataset at $NC = 2$ also depth was relevant, pointing in the same direction as high latitudes and associated to cluster 2. At $NC = 3$ (Fig. 15) depth did distinguish cluster 3 from the others, while latitude discriminated clusters 3 and 1 from cluster 2, associated to high longitudes. MEM1 (Fig. 13) and MEM2 were also associated to cluster 2, while MEM5 was associated to cluster 1.

It is also clear from the biplots of Fig. 14 and 15 that the much greater variation in the GEN.OTHO data is responsible for the projection of individuals around the intersection of axes, with little or no cluster structure emerging. On the other hand, the reduced variation and spatial coverage of the ALL dataset caused a better discrimination among the clusters (Fig. 15), with projection of individuals forming a triangle shaped form: the individuals projected at the tips are those more strongly associated to each cluster, while those in between are individuals associated with high memberships to two or even all three clusters.

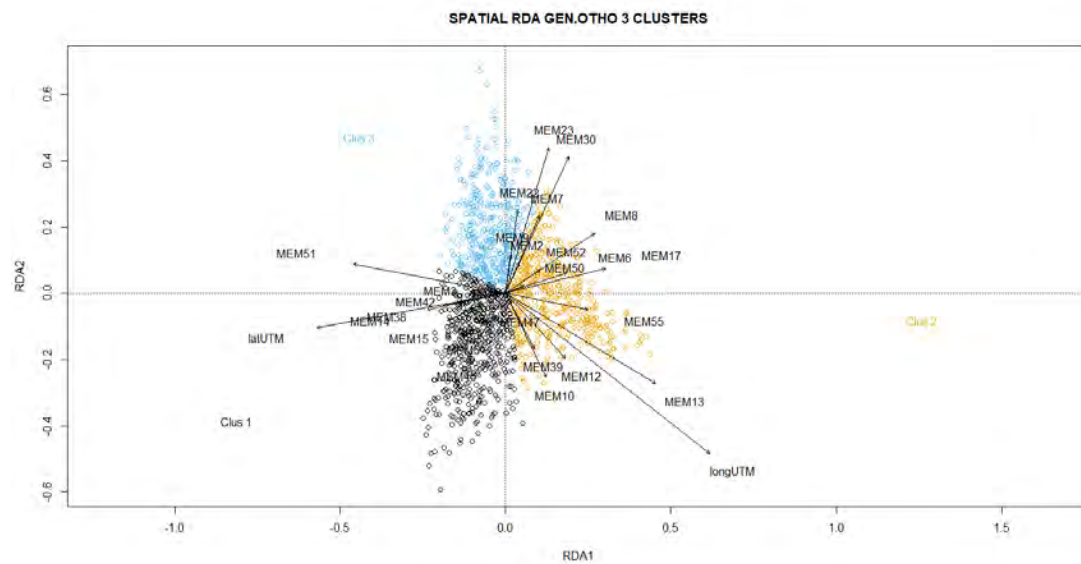


Figure 14: Spatial model for GEN.OTHO dataset at NC = 3. Explained variance is 0.30 and 0.05 respectively on the first and second constrained axis.

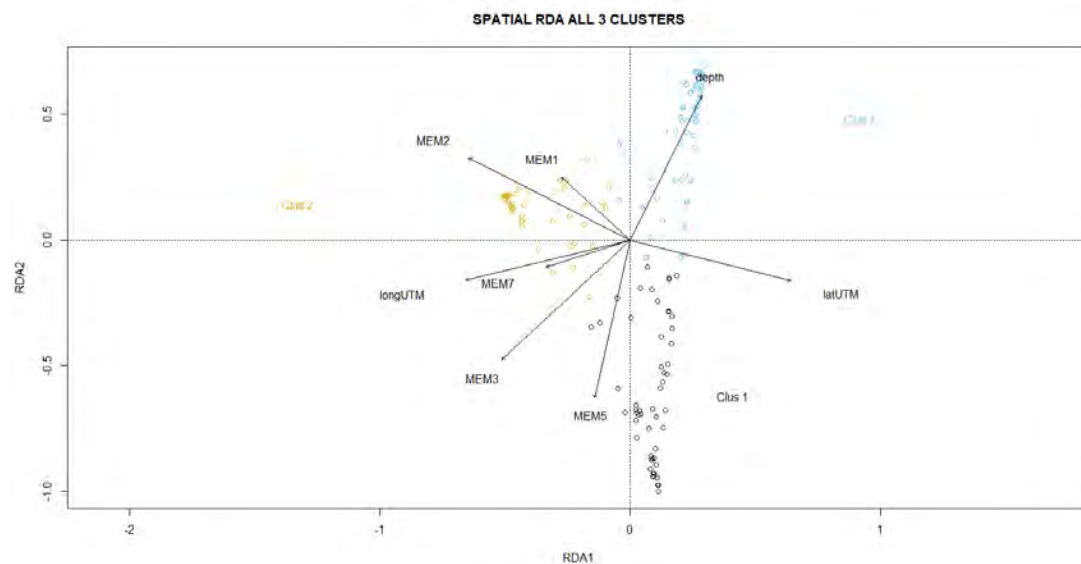


Figure 15: Spatial model for ALL dataset at NC = 3. Explained variance is 0.37 and 0.08 respectively on the first and second constrained axis.

From the spatial analysis we can conclude that the spatial distribution of GEN.OTHO clusters on one hand, and that of MICRO and ALL clusters on the other hand, are the ones better explained by spatial components at NC = 2 and NC = 3. In fact, the higher spatial resolution of the GEN.OTHO cluster favors a division in three populations: one, and clearly separated is

the Eastern Mediterranean (subarea 22b showed some traits common to FKM1 too). The other two are less well solved, but we may hypothesize that one is the westernmost, possibly related to Atlantic hake populations, and extending along the Alboran and Balearic sea and along the African coast up to the Sicily channel. The other is more characterizing the central-northern part of the Mediterranean, in particular along the Italian coasts, including the Adriatic Sea. The spatial coverage of the ALL dataset is not enough to adequately solve these three populations: while the eastern is clearly separated, the central is limited to the Adriatic and Ionian only, while all the rest is brought together into one western Mediterranean cluster. In fact, NC = 2 is the clusterization spatially better described for the ALL dataset. In this case, there is a neat distinction between a population of the Eastern Mediterranean (including 20a), and another of the Western Mediterranean (including the Adriatic 18a and Strait of Sicily 16B). This result also reinforces the idea that the possible two populations of the Western and Central Mediterranean are more intermixed and closer one to the other, than the one in the Eastern Mediterranean.

Environmental analysis

Since we computed the medians of the environmental variables over three different time intervals (1 year, 3 years, 5 years), and since we extracted environmental variables from the euphotic zone and from bottom, we calculated separate RDA for each of these sets of explanatory variables on each dataset of the original data (GEN, OTHO, MICRO), including their combinations (Table 20).

Table 20: Adjusted R^2 for each dataset and their combinations of *Merluccius merluccius*, calculated for RDA models with environmental variables as explanatory variables. Different models were built for each set of explanatory variables: bottom and euphotic zone (surface), and the three time intervals considered (5 years, 3 years, 1 year). Results are reported for each original dataset, and the clusterizations from 2 to 6 clusters.

		GENETIC																	
		INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			6 CLUSTERS		
		y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom		0.04	0.04	0.04	0.10	0.10	0.10	0.02	0.02	0.02	0.01	0.01	0.01	0.03	0.03	0.03	0.08	0.08	0.08
surface		0.05	0.05	0.05	0.10	0.10	0.10	0.02	0.02	0.02	0.01	0.01	0.01	0.03	0.02	0.02	0.08	0.09	0.08
		OTOLITH SHAPE																	
		INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			6 CLUSTERS		
		y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom		0.07	0.06	0.06	0.12	0.11	0.11	0.08	0.07	0.07	0.07	0.06	0.07	0.06	0.06	0.06	0.06	0.05	0.05
surface		0.11	0.10	0.13	0.11	0.10	0.09	0.08	0.07	0.07	0.07	0.06	0.07	0.06	0.06	0.06	0.06	0.05	0.06
		OTOLITH MICROCHEMISTRY																	
		INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			6 CLUSTERS		
		y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom		0.22	0.22	0.20	0.43	0.41	0.35	0.39	0.37	0.33	0.31	0.30	0.27	0.26	0.24	0.22	0.25	0.24	0.22
surface		0.20	0.19	0.21	0.38	0.37	0.39	0.36	0.34	0.35	0.28	0.27	0.29	0.22	0.22	0.24	0.22	0.22	0.22
		GENETIC + OTOLITH SHAPE																	
		INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			6 CLUSTERS		
		y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom		0.05	0.05	0.05	0.19	0.19	0.17	0.21	0.22	0.20	0.21	0.21	0.20	0.20	0.20	0.19	0.13	0.13	0.13
surface		0.05	0.05	0.05	0.21	0.20	0.18	0.24	0.25	0.22	0.24	0.24	0.22	0.23	0.23	0.21	0.15	0.15	0.14

OTOLITH SHAPE + MICROCHEMISTRY																		
	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			6 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.20	0.20	0.19	0.49	0.47	0.43	0.45	0.44	0.39	0.37	0.36	0.33	0.32	0.32	0.29	0.32	0.32	0.29
surface	0.19	0.19	0.19	0.47	0.47	0.48	0.44	0.43	0.43	0.35	0.35	0.36	0.30	0.31	0.32	0.29	0.30	0.31
ALL																		
	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			6 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.12	0.12	0.12	0.59	0.58	0.55	0.48	0.47	0.44	0.42	0.41	0.39	0.35	0.33	0.32	0.30	0.30	0.29
surface	0.12	0.12	0.12	0.58	0.57	0.60	0.46	0.46	0.47	0.40	0.41	0.42	0.32	0.34	0.34	0.28	0.29	0.29

In Annex 1 the median values over hake range of each of the extracted variables are presented for euphotic zone and bottom layer.

In general, the explanatory power of environmental variables decreased with decreasing time interval over which they were computed, but differences among the tree time intervals were never very high. Hake individuals sampled in the present study were around 1 year old, thus it might be expected that the medians over 1 year would be better matched to the genetic or otolith data. Nevertheless, we should take into account also the migrations that the specimen might have performed in search for food or because of ontogenic changes. Such movements are unknown and so are the actual characteristics that each specimen has experienced during its life. It seems that the longest time interval the one arguably more stable and more related to long term dynamics, is the one that is best related to the genetic and otolith shape and microchemistry characteristics. Thus, the relationships with the 5 years time interval were preferably commented.

The bottom dataset of environmental variables almost always explained as much or more than the euphotic zone dataset of environmental variables. The notable exception was the case of the GEN.OTHO dataset, when at all NC euphotic zone variables always explained more than the bottom ones. Furthermore, for all datasets the explanatory power of the environmental variables decreased with increasing NC, but for the GEN.OTHO dataset, where the maximum explanatory power was observed for NC = 3 and NC = 4, being lower both at lower and at higher NC. Given the much lower variability due to the lower number of both, sampled individuals and sampling locations, MICRO dataset, as well as the OTHO.MICRO and the ALL datasets, showed very high fractions of explained variance by the set of environmental variables, reaching around 60% in the ALL dataset for NC = 2. The explained variability of cluster distribution in the GEN and OTHO dataset was much lower (max around 10% in both cases). Nevertheless, two facts stand out: in the ALL dataset we were able to explain more than in the MICRO dataset, even though the addition of all GEN and OTHO variables (while at the same time losing ca. one quarter of individuals) should have added much variance to the MICRO variables. Notably, by themselves, the variances of GEN and OTHO are not well explained by environmental variables. The second outstanding fact is that by combining the GEN dataset and the OTHO dataset, all explained variances for all NC are higher than those of the two original datasets. These two facts mean that the combination of datasets, notwithstanding the differences in the variables, is a promising

strategy since it allows more robust clusterization results, which are also better related to explanatory variables (environmental variables in this case, but also compare Table 19). Thus, the GEN.OTHO results should be preferably considered, and the ALL results are interesting, but hindered in our case by the comparatively low number of individuals in the MICRO dataset.

For GEN.OTHO and ALL at NC = 2 and NC = 3 we built RDA models using first all variables of the environmental dataset that gave the best results as judged by the adjusted R^2 : for the GEN.OTHO this was the dataset with medians of the euphotic zone over 5 years, for ALL the dataset with medians at bottom over 5 years (for NC = 2 the dataset over 1 year was the one explaining more of the ALL dataset variance, but we still used the dataset over 5 years for better consistency with NC = 3 and with the GEN.OTHO dataset). Then, we used forward selection to select a subset of variables explaining as much as possible of the variance of each dataset. A double-stopping criterion (Blanchet & Legendre, 2008) was used in forward selection and the significance was tested after 999 permutations.

The whole environmental model for GEN.OTHO data (Fig. 16) revealed the distinction between warmer and saltier surface waters of the Eastern Mediterranean, and the colder, less saline waters, but more oxygenated and richer in inorganic nutrients and chlorophyll of the Western Mediterranean, including the Adriatic Sea and part of the Central Mediterranean. Interestingly, the primary production points in the opposite direction as the chlorophyll concentration, a fact worth more investigation. At NC = 3 the cluster 1 is strongly associated to high levels of pH, sharing high salinity with cluster 2, while cluster 3 is associated to the highest levels of nutrient and chlorophyll. High primary productivity is also characterizing cluster 3 and more in general the gradient along the second axis. It has to be noted, that the second RDA explained only 3% of the total variance, compared to the 23% explained by the first axis. Thus, the prevalent distinction is still the one separating the Eastern Mediterranean cluster 2 from the other two clusters. Subject to forward selection the selected variables at NC = 2 were surface temperature, primary production, velocity, oxygen, phosphate, and pH (Table 21). At NC = 3 these same variables were joined by DIC, salinity, nitrates and phytoplankton biomass to better distinguish between the three clusters (Fig. 18).

For the ALL dataset the only constrained axis explaining 63% of the total variance is a gradient from warm and saline bottom waters to waters rich in inorganic nutrients. At NC = 3 the main gradient (38% of total variance) separated between cluster 1 and 3 on one side, associated to low salinity and temperature, but high nutrients, and cluster 2 on the other side, associated to high temperature and salinity and low nutrients. The second gradient (14% of total variance) mostly separated cluster 1, corresponding to Adriatic-Ionian waters, showing high relationships with bottom ammonia, dissolved oxygen and pH, from the other two clusters. The forward selection procedure (Table 21) did select temperature, salinity, chlorophyll, velocity, and pH (associated to cluster 1), and ammonium (associated to cluster 2) at NC = 2 (Fig. 17). At NC = 3, chlorophyll was not significant, but instead dissolved oxygen, nutrients (bottom nitrates and phosphates), DIC and alkalinity were selected along with temperature, salinity, velocity and pH (Table 21, Fig. 19).

The principal physical variables temperature and salinity are almost always selected, as well as some variables describing the nutrient enrichment and productivity characteristics (either chlorophyll, phytoplankton biomass or primary production). pH also seems an important parameter in defining the water types of the Mediterranean. Interestingly, also velocity is always selected and though it has a low marginal explanatory power, it brings information that no other variables is providing.

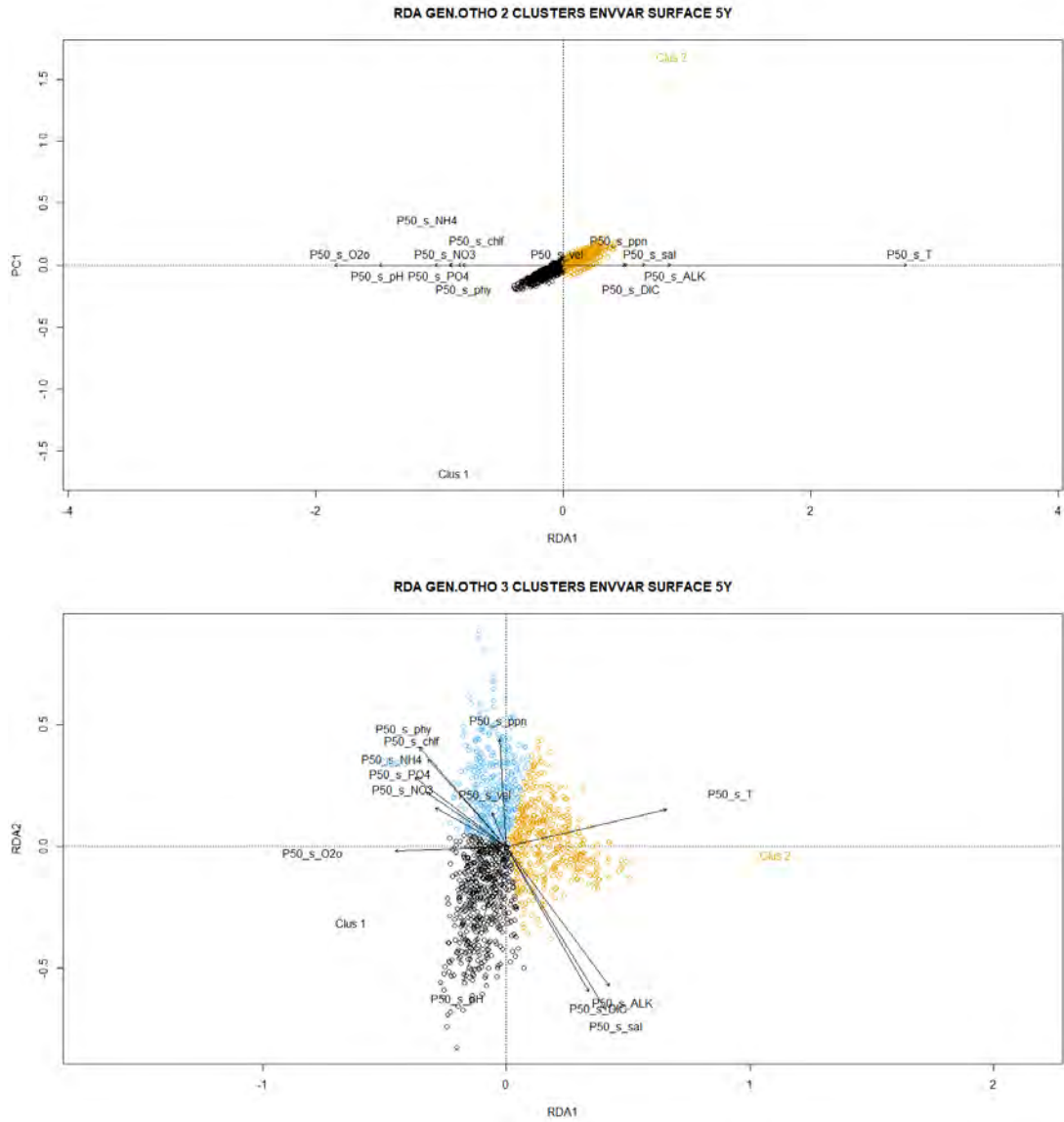


Figure 16: Whole surface environmental models for GEN.OTHO dataset at NC = 2 (above) and at NC = 3 (below). Explained variance is 0.22 on the only constrained axis at NC = 2, and 0.23 and 0.03 respectively on the first and second constrained axis at NC = 3.

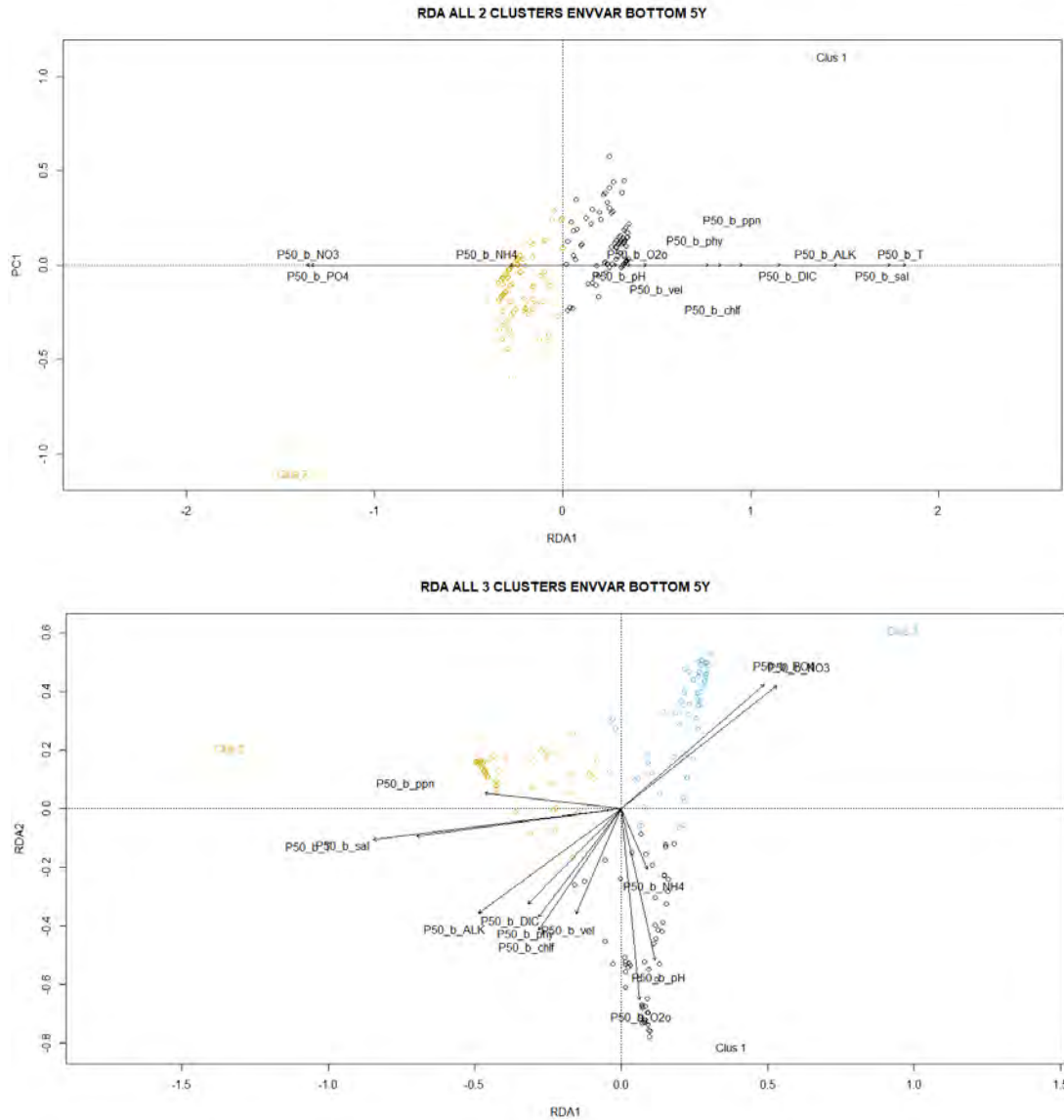


Figure 17: Whole bottom environmental models for ALL dataset at NC = 2 (above) and at NC = 3 (below). Explained variance is 0.63 on the only constrained axis at NC = 2, and 0.38 and 0.14 respectively on the first and second constrained axis at NC = 3.

Table 21: Forward selection on environmental parameters for GEN.OTHO and ALL at NC = 2 and NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

variables	AdjR2Cum	p-value
GEN.OTHO 2 CLUSTERS		
P50_s_T	0.13	0.001
P50_s_PO4	0.15	0.001
P50_s_ppn	0.17	0.001
P50_s_O2o	0.18	0.001
P50_s_vel	0.19	0.001
P50_s_pH	0.20	0.002
GEN.OTHO 3 CLUSTERS		
P50_s_T	0.13	0.001
P50_s_ppn	0.19	0.001

P50_s_sal	0.20	0.001
P50_s_PO4	0.21	0.001
P50_s_O2o	0.22	0.001
P50_s_vel	0.23	0.001
P50_s_NO3	0.24	0.006
P50_s_DIC	0.24	0.005
P50_s_pH	0.24	0.002
P50_s_phy	0.25	0.030
ALL 2 CLUSTERS		
P50_b_T	0.33	0.001
P50_b_sal	0.34	0.023
P50_b_NH4	0.45	0.001
P50_b_chlf	0.47	0.008
P50_b_vel	0.50	0.004
P50_b_pH	0.51	0.032
ALL 3 CLUSTERS		
P50_b_T	0.22	0.001
P50_b_O2o	0.27	0.001
P50_b_vel	0.33	0.001
P50_b_NO3	0.36	0.001
P50_b_PO4	0.38	0.001
P50_b_sal	0.39	0.018
P50_b_DIC	0.40	0.015
P50_b_ALK	0.43	0.001
P50_b_pH	0.44	0.006

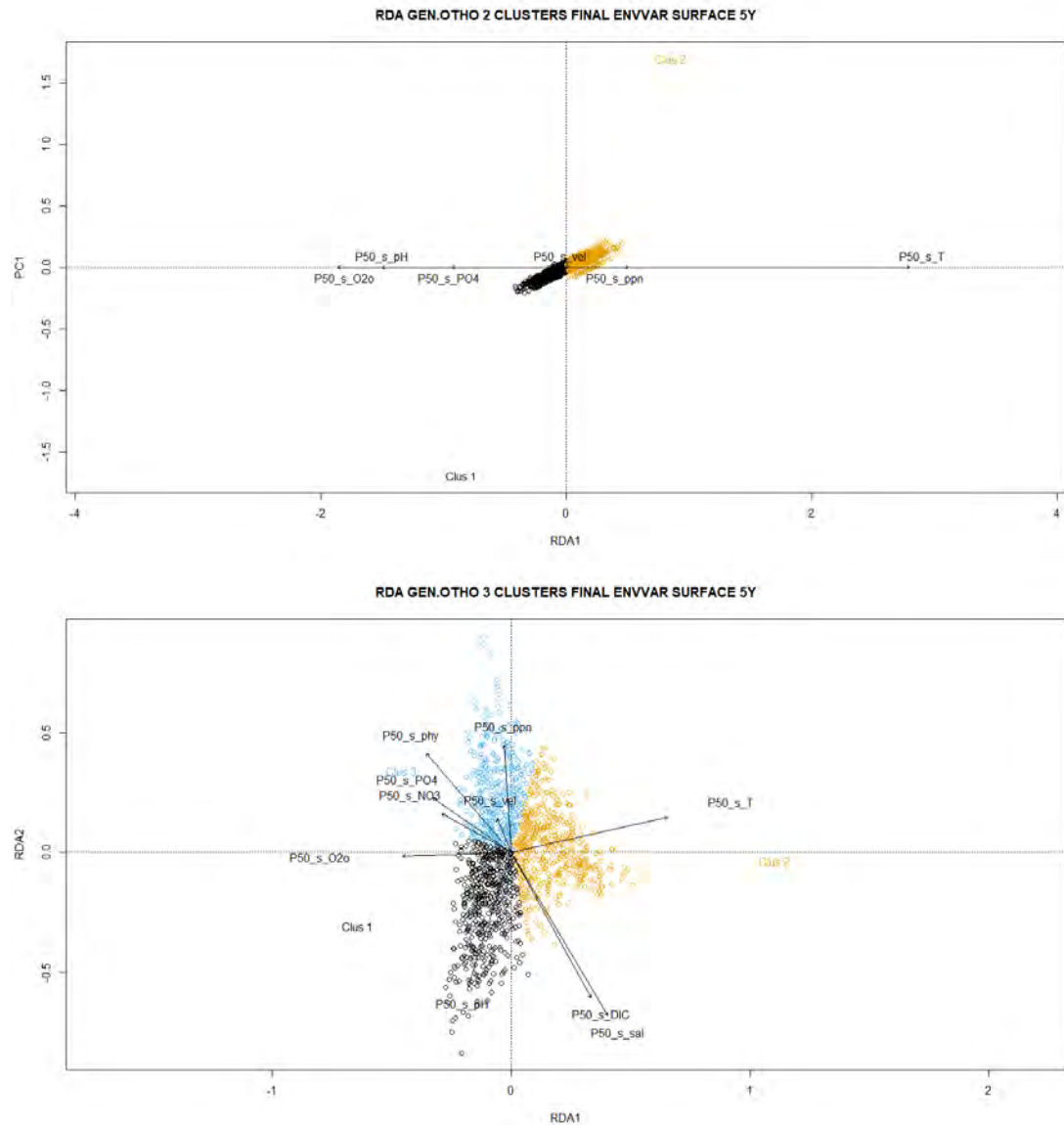


Figure 18: Significant bottom environmental models for ALL dataset at NC = 2 (above) and at NC = 3 (below). Explained variance is 0.20 on the only constrained axis at NC = 2, and 0.22 and 0.03 respectively on the first and second constrained axis at NC = 3.

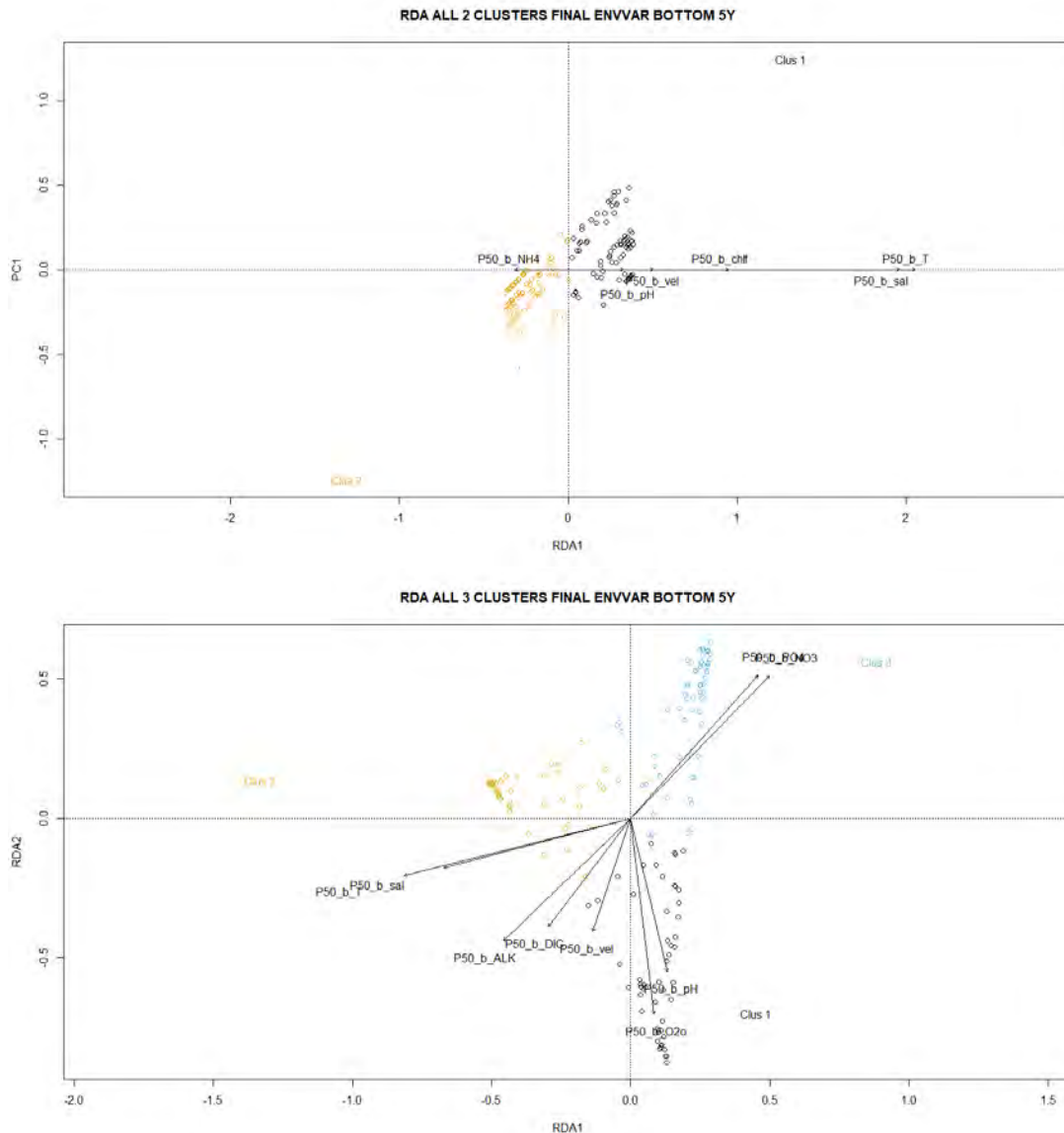


Figure 19: Significant bottom environmental models for ALL dataset at NC = 2 (above) and at NC = 3 (below). Explained variance is 0.53 on the only constrained axis at NC = 2, and 0.36 and 0.11 respectively on the first and second constrained axis at NC = 3.

Variation partitioning and best model selections

The variation partitioning procedure was applied only to the GEN.OTHO and ALL dataset for NC = 2 and 3 (Tables 22-23), since these were the datasets better related to spatial and environmental variables. As explanatory variables all groups of variables were used, i.e., depth (if significant), longitude and latitude (indicated as LongLat), significant MEMs and the whole group of environmental variables (euphotic zone medians over 5 years for the GEN.OTHO dataset and bottom medians over 5 years for the ALL dataset).

For the GEN.OTHO dataset three groups of variables were compared (depth was not significant for either NC) (Fig. 20). The group of MEM variables was the one with the highest explanatory power in both cases. It also had the highest conditional explanatory power when either geographic coordinate or the environmental variables, or both, were taken into account (Table 22). Thus, the MEMs being able to model special features at all scales do

bring some additional explanation which is not shared with the other two groups of variables. Nevertheless, the shared effect between the environmental variables and the MEMs is relevant (6% and 5 % respectively at NC =2 and NC =3) as it is the shared effect among all three groups of variables (12% and 16% respectively at NC =2 and NC =3). These shared effects cannot be attributed to one single group of variables. The environmental variables are strongly correlated with both, the geographic coordinates, and the MEMs, which is not a surprise: the physical, chemical and biological features of the Mediterranean are strongly spatially structured as a consequence of the spatial structuring of the processes driving these characteristics. The additional spatial information brought by MEMs can be attributed to an imperfect description of the environmental characteristics by the Copernicus models (too coarse resolution, some processes not accounted for, etc.), but also to additional factors shaping the spatial organization of the hake population, such as the availability of suitable habitats, autocorrelation, existence of discontinuities, dispersal mechanisms, inter and intra-species interactions, including interactions with fishery. It has to be noted also that the amount of the unaccounted variance (residuals) is high, due to the high variability between individuals, not related neither to the environmental variables used in the present study, nor to the spatial features of the Mediterranean.

Table 22: Variation partitioning for GEN.OTHO dataset of *M. merluccius* at NC = 2 and NC = 3. + indicates the total variation explained by the groups of variables; | indicates the variation explained by the first group of variables conditioned to the second; U indicates the fraction of variance explained jointly by the two groups of variables. Compare to Fig. 14 for a graphical representation.

GEN.OTHO	2 CLUSTERS	3 CLUSTERS
Explanatory group of variables	Adjusted R ²	Adjusted R ²
<i>LongLat</i>	0.15	0.20
<i>MEM</i>	0.29	0.30
<i>EnvVar Surf 5y</i>	0.21	0.24
<i>LongLat + MEM</i>	0.31	0.33
<i>LongLat + EnvVar Surf 5y</i>	0.21	0.26
<i>MEM + EnvVar Surf 5y</i>	0.32	0.34
<i>LongLat + MEM + EnvVar Surf 5y</i>	0.32	0.34
<i>LongLat EnvVar Surf 5y</i>	0.01	0.01
<i>LongLat MEM</i>	0.02	0.03
<i>MEM EnvVar Surf 5y</i>	0.11	0.09
<i>MEM LongLat</i>	0.16	0.13
<i>EnvVar Surf 5y LongLat</i>	0.07	0.06
<i>EnvVar Surf 5y MEM</i>	0.03	0.03
<i>LongLat MEM + EnvVar Surf 5y</i>	0.00	0.00
<i>MEM LongLat + EnvVar Surf 5y</i>	0.11	0.08
<i>EnvVar Surf 5y LongLat + MEM</i>	0.01	0.01
<i>LongLat U MEM</i>	0.00	0.01
<i>MEM U EnvVar Surf 5y</i>	0.06	0.05
<i>LongLat U EnvVar Surf 5y</i>	0.02	0.02
<i>LongLat U MEM U EnvVar Surf 5y</i>	0.12	0.16
<i>Residuals</i>	0.68	0.66

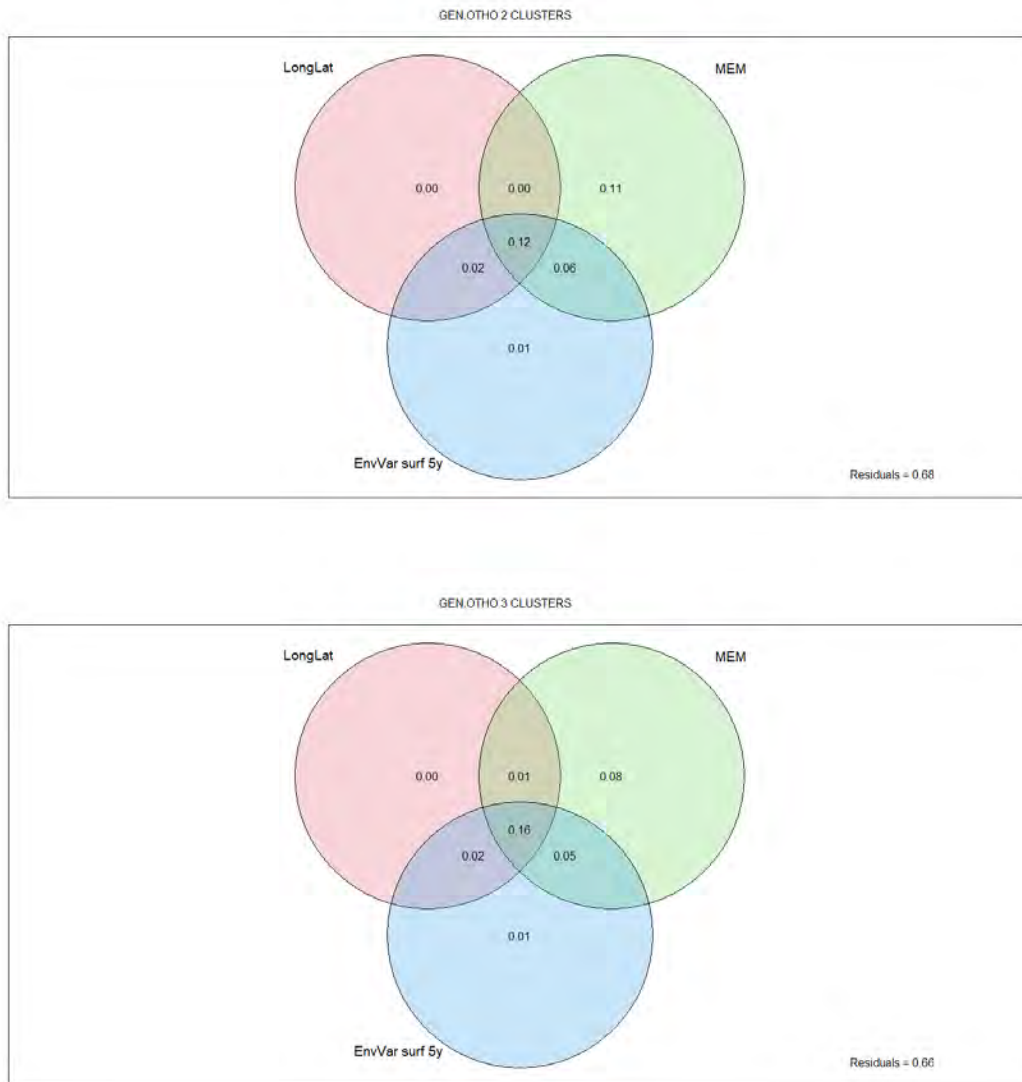


Figure 20: Graphical representation of the variation partitioning of GEN.OTHO clusterizations of *M. merluccius* at 2 and 3 clusters among three groups of variables: LongLat, MEMs and EnvVar Surf 5y. Negative or zero fractions are not displayed. See Table 22 for more details.

The variation partitioning of the ALL dataset revealed that in this case the group of environmental variables is the one showing the highest explanatory power and the highest conditional explanatory power when accounting for the other groups of variables (Fig. 21). In particular, the environmental variables shared very few of their explanatory power with depth, half of it with geographic coordinates, and the majority with MEMs (Table 23). Thus, the bottom environmental variables were able to represent much of the characteristics of the Mediterranean, including its spatial features. The much simpler sampling design in the MICRO and ALL dataset, with few far away sampling location, is responsible for a lower power of MEMs in correctly representing all-scale spatial features. The fraction of variance jointly explained by the four datasets was relatively high for NC = 2, but much lower for NC = 3. The four groups of variables, in particular all the possible combinations including the set of bottom environmental variables, were able to explain the majority of the variance of the ALL dataset, even though at NC = 3 still half of the variance remained unexplained.

Table 23: Variation partitioning for ALL dataset of *M. merluccius* at NC = 2 and NC = 3. + indicates the total variation explained by the groups of variables; | indicates the variation explained by the first group of variables conditioned to the second; U indicates the fraction of variance explained jointly by the two groups of variables. Compare to Fig. 15 for a graphical representation.

ALL	2 CLUSTERS	3 CLUSTERS
Explanatory group of variables	Adjusted R²	Adjusted R²
<i>Depth</i>	0.08	0.06
<i>LongLat</i>	0.31	0.24
<i>MEM</i>	0.51	0.36
<i>EnvVar Bot 5y</i>	0.59	0.48
<i>Depth + LongLat</i>	0.31	0.29
<i>Depth + MEM</i>	0.52	0.36
<i>Depth + EnvVar Bot 5y</i>	0.59	0.48
<i>LongLat + MEM</i>	0.52	0.43
<i>LongLat + EnvVar Bot 5y</i>	0.59	0.49
<i>MEM + EnvVar Bot 5y</i>	0.60	0.49
<i>Depth + LongLat + MEM</i>	0.53	0.43
<i>Depth + LongLat + EnvVar Bot 5y</i>	0.60	0.49
<i>Depth + MEM + EnvVar Bot 5y</i>	0.60	0.49
<i>LongLat + MEM + EnvVar Bot 5y</i>	0.60	0.49
<i>Depth + LongLat + MEM + EnvVar Bot 5y</i>	0.60	0.49
<i>Depth LongLat</i>	0.00	0.04
<i>Depth MEM</i>	0.01	0.00
<i>Depth EnvVar Bot 5y</i>	-0.00	0.00
<i>LongLat Depth</i>	0.23	0.23
<i>LongLat MEM</i>	0.01	0.06
<i>LongLat EnvVar Bot 5y</i>	-0.00	0.01
<i>MEM Depth</i>	0.44	0.30
<i>MEM LongLat</i>	0.21	0.18
<i>MEM EnvVar Bot 5y</i>	0.00	0.01
<i>EnvVar Bot 5y Depth</i>	0.51	0.42
<i>EnvVar Bot 5y LongLat</i>	0.29	0.24
<i>EnvVar Bot 5y MEM</i>	0.09	0.12
<i>Depth MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>Depth LongLat + EnvVar Bot 5y</i>	0.00	0.00
<i>Depth LongLat + MEM</i>	0.00	0.00
<i>LongLat MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat Depth + EnvVar Bot 5y</i>	0.00	0.01
<i>LongLat Depth + MEM</i>	0.01	0.06
<i>MEM Depth + EnvVar Bot 5y</i>	0.00	0.00
<i>MEM LongLat + EnvVar Bot 5y</i>	0.00	0.00
<i>MEM Depth + LongLat</i>	0.22	0.14
<i>EnvVar Bot 5y LongLat + MEM</i>	0.07	0.06
<i>EnvVar Bot 5y Depth + MEM</i>	0.08	0.12
<i>EnvVar Bot 5y Depth + LongLat</i>	0.28	0.20
<i>Depth LongLat + MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat Depth + MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>MEM Depth + LongLat + EnvVar Bot 5y</i>	0.00	-0.00
<i>EnvVar Bot 5y Depth + LongLat + MEM</i>	0.07	0.06
<i>Depth U LongLat</i>	0.00	-0.00

LongLat \cup MEM	0.00	0.01
Depth \cup MEM	0.00	-0.00
Depth \cup EnvVar Bot 5y	0.00	0.00
LongLat \cup EnvVar Bot 5y	0.01	0.06
MEM \cup EnvVar Bot 5y	0.21	0.14
Depth \cup LongLat \cup EnvVar Bot 5y	0.01	0.00
Depth \cup LongLat \cup MEM	0.00	0.00
LongLat \cup MEM \cup EnvVar Bot 5y	0.22	0.16
Depth \cup MEM \cup EnvVar Bot 5y	0.00	0.04
Depth \cup LongLat \cup MEM \cup EnvVar Bot 5y	0.08	0.02
Residuals	0.40	0.51

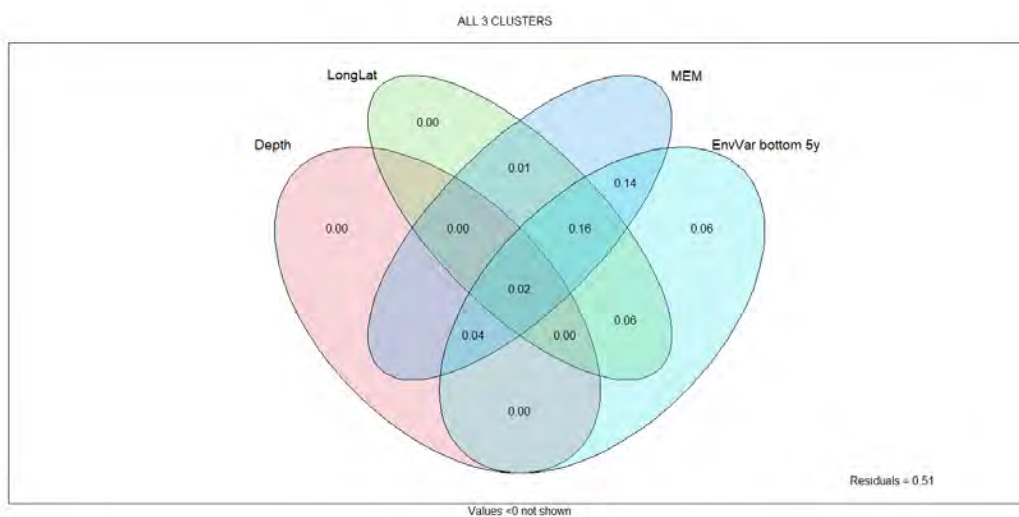
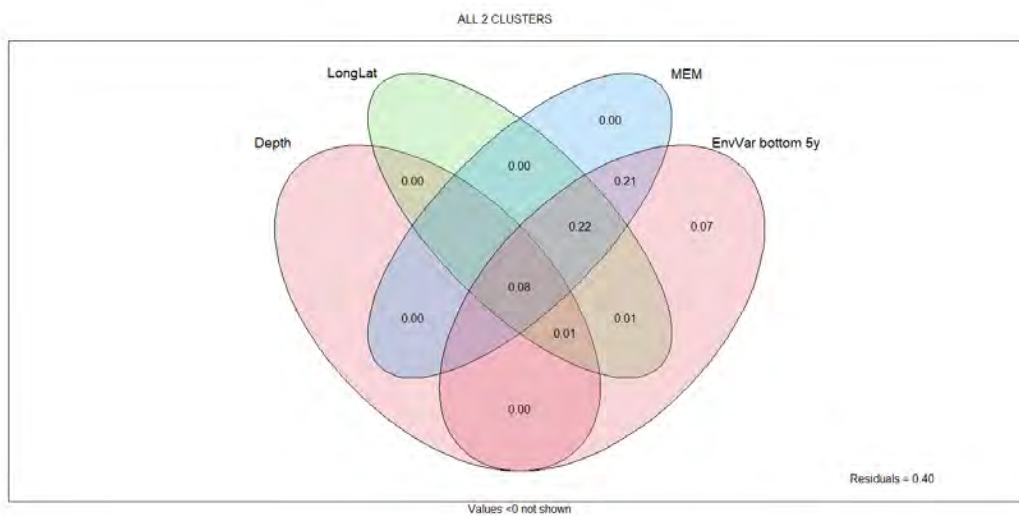


Figure 21: Graphical representation of the variation partitioning of ALL dataset of *M. merluccius* at 2 and 3 clusters four groups of variables (Depth, LongLat, MEM, EnvVar Bot 5y). Negative or zero fractions are not displayed. See Table 23 for more details.

The whole models for GEN.OTHO and ALL at NC = 2 and 3 were subjected to the forward selection procedure with double stopping criterion in order to obtain a parsimonious final model. Results were very different for the GEN.OTHO and the ALL model (Table 24).

In the first case, many different variables were selected as significant (22 and 24 respectively for 2 and 3 clusters), in particular from the set of MEMs (Fig. 16). Considering that MEMs and environmental variables were highly correlated explains why in the forward selection procedure so few environmental variables were selected: ammonium, pH, velocity and phytoplankton biomass at NC = 2, and nitrates, pH velocity and primary production at NC = 3. In both cases thus we have descriptors of nitrogen forms, descriptors related to production, and pH and velocity. These variables much probably show very specific spatial structures that cannot be captured by spatial components. As for the MEMs selected, 10 are the same in the GEN.OTHO model for NC = 2 and NC = 3: mostly large scale (MEM2, MEM3, MEM7, MEM8, MEM10) and medium scale (MEM14, MEM15, MEM16, MEM22); only a few are small-scale components (MEM30, MEM42). The other MEMs selected in the two models are mostly medium-scale components at NC = 2 and small-scale components at NC = 3. Thus, the general spatial pattern is the same, but with more clusters the spatial distribution needs more detailed features to be described by small-scale MEMs.

The fact that so many MEMs and so few environmental variables were selected for the final parsimonious models of the GEN.OTHO dataset in any case should not be interpreted in the sense of a lesser importance of the environmental variables compared to the spatial component. In fact, the forward selection procedure is a “blind” statistical procedure, based solely on the additional explanatory power brought by variables progressively included in the model after the first one, with the highest marginal explanatory power, has been identified. In case of several highly correlated variables, only one is selected to fulfill the procedures requirements, without any judgment of the quality of the information provided by the selected variable. The relationships between the spatial and environmental variables were already investigated in the paragraphs above, as well as the models of GEN.OTHO and ALL datasets based only on environmental variables.

Finally, for both NC, a geographic coordinate was selected in the best parsimonious model: latitude at NC = 2 and longitude at NC = 3. Given the geometry of the Mediterranean and the prevalent spatial gradient between the Eastern and the Western basins, the two coordinates are almost interchangeable: the sampling locations of the Eastern basin had higher longitude values and lower latitude values than the sampling locations in the Western Mediterranean.

The forward selection procedure applied on the four groups of variables for the ALL dataset did select only a handful of variables (4 and 7 respectively) for NC = 2 and NC = 3 (Fig. 23). Two MEMs (MEM6 and MEM7) and two environmental variables (median bottom velocity and median bottom DIC) were still able to explain the majority of the variance of the ALL dataset at NC=2. At NC = 3, along MEM7, MEM1 was selected, and along velocity bottom temperature, ammonium, salinity, and pH were selected. It means that the ALL dataset clusters spatial distribution is much more defined and can be described by some large and medium scale spatial components, together with a few environmental variables (some of them selected also for the GEN.OTHO dataset). At the distance at which samples were

collected for the MICRO dataset, the differences in spatial and environmental characteristics of the habitats of hake populations are robust enough to be correctly described by a reduced number of variables.

The inspection of the residuals of the final models did not reveal recognizable patterns of variance left in the data (results not shown), but rather randomly dispersed residuals. Thus, we can say that these models, while still missing important fractions of the total variance of the datasets, do take account of the majority of their structured variance.

Table 24: Forward selection on all parameters (spatial and environmental) for GEN.OTHO and ALL datasets of M. merluccius at NC = 2 and NC = 3. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

variables	AdjR2Cum	p-value
GEN.OTHO 2 CLUSTERS		
latUTM	0.13	0.001
MEM3	0.19	0.001
MEM39	0.21	0.001
MEM11	0.22	0.001
MEM15	0.24	0.001
MEM42	0.24	0.001
MEM25	0.25	0.001
P50_s_NH4	0.26	0.001
MEM7	0.27	0.001
MEM19	0.28	0.001
MEM30	0.28	0.006
MEM14	0.29	0.001
MEM8	0.29	0.006
MEM10	0.29	0.004
MEM9	0.30	0.006
MEM16	0.30	0.003
P50_s_vel	0.30	0.013
P50_s_pH	0.31	0.02
MEM2	0.31	0.028
P50_s_phy	0.31	0.01
MEM26	0.31	0.026
MEM22	0.32	0.04
GEN.OTHO 3 CLUSTERS		
longUTM	0.15	0.001
P50_s_pH	0.20	0.001
MEM7	0.22	0.001
MEM13	0.23	0.001
MEM51	0.25	0.001
MEM8	0.26	0.001
MEM14	0.26	0.001
MEM30	0.27	0.001
MEM6	0.28	0.001
MEM10	0.29	0.001
P50_s_vel	0.29	0.001
MEM46	0.30	0.001
MEM2	0.31	0.001

MEM23	0.31	0.004
MEM47	0.31	0.002
MEM3	0.32	0.005
MEM38	0.32	0.007
MEM17	0.32	0.012
MEM15	0.32	0.003
MEM55	0.33	0.004
P50_s_NO3	0.33	0.006
P50_s_ppn	0.33	0.017
MEM42	0.33	0.013
MEM22	0.33	0.013
ALL 2 CLUSTERS		
MEM6	0.38	0.001
MEM7	0.48	0.001
P50_b_vel	0.55	0.001
P50_b_DIC	0.56	0.028
ALL 3 CLUSTERS		
P50_b_T	0.22	0.001
MEM1	0.30	0.001
MEM7	0.35	0.001
P50_b_NH4	0.40	0.001
P50_b_sal	0.42	0.008
P50_b_vel	0.43	0.035
P50_b_pH	0.44	0.048

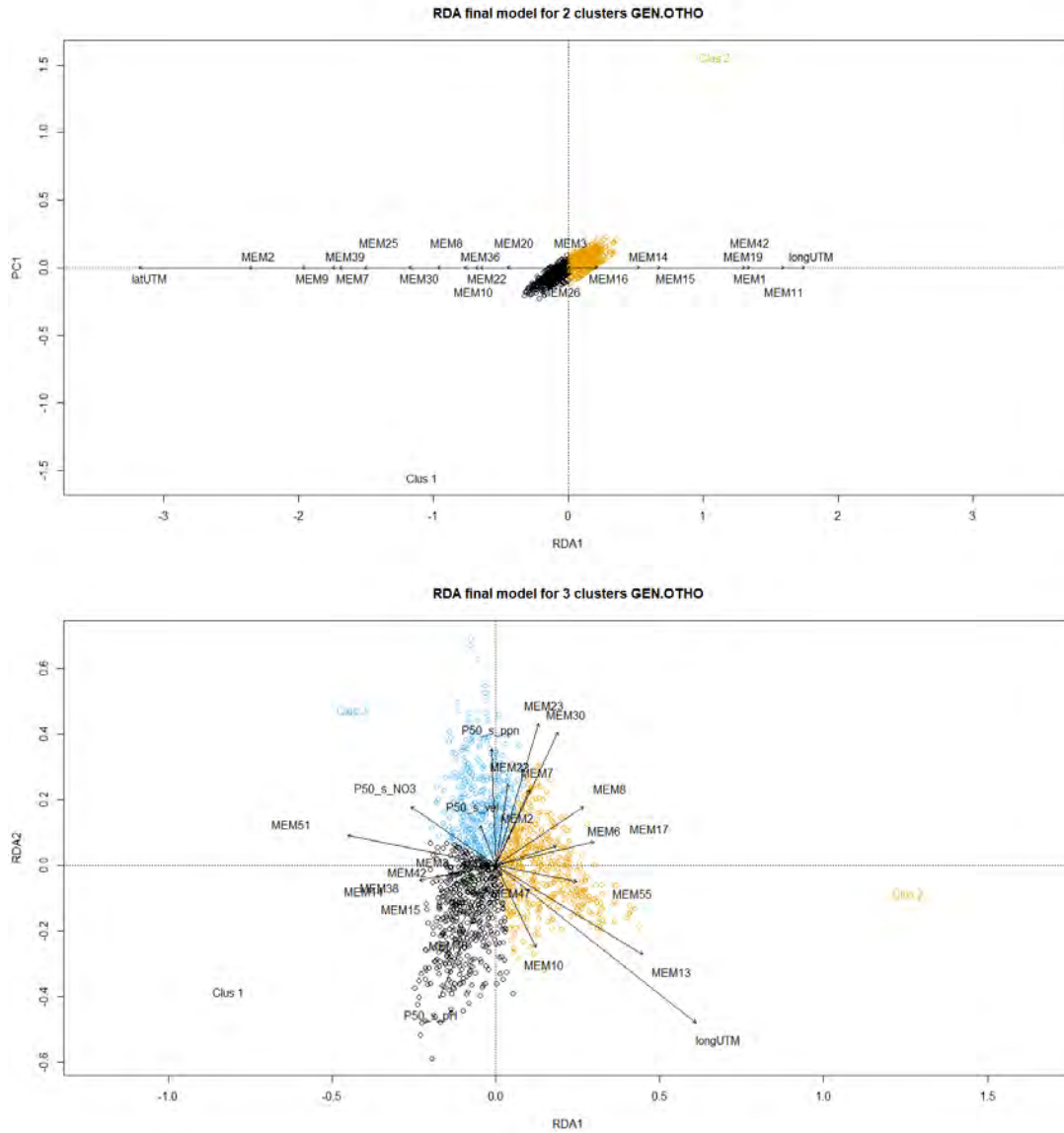


Figure 22: Final parsimonious models for GEN.OTHO dataset of *M. merluccius* at NC = 2 (above) and at NC = 3 (below). Explained variance is 0.33 on the only constrained axis at NC = 2, and 0.30 and 0.05 respectively on the first and second constrained axis at NC = 3.

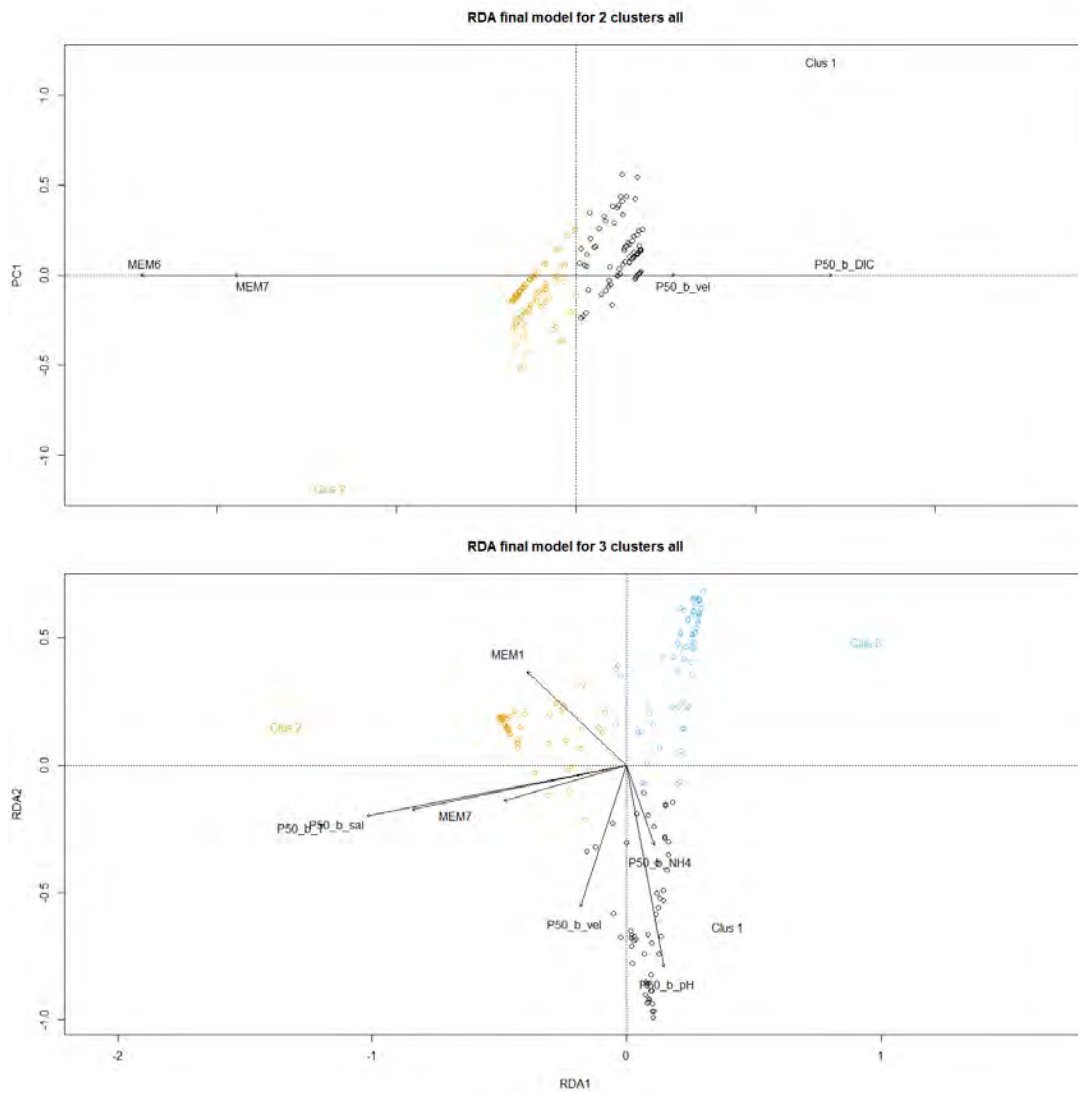


Figure 23: Final parsimonious models for ALL dataset of *M. merluccius* at NC = 2 (above) and at NC = 3 (below). Explained variance is 0.57 on the only constrained axis at NC = 2, and 0.38 and 0.09 respectively on the first and second constrained axis at NC = 3.

Conclusions

The analysis on the hake data revealed notable differences among the datasets. The GEN and OTHO datasets were those with the highest numbers of individuals analyzed and representing a high number of sampling locations and subareas. Nevertheless, they were also those with the highest variance due not only to the number of variables, but also to their inherent complexity. In particular, there is a high variability of the genotype between the individuals (see also D.1.5.1 and D.1.5.2). Much of this variance cannot be explained by the spatial and environmental variables used in the present study.

The MICRO dataset and the combinations of datasets in which MICRO variables were included was heavily limited from the point of view of the spatial coverage. Care should be taken when extrapolating results from a reduced spatial coverage to a wider area. The comparison between the MICRO dataset results and the GEN and OTHO dataset results, with its differences as to the clusterization of the individuals of the same GSA subareas, stand as a

warning. Thus, while there were some general similarities in the patterns observed in MICRO, GEN and OTHO data, there were also some differences that it is hard to reconcile.

The GEN and OTHO datasets did not exhibit a clearly recognizable structure in different clusters. The individual data seem to be distributed along a continuum, and many records were associated to several clusters. In the case of the MICRO data, the structure was much clearer, with several individuals associated strongly to one of the clusters, and a sparse number of individuals associated with similar memberships to two or more clusters. Most probably this was more the result of the sampling design of MICRO dataset, with few sampling locations wide apart one from the other, than of a supposedly better discrimination power of the MICRO variables. Nevertheless, a final word on this could be given only if the spatial coverage of the MICRO dataset would be the same as those of the GEN or OTHO datasets.

The combination of datasets as attempted in this Task is nevertheless a promising way. In fact, by combining GEN and OTHO data together, the relations with both the spatial and the environmental variables increased, as judged by their respective explanatory power on the GEN, the OTHO and the GEN.OTHO dataset. The same happened when adding GEN and OTHO to the MICRO dataset: the best explanatory power was reached for the ALL dataset. This is not an obvious result. Combining datasets means also increasing the total variance due to biases, errors and randomness in each dataset: the combined datasets have more variables and are thus more complex than each single dataset is. Thus, the observed increase in patterns interpretable by spatial and/or environmental variables when combining dataset indicates that together these different variables might better distinguish the patterns in the hake populations of the Mediterranean.

The spatial variables used in the analysis of hake data proved to be adequate to describe most of the spatially structured features of the Mediterranean, a basin showing strong spatial organization. Such spatial structure is found also in the environmental variables (temperature, salinity, nutrients, etc.), because of the spatial structuring of the processes affecting them. Thus, the spatial variables and the environmental ones shared much of their respective explanatory power. For smaller scales where spatial features are due to local processes, such as spatial autocorrelation, migrations, dispersal, etc., the MEMs proved more efficient in modelling than environmental variables computed on a fixed grid biogeochemical model.

The environmental variables used in the present study were medians over three time intervals (1, 3, and 5 years). There was no big difference in the explanatory power of the median environmental factors calculated over the three time intervals considered, and neither between the bottom and the euphotic zone environmental datasets. This was also due to the depth range of the hake and the approach used in the computation of the medians (median over the last two levels of the model for bottom value; median over the entire euphotic zone for the euphotic layer): in many coastal areas the values computed for the bottom and the euphotic zone were overlapping. The choice of considering averaged values over long period of time seem reasonable when analyzing combined datasets of different variables. A refinement of the analysis performed in this task could be to explore

also the other extracted values (5th and 95th percentiles of environmental variables), and additional time intervals. Notwithstanding the obvious inaccuracies of a whole Mediterranean 3D biogeochemical and physical model, the significant relationships found between the environmental datasets and the genetic, otolith shape and otolith microchemistry data, clearly demonstrate the validity of this approach. Model simulations are the only product by which the salient characteristics of the waters at a basin-scale can be described, summarized, and put into relations with data sampled over a discrete sampling grid.

From the results of the analysis performed in this Task, we can conclude that there is some differentiation in the hake populations of the Mediterranean. There is a difference between the Eastern Mediterranean hake populations and the populations of the rest of the basin; as well as some indications for a further differentiation in the Western-Central-Adriatic population. The observed fuzziness might be due to the mobility of the species in its different life phases, and to the absence of strong discontinuities in its habitat. These results come predominantly from the analysis of the two more robust datasets: the combination of the GEN and OTHO datasets, with the highest spatial coverage for the highest number of variables, and the ALL dataset, i.e., the combination of all three datasets, even though on a reduced sampling grid. The indications from these two study cases are consistent at NC = 2 but show some differences at NC = 3. Due to the lower spatial coverage of the ALL dataset a conclusive indication cannot be reached, and at this NC the GEN.OTHO results should be considered

Aristaeomorpha foliacea

Data

Data for *Aristaeomorpha foliacea* were made up by 20 Principal Components (PCA) extracted from the genetic data generated in WP1 (D.1.5.1).

All PCAs showed different means and standard deviations and to prevent variables with highest means and standard deviations to unduly influence the analyses, all variables were standardized before the analysis by subtracting the mean and dividing by the standard deviation.

The number of individuals with complete records for genetic PCAs was 770 sampled in 40 different locations (Fig. 24).

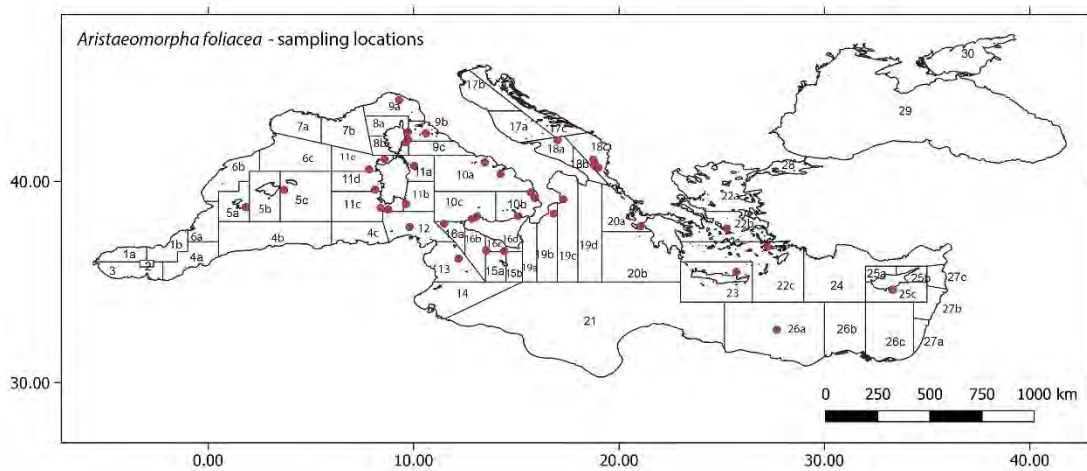


Figure 24: Sampling locations for *Aristaeomorpha foliacea*.

Results and discussion

Fuzzy clustering

For *A. foliacea*, the fuzzy k-means algorithm was calculated for NC = 2 to 10, with 100 random initializations and a parameter of fuzziness $m = 1.1$, using all provided genetic PCAs. The 6 cluster validity indexes applied did give divergent results. In fact, PC and PE recommended NC = 2 as the best number of clusters, MPC and SIL.F NC = 10 and Sil and XB NC = 9. If anything, such results might indicate that there is no clear genetic differentiation in the *A. foliacea* data.

The FKM membership grades of individuals grouped by GFCM subregion (Table 25) at NC = 2 showed a separation between the Western Mediterranean samples on one side, and the Eastern, Central and Adriatic samples on the other side. Such division remains stable up to NC = 4 clusters, and only at NC = 5 did the Adriatic samples split from the Eastern and Central. Increasing the number of clusters did reduce the average membership grade of each GFCM area in each cluster, as expected.

From the point of view of the average membership grades per GSA subarea (Table 26), it does not seem to be a spatial organization in the FKM results. Some neighboring GSA subareas were assigned to different clusters, and vice-versa very distant GSA subareas shared the same cluster (Fig. 25). Thus, the clusterization did not produce clusters that have a clear structure in space.

Table 25: Spatial interpretation of fuzzy clusters results (NC =2, 3, 4) for the *A. foliacea* genetic PCAs considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

ARS	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
CODE_GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.516	0.484	0.318	0.346	0.336	0.238	0.257	0.256	0.249
Center	0.452	0.548	0.392	0.287	0.322	0.314	0.246	0.211	0.230
Adriatic	0.498	0.502	0.345	0.321	0.334	0.276	0.217	0.257	0.250
East	0.472	0.528	0.347	0.314	0.339	0.276	0.211	0.245	0.268

Table 26: Spatial interpretation of fuzzy clusters results (NC =2, 3, 4) for the *A. foliacea* genetic PCAs considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

ARS	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
CODE_AREA	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
5a	0.530	0.470	0.291	0.371	0.338	0.221	0.198	0.290	0.290
5c	0.593	0.407	0.240	0.422	0.339	0.148	0.261	0.354	0.236
8a	0.477	0.523	0.372	0.297	0.331	0.295	0.267	0.183	0.254
8b	0.494	0.506	0.348	0.364	0.288	0.268	0.239	0.282	0.212
9a	0.509	0.491	0.309	0.314	0.376	0.243	0.260	0.222	0.275
9b	0.568	0.432	0.266	0.357	0.376	0.174	0.275	0.267	0.284
9c	0.448	0.552	0.438	0.327	0.235	0.363	0.206	0.260	0.171
10a	0.698	0.302	0.149	0.515	0.336	0.087	0.341	0.352	0.220
10b	0.569	0.431	0.266	0.406	0.328	0.178	0.314	0.276	0.233
10c	0.449	0.551	0.381	0.295	0.324	0.316	0.208	0.222	0.255
11a	0.410	0.590	0.433	0.291	0.276	0.318	0.245	0.218	0.218
11b	0.465	0.535	0.368	0.318	0.314	0.296	0.242	0.241	0.221
11c	0.570	0.430	0.266	0.346	0.389	0.188	0.288	0.235	0.289
11d	0.456	0.544	0.380	0.343	0.277	0.280	0.260	0.282	0.178
11e	0.467	0.533	0.353	0.315	0.332	0.269	0.240	0.257	0.234
12	0.514	0.486	0.349	0.309	0.342	0.264	0.289	0.203	0.244
13	0.340	0.660	0.499	0.205	0.296	0.397	0.269	0.144	0.190
16a	0.382	0.618	0.430	0.262	0.308	0.347	0.275	0.178	0.200
16c	0.459	0.541	0.357	0.290	0.354	0.279	0.238	0.204	0.278
16b	0.506	0.494	0.337	0.301	0.363	0.262	0.275	0.198	0.265
18a	0.511	0.489	0.300	0.334	0.365	0.239	0.251	0.257	0.253
18c	0.476	0.524	0.417	0.300	0.283	0.336	0.162	0.257	0.245
19b	0.384	0.616	0.484	0.309	0.207	0.429	0.186	0.252	0.132
19c	0.453	0.547	0.406	0.341	0.253	0.349	0.146	0.333	0.173
20a	0.466	0.534	0.372	0.279	0.349	0.302	0.226	0.213	0.259
22b	0.556	0.444	0.292	0.433	0.276	0.211	0.223	0.361	0.205
22c	0.377	0.623	0.396	0.199	0.406	0.326	0.215	0.127	0.332
23	0.444	0.556	0.364	0.350	0.286	0.279	0.239	0.296	0.185

25c	0.498	0.502	0.321	0.307	0.371	0.255	0.197	0.226	0.323
26a	0.455	0.545	0.410	0.299	0.291	0.353	0.197	0.256	0.194

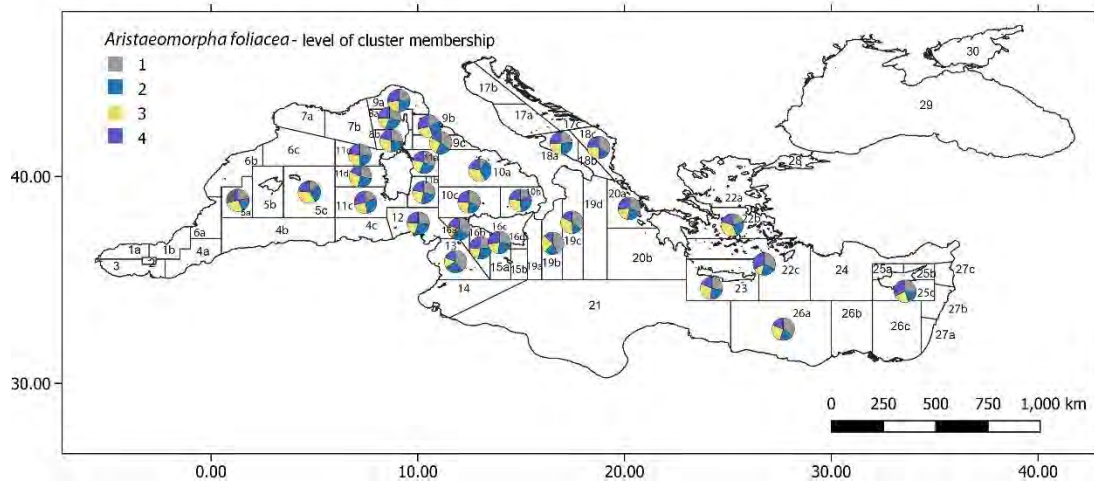
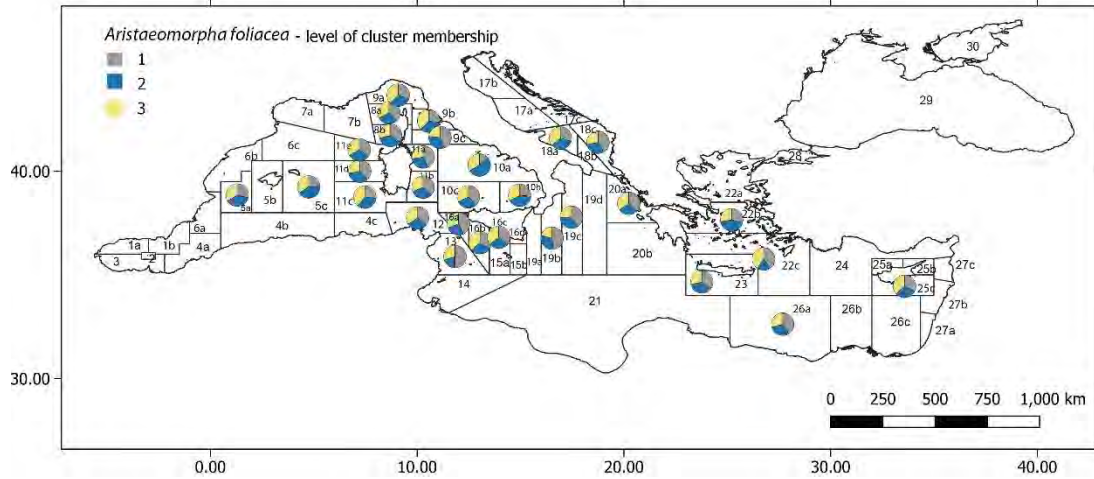
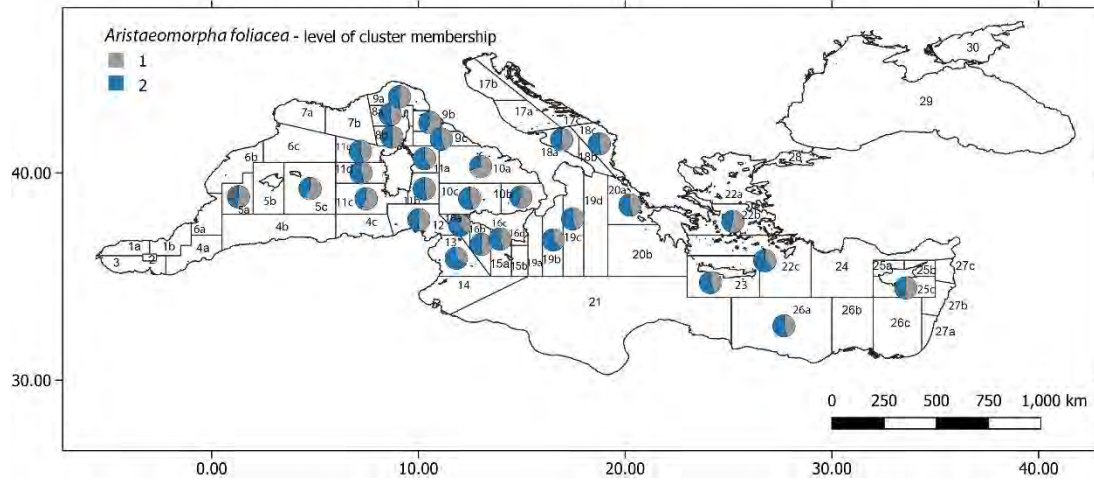


Figure 25: Pie charts with average membership grades per GSA subarea for *A. foliacea* genetic dataset at NC = 2 (upper panel), NC = 3 (middle panel), and NC = 4 (lower panel).

Spatial analysis

The spatial analysis was performed taking into account the possible explanatory power on the fuzzy membership grades of depth, metric geographic coordinates (longitude and latitude), and of possible significant MEMs (Table 27).

Table 27: Spatial analysis results for A. foliacea. The explanatory power of each spatial variable or group of variables is reported as adjusted R². All reported models with depth and LongLat were tested for significance with an ANOVA and “ns” indicates those that were not significant. The significance of MEMs was tested with a forward selection procedure and a double-stopping criterion. Only significant MEMs are reported. For MEMs also the number of significant MEMs, their names and the winner spatial weighting matrix are reported.

ARS	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	ns	ns	0.00	4	MEM13, MEM16, MEM18, MEM17	Relative_Down_5
2 CLUSTERS	ns	0.00	0.04	6	MEM8, MEM5, MEM9, MEM6, MEM16, MEM11	Delaunay_Up_0.9
3 CLUSTERS	ns	ns	0.02	3	MEM5, MEM9, MEM16	Delaunay_Up_0.7
4 CLUSTERS	ns	ns	0.01	5	MEM13, MEM9, MEM15, MEM5, MEM14	Gabriel_Up_0.7
5 CLUSTERS	ns	ns	0.00	2	MEM15, MEM17	MST_Down_4
6 CLUSTERS	ns	ns	0.00	1	MEM6	Relative_Up_0.5
7 CLUSTERS	ns	ns	0.00	2	MEM14, MEM10	Gabriel_Up_0.1
8 CLUSTERS	ns	ns	ns			
9 CLUSTERS	ns	ns	0.00	1	MEM5	Gabriel_Up_0.3
10 CLUSTERS	ns	ns	0.00	2	MEM9, MEM15	MST_Up_0.5

As it can be seen from Table 27 *A. foliacea* genetic data do not seem to be spatially organized. Depth was never a significant explanatory variable, nor longitude and latitude, with the exception for 2 clusters where they explained a very low amount of variance. So did also MEMs: while at least some MEMs were always selected as significant, but for NC = 8, the amount of variance explained was very close to 0 in all cases. Only at NC = 2 the adjusted R² reached a value of 0.04, at NC = 3 was 0.02, and at NC = 4 was 0.01.

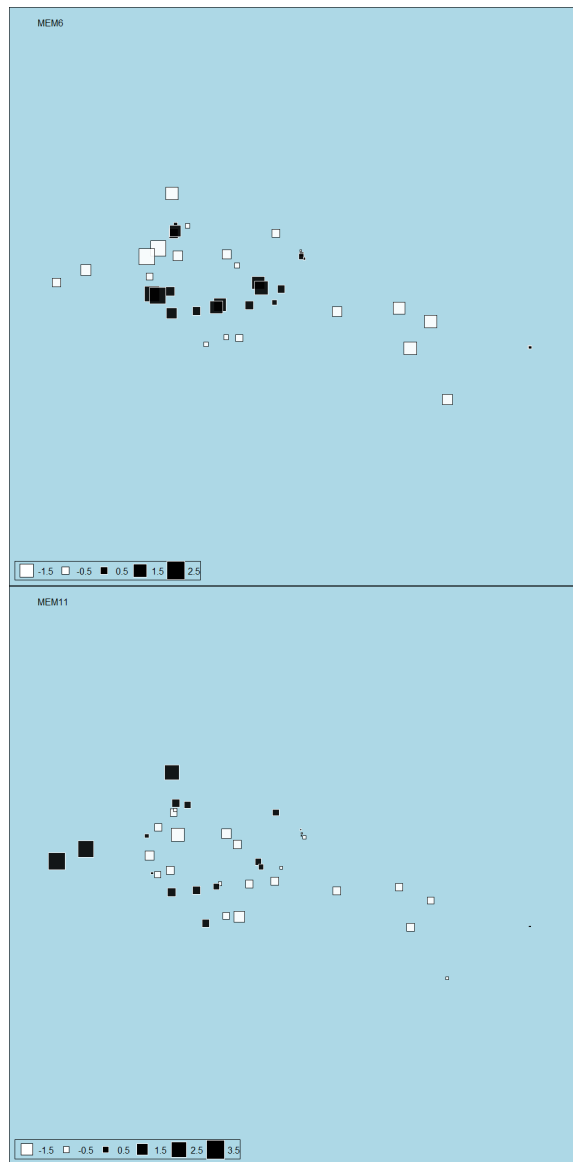
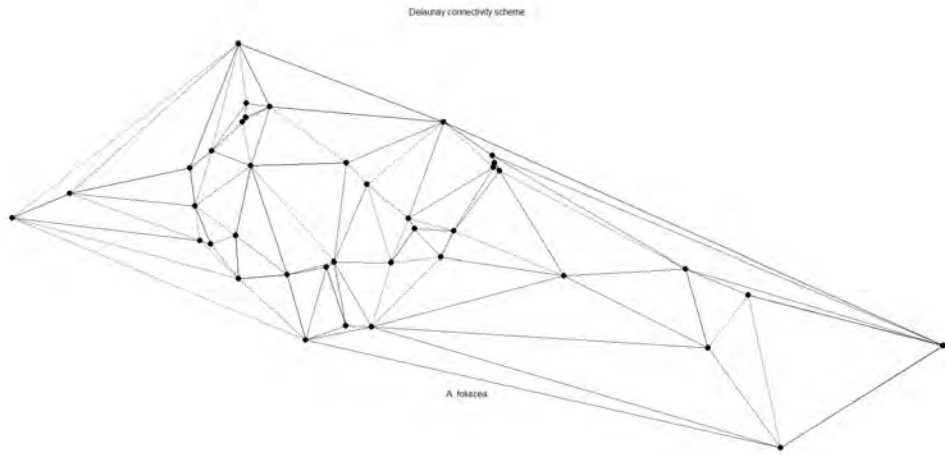


Figure 26: The Delaunay connectivity scheme for *A. foliaceae* sampling locations (upper panel), and two MEMs that resulted significant in explaining its genetic variation at $NC = 2$ (see text): MEM6 (middle panel) and MEM11 (lower panel).

Environmental analysis

Using bottom and euphotic zone (“surface”) median values of environmental variables at all three time intervals (Table 28) did not give interesting results. In fact, the explanatory power of the whole set of 13 environmental variables was always low, and in the vast majority of cases non-significant. There was not much difference when using the three different time intervals, but the bottom dataset showed better relationships with *A. foliaceae* dataset than the euphotic zone one. The best results were obtained for the clusterization with NC = 2 with the bottom 3 years dataset that explained an adjusted $R^2 = 0.04$. A forward selection selected median bottom velocity and nitrate concentration as the only two significant environmental variables.

Table 28: Adjusted R^2 for *A. foliaceae*, calculated for RDA models with environmental variables as explanatory variables. Different models were built for each set of explanatory variables: bottom and surface (euphotic zone), and the three time intervals considered (5 years, 3 years, 1 year). Results are reported for each original dataset, and the clusterizations from 2 to 10 clusters.

ARS	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.00	ns	ns	0.03	0.04	0.03	0.01	0.01	0.01	0.01	0.01	0.01	ns	ns	0.00
surface	0.00	0.00	ns	ns	0.02	0.01	0.01	0.01	ns	0.01	0.01	ns	ns	ns	ns
	6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
surface	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

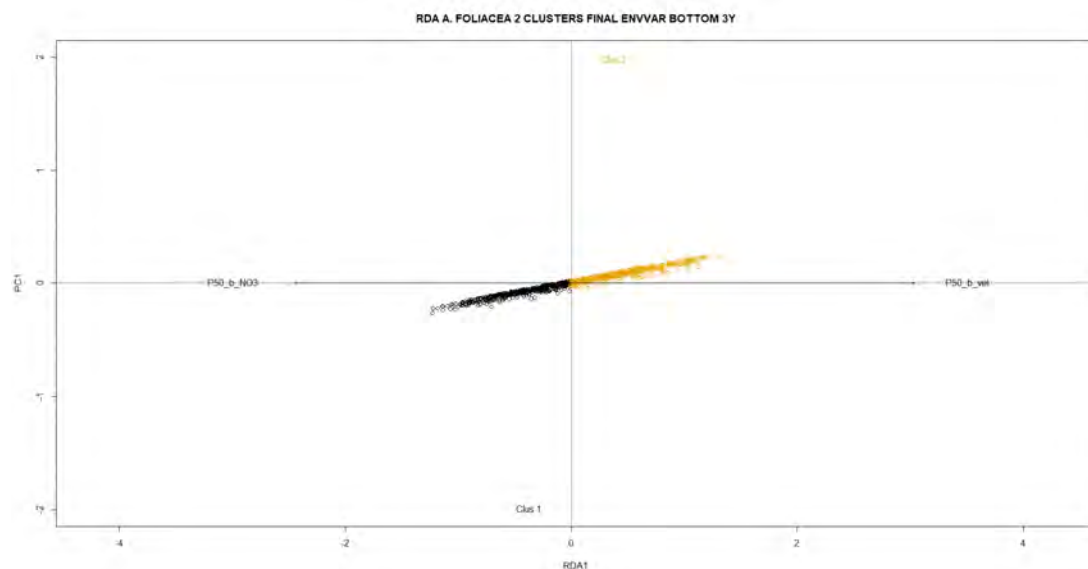


Figure 27: The final model with significant environmental variables explaining 0.03 of *A. foliaceae* genetic variance on the only constrained axis.

Variation partitioning and best model selection

A final model was built by combining the significant MEMs and the set of median bottom environmental variables over 3 years. Variation partitioning was applied on these two

groups of variables (Fig. 28): much of the explanatory power was shared between significant MEMs and the set of environmental variables used.

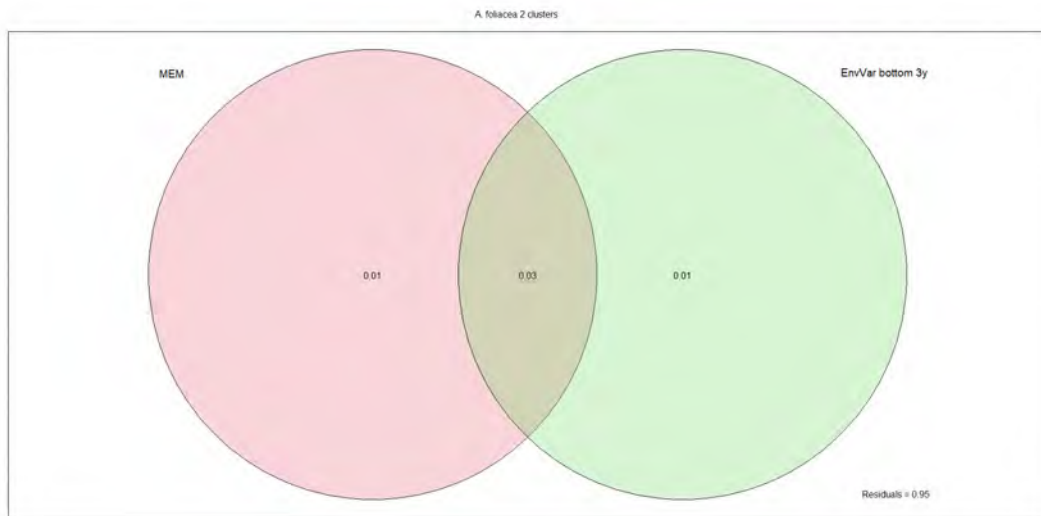


Figure 28: Variation partitioning with the explanatory power of the selected MEM and of the set of bottom environmental variables at time interval = 3 years.

The final model was also subject to forward selection and did select only three variables: two MEMs (MEM6 and MEM11) (Fig. 26) and bottom velocity. Bottom velocity was related to cluster 2, while the two MEMs were related to high cluster 1 membership grades (Fig. 29).

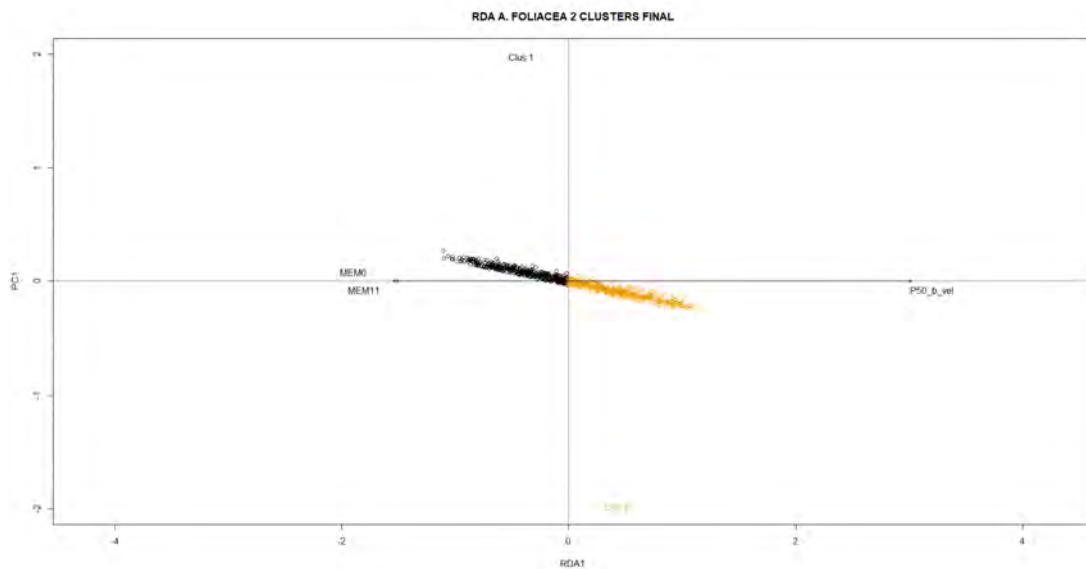


Figure 29: The final model with significant spatial and environmental variables explaining 0.04 of *A. foliacea* genetic variance on the only constrained axis.

Conclusions

The genetic PCAs of *A. foliacea* showed a very low differentiation in different groups, as revealed already by the internal cluster validity indexes applied. Furthermore, also the external validation procedures, i.e., the spatial and environmental analysis, showed that the possible sub-populations identified with the fuzzy k-means were not related neither to spatial nor to environmental features. In fact, at NC = 2 to NC = 4 it seems that the groups were not organized spatially. This was confirmed by the low explanatory power of MEMs, and no statistically significant effect of depth or longitude and latitude. The explanatory power of the environmental variables was also low and was significant, if at all, only for lower NC values (2 to 4).

All this points to the fact that for *A. foliacea* in the Mediterranean based on the data of the present study and with the methods here applied, no different stocks could have been identified. This might be due to the absence of discontinuities in the habitat of this species or to its capacity to overcome such discontinuities with an efficient dispersal thanks to the convective currents in its depth range.

Aristeus antennatus

Data

Data for *Aristeus antennatus* were made up by 200 Principal Components (PCA) extracted from the genetic data generated in WP1 D.1.5.1).

All PCAs showed different means and standard deviations and to prevent variables with highest means and standard deviations to unduly influence the analyses, all variables were standardized before the analysis by subtracting the mean and dividing by the standard deviation.

The number of individuals with complete records for genetic PCAs was 832 sampled in 45 different locations (Fig. 29).

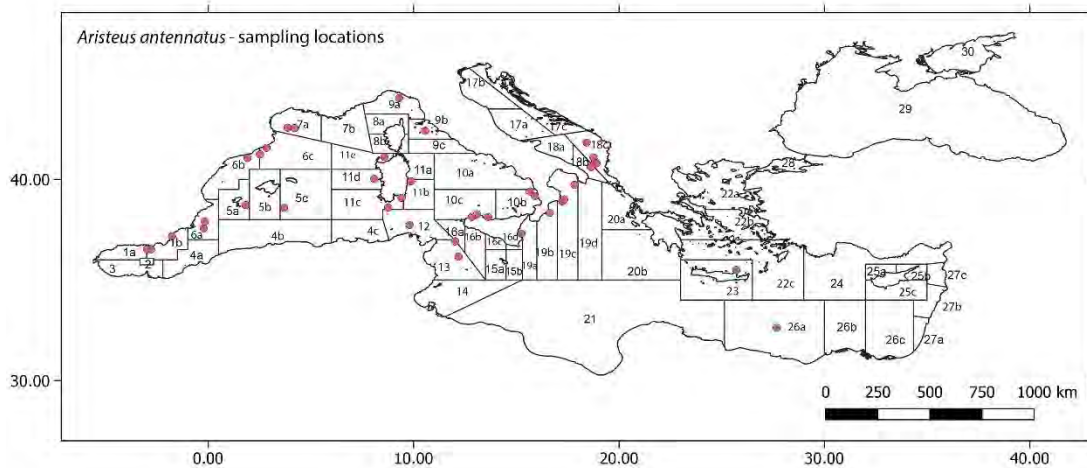


Figure 29: Sampling locations for *Aristeus antennatus*.

Results and discussion

Fuzzy clustering

The fuzzy k-means ($m=1.1$, 100 random initializations) was applied on the first 20 PCAs: it emerged from several trials performed that an only slightly larger number of PCAs (such as, e.g., 25) would produce equally distributed cluster membership grades for any NC and already with a parameter of fuzziness $m = 1.1$ ($m = 1$ is the value for which a “hard” partition is obtained, and the parameter of fuzziness can be the greater the stronger is the structuring in different groups in the data).

The 6 internal cluster validity indexes were again in strong disagreement: the PC and PE measures indicated in $NC = 2$ the best cluster number, while all others indicated that $NC = 10$ should be the choice. The fact that the two extremes were favored indicates that there is no genetic diversification in the data that could be detected by the clustering.

When looking at the spatial distribution in function of GFCM subregions (Table 29) it seems that no clear pattern emerged, in particular considering the Western and Central part of the

Mediterranean. Only the Eastern part seems consistently different from the others, but it is based on two GSA subregions only (23 and 26a).

Table 29: Spatial interpretation of fuzzy clusters results (NC = 2 and 3) for the *A. antennatus* genetic PCAs considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

ARA	2 CLUSTERS		3 CLUSTERS		
CODE_GFCM	FKM1	FKM2	FKM1	FKM2	FKM3
West	0.50	0.50	0.33	0.33	0.34
Center	0.49	0.51	0.34	0.33	0.33
Adriatic	0.49	0.51	0.34	0.36	0.30
East	0.42	0.58	0.44	0.24	0.32

The inspection of the average membership grades at the level of GSA subregions for NC = 2 and NC = 3 also did not show a clear spatial pattern. Many neighboring subregions had different cluster preferences, while some clusters (e.g., FKM2 at NC = 2) did characterize both extremes of the Mediterranean basin. Furthermore, in many cases (Fig. 31), the membership grades were equally distributed between all clusters.

Table 30: Spatial interpretation of fuzzy clusters results (NC = 2 and 3) for the *A. antennatus* genetic PCAs considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

ARA	2 CLUSTERS		3 CLUSTERS		
CODE_AREA	FKM1	FKM2	FKM1	FKM2	FKM3
1b	0.49	0.51	0.36	0.24	0.40
5a	0.55	0.45	0.29	0.38	0.33
5c	0.51	0.49	0.33	0.34	0.33
6a	0.48	0.52	0.32	0.31	0.36
6b	0.50	0.50	0.33	0.34	0.33
6c	0.44	0.56	0.38	0.30	0.31
7a	0.45	0.55	0.39	0.31	0.31
9a	0.48	0.52	0.37	0.32	0.31
9b	0.61	0.39	0.23	0.34	0.42
10b	0.51	0.49	0.35	0.36	0.29
10c	0.54	0.46	0.29	0.38	0.33
11b	0.53	0.47	0.29	0.24	0.47
11c	0.51	0.49	0.30	0.36	0.34
11d	0.49	0.51	0.36	0.32	0.32
11e	0.49	0.51	0.35	0.34	0.31
12	0.46	0.54	0.36	0.38	0.25
13	0.48	0.52	0.38	0.33	0.29
18b	0.48	0.52	0.35	0.33	0.32
18c	0.51	0.49	0.34	0.39	0.27
19a	0.49	0.51	0.35	0.41	0.24
19b	0.51	0.49	0.30	0.34	0.36
19c	0.49	0.51	0.36	0.27	0.37
19d	0.52	0.48	0.33	0.31	0.36
23	0.46	0.54	0.45	0.25	0.30
26a	0.41	0.59	0.44	0.24	0.32

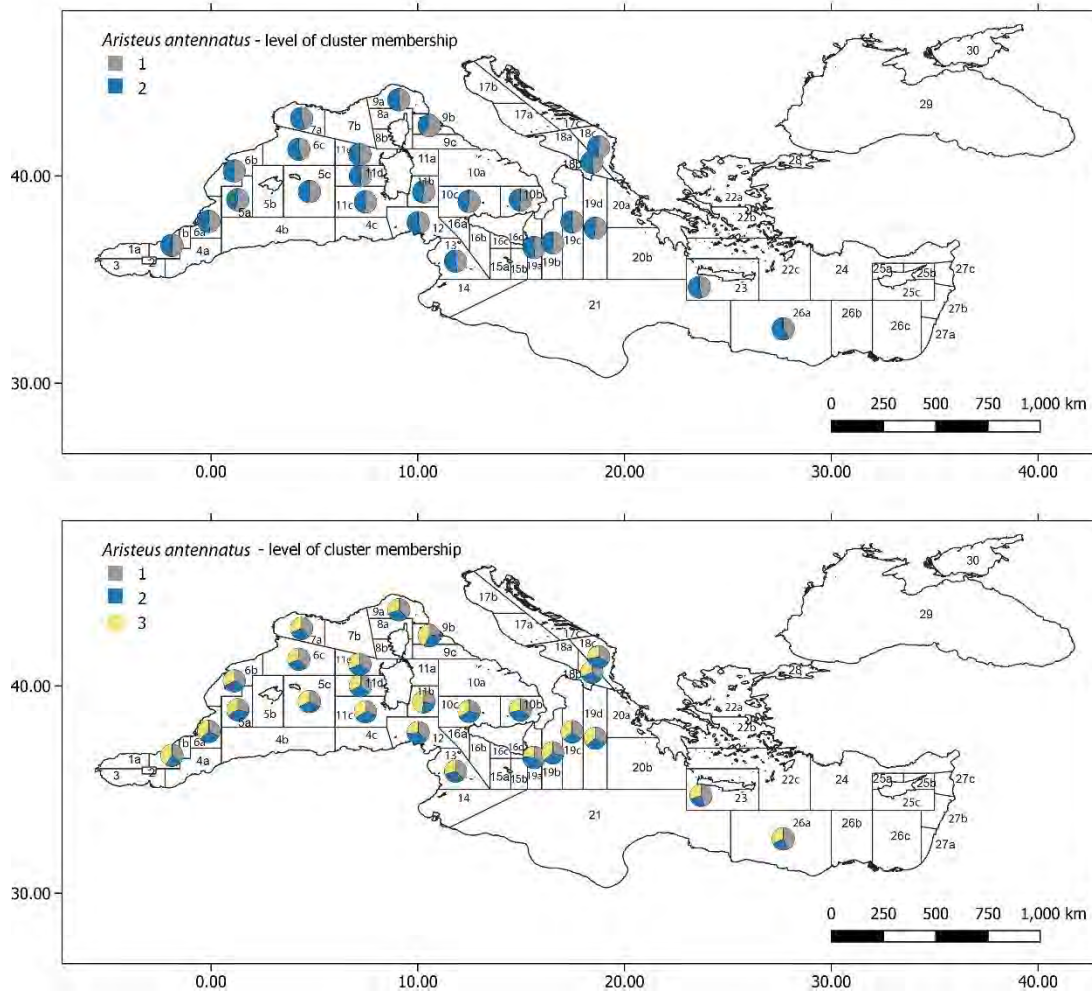


Figure 31: Pie charts with average membership grades per GSA subarea for *A. antennatus* genetic dataset at NC = 2 (upper panel) and NC = 3 (lower panel).

Spatial analysis

For the spatial analysis depth, longitude and latitude (metric coordinates) and selected MEMs were used (Table 31). The explanatory power of all of these on the fuzzy k-means membership grades obtained from genetic PCAs, as well as on the original genetic PCAs, was extremely low, never exceeding 0.01 (explanatory power of significant MEMs for NC = 2 and NC = 3). This confirmed that the genetic variability in *A. antennatus* does not have a spatial organization, which in turn means that there are no different population of this species in the Mediterranean.

Table 31: Spatial analysis results for *A. antennatus*. The explanatory power of each spatial variable or group of variables is reported as adjusted R^2 . All reported models with depth and LongLat were tested for significance with an ANOVA and "ns" indicates those that were not significant. The significance of MEMs was tested with a forward

selection procedure and a double-stopping criterion. Only significant MEMs are reported. For MEMs also the number of significant MEMs, their names and the winner spatial weighting matrix are reported.

ARA	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	0.00	0.00	4	MEM5, MEM13, MEM4, MEM16	MST_Up_0.5
2 CLUSTERS	ns	ns	0.01	2	MEM2, MEM5	Delaunay_Up_0.6
3 CLUSTERS	ns	ns	0.01	4	MEM7, MEM2, MEM10, MEM6	Delaunay_Down_10
4 CLUSTERS	ns	ns	0.00	1	MEM16	Delaunay_Up_1
5 CLUSTERS	ns	ns	0.00	1	MEM18	Relative_Up_1
6 CLUSTERS	ns	0.00	ns			
7 CLUSTERS	ns	0.00	0.00	1	MEM2	Delaunay_Up_0.1
8 CLUSTERS	ns	ns	ns			
9 CLUSTERS	0.00	0.00	ns			
10 CLUSTERS	0.00	0.00	0.00	2	MEM8, MEM4	Delaunay_Down_2

Environmental analysis

The bottom and euphotic zone (“surface”) median values of environmental variables at all three time intervals (Table 32) were not related to the distribution of *A. antennatus* in populations in the Mediterranean. The highest adjusted $R^2 = 0.01$ was obtained for NC = 2 with bottom medians over 3 years and for NC = 7 with bottom medians over 5 years. Euphotic zone medians were always non-significant but for those over 3 years at NC = 3. Thus, not only the distribution of *A. antennatus* populations does not have a spatial structure, but it is also not related to the environmental variables used in the present study.

Table 32: Adjusted R^2 for *A. antennatus*, calculated for RDA models with environmental variables as explanatory variables. Different models were built for each set of explanatory variables: bottom and surface (euphotic zone), and the three time intervals considered (5 years, 3 years, 1 year). Results are reported for each original dataset, and the clusterizations from 2 to 10 clusters.

ARA	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	ns	ns	ns	ns	ns	ns	ns	0.01	ns	ns	ns	ns	ns	ns	ns
surface	ns	ns	ns	ns	ns	ns	ns	0.01	ns	ns	ns	ns	ns	ns	ns
	6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	ns	ns	0.00	0.01	0.00	0.00	ns	ns	ns	ns	ns	0.00	ns	ns	0.00

surface	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
---------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Variation partitioning and best model selection

Given the low and mostly non-significant explanatory power of the models of *A. antennatus* population distributions built with either the spatial or the environmental variables, no attempt was made with the variation partitioning procedure and with the construction of a best final model.

Conclusions

Based on the data of the present study and the methods applied here, the populations of *A. antennatus* in Mediterranean do not show a genetic structure, neither when judging from internal validity indexes, nor when considering the results of the spatial and environmental analysis. Complex patterns of dispersal corridors or discontinuities in the habitat of *A. antennatus* might cause complex patterns of distribution. The MEMs should have been able to detect any spatial pattern, at local (small) scale or at a regional (large) scale, notwithstanding the complexity of the drivers causing it. Thus, the most probable hypothesis based on the genetic PCAs provided and the methods applied here is that *A. antennatus* in the Mediterranean cannot be divided in different stocks.

Nephrops norvegicus

Data

The genetic dataset for *Nephrops norvegicus* was made up by 20 Principal Components (PCA) extracted from the genetic data generated in WP1 (D.1.5.2).

All PCAs showed different means and standard deviations and to prevent variables with highest means and standard deviations to unduly influence the analyses, all variables were standardized before the analysis by subtracting the mean and dividing by the standard deviation.

The number of individuals with complete records for genetic PCAs was 885 sampled in 69 different locations (Fig. 32).

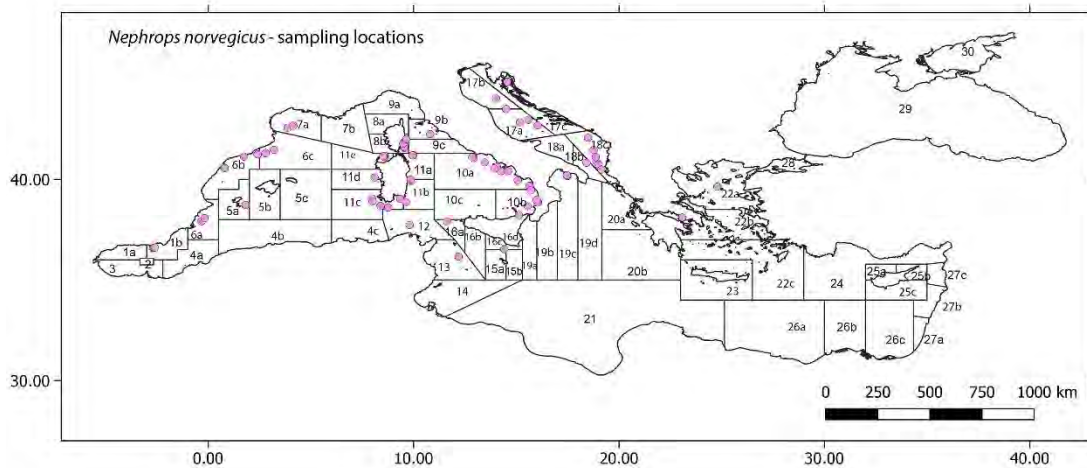


Figure 32: Sampling locations for *Nephrops norvegicus*.

Results and discussion

Fuzzy clustering

The fuzzy k-means was applied on the 20 genetic PCAs with $m=1.1$ and 100 random initializations. The 6 internal cluster validity indexes were in strong disagreement: the PC, PE, and SIL measures indicated in $NC = 2$ the best cluster number, while MPC, SIL.F and XB indexes indicated in $NC = 10$ the best cluster number. This result might indicate that the genetic diversification in *N. norvegicus* is not very strong.

Checking the results for $NC = 2$ to $NC = 10$, it seems that for the lower and the higher end the spatial organization is less clear than for the $NC = 4$ to $NC = 7$. Thus, these results are presented in Tables 33-34 aggregated per GFCM subregion, and in Tables 35-36 aggregated per GSA subarea.

The spatial distribution in function of GFCM subregions did not give a clear picture. In fact, this could have been anticipated, since the number of clusters explored is equal ($NC = 4$) or greater ($NC = 5-7$) than the number of GFCM subregions. The West and Center subregions

are those more dispersed among the clusters at the different NC, while the Adriatic and the East subregions show a clear preference for a specific cluster up to NC = 6 (Adriatic) and up to NC = 7 (East). The interpretation of these results for the Eastern basins has to take into account that only two subareas were sampled in this subregion. We can infer that the fragmentation of the populations is greater in the Western and the Central basin of the Mediterranean, while the Adriatic shows a consistent homogeneity of the population of *N. norvegicus*.

Table 33: Spatial interpretation of fuzzy clusters results (NC = 4-5) for the *N. norvegicus* genetic PCAs considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

NEP	4 CLUSTERS				5 CLUSTERS				
CODE_GFCM	FKM1	FKM2	FKM3	FKM4	FKM1	FKM2	FKM3	FKM4	FKM5
West	0.31	0.08	0.30	0.31	0.24	0.24	0.24	0.23	0.05
Center	0.17	0.31	0.28	0.25	0.16	0.25	0.21	0.12	0.26
Adriatic	0.12	0.60	0.15	0.13	0.10	0.13	0.11	0.11	0.55
East	0.14	0.56	0.16	0.13	0.12	0.13	0.11	0.13	0.52

Table 34: Spatial interpretation of fuzzy clusters results (NC = 6-7) for the *N. norvegicus* genetic PCAs considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

NEP	6 CLUSTERS						7 CLUSTERS						
CODE_GFCM	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6	FKM7
West	0.19	0.20	0.19	0.20	0.19	0.03	0.18	0.05	0.18	0.18	0.18	0.03	0.19
Center	0.17	0.16	0.21	0.15	0.09	0.23	0.15	0.20	0.15	0.19	0.07	0.11	0.13
Adriatic	0.11	0.09	0.11	0.09	0.09	0.50	0.06	0.35	0.09	0.08	0.07	0.30	0.07
East	0.09	0.11	0.12	0.10	0.11	0.47	0.05	0.11	0.05	0.08	0.06	0.60	0.06

A look at the individual membership grades averaged over GSA subareas (Tables 35-36) does reveal interesting facts. First, the membership grades tend to indicate clear preferences and clear “dislikes” for specific clusters for many subareas. That is to say that the differentiation among at least some clusters is quite clear and up to a certain point it follows a spatial organization. The maximum membership grade is generally much higher than the value of $1/NC$ for each NC. This is evident even at NC = 7 where the winner membership grades oscillate between 0.22 and 0.60, being thus always much higher than $1/7 = 0.14$. Such results support the hypothesis that *N. norvegicus* population in the Mediterranean shows a clear genetic differentiation in sub-populations. In many cases, two or more clusters show similar membership grades, but other two or more show a much lower value. Considering these results, it may be possible to build a matrix of relationships between the different sub-populations in the GSA subareas.

The most solid result (Fig. 33-34) is the separation of the Adriatic populations from the others. They show high membership grades for a specific cluster together with subareas 19c and 22a up to NC = 6. At NC = 7 they split into two clusters, subareas 17b, 18b, 18c and 19c in FKM2, subareas 17a, 17c and 22a in FKM6. This would point out to the possible coexistence of slightly different sub-populations of *N. norvegicus* in the Adriatic. Nevertheless, the spatial density of sampling locations it is not detailed enough to make further hypotheses on this.

Subarea 20a showed a much different behavior from the rest of the Adriatic-Ionian-Eastern Mediterranean subareas. It was usually associated to some Western and Central subareas. The sampling location in this case was deep inside the Corinth Gulf, thus in conditions of great isolation from the rest of the waters of the Mediterranean, but potentially connected to the Aegean Sea through the Corinth Canal. This might explain the difference in its *N. norvegicus* population compared to the neighboring areas. Because of its position for this sampling location the environmental variables were not extracted, thus its membership grades will not be used in the environmental RDA.

As for the subareas of the Western Mediterranean, they show much more dispersed membership grades among all those clusters that are not the “Adriatic” one. In particular, at NC = 4 there seem to be 3 sub-populations intermixing in the Western Mediterranean (Fig. 33). FKM3 is mostly present along the Spanish coast, reaching Sardinia and subarea 13 on the African coastline. Cluster FKM4 is present from the Balearic Islands to the French coast (7a), southern Sardinia and Tunisia, and Italian coastline (9c and 10a). FKM1 is mostly limited to subareas between Corsica and Sardinia and the Italian coastline (9b in the North and 10b in the South). At increasing NC, the spatial distribution of the clusters involving Western Mediterranean subareas is harder to follow and the differences in membership grades among the clusters reduce progressively.

While the most solid result is the separation of the Adriatic from the Western Mediterranean, it has to be noted that this happens only at NC = 4, not before. This means that the three clusters identified in the Western Mediterranean show a level of differentiation that is comparable to the one shown by the Adriatic samples. Thus, while the spatial distribution of the three clusters in the Western Mediterranean seems less easily interpretable, perhaps because of a sparser sampling design in this area not able to adequately capture this distribution, all three these clusters should be considered if the Adriatic one is considered.

Table 35: Spatial interpretation of fuzzy clusters results (NC = 4-5) for the *N. norvegicus* genetic PCAs considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

NEP	4 CLUSTERS				5 CLUSTERS				
CODE_AREA	FKM1	FKM2	FKM3	FKM4	FKM1	FKM2	FKM3	FKM4	FKM5
1b	0.25	0.13	0.36	0.25	0.18	0.30	0.21	0.23	0.09
5a	0.18	0.10	0.31	0.40	0.20	0.25	0.35	0.14	0.07
6a	0.31	0.09	0.32	0.28	0.24	0.28	0.24	0.18	0.06
6b	0.29	0.08	0.34	0.29	0.21	0.30	0.23	0.21	0.05
6c	0.20	0.07	0.38	0.36	0.19	0.31	0.29	0.17	0.03
7a	0.33	0.08	0.25	0.34	0.29	0.19	0.29	0.18	0.05
8b	0.40	0.07	0.20	0.33	0.28	0.13	0.25	0.30	0.03
9b	0.35	0.07	0.31	0.28	0.27	0.19	0.20	0.31	0.03
9c	0.29	0.08	0.30	0.34	0.20	0.21	0.23	0.32	0.04
10a	0.29	0.09	0.26	0.36	0.24	0.19	0.28	0.24	0.05
10b	0.38	0.07	0.29	0.25	0.26	0.24	0.21	0.24	0.04
11a	0.35	0.09	0.30	0.26	0.24	0.23	0.20	0.29	0.05
11b	0.18	0.09	0.24	0.50	0.16	0.18	0.31	0.31	0.05

11c	0.31	0.08	0.29	0.32	0.29	0.23	0.27	0.17	0.04
11d	0.29	0.08	0.35	0.28	0.21	0.27	0.20	0.28	0.04
11e	0.38	0.09	0.26	0.27	0.30	0.20	0.18	0.27	0.05
12	0.15	0.06	0.31	0.48	0.15	0.27	0.42	0.13	0.04
13	0.27	0.06	0.44	0.23	0.23	0.41	0.17	0.16	0.03
16c	0.21	0.12	0.32	0.36	0.21	0.23	0.32	0.15	0.08
17a	0.14	0.61	0.15	0.11	0.10	0.14	0.11	0.11	0.53
17b	0.12	0.59	0.19	0.11	0.08	0.16	0.10	0.12	0.54
17c	0.09	0.70	0.11	0.10	0.08	0.10	0.08	0.08	0.66
18b	0.12	0.64	0.11	0.13	0.11	0.09	0.12	0.08	0.60
18c	0.15	0.53	0.17	0.15	0.11	0.15	0.12	0.16	0.45
19c	0.10	0.60	0.13	0.17	0.11	0.11	0.16	0.08	0.54
20a	0.16	0.18	0.48	0.18	0.13	0.44	0.16	0.17	0.10
22a	0.14	0.56	0.16	0.13	0.12	0.13	0.11	0.13	0.52

Table 36: Spatial interpretation of fuzzy clusters results (NC = 6-7) for the *N. norvegicus* genetic PCAs considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

NEP	6 CLUSTERS						7 CLUSTERS						
CODE_AREA	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6	FKM1	FKM2	FKM3	FKM4	FKM5	FKM6	FKM7
1b	0.16	0.16	0.25	0.16	0.20	0.07	0.15	0.06	0.14	0.23	0.17	0.12	0.13
5a	0.16	0.20	0.20	0.31	0.09	0.05	0.20	0.10	0.14	0.16	0.07	0.05	0.27
6a	0.24	0.18	0.19	0.21	0.15	0.03	0.16	0.04	0.22	0.19	0.13	0.06	0.21
6b	0.24	0.17	0.23	0.17	0.16	0.03	0.16	0.05	0.23	0.23	0.16	0.02	0.16
6c	0.18	0.18	0.25	0.24	0.14	0.02	0.17	0.03	0.16	0.25	0.13	0.03	0.23
7a	0.21	0.19	0.15	0.26	0.15	0.03	0.17	0.05	0.21	0.13	0.15	0.03	0.26
8b	0.12	0.23	0.11	0.24	0.27	0.02	0.21	0.04	0.12	0.10	0.26	0.02	0.24
9b	0.15	0.23	0.17	0.17	0.26	0.02	0.23	0.04	0.14	0.16	0.25	0.01	0.17
9c	0.21	0.14	0.18	0.19	0.25	0.02	0.14	0.03	0.21	0.17	0.24	0.02	0.17
10a	0.21	0.19	0.13	0.24	0.20	0.04	0.18	0.04	0.20	0.12	0.19	0.05	0.22
10b	0.21	0.21	0.20	0.14	0.21	0.03	0.22	0.04	0.20	0.19	0.21	0.02	0.12
11a	0.19	0.16	0.19	0.17	0.25	0.03	0.14	0.05	0.19	0.19	0.24	0.03	0.16
11b	0.09	0.26	0.19	0.19	0.25	0.03	0.31	0.03	0.06	0.20	0.22	0.03	0.14
11c	0.22	0.21	0.19	0.23	0.13	0.02	0.20	0.06	0.21	0.17	0.13	0.01	0.22
11d	0.18	0.19	0.21	0.15	0.23	0.03	0.18	0.05	0.16	0.22	0.21	0.03	0.14
11e	0.12	0.26	0.19	0.17	0.23	0.04	0.23	0.07	0.12	0.17	0.22	0.03	0.16
12	0.28	0.14	0.27	0.20	0.09	0.02	0.15	0.03	0.28	0.21	0.06	0.08	0.18
13	0.16	0.28	0.32	0.10	0.12	0.02	0.28	0.02	0.13	0.33	0.11	0.04	0.08
16c	0.22	0.20	0.16	0.27	0.11	0.04	0.18	0.03	0.19	0.15	0.10	0.07	0.28
17a	0.08	0.11	0.11	0.08	0.10	0.52	0.09	0.24	0.07	0.11	0.08	0.34	0.07
17b	0.10	0.07	0.16	0.08	0.10	0.48	0.05	0.46	0.07	0.11	0.08	0.19	0.05
17c	0.08	0.07	0.12	0.06	0.06	0.62	0.04	0.35	0.05	0.05	0.03	0.45	0.03
18b	0.10	0.09	0.09	0.11	0.06	0.55	0.06	0.34	0.08	0.06	0.04	0.32	0.09
18c	0.16	0.10	0.10	0.10	0.14	0.40	0.07	0.29	0.13	0.07	0.09	0.28	0.06
19c	0.12	0.11	0.09	0.13	0.07	0.49	0.08	0.42	0.09	0.07	0.04	0.18	0.11

20a	0.26	0.12	0.39	0.11	0.06	0.07	0.09	0.09	0.24	0.34	0.05	0.08	0.11
22a	0.09	0.11	0.12	0.10	0.11	0.47	0.05	0.11	0.05	0.08	0.06	0.60	0.06

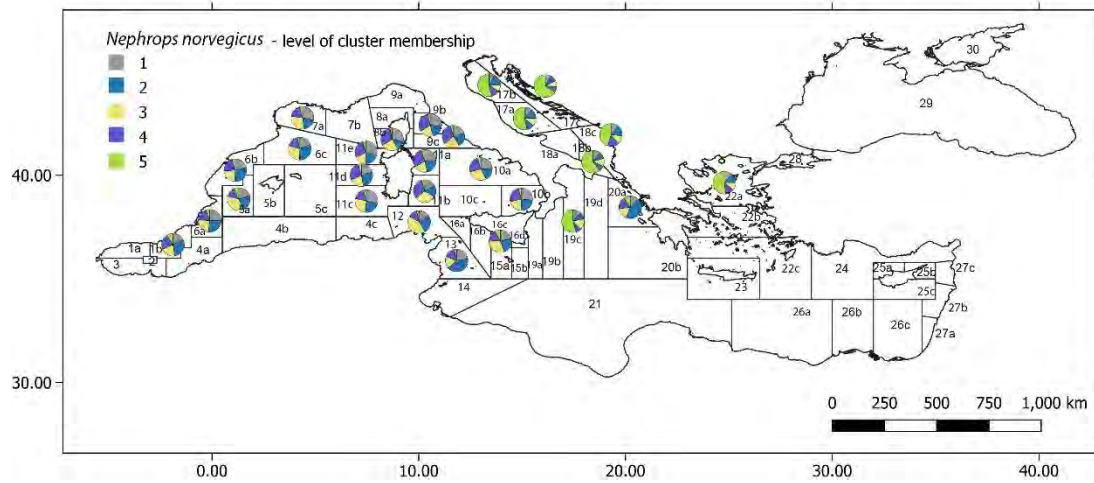
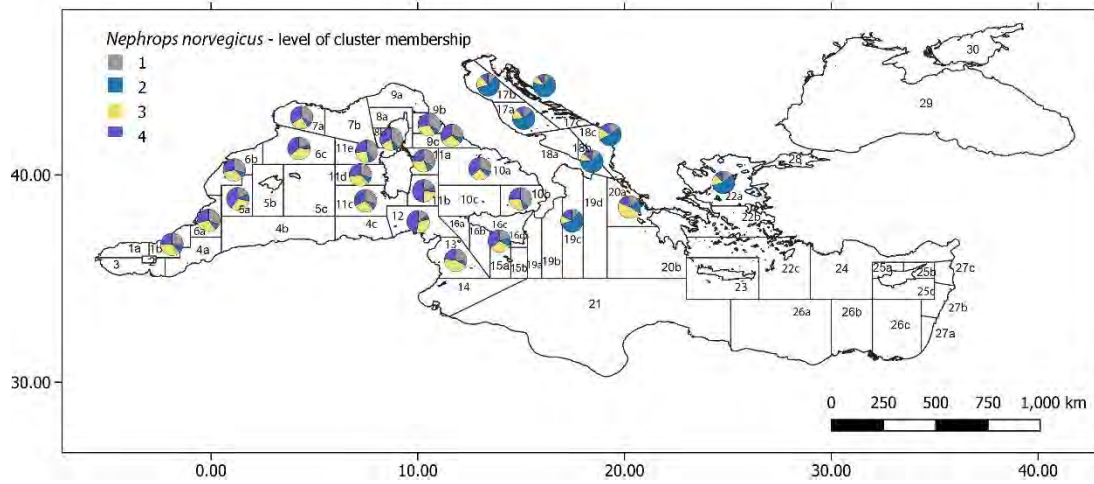


Figure 33: Pie charts with average membership grades per GSA subarea for *N. norvegicus* genetic dataset at NC = 4 (upper panel) and NC = 5 (lower panel).

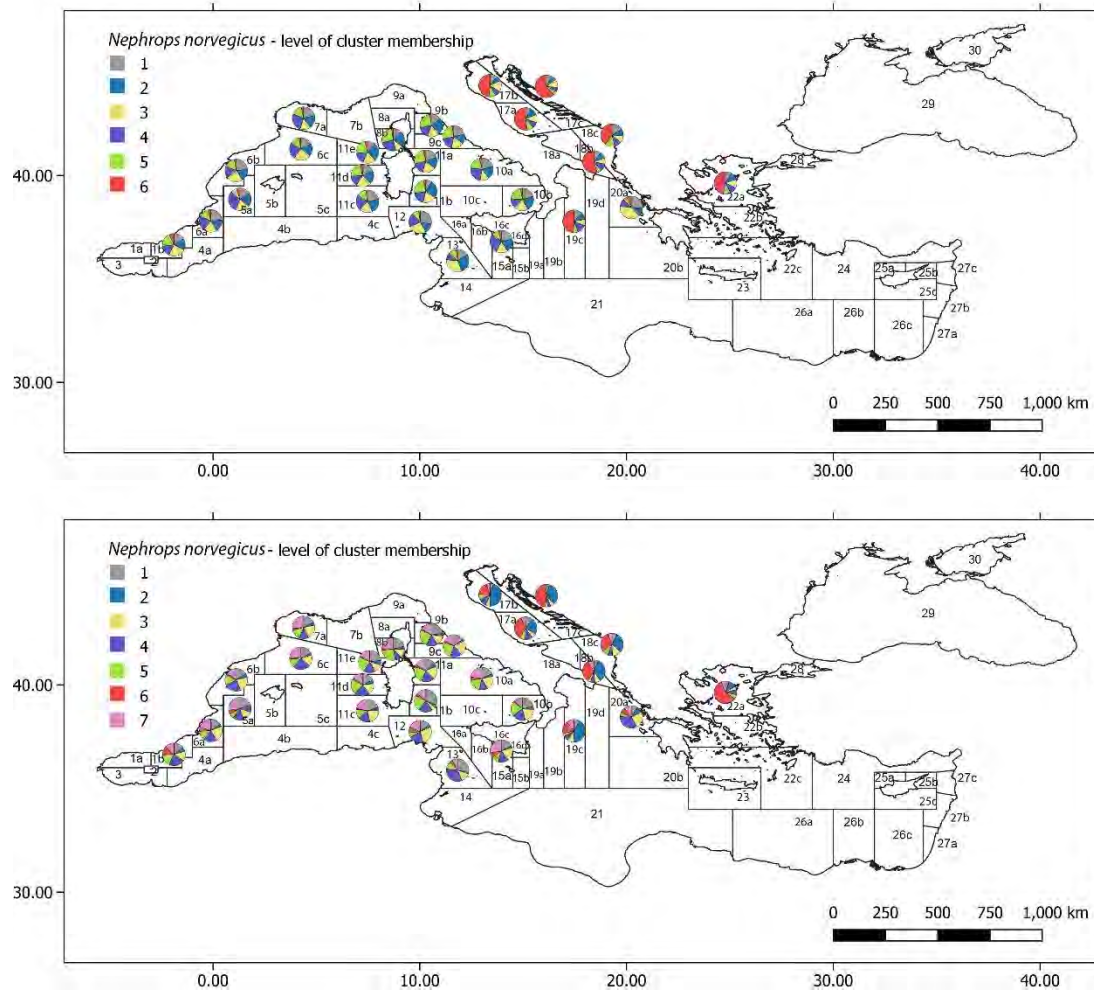


Figure 34: Pie charts with average membership grades per GSA subarea for *N. norvegicus* genetic dataset at NC = 6 (upper panel) and NC = 7 (lower panel).

Spatial analysis

The spatial variables used did show high explanatory power and statistically very significant relationships with the distribution of the fuzzy membership grades of *N. norvegicus* (Table 37).

The adjusted R^2 of depth was highest for NC = 4 (0.09) and NC = 5 (0.07), decreasing slowly at higher NC and being low or non-significant for the original data and NC = 2-3.

The large-scale linear gradient modelled by the metric coordinates of the sampling points was always significant, but for NC = 2 and NC = 3. The combined effect of longitude and latitude was much higher for NC = 4 and NC = 5 (Table 37), as already happened with the explanatory power of depth. Obviously, a great part in this is due to the difference in the membership grades of the Adriatic-Ionian-Aegean compared to the Western and part of the Central Mediterranean basin.

The MEMs were computed on the detrended data (as in the case of the hake) in order to obtain less MEMs and related to smaller scale spatial patterns. Results were in line with

those obtained with depth and with the geographic coordinates. The highest explanatory power of the selected MEMs was for NC = 4 and NC = 5, reaching respectively one quarter and one fifth of the variance of *N. norvegicus* genetic PCAs. Interesting also that there is consistence in the best spatial matrix and in the number of MEMs selected among the different NC (Table 37). For instance, NC = 4 and 5 were described both by 17 MEMs built based on the Relative connectivity scheme (Fig.) and with the concave up weighting function. At NC = 6 there were also 17 MEMs selected, mostly overlapping with the MEMs selected at NC = 4 and 5 but based on the Gabriel’s connectivity scheme and the concave down weighting function. This spatial matrix was also the one selected for all higher NC, where the number of MEMs increased to 21 (22 for NC = 10), being equal to the number of MEMs selected for the original dataset, also built with the Gabriel’s connectivity scheme and the concave down function. As already mentioned, many MEMs were selected repeatedly for different NC and with different spatial matrixes: some of them are the large-scale MEM1 (Fig. 36), MEM3 (Fig. 37), MEM6 (Fig. 38), the medium scale MEM12 (Fig. 39), and the small scale MEM22 (Fig. 40) and MEM30 (Fig. 41). Nevertheless, care should be taken when interpreting MEMs obtained from different spatial matrixes, since they might represent very different spatial structures: thus, a visual inspection is always necessary (Fig. 36). All this shows that the MEMs consistently modelled a quite clear spatial pattern.

Table 37: Spatial analysis results for N. norvegicus. The explanatory power of each spatial variable or group of variables is reported as adjusted R². All reported models with depth and LongLat were tested for significance with an ANOVA and “ns” indicates those that were not significant. The significance of MEMs was tested with a forward selection procedure and a double-stopping criterion. Only significant MEMs are reported. For MEMs also the number of significant MEMs, their names and the winner spatial weighting matrix are reported.

NEP	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.01	0.02	0.05	21	MEM3, MEM6, MEM11, MEM7, MEM9, MEM15, MEM12, MEM22, MEM24, MEM13, MEM8, MEM16, MEM10, MEM27, MEM19, MEM17, MEM4, MEM32, MEM29, MEM30, MEM20	Gabriel_Down_4
2 CLUSTERS	ns	ns	0.02	3	MEM12, MEM27, MEM16	Delaunay_Down_10
3 CLUSTERS	0.01	ns	0.02	5	MEM27, MEM2, MEM20, MEM14, MEM12	Delaunay_Up_0.6
4 CLUSTERS	0.09	0.14	0.25	17	MEM6, MEM1, MEM13, MEM11, MEM29, MEM17,	Relative_Up_0.4

					MEM4, MEM30, MEM5, MEM32, MEM7, MEM22, MEM3, MEM20, MEM28, MEM18, MEM12	
5 CLUSTERS	0.07	0.11	0.20	17	MEM1, MEM6, MEM12, MEM27, MEM10, MEM2, MEM29, MEM16, MEM30, MEM5, MEM20, MEM9, MEM4, MEM18, MEM3, MEM22, MEM8	Relative_Up_0.3
6 CLUSTERS	0.05	0.08	0.13	17	MEM3, MEM11, MEM15, MEM6, MEM9, MEM10, MEM18, MEM21, MEM1, MEM7, MEM32, MEM17, MEM28, MEM24, MEM30, MEM19, MEM4	Gabriel_Down_6
7 CLUSTERS	0.04	0.06	0.13	21	MEM3, MEM11, MEM15, MEM28, MEM6, MEM9, MEM19, MEM30, MEM12, MEM22, MEM21, MEM18, MEM24, MEM32, MEM27, MEM1, MEM7, MEM17, MEM8, MEM13	Gabriel_Down_9
8 CLUSTERS	0.03	0.05	0.11	21	MEM3, MEM11, MEM15, MEM6, MEM9, MEM27, MEM13, MEM22, MEM19, MEM28, MEM30, MEM10, MEM32, MEM1, MEM18, MEM21, MEM7, MEM24, MEM4, MEM17, MEM33	Gabriel_Down_6

9 CLUSTERS	0.03	0.04	0.10	21	MEM3, MEM11, MEM15, MEM6, MEM28, MEM9, MEM19, MEM29, MEM24, MEM10, MEM32, MEM1, MEM30, MEM18, MEM21, MEM27, MEM22, MEM13, MEM12, MEM4, MEM7	Gabriel_Down_9
10 CLUSTERS	0.02	0.03	0.10	22	MEM3, MEM11, MEM6, MEM10, MEM1, MEM9, MEM15, MEM22, MEM33, MEM32, MEM7, MEM5, MEM24, MEM25, MEM13, MEM17, MEM8, MEM14, MEM28, MEM30, MEM27, MEM2	Gabriel_Down_3

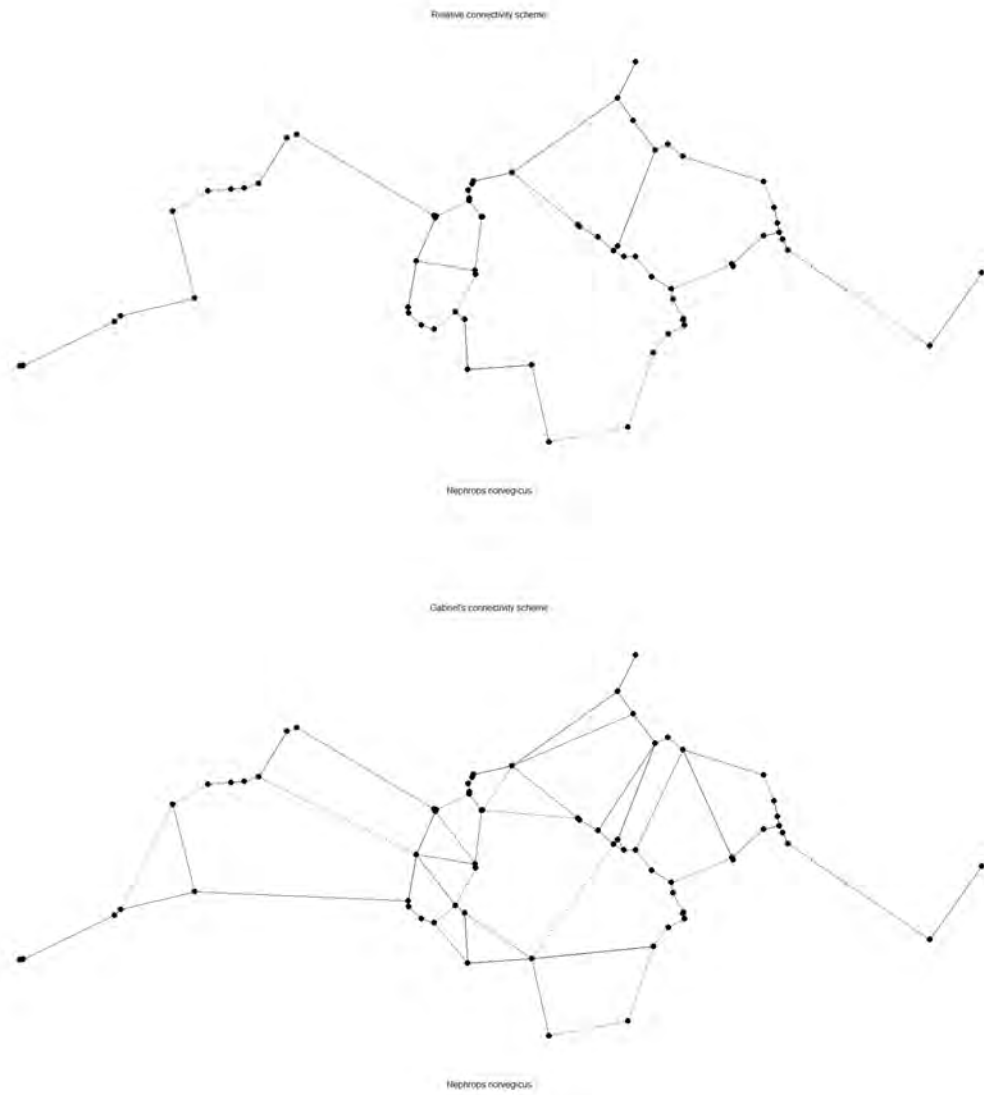


Figure 35: The relative neighborhood graph connectivity scheme (upper panel) and the Gabriel's graph connectivity scheme (lower panel) for sampling locations of *N. norvegicus*.

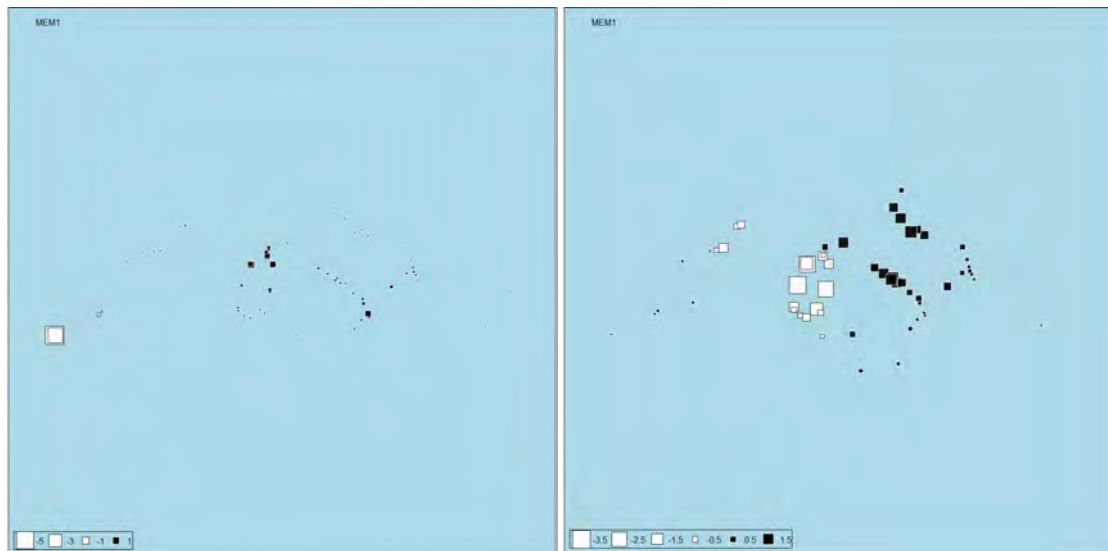


Figure 36: MEM1 for *N. norvegicus* built on the relative neighborhood graph connectivity scheme (left panel) and on the Gabriel's graph connectivity scheme (right panel).

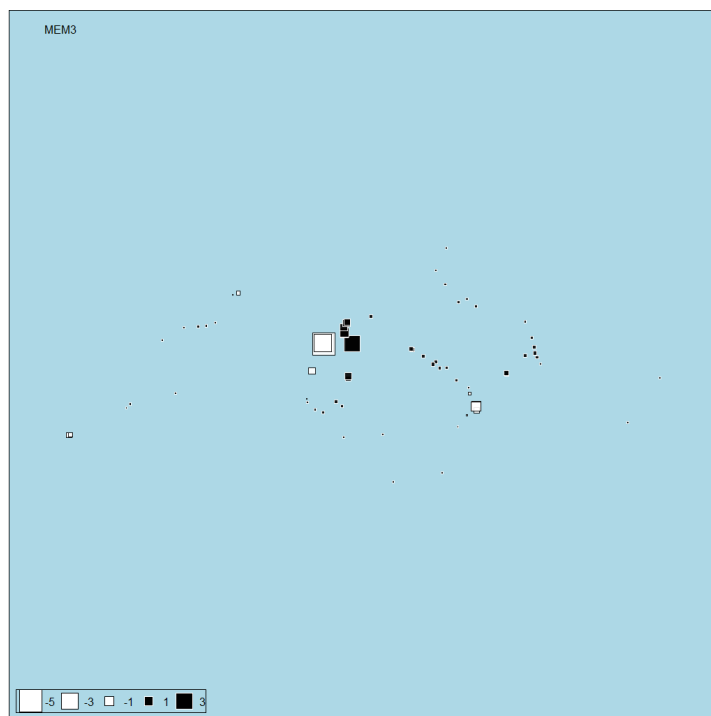


Figure 37: MEM3 for *N. norvegicus* built on the relative neighborhood graph connectivity scheme.

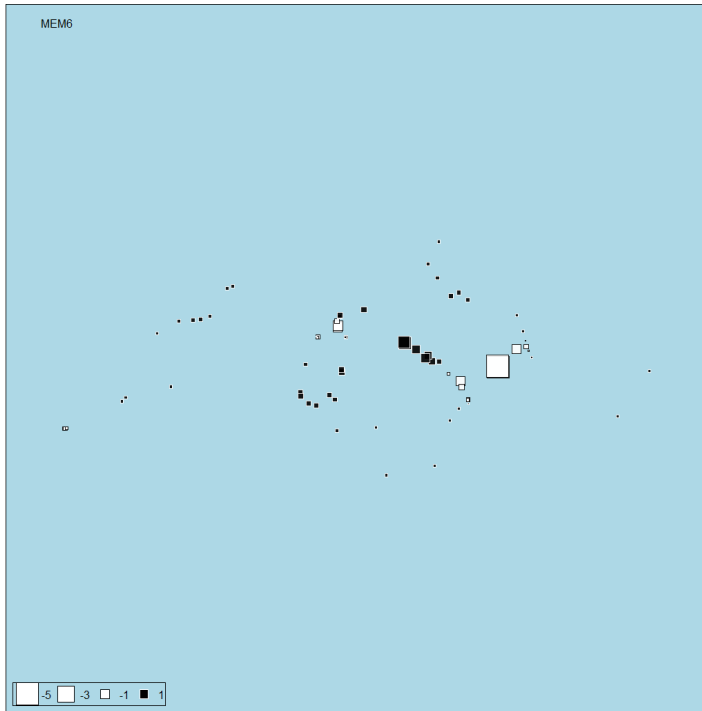


Figure 38: MEM6 for *N. norvegicus* built on the relative neighborhood graph connectivity scheme.

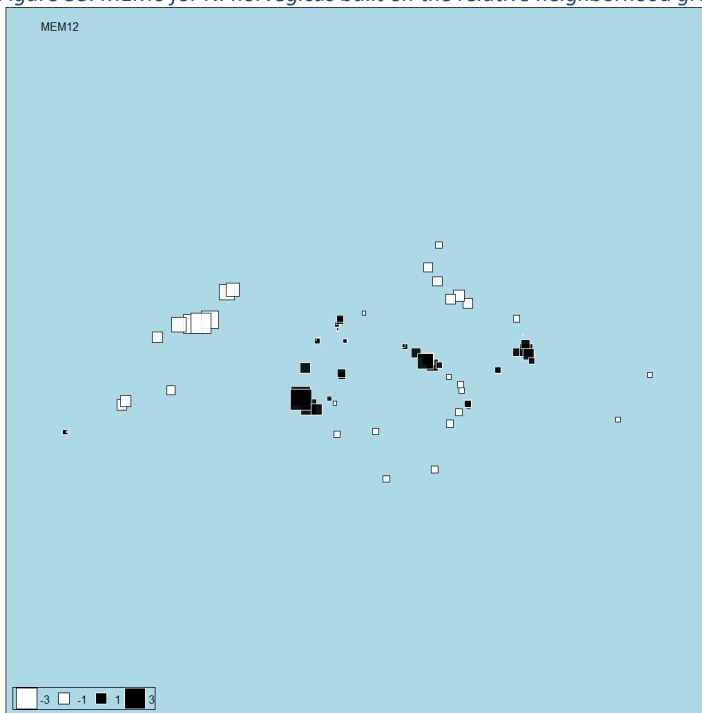


Figure 39: MEM12 for *N. norvegicus* built on the relative neighborhood graph connectivity scheme.

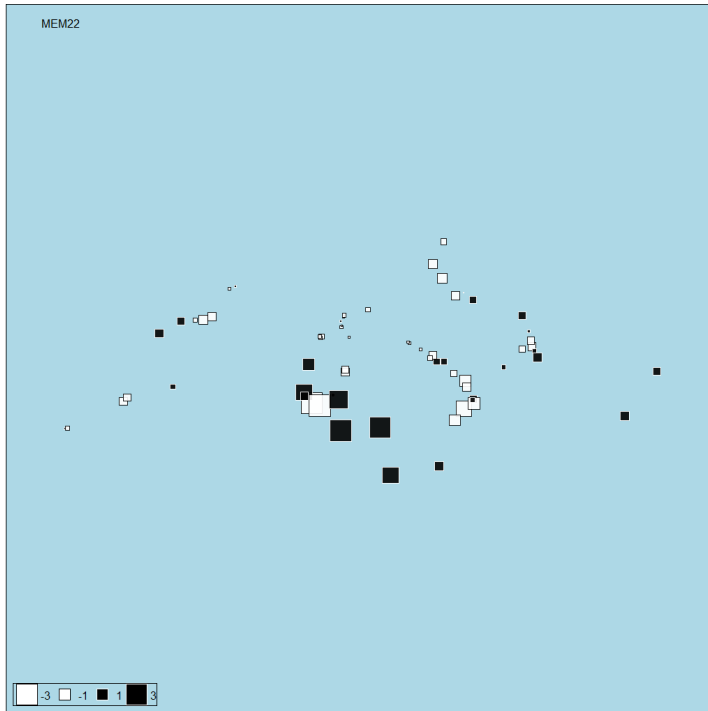


Figure 40: MEM22 for *N. norvegicus* built on the relative neighborhood graph connectivity scheme.

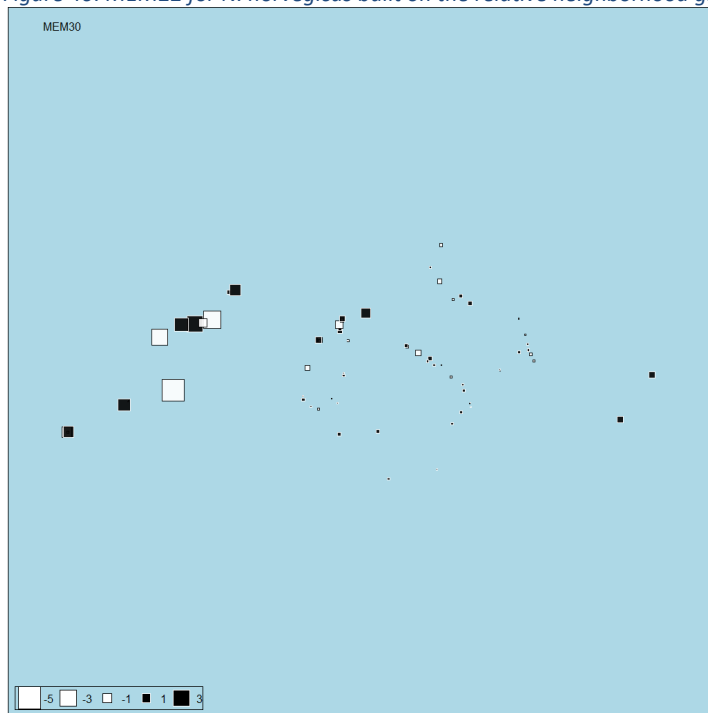


Figure 41: MEM30 for *N. norvegicus* built on the relative neighborhood graph connectivity scheme.

The results of the spatial analysis clearly point out that $NC = 4$ and $NC = 5$ are those clusterizations that have the most interpretable spatial structure and should be taken into account preferably in the following procedures. Thus, we built a complete spatial model for $NC = 4$ and $NC = 5$ with all significant variables and subjected them to the forward selection procedure with double stopping criterion in order to select the best parsimonious spatial model. For $NC = 4$, the final spatial model retained 18 out of 20 spatial variables and the

total adjusted R^2 was 0.30 on 3 RDA axes (Table 38). As it can be clearly seen the main differentiation is between cluster 2 (the “Adriatic” one) and the other clusters (Fig. 42): this differentiation is explained by the main gradient along the first RDA axis. The second RDA axis mostly explained the difference between the remaining three clusters, with clusters 1 and 3 at the extremes of that gradient and cluster 4 in between. The final spatial model for NC = 5 did also retain most of the variables, 17 on 20, and the total adjusted R^2 was 0.23 on 4 RDA axes (Table 38). The biplot shows a very similar situation as for NC = 4, just with cluster 1 mixed with individuals belonging to cluster 4 (Fig. 42). The fact that the spatial models are not able to adequately explain the difference in the “non-Adriatic” clusters does not mean that these are not different, but just that the spatial variables used here are not related to this difference. Again, other complex spatial dynamics, perhaps related to dispersal connectivity and the presence of discontinuities in the habitats, might be hypothesized, but not explored. It has to be noted also that the MEMs are derived on the actual sampling design. If this is not sufficiently detailed for some spatial patterns to emerge, also the MEMs will not be able to correctly detect them.

Table 38: Forward selection on spatial components for N. norvegicus genetic PCAs at NC = 4 and NC = 5. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

NEP 4 CLUSTERS			NEP 5 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
longUTM	0.13	0.001	longUTM	0.09	0.001
depth	0.19	0.001	depth	0.15	0.001
MEM6	0.22	0.001	MEM6	0.16	0.001
MEM1	0.23	0.001	MEM27	0.17	0.001
latUTM	0.25	0.001	MEM20	0.18	0.001
MEM5	0.26	0.001	MEM1	0.19	0.001
MEM13	0.27	0.001	MEM12	0.19	0.001
MEM4	0.27	0.002	latUTM	0.20	0.001
MEM11	0.28	0.001	MEM5	0.21	0.001
MEM7	0.28	0.001	MEM2	0.21	0.001
MEM30	0.29	0.003	MEM30	0.21	0.006
MEM17	0.29	0.003	MEM22	0.22	0.005
MEM22	0.29	0.006	MEM10	0.22	0.006
MEM3	0.29	0.009	MEM18	0.22	0.011

MEM32	0.30	0.007	MEM9	0.22	0.023
MEM29	0.30	0.005	MEM4	0.23	0.027
MEM12	0.30	0.023	MEM3	0.23	0.028
MEM28	0.30	0.033			

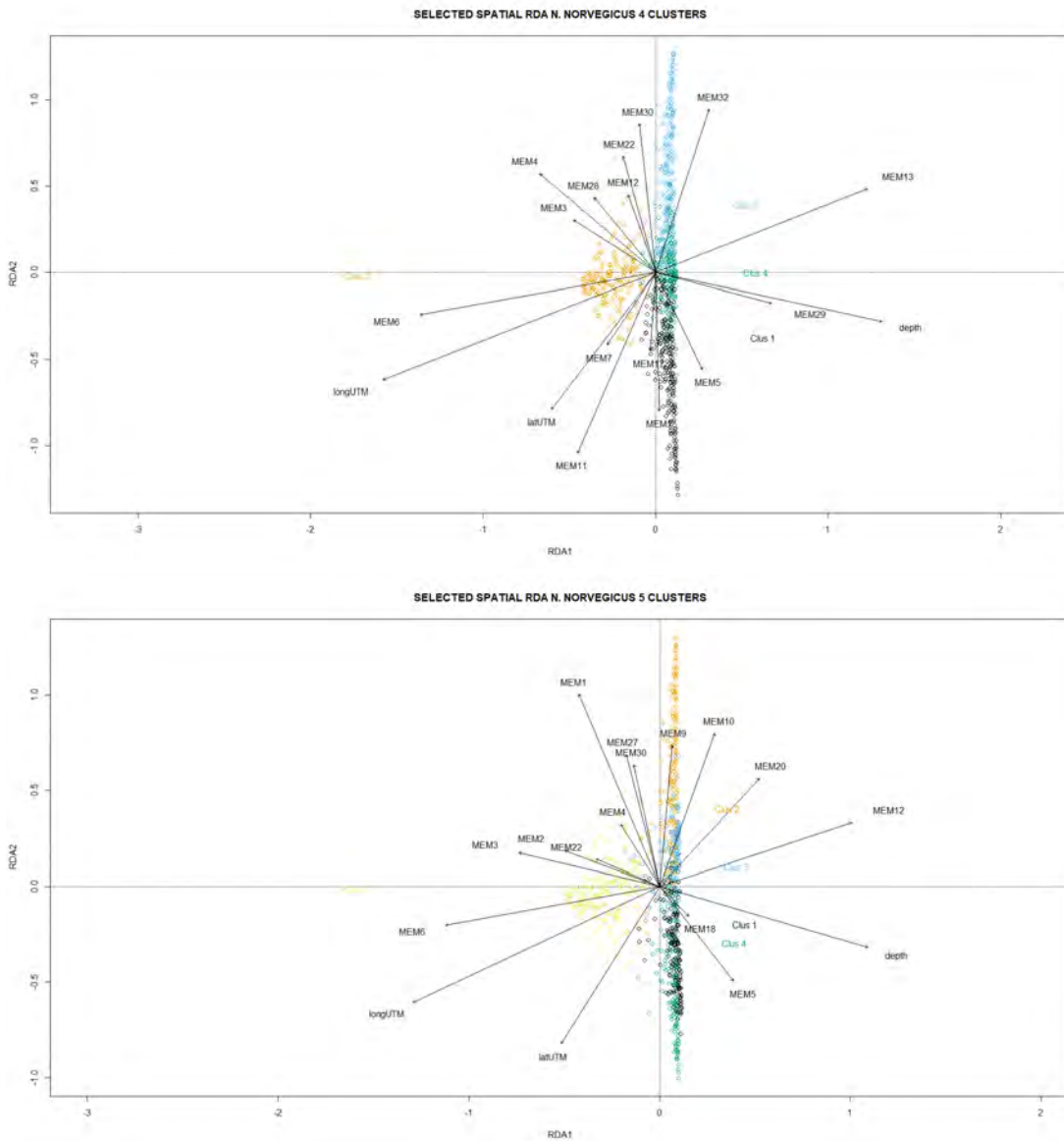


Figure 42: Final spatial RDA models for *N. norvegicus* at NC = 4 (upper panel) and NC = 5 (lower panel). Explained variance on the first and the second RDA axes was respectively 0.29 and 0.02 for NC = 4, and 0.21 and 0.02 for NC = 5.

Environmental analysis

Individuals of the GSA subarea 20a were sampled deep in the Gulf of Corinth, where the Copernicus models are not defined. Thus, no environmental variables could have been

extracted for this sampling location and these individuals were removed from the environmental analysis and from any further procedure where environmental variables were used (i.e., variation partitioning and best model selection).

The RDA models built for genetic PCAs of *N. norvegicus* did not show any significant difference between the explanatory power of bottom and of euphotic zone median variables (Table 39). Thus, the bottom variables were used as the best indicator of the habitat of this bottom dwelling species. As for the spatial models, the environmental dataset did not explain the membership grades distribution at NC = 2 or NC = 3 but showed the best relationships with the distribution of membership grades at NC = 4 (adjusted $R^2 = 0.29$ for the bottom medians over 5 years) and NC = 5 (adjusted $R^2 = 0.22$ for the bottom medians over 5 years), decreasing at increasing NC.

*Table 39: Adjusted R^2 for *N. norvegicus*, calculated for RDA models with environmental variables as explanatory variables. Different models were built for each set of explanatory variables: bottom and surface (euphotic zone), and the three time intervals considered (5 years, 3 years, 1 year). Results are reported for the clusterizations from 2 to 10 clusters.*

NEP	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.05	0.05	0.04	ns	ns	ns	0.01	0.01	ns	0.29	0.28	0.28	0.22	0.22	0.22
surface	0.05	0.05	0.05	0.02	0.02	ns	0.02	0.02	0.01	0.29	0.28	0.28	0.22	0.22	0.21
	6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.17	0.17	0.17	0.15	0.14	0.13	0.12	0.11	0.11	0.10	0.10	0.09	0.09	0.09	0.08
surface	0.17	0.17	0.16	0.15	0.15	0.14	0.13	0.12	0.12	0.11	0.11	0.10	0.10	0.10	0.09

The ensemble of 13 median bottom variables over 5 years was subjected to forward selection in order to select the best environmental model for NC = 4 and NC = 5 (Table 40). For NC = 4 the model selected 8 variables explaining an adjusted $R^2 = 0.28$ on three constrained axes (Fig. 43). Almost all this variance was explained on the first and only significant axis, while the other axes were not significant and did not explain anything. Thus, from the biplot only the clear difference between the “Adriatic” cluster and the other clusters can be commented. At NC = 5, there were 4 constrained axes, of which the first two significant, nevertheless again almost all explained variance was explained by the first axis. The variables selected in this second case were 7, all already selected for the model at NC = 4. Thus, we can say that the environmental variables can explain the difference between the “Adriatic” cluster and the other clusters, yet not the difference among these latter. The variables selected as the best environmental predictors explaining the difference between the “Adriatic” cluster and the others are bottom PO₄, pH, dissolved oxygen, temperature, velocity, phytoplankton biomass and chlorophyll concentration (this latter only for NC = 4).

Table 40: Forward selection on environmental variables at bottom over 5 years for *N. norvegicus* genetic PCAs at NC = 4 and NC = 5. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

NEP 4 CLUSTERS			NEP 5 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_b_pH	0.23	0.001	P50_b_pH	0.17	0.001
P50_b_chlf	0.24	0.001	P50_b_phy	0.18	0.001
P50_b_O2o	0.26	0.001	P50_b_O2o	0.20	0.001
P50_b_vel	0.27	0.001	P50_b_vel	0.21	0.001
P50_b_PO4	0.27	0.017	P50_b_PO4	0.21	0.001
P50_b_phy	0.28	0.001	P50_b_T	0.21	0.002
P50_b_T	0.28	0.027			

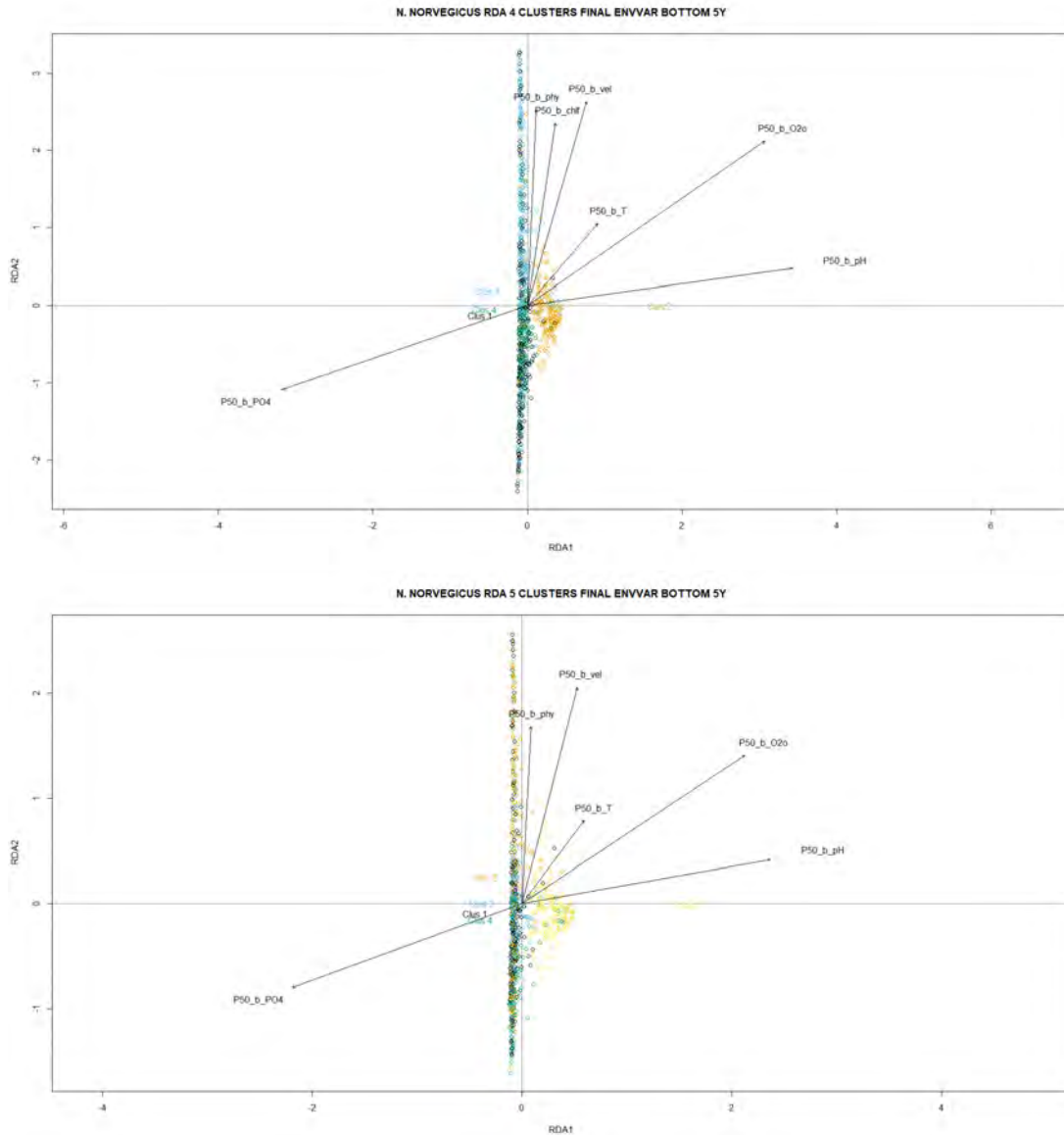


Figure 43: Final environmental RDA models for *N. norvegicus* at NC = 4 (upper panel) and NC = 5 (lower panel). Explained variance on the first and only significant RDA axis at NC = 4 was 0.28; explained variance on the first and second significant RDA axes at NC = 5 was respectively 0.21 and 0.01.

Variation partitioning and best model selection

The spatial variables (depth, longitude, latitude, selected MEMs) and the environmental variables were used in the variation partitioning procedure to show the mutual relationships among them (Table 41). The dataset of environmental variables had the highest explanatory power at NC = 4 and together with the selected MEMs explained as much of the distribution of fuzzy membership grades of *N. norvegicus* as possible. The contributions of depth and of the geographic coordinates was fully covered by the other two groups of variables. Nevertheless, when comparing the contribution of MEMs and of environmental variables to the three groups, MEMs showed a higher marginal explanatory power. This was because they were less related to longitude and latitude compared to the set of environmental variables. At NC = 5 the results were similar but for the fact that longitude and latitude showed a much higher shared explanatory power with the MEMs selected in this case. This

is easily understandable since the MEMs selected at NC = 5 were all large-scale MEMs, thus arguably more related to large-scale gradients described also by the geographic coordinates. In both cases the fraction of residuals is rather high, but this can be ascribed to the large part of randomness in the genetic data, not associated to structured space or environmental features of the habitat of *N. norvegicus*.

Table 41: Variation partitioning for the *N. norvegicus* genetic dataset at NC = 4 and NC = 5. + indicates the total variation explained by the groups of variables; | indicates the variation explained by the first group of variables conditioned to the second; U indicates the fraction of variance explained jointly by the two groups of variables. Compare to Fig. 44 for a graphical representation.

NEPHROPS NORVEGICUS	4 CLUSTERS	5 CLUSTERS
Explanatory group of variables	Adjusted R²	Adjusted R²
Depth	0.09	0.07
LongLat	0.14	0.11
MEM	0.26	0.21
EnvVar Bot 5y	0.29	0.22
Depth + LongLat	0.21	0.16
Depth + MEM	0.26	0.21
Depth + EnvVar Bot 5y	0.29	0.22
LongLat + MEM	0.31	0.23
LongLat + EnvVar Bot 5y	0.29	0.22
MEM + EnvVar Bot 5y	0.32	0.24
Depth + LongLat + MEM	0.31	0.23
Depth + LongLat + EnvVar Bot 5y	0.29	0.22
Depth + MEM + EnvVar Bot 5y	0.32	0.24
LongLat + MEM + EnvVar Bot 5y	0.32	0.24
Depth + LongLat + MEM + EnvVar Bot 5y	0.32	0.24
Depth LongLat	0.06	0.00
Depth MEM	0.00	0.00
Depth EnvVar Bot 5y	0.00	0.00
LongLat Depth	0.12	0.00

<i>LongLat MEM</i>	0.05	0.00
<i>LongLat EnvVar Bot 5y</i>	0.00	0.03
<i>MEM Depth</i>	0.17	0.02
<i>MEM LongLat</i>	0.16	0.02
<i>MEM EnvVar Bot 5y</i>	0.03	0.08
<i>EnvVar Bot 5y Depth</i>	0.20	0.01
<i>EnvVar Bot 5y LongLat</i>	0.14	0.04
<i>EnvVar Bot 5y MEM</i>	0.06	0.07
<i>Depth MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>Depth LongLat + EnvVar Bot 5y</i>	0.00	0.00
<i>Depth LongLat + MEM</i>	0.00	0.00
<i>LongLat MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat Depth + EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat Depth + MEM</i>	0.05	0.03
<i>MEM Depth + EnvVar Bot 5y</i>	0.03	0.02
<i>MEM LongLat + EnvVar Bot 5y</i>	0.03	0.02
<i>MEM Depth + LongLat</i>	0.11	0.08
<i>EnvVar Bot 5y LongLat + MEM</i>	0.01	0.01
<i>EnvVar Bot 5y Depth + MEM</i>	0.06	0.04
<i>EnvVar Bot 5y Depth + LongLat</i>	0.08	0.07
<i>Depth LongLat + MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat Depth + MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>MEM Depth + LongLat + EnvVar Bot 5y</i>	0.03	0.02
<i>EnvVar Bot 5y Depth + LongLat + MEM</i>	0.01	0.01
<i>Depth ∪ LongLat</i>	0.00	0.00
<i>LongLat ∪ MEM</i>	0.00	0.00

<i>Depth UMEM</i>	0.00	0.00
<i>Depth UEnvVar Bot 5y</i>	0.00	0.00
<i>LongLat UEnvVar Bot 5y</i>	0.05	0.03
<i>MEM UEnvVar Bot 5y</i>	0.08	0.06
<i>Depth ULongLat UEnvVar Bot 5y</i>	0.00	0.00
<i>Depth ULongLat UMEM</i>	0.00	0.00
<i>LongLat UMEM UEnvVar Bot 5y</i>	0.07	0.06
<i>Depth UMEM UEnvVar Bot 5y</i>	0.06	0.04
<i>Depth ULongLat UMEM UEnvVar Bot 5y</i>	0.03	0.02
<i>Residuals</i>	0.68	0.76



Figure 44: Graphical representation of the variation partitioning of the *N. norvegicus* genetic dataset at 4 and 5 clusters among four groups of variables (Depth, LongLat, MEM, Env/Var Bot 5y). Negative or zero fractions are not displayed. See Table 39 for more details.

The whole group of variables was subjected to the forward selection procedure in order to build the best parsimonious RDA models for NC = 4 and NC = 5 (Table 42). At NC = 4, fourteen variables were selected explaining together an adjusted $R^2 = 0.32$ on a total of 3 RDA axis (of which only the first two significant). Among the selected variables, there were longitude and latitude, pH, chlorophyll concentration, dissolved oxygen and temperature, and 8 MEMs of spanning all scales (large – MEM1, MEM3, MEM6; medium – MEM11, MEM12, MEM13; small – MEM30, MEM32). At NC = 5, only 11 variables were selected explaining together an adjusted $R^2 = 0.24$ on a total of 4 RDA axis (of which only the first two significant). In this case only latitude was selected, together with 5 environmental variables (pH, phytoplankton biomass, temperature, dissolved oxygen and PO₄: with the exception of PO₄ all already selected at NC = 4 or anyway related to the same processes, as in the case of phytoplankton biomass and chlorophyll concentrations), and 5 large scale MEMs (MEM3,

MEM6, MEM8, MEM9, MEM10). Clearly, the much complex spatial structure of cluster membership grades for NC = 5 cannot be adequately described, as it is evident also from the biplot (Fig. 45) where the samples with highest membership for clusters 1 and 3 are intermixed with those of clusters 4 and 2. The biplot of the model for 4 clusters on the other hand, shows a better discrimination between the individuals with high memberships for different clusters also in the “non-Adriatic” part of the plot. Nevertheless, it has to be reminded the much lower discriminatory power of the differences between clusters 1, 3 and 4, than the discriminatory power between these three clusters and cluster 2.

Table 42: Forward selection on the final models for N. norvegicus genetic PCAs at NC = 4 and NC = 5. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

NEP 4 CLUSTERS			NEP 5 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_b_pH	0.23	0.001	P50_b_pH	0.17	0.001
MEM6	0.25	0.001	MEM6	0.19	0.001
latUTM	0.26	0.001	latUTM	0.19	0.001
MEM11	0.27	0.001	MEM10	0.20	0.001
MEM1	0.28	0.001	P50_b_phy	0.21	0.001
P50_b_chlf	0.28	0.001	P50_b_O2o	0.22	0.001
P50_b_O2o	0.29	0.001	MEM8	0.22	0.001
MEM32	0.30	0.002	MEM9	0.23	0.002
longUTM	0.30	0.003	MEM3	0.23	0.031
MEM30	0.30	0.003	P50_b_PO4	0.23	0.033
P50_b_T	0.31	0.007	P50_b_T	0.24	0.001
MEM13	0.31	0.003			
MEM3	0.31	0.012			
MEM12	0.32	0.011			

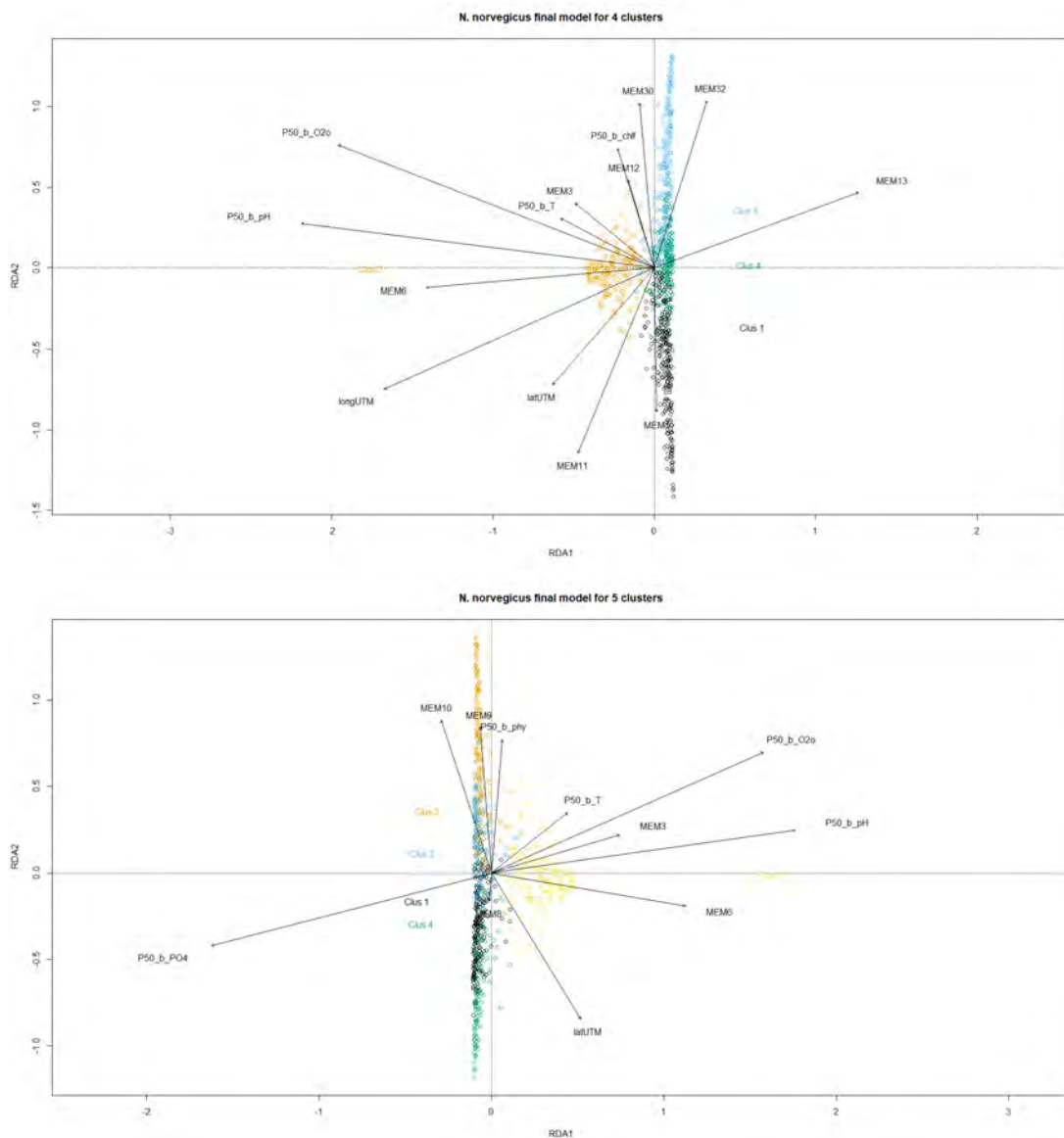


Figure 45: Best final parsimonious RDA models for *N. norvegicus* at NC = 4 (upper panel) and NC = 5 (lower panel). Explained variance on the first and second significant RDA axes was 0.30 and 0.02 at NC = 4, and 0.22 and 0.02 at NC = 5.

Conclusions

The applied internal quality clustering indexes indicated that divergence in the *N. norvegicus* genetic PCA data is not very strong. Nevertheless, the ensuing clusters, in particular at NC = 4, were strongly related to spatial and environmental features of the Mediterranean, supporting the case for different stocks.

The fuzzy k-means applied to the PCA derived on genetic data allowed to detect the discontinuities in the data and to inspect their inherent fuzziness. Hard edge discontinuities are rare in the real world, where usually gradual changes along a gradient are more common. This is more plausible for genetic data of potential sub-population of the same species living in different areas of the same basin and linked by dispersal through currents

advection. Thus, FKM on one side permits to detect differences in the data, if these are strong enough; on the other hand, FKM does not impose a hard-edge discontinuity where most probably does not exist, allowing to relate this fuzziness to the drivers that can cause it. The advantage over other hard clustering methods is also that FKM membership grades are continuous variables, thus common gradient analysis methods, such as RDA, can be used to relate their distribution to other factors: with hard clustering method, DA (Discriminant Analysis) should be used.

The procedure of first computing the PCAs from genetic data and then using them in FKM to detect possible different clusters, is especially useful when the clusterization is not known a priori from a different source of information. In that case, if an a priori clusterization does exist, this can be explained by applying a DA procedure using either internal information (i.e., related to the genetic data), or external information (i.e., related to factors that may have caused the observed genetic structuring). But if such a priori knowledge does not exist, besides the internal validation with cluster validity indexes, only external validation procedure may be applied (spatial variables, environmental variables, variable on competition, predation, etc.).

The use of RDA models to relate spatial and environmental variables to the observed discontinuities in the genetic PCAs can be seen as an external validation of the clustering results. If a sub-population does exist, much probably is spatially isolated from other sub-populations (otherwise genetic traits specific to it would not have emerged and most of all would not have persisted in time), and perhaps experiencing different environmental conditions. With the procedures proposed here the internal validation of clustering results, i.e., the validation based on the characteristics of the data on which the clusterization was performed, should be done on the PCAs. Each PCA represent one independent axis of variation of the original data and can be interpreted in terms of the original data. The internal validation was not performed in this Task since the PCAs were produced by the other research groups involved in the project and used as they were provided.

On the other hand, the validation through the spatial components and the environmental variables for *N. norvegicus* genetic data was consistent and solid and clearly indicated that at least 4 stocks can be identified, and their habitats described in terms of their spatial structuring and environmental conditions. While the genetic diversification among these stocks might not be very strong, as indicated by the internal validity cluster indexes application, the external validation procedures confirmed the existence of different stocks. These results should not be seen as a contradiction. In fact, stocks may be occasionally connected to others, thus increasing the mixing and decreasing the genetic diversification, but still inhabit different areas and experiencing different environmental conditions. Such result could not be obtained for less than 4 clusters, indicating that they share the same level of diversification in terms of the genetic data. The sampling design was not detailed enough to disentangle possible diversifications internal to the Adriatic Sea, nor to clearly describe the diversification among the three Western Mediterranean clusters in terms of space or environmental characteristics.

Parapenaeus longirostris

Data

The genetic dataset for *Parapenaeus longirostris* was made up by 50 Principal Components (PCA) extracted from the genetic data generated in WP1 (D.1.5.2).

All PCAs showed different means and standard deviations and to prevent variables with highest means and standard deviations to unduly influence the analyses, all variables were standardized before the analysis by subtracting the mean and dividing by the standard deviation.

The number of individuals with complete records for genetic PCAs was 748 sampled in 51 different locations (Fig. 46).

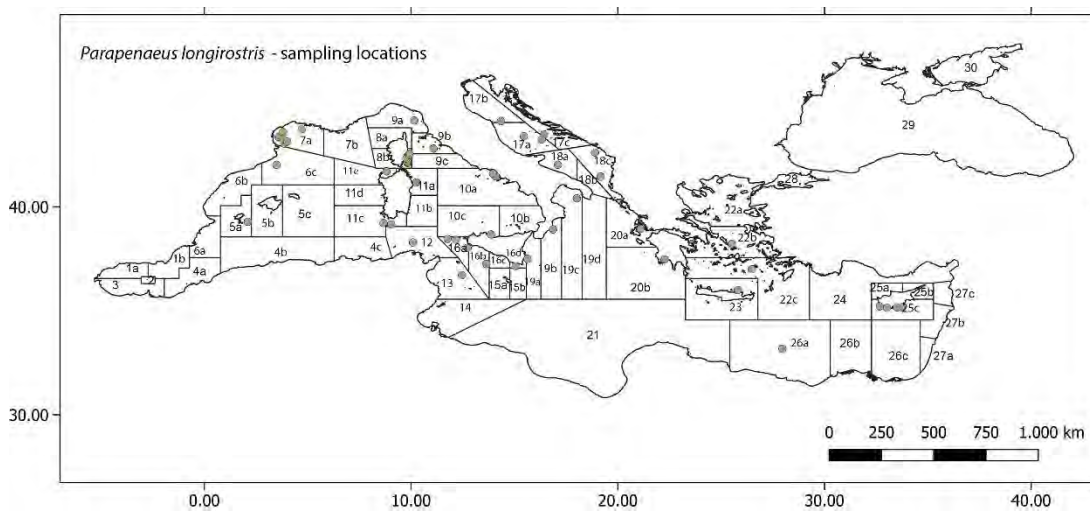


Figure 46: *Parapenaeus longirostris* sampling locations in the Mediterranean Sea.

Results and discussion

Fuzzy clustering

Fuzzy k-means was applied on 20 PCAs with $m = 1.1$ and 100 random initializations. The number of genetic PCAs to use was determined empirically by trying different numbers and checking the fuzziness obtained with the same fuzziness parameter. It came out that the less PCAs were used, the less fuzziness would be in the results: obviously, limiting the number of PCAs also the amount of genetic information provided would be limited. Thus, we used as much PCAs as possible, but conditioned to the fact that a clear cluster structure should emerge. It turned out that such optimal number was again 20 PCAs. A more quantitative approach would see the use of the explained variance per PCA axis to choose the number of PCAs needed to correctly reconstruct the general characteristics of the original genetic dataset.

The six internal quality indexes applied to the clustering results for $NC = 2$ to 10 (Table 43) were unanimous in indicating in $NC = 2$ the best division in clusters. Only MPC did indicate the highest NC as the best choice. It means that the genetic diversification in *P. longirostris* is

very clear and no doubt at least two genetically different sub-populations do exist in the Mediterranean.

Table 43: Results of the internal cluster validity indexes applied to *P. longirostris* fuzzy clusters. See Table 4 for the optimal value of each index.

DPS	PC	PE	MPC	SIL	SIL.F	XB
2 CLUSTERS	0.71	0.45	0.41	0.09	0.12	8.51
3 CLUSTERS	0.60	0.67	0.40	0.05	0.08	11.91
4 CLUSTERS	0.55	0.82	0.40	0.00	0.03	11.79
5 CLUSTERS	0.50	0.96	0.38	-0.03	-0.01	14.00
6 CLUSTERS	0.50	1.02	0.40	-0.03	-0.02	12.99
7 CLUSTERS	0.49	1.07	0.40	-0.02	0.00	9.99
8 CLUSTERS	0.48	1.12	0.41	-0.09	-0.07	10.66
9 CLUSTERS	0.48	1.13	0.42	-0.11	-0.08	11.14
10 CLUSTERS	0.50	1.12	0.45	-0.12	-0.09	10.38

When looking at the membership grades averaged per GFCM subregion (Table 44) at NC = 2 the Western and Centre subregions are grouped together, while the Adriatic and the Eastern are grouped together. At NC = 3 the situation is less clear, with only the East being clearly associated to one cluster, while the other three subregions show the maximum membership for the same cluster, and one cluster does not have any subregions being associated preferentially to it. This results thus confirms that at NC = 2 the clusterization gave results more easily interpretable than at CN = 3.

Table 44: Spatial interpretation of fuzzy clusters results (NC = 2 and 3) for the *P. longirostris* genetic PCAs considering GFCM subregions. Highlighted in green the highest average membership values for each subregion.

DPS	2 CLUSTERS		3 CLUSTERS		
CODE_GFCM	FKM1	FKM2	FKM1	FKM2	FKM3
West	0.67	0.33	0.41	0.17	0.42
Center	0.55	0.45	0.30	0.25	0.44
Adriatic	0.36	0.64	0.15	0.41	0.45
East	0.24	0.76	0.07	0.53	0.40

On the other hand, when averaging the membership grades per GSA subarea the result is quite clear, with the majority of subareas clearly associated to one cluster. At NC = 2 there is a separation between the Western Mediterranean basin and the Eastern one. The discontinuity between the two sub-populations is in the Ionian Sea: the subarea 19a has a higher membership for cluster 1, subarea 19b is equally distributed between the two, subarea 19c has a higher membership for cluster 2 (Fig. 47). Some subareas show a behavior that is not in line with this broad picture, i.e., subareas 17c and 22c: in such cases a more detailed inspection into the possible causes should be made, in particular regarding possible missing genetic information, lower number of individuals sampled, etc. At NC = 3 these two subareas split into a new cluster where they were joined by some subareas from all basins, but generally showing more fuzzy membership grades.

From all above the results for *P. longirostris* clearly show the existence of two neatly separated sub-populations living one in the Western and most of the Central

Mediterranean, and the other in the Eastern basin and the Adriatic Sea. Spatial and environmental analysis may further reinforce this outcome. The general pattern of current-mediated dispersal in the depth range of this species and the complex bathymetric and morphological features of the area of the Strait of Sicily-Strait of Messina may be responsible for the observed separation of *P. longirostris* population into two sub-populations.

Table 45: Spatial interpretation of fuzzy clusters results (NC =2 and 3) for the *P. longirostris* genetic PCAs considering GSA subareas. Highlighted in green the highest average membership values for each subarea.

DPS CODE_AREA	2 CLUSTERS		3 CLUSTERS		
	FKM1	FKM2	FKM1	FKM2	FKM3
5a	0.81	0.19	0.56	0.07	0.37
6c	0.70	0.30	0.34	0.11	0.55
7a	0.69	0.31	0.45	0.15	0.39
8b	0.56	0.44	0.36	0.29	0.35
9a	0.61	0.39	0.39	0.18	0.43
9b	0.71	0.29	0.44	0.14	0.42
10a	0.60	0.40	0.43	0.25	0.32
10c	0.53	0.47	0.36	0.27	0.36
11a	0.53	0.47	0.20	0.26	0.54
11c	0.78	0.22	0.48	0.08	0.44
11e	0.69	0.31	0.40	0.10	0.50
12	0.84	0.16	0.52	0.05	0.43
13	0.77	0.23	0.47	0.10	0.43
16a	0.81	0.19	0.40	0.04	0.56
16b	0.70	0.30	0.44	0.12	0.44
17a	0.20	0.80	0.12	0.36	0.53
17b	0.29	0.71	0.14	0.50	0.37
17c	0.63	0.37	0.10	0.14	0.76
18a	0.30	0.70	0.14	0.48	0.38
18c	0.42	0.58	0.20	0.31	0.49
19a	0.53	0.47	0.28	0.19	0.54
19b	0.50	0.50	0.24	0.29	0.47
19c	0.46	0.54	0.30	0.33	0.37
20a	0.38	0.62	0.17	0.31	0.51
20b	0.26	0.74	0.09	0.60	0.30
22b	0.28	0.72	0.08	0.51	0.41
22c	0.56	0.44	0.16	0.18	0.65
23	0.18	0.82	0.08	0.61	0.32
25c	0.16	0.84	0.06	0.67	0.27
26a	0.17	0.83	0.06	0.47	0.47

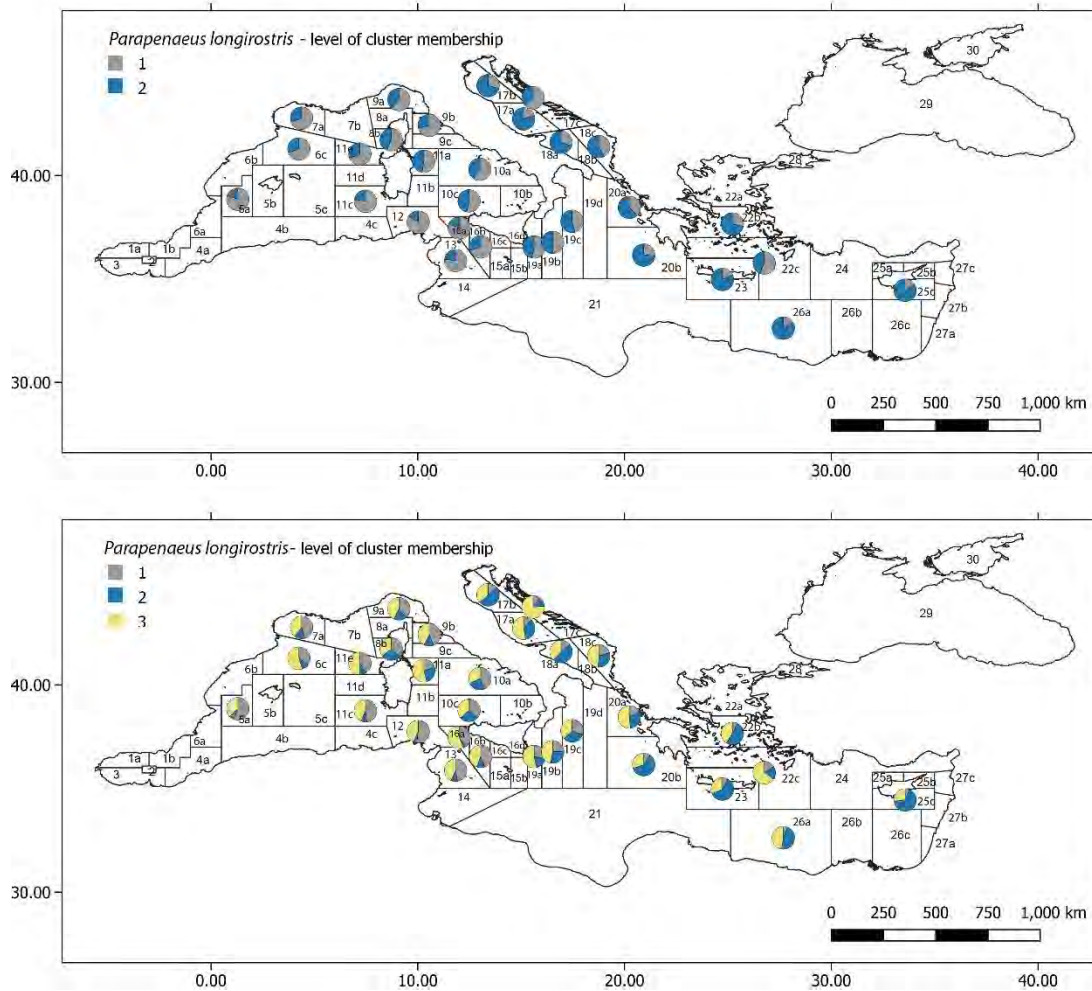


Figure 47: Pie charts with average membership grades per GSA subarea for *P. longirostris* genetic dataset at NC = 2 (upper panel) and NC= 3 (lower panel).

Spatial analysis

The spatial components did show a high explanatory power in particular for NC = 2 and NC = 3. Depth and geographic coordinates showed a clear decrease in explanatory power from 2 to 3 clusters. The explanatory power remained the same between these two cluster numbers for significant MEMs, but 9 against 5 MEMs were necessary to model the distribution of fuzzy membership grades at NC = 3 compared to NC = 2. A pattern could be observed also in the selected spatial matrixes (Fig. 48), changing with increasing NC, and with the number of selected MEMs, in general increasing with the increasing NC. All this shows that the results obtained on the genetic PCAs of *P. longirostris* are rather consistent and robust.

Table 46: Spatial analysis results for *P. longirostris*. The explanatory power of each spatial variable or group of variables is reported as adjusted R^2 . All reported models with depth and LongLat were tested for significance with an ANOVA and "ns" indicates those that were not significant. The significance of MEMs was tested with a forward

selection procedure and a double-stopping criterion. Only significant MEMs are reported. For MEMs also the number of significant MEMs, their names and the winner spatial weighting matrix are reported.

DPS	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.01	0.03	0.02	8	MEM15, MEM2, MEM24, MEM4, MEM20, MEM7, MEM19, MEM1	Relative_Up_0.3
2 CLUSTERS	0.06	0.28	0.10	5	MEM23, MEM15, MEM21, MEM24, MEM20	Relative_Up_0.2
3 CLUSTERS	0.03	0.14	0.10	9	MEM21, MEM15, MEM4, MEM24, MEM23, MEM7, MEM2, MEM20, MEM5	Relative_Up_0.3
4 CLUSTERS	0.02	0.09	0.06	7	MEM21, MEM13, MEM24, MEM15, MEM23, MEM4, MEM14	MST_Up_0.1
5 CLUSTERS	0.02	0.07	0.04	5	MEM21, MEM13, MEM24, MEM6, MEM15	MST_Up_0.1
6 CLUSTERS	0.02	0.05	0.05	9	MEM20, MEM13, MEM15, MEM5, MEM24, MEM6, MEM16, MEM14, MEM4	MST_Down_2
7 CLUSTERS	0.02	0.05	0.05	10	MEM20, MEM13, MEM24, MEM14, MEM4, MEM15, MEM18, MEM5, MEM2, MEM16	MST_Down_2
8 CLUSTERS	0.01	0.04	0.05	11	MEM19, MEM18, MEM24, MEM3, MEM15, MEM7, MEM4, MEM10, MEM21, MEM11, MEM6	MST_Up_0.6
9 CLUSTERS	0.01	0.04	0.05	11	MEM19, MEM4, MEM24, MEM18, MEM21, MEM7,	MST_Up_0.6

					MEM15, MEM10, MEM11, MEM5, MEM6	
10 CLUSTERS	0.01	0.03	0.04	12	MEM24, MEM19, MEM4, MEM21, MEM18, MEM7, MEM8, MEM11, MEM5, MEM6, MEM15, MEM22	MST_Up_0.6

The spatial components (depth, the metric geographic coordinates and the selected significant MEMs) were subjected to forward selection with a double-stopping criterion in order to build the best parsimonious spatial model (Table 47). At NC = 3 both longitude and latitude were selected together with some medium to small scale MEMs (MEM23, MEM15, MEM21, MEM24) (Fig. 49-52). At NC = 3, with a lower overall adjusted R² (0.33. for NC = 2 and 0.18 for NC = 3) other 3 large-scale MEMs (MEM4, MEM5, MEM7) were selected along with the same variables as for NC = 2. Thus, the spatial distribution at 3 clusters is less clear than for 2 clusters and needs additional MEMs to be modelled.

Table 47: Forward selection on the spatial components for P. longirostris genetic PCAs at NC = 2 and NC = 3. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

DPS 2 CLUSTERS			DPS 3 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
longUTM	0.25	0.001	longUTM	0.13	0.001
latUTM	0.28	0.001	latUTM	0.14	0.001
MEM23	0.30	0.001	MEM21	0.15	0.001
MEM15	0.31	0.001	MEM15	0.16	0.003
MEM21	0.33	0.001	MEM7	0.16	0.004
MEM24	0.33	0.002	MEM24	0.17	0.002
			MEM23	0.17	0.005
			MEM5	0.18	0.007
			MEM4	0.18	0.017

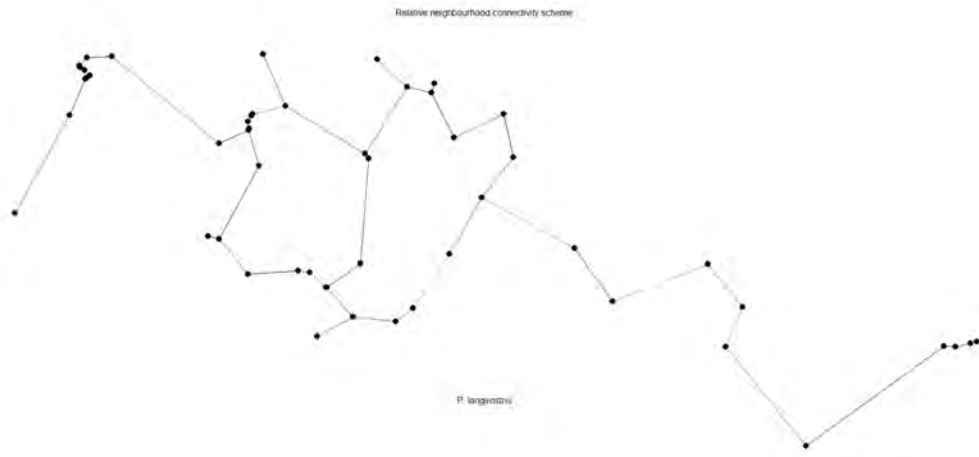


Figure 48: The relative neighborhood graph connectivity scheme for *P. longirostris* sampling locations.

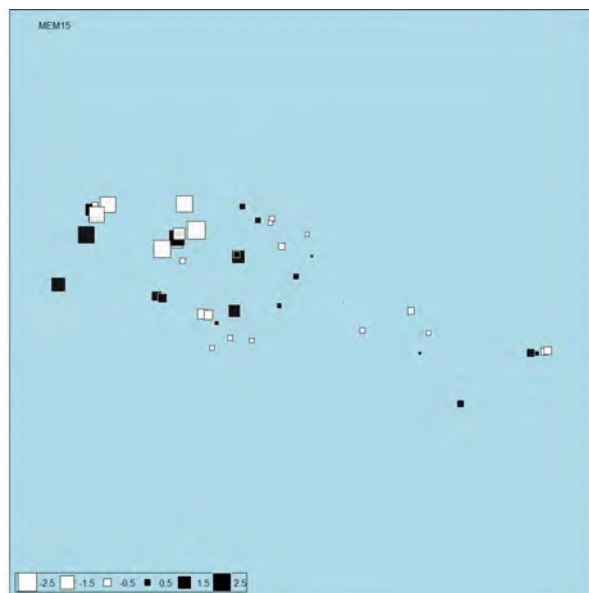


Figure 49: MEM15 built on the relative neighborhood graph connectivity scheme for *P. longirostris*.

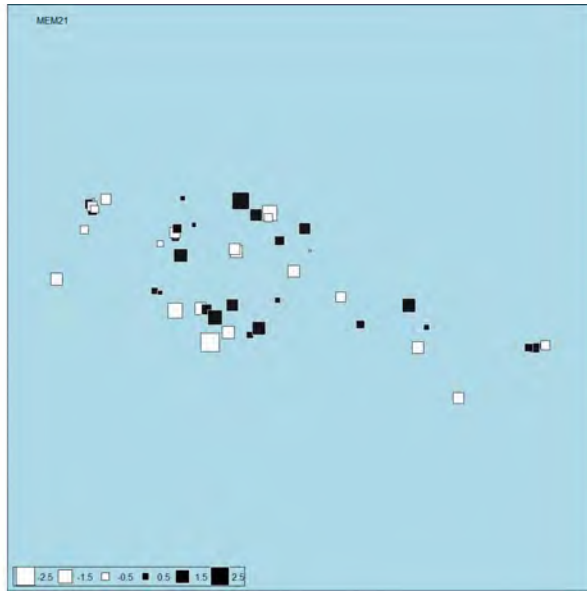


Figure 50: MEM21 built on the relative neighborhood graph connectivity scheme for *P. longirostris*.

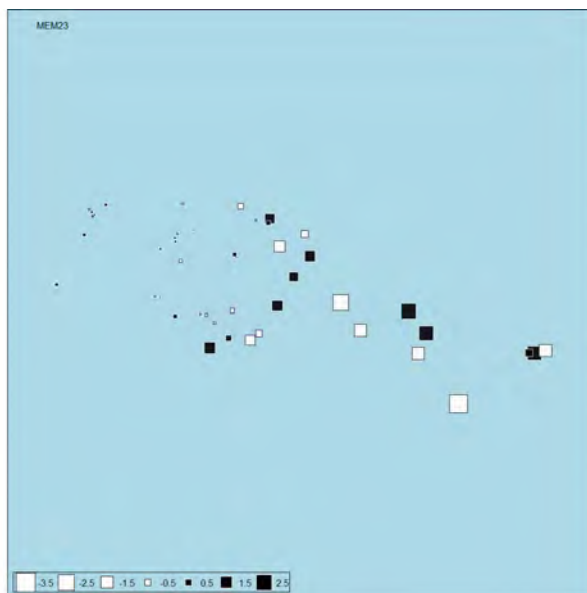


Figure 51: MEM23 built on the relative neighborhood graph connectivity scheme for *P. longirostris*.

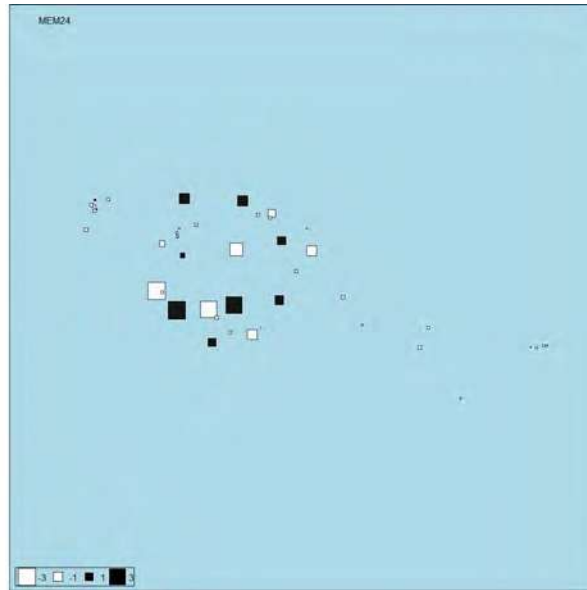


Figure 52: MEM24 built on the relative neighborhood graph connectivity scheme for *P. longirostris*.

The final spatial models (Fig. 53) showed cluster 2 associated to high longitudes and MEM21 and MEM24, while cluster 1 was associated to low longitudes and MEM15 and MEM23. At NC = 3 the situation is similar, with the difference between these two clusters still explained mostly on the first gradient axis, while the second gradient tried to separate the cluster 1 and 2 from cluster 3. Considering also the much lower variance explained by this second RDA axis, the division in 3 clusters does not seem justified by the spatial models.

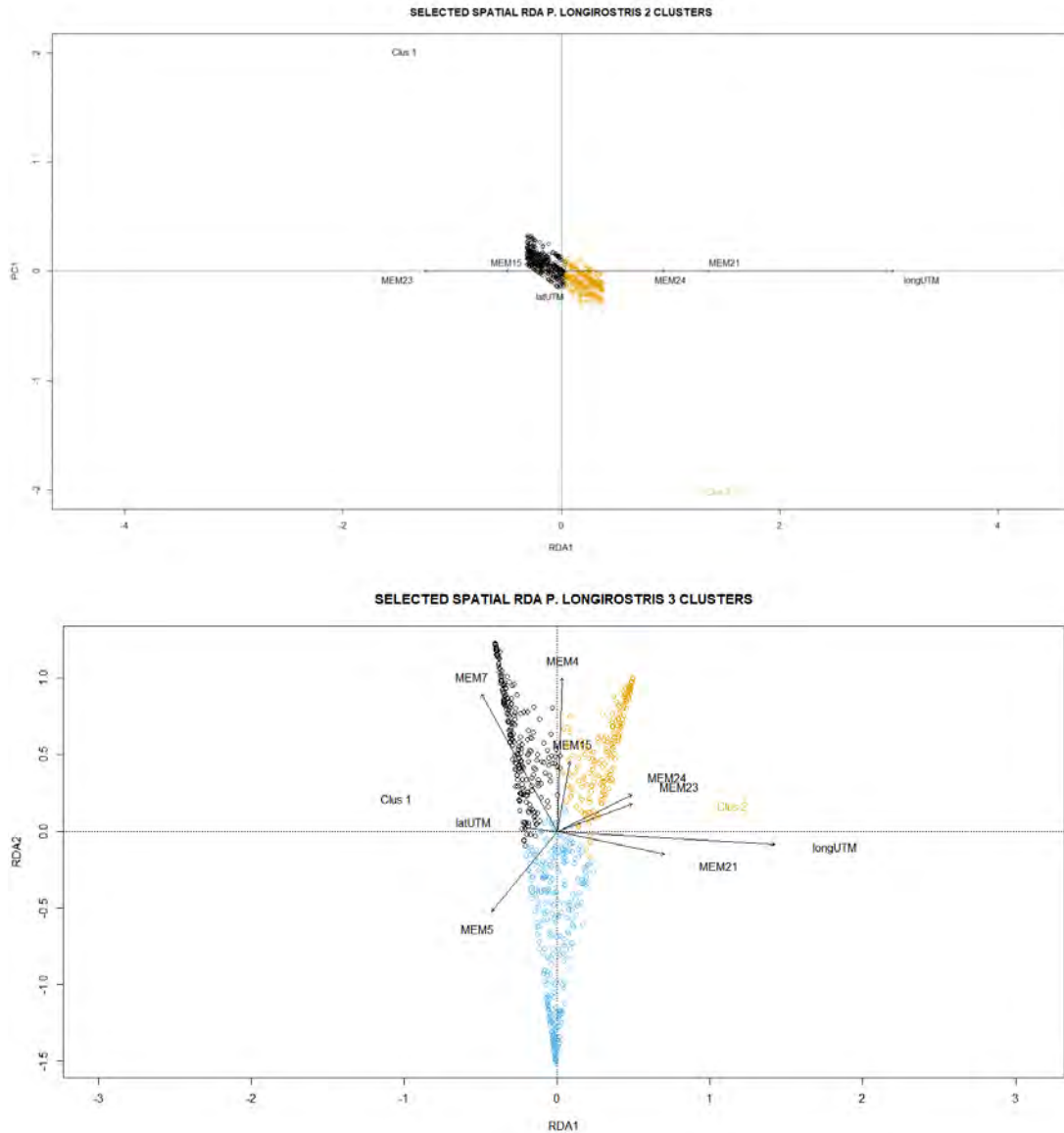


Figure 53: Best spatial parsimonious RDA models for *P. longirostris* at NC = 2 (upper panel) and NC = 3 (lower panel). Explained variance on the first and only RDA axis was 0.34 for NC = 2 and was 0.18 and 0.02 on the first and second RDA axis at NC = 3.

Environmental analysis

The environmental models were built by explaining the fuzzy k-means membership grades with the dataset of 13 median environmental variables extracted for two depths (bottom and euphotic zone) and three time intervals (5 years, 3 years, 1 year). The dataset calculated over the euphotic zone proved almost always better in explaining genetic PCAs of *P. longirostris* than the bottom dataset. The best result was obtained for NC = 2 with the “surface” medians over 1 year (Table 48). At NC = 3 the adjusted R^2 were also still high but decreased progressively with increasing NC.

Table 48: Adjusted R^2 for *P. longirostris*, calculated for RDA models with environmental variables as explanatory variables. Different models were built for each set of explanatory variables: bottom and surface (euphotic zone), and the three time intervals considered (5 years, 3 years, 1 year). Results are reported for the clusterizations from 2 to 10 clusters.

DPS	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.02	0.03	0.03	0.25	0.25	0.24	0.14	0.14	0.14	0.09	0.09	0.09	0.07	0.07	0.07
surface	0.03	0.03	0.04	0.31	0.31	0.32	0.17	0.18	0.18	0.11	0.11	0.12	0.08	0.09	0.09
	6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.03	0.04	0.04
surface	0.06	0.07	0.07	0.06	0.06	0.06	0.05	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04

The set of median euphotic zone environmental variables over 1 year was subjected to a forward selection procedure with a double-stopping criterion to build the parsimonious environmental model (Table 49) explaining the variance of the membership grades of *P. longirostris* for NC = 2 and NC = 3. Three variables were selected in both cases, i.e., surface alkalinity, temperature and velocity. The other two variables were different, since at NC = 2 two nutrient concentrations were selected (PO₄ and NO₃), while at NC = 3 two variables related to primary production were selected (chlorophyll concentration and phytoplankton biomass). These two pairs of variables seem interchangeable (Fig. 54), since at NC = 2 and at NC = 3 they are related to cluster 1 and together with velocity of the euphotic zone define one of the extremes of the main gradient. Thus, the differences of the Western and of the Eastern Mediterranean basins in terms of physical and biogeochemical variables is best related to the observed differences in the genetic PCAs.

Table 49: Forward selection on environmental variables at surface over 1 years for *P. longirostris* genetic PCAs at NC = 2 and NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

DPS 2 CLUSTERS			DPS 3 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_s_ALK	0.27	0.001	P50_s_ALK	0.14	0.001
P50_s_T	0.28	0.002	P50_s_vel	0.15	0.001
P50_s_vel	0.29	0.005	P50_s_T	0.15	0.029
P50_s_PO4	0.29	0.028	P50_s_chlf	0.15	0.037
P50_s_NO3	0.30	0.004	P50_s_phy	0.17	0.001

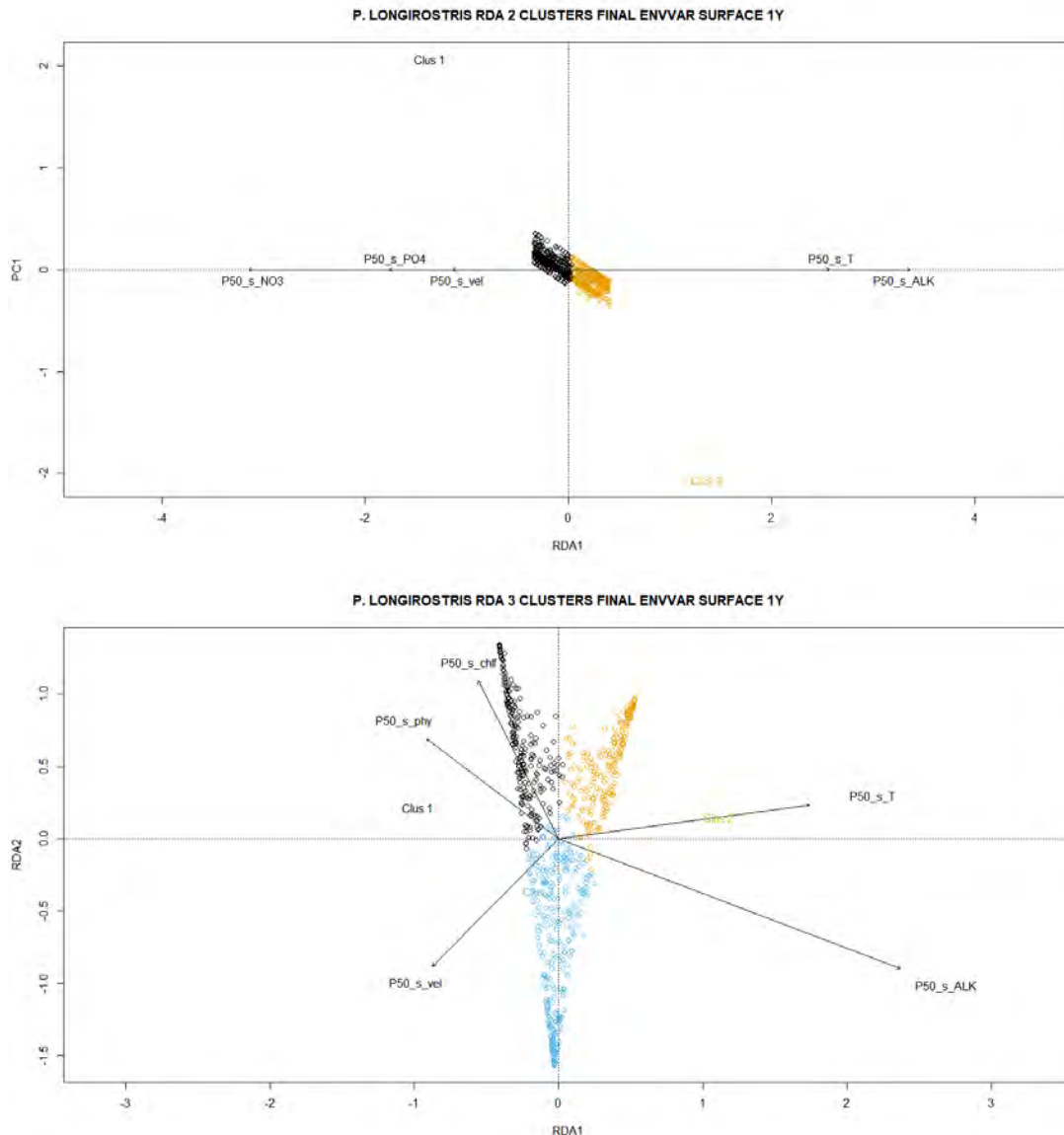


Figure 54: Final environmental RDA models for *P. longirostris* at NC = 2 (upper panel) and NC = 3 (lower panel). Explained variance on the first and only RDA axis at NC = 2 was 0.30; explained variance on the first and second significant RDA axes at NC = 3 was respectively 0.16 and 0.01.

Variation partitioning and best model selection

The four groups of variables (depth, geographic coordinates, significant MEMs, median environmental variables over 1 year in the euphotic zone) were used in the variation partitioning procedure in order to explore their mutual relationships and explanatory power on genetic PCAs of *P. longirostris* (Table 50). The large-scale pattern separating the Western and Eastern Mediterranean *P. longirostris* sub-populations reflected in the large explanatory power of the geographic coordinates, i.e., the variables used to model this same large-scale gradient. The environmental variables had the highest explanatory power and were able to represent also the longitude-latitude gradient when considered as a group (Fig. 55). Nevertheless, the MEMs were the only that brought some additional explanatory power when compared to all three other groups. Thus, while the general gradient is best represented by the geographic coordinates and the environmental variables, it is the

addition of some medium and small scales MEMs that can increase the explained variance of the distribution of *P. longirostris* membership grades at NC = 2 and NC = 3.

Table 50: Variation partitioning for the *P. longirostris* genetic dataset at NC = 2 and NC = 3. + indicates the total variation explained by the groups of variables; | indicates the variation explained by the first group of variables conditioned to the second; U indicates the fraction of variance explained jointly by the two groups of variables. Compare to Fig. 55 for a graphical representation.

PARAPENAEUS LONGIROSTRIS	2 CLUSTERS	3 CLUSTERS
Explanatory group of variables	Adjusted R²	Adjusted R²
Depth	0.06	0.03
LongLat	0.28	0.14
MEM	0.10	0.10
EnvVar Surf 1y	0.32	0.18
Depth + LongLat	0.29	0.15
Depth + MEM	0.12	0.12
Depth + EnvVar Surf 1y	0.32	0.18
LongLat + MEM	0.34	0.19
LongLat + EnvVar Surf 1y	0.32	0.18
MEM + EnvVar Surf 1y	0.34	0.20
Depth + LongLat + MEM	0.34	0.19
Depth + LongLat + EnvVar Surf 1y	0.32	0.18
Depth + MEM + EnvVar Surf 1y	0.34	0.20
LongLat + MEM + EnvVar Surf 1y	0.34	0.20
Depth + LongLat + MEM + EnvVar Surf 1y	0.34	0.20
Depth LongLat	0.01	0.00
Depth MEM	0.02	0.02
Depth EnvVar Surf 1y	0.00	0.00
LongLat Depth	0.24	0.12
LongLat MEM	0.24	0.09
LongLat EnvVar Surf 1y	0.00	0.00

MEM Depth	0.06	0.09
MEM LongLat	0.06	0.04
MEM EnvVar Surf 1y	0.03	0.03
EnvVar Surf 1y Depth	0.26	0.15
EnvVar Surf 1y LongLat	0.03	0.03
EnvVar Surf 1y MEM	0.24	0.10
Depth MEM + EnvVar Surf 1y	0.00	0.00
Depth LongLat + EnvVar Surf 1y	0.00	0.00
Depth LongLat + MEM	0.00	0.00
LongLat MEM + EnvVar Surf 1y	0.00	0.00
LongLat Depth + EnvVar Surf 1y	0.00	0.00
LongLat Depth + MEM	0.22	0.07
MEM Depth + EnvVar Surf 1y	0.02	0.03
MEM LongLat + EnvVar Surf 1y	0.03	0.02
MEM Depth + LongLat	0.04	0.04
EnvVar Surf 1y LongLat + MEM	0.00	0.01
EnvVar Surf 1y Depth + MEM	0.22	0.08
EnvVar Surf 1y Depth + LongLat	0.02	0.03
Depth LongLat + MEM + EnvVar Surf 1y	0.00	0.00
LongLat Depth + MEM + EnvVar Surf 1y	0.00	0.00
MEM Depth + LongLat + EnvVar Surf 1y	0.02	0.02
EnvVar Surf 1y Depth + LongLat + MEM	0.00	0.01
Depth \cup LongLat	0.00	0.00
LongLat \cup MEM	0.00	0.00
Depth \cup MEM	0.00	0.00
Depth \cup EnvVar Surf 1y	0.00	0.00

LongLat \cup EnvVar Surf 1y	0.22	0.07
MEM \cup EnvVar Surf 1y	0.02	0.01
Depth \cup LongLat \cup EnvVar Surf 1y	0.02	0.02
Depth \cup LongLat \cup MEM	0.00	0.00
LongLat \cup MEM \cup EnvVar Surf 1y	0.02	0.05
Depth \cup MEM \cup EnvVar Surf 1y	0.01	0.00
Depth \cup LongLat \cup MEM \cup EnvVar Surf 1y	0.03	0.01
Residuals	0.66	0.80

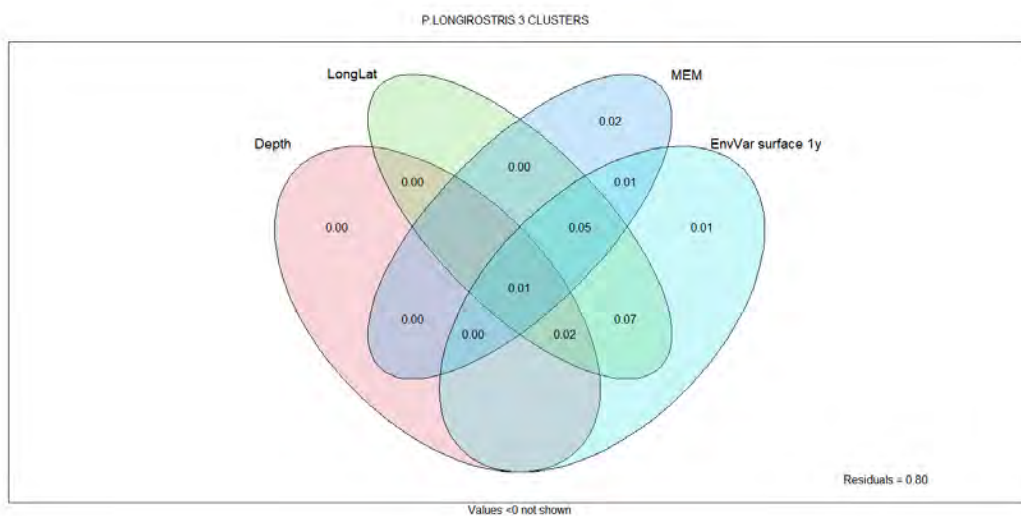
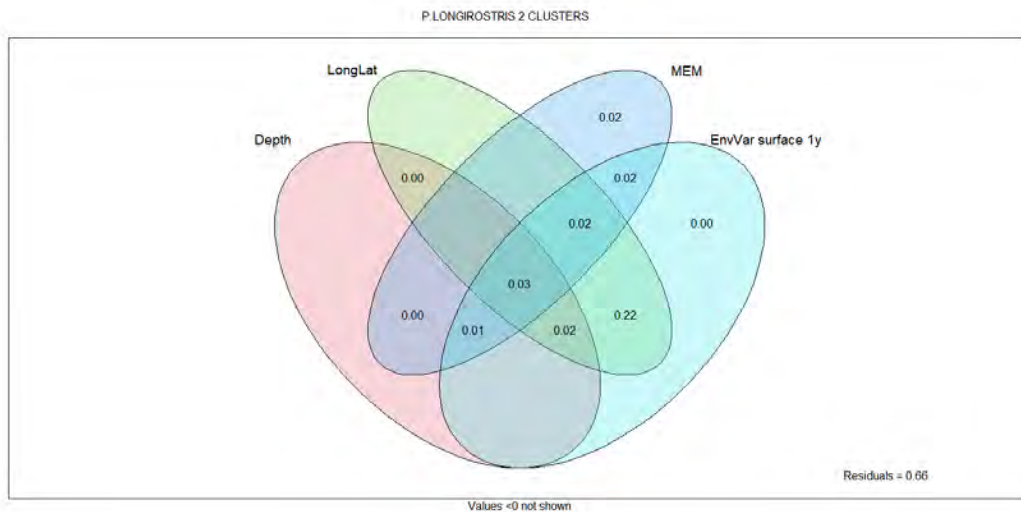


Figure 55: Graphical representation of the variation partitioning of the *P. longirostris* genetic dataset at 2 and 3 clusters among four groups of variables (Depth, LongLat, MEM, EnvVar Surf 1y). Negative or zero fractions are not displayed. See Table 55 for more details.

All these four groups of variables were subjected to a forward selection procedure with a double-stopping criterion in order to select the best model for *P. longirostris* fuzzy membership grades (Table 51). At NC = 2 the final model had an adjusted $R^2 = 0.32$ with 5 significant variables: surface alkalinity and NH_4 concentrations, longitude, depth and MEM23. At NC = 3 the adjusted R^2 was 0.19, but obtained with 11 variables: surface alkalinity, velocity, salinity, longitude and latitude, and 6 large and medium scale MEMs (MEM2, MEM4, MEM15, MEM20, MEM21, MEM24). Again, this supports the choice of NC = 2 as the best grouping of *P. longirostris* genetic data. The resulting biplots of the best models are shown in Fig. 56 and do not add much more than the spatial and the environmental models described above, but for the depth at NC = 2 characterizing with high values cluster 1 (i.e., the Western Mediterranean cluster).

Table 51: Forward selection on the final models for *P. longirostris* genetic PCAs at NC = 2 and NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

DPS 2 CLUSTERS			DPS 3 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_s_ALK	0.27	0.001	P50_s_ALK	0.14	0.001
MEM23	0.29	0.001	P50_s_vel	0.15	0.001
longUTM	0.30	0.001	MEM24	0.16	0.004
Depth	0.32	0.001	longUTM	0.16	0.003
P50_s_NH4	0.32	0.001	MEM20	0.17	0.009
			MEM2	0.17	0.01
			MEM4	0.18	0.001
			P50_s_sal	0.18	0.006
			MEM21	0.18	0.016
			MEM15	0.19	0.028
			latUTM	0.19	0.049

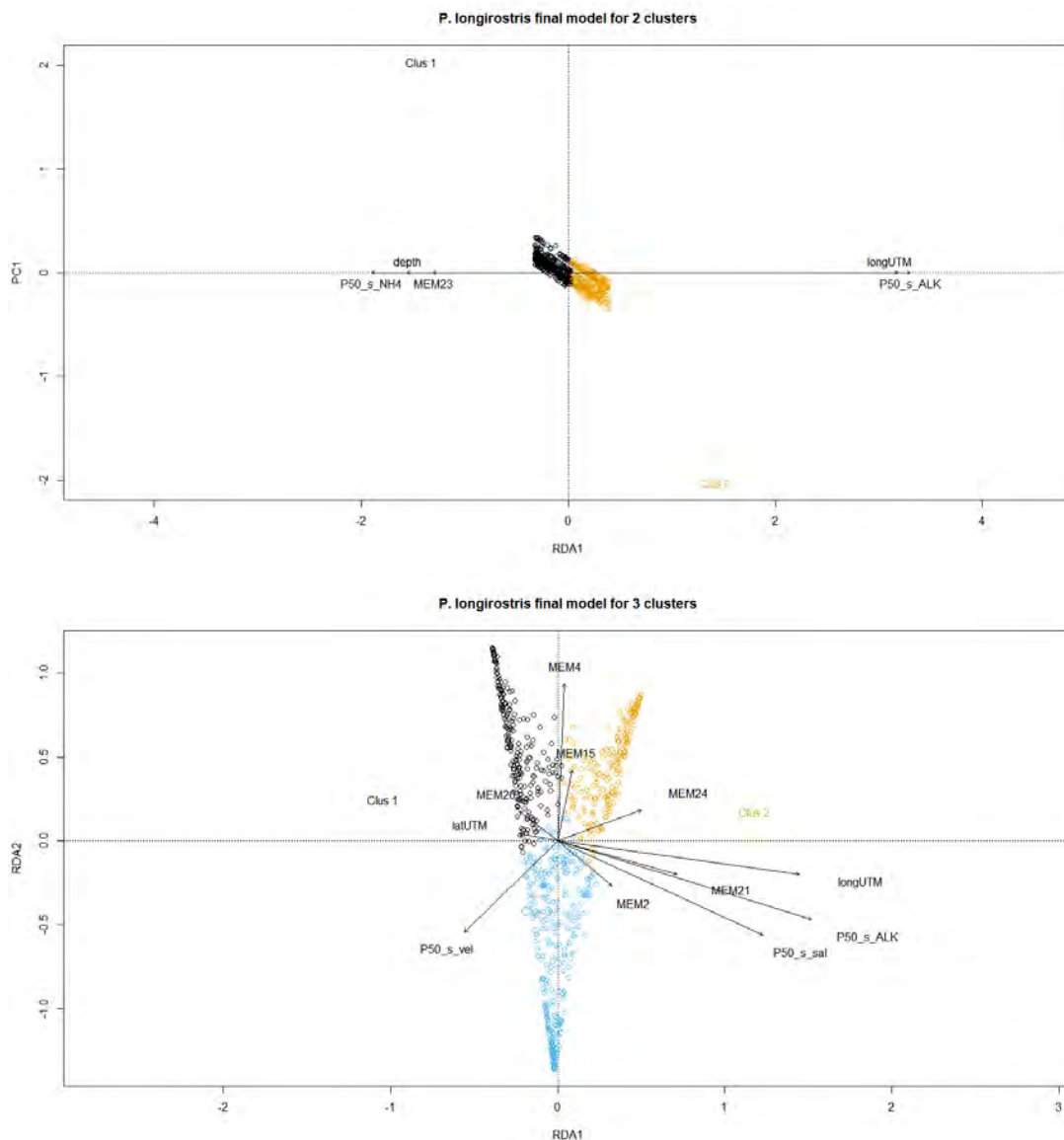


Figure 56: Best final parsimonious RDA models for *P. longirostris* at NC = 2 (upper panel) and NC = 3 (lower panel). Explained variance on the first and only RDA axis was 0.33 at NC = 2, and 0.18 and 0.02 on the first and the second RDA axes at NC = 3.

Conclusions

The analysis of the *P. longirostris* genetic PCAs led to very interesting results. The results of the internal and external validation procedure were in strong agreement. Thus, for this species to a clear genetic diversification in sub-populations corresponds a clear spatial and environmental diversification of their habitats, i.e.: different stocks do exist in the Mediterranean for *P. longirostris*.

It was clear already from the internal cluster validity indexes applied that there is a clear separation in the genetic data in two sub-population. The fuzzy membership grades obtained showed clear preferences for one of the clusters in most cases. By aggregating the data on the basis of GSA subareas still one cluster clearly prevailed over the others at NC = 2, while at NC = 3 the situation started to be fuzzier.

The external validation procedures applied here, i.e., the spatial and environmental models built, did confirm the results of the fuzzy k-means: two stock of *P. longirostris* do co-exist in the Mediterranean, one found mostly in the Western and Central Mediterranean, the other in the Adriatic and in the Eastern Mediterranean. The discontinuity between the two seem located in the 19b subarea of the Ionian. Possible causes of such diversification might be the current patterns influencing dispersal in the depth range of this species, and the complex bathymetry and morphology characterizing especially the area where the discontinuity between the two sub-populations does occur, i.e., between the Strait of Sicily and the Ionian Sea.

Mullus barbatus

Data

Data for *Mullus barbatus* were made up by 20 Principal Components (PCA) (Table 5) extracted from the genetic data in WP1 (D.1.5.2); 6 PCA (Table 5) extracted from otolith shape data (D.2.4); and concentrations of 16 chemical elements from the core region of the otoliths and of 16 chemical elements from the edge region of the otoliths (Table 52), as provided by the microchemistry working group (D.4.5).

All variables showed different means and standard deviations, both inside each dataset and across different datasets. To prevent variables with highest means and standard deviations to unduly influence the analyses, all variables were standardized before the analysis by subtracting the mean and dividing by the standard deviation.

The number of individuals with complete records for each of the three datasets was different, as it was the number of sampling location, i.e., the spatial coverage of each dataset (Table 52). Otolith shape data had the highest number of individuals with complete records and the widest spatial coverage. Genetic data had also a rather high spatial coverage but with much less individuals sampled. Finally, the otolith microchemistry data had the lowest number of individuals analyzed and presented the lowest spatial coverage (Fig. 57).

As for the hake, we also considered the following combinations of the datasets produced respectively by the genetic, otolith shape and otolith microchemistry working groups: a combination of all three datasets; a combination of the genetic and otolith shape variables; and a combination of the otolith shape and otolith microchemistry data. The combination of all three datasets produced the highest number of variables to include in the analysis, but its spatial coverage was very limited by that of otolith microchemistry data. The rationale for the combination of the genetic and otolith shape was that both datasets have a wide spatial coverage. The combined, genetic-otolith shape dataset, also shows quite a high spatial coverage, but a much-reduced number of individuals if compared to that of the otolith shape dataset. The rationale for the combination of the otolith shape and otolith microchemistry data was that both refer to characteristics of the otoliths. Thus, their combination might give us more insight into possible diversification among populations based on the otoliths, while again its spatial coverage was limited by that of otolith microchemistry data. In Table 52 there also the acronyms by which each dataset or combination of datasets will be referred to from now on.

Table 52: Datasets used in the analysis for Mullus barbatus.

DATASET	ACRONYM	N VARIABLES	TYPE OF VARIABLE	COMPLETE RECORDS	SAMPLING LOCATIONS
GENETIC	GEN	20	Principal components	771	68
OTOLITH SHAPE	OTHO	6	Principal components	1756	80

MICROCHEMISTRY	MICRO	32	Element concentrations in otolith core and edge	250	25
GENETIC + OTOLITH SHAPE	GEN.OTHO	26	Principal components	624	64
OTOLITH SHAPE + MICROCHEMISTRY	OTHO.MICRO	38	Principal components + Element concentrations in otolith core and edge	239	25
ALL (GENETIC + OTOLITH SHAPE + OTOLITH MICROCHEMISTRY)	ALL	58	Principal components + Element concentrations in otolith core and edge	130	20

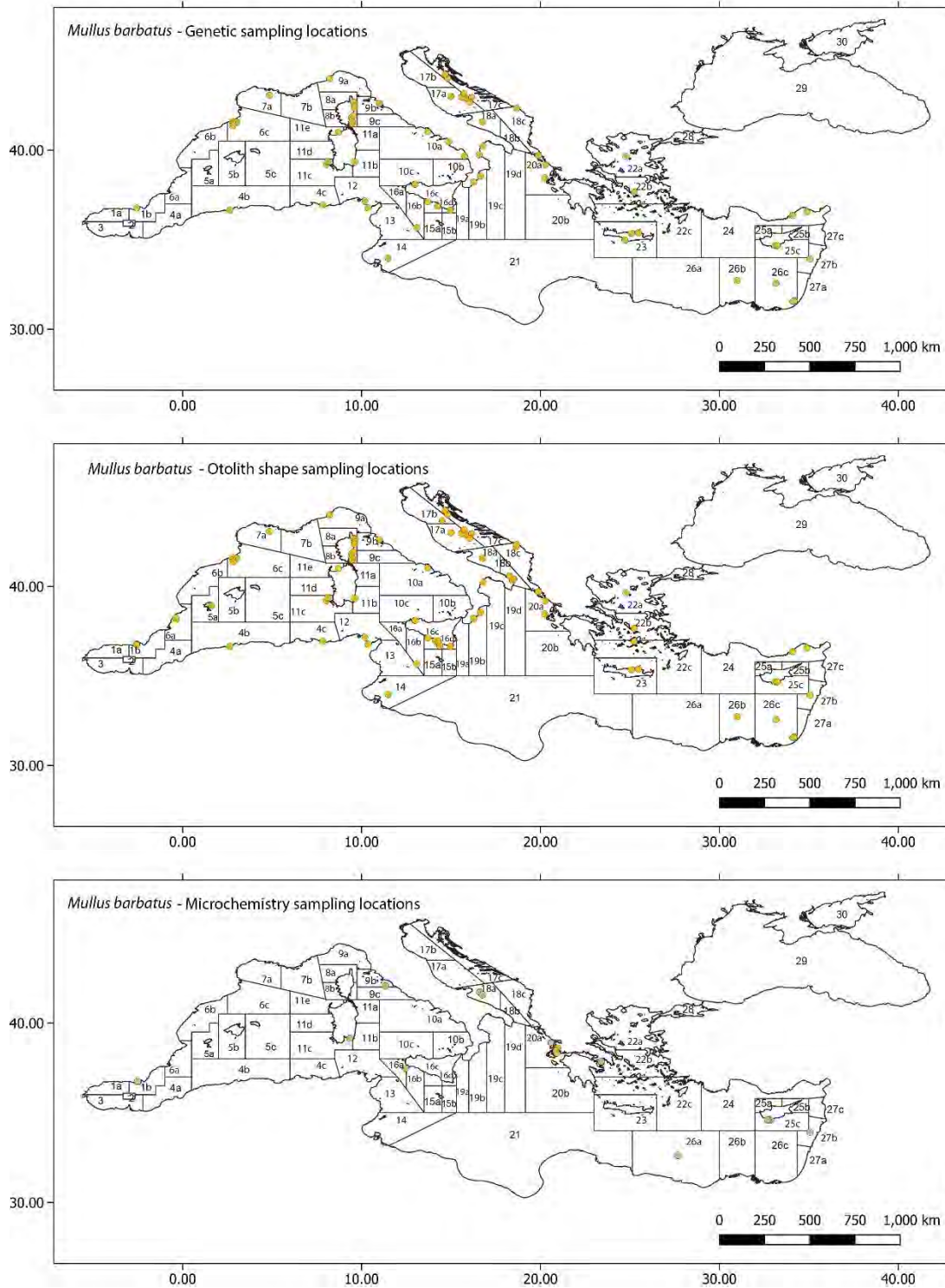


Figure 57: distribution of sampling location for *Mullus barbatus* genetic data (top), otolith shape data (middle), otolith microchemistry data (bottom).

Results and discussion

Fuzzy clustering

The full results of the internal indexes for cluster validity for *Mullus barbatus* are not presented, but in Table 53 these results are summarized reporting how many times each

number of clusters was chosen as the optimal clusterization or the second optimal clusterization by the six internal cluster validity indexes applied.

The applied internal cluster validity indexes gave quite different results, both among them and across the different datasets of *Mullus barbatus* analyzed. For some datasets the situation is very unclear (e.g., GEN and GEN.OTHO datasets), while for some others there was unanimity between cluster validity indexes (MICRO, OTHO.MICRO and ALL datasets). In this latter case the lower NC were favored, while for GEN, OTHO and GEN.OTHO some indexes indicated low NC, while others indicated high NC (i.e., 7, 9 or 10). That the extremes of the NC considered are favored by the six clusters validity indexes might indicate that no clear structure is present in the dataset examined. Thus, it seems that the different datasets show a different internal structure, possibly related to different characteristics of *Mullus barbatus* individuals. To apply the majority vote criterion, as done for the hake, is also difficult due to the variation of the results. The external validation, as will be performed later on with the environmental and spatial variables, might add clarity to the results of the internal validation with the cluster validity indexes.

Table 53: The number of times each NC for each dataset of *Mullus barbatus* was selected by the six internal cluster validity indexes applied as the optimal number of clusters or the second optimal number of clusters.

SELECTED	GEN		OTHO		MICRO		GEN.OTHO		OTHO.MICRO		ALL	
	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd
NC = 2	2		2		6		3		5	1	6	
NC = 3		2		2		4		2	1	3		2
NC = 4				1								1
NC = 5												
NC = 6				3								
NC = 7	1	1	4									
NC = 8										1		
NC = 9		2					1	3				
NC = 10	3	1				2	2	1		1		3

Considering the average membership grades per GFCM subregions for the GEN dataset (Table 54) at NC = 2 there seem to be a division between the Eastern basin and the other three subregions. Nevertheless, care should be taken when considering these results since the number of individuals analyzed and the number of GSA subareas explored was lower for the Eastern Mediterranean. At NC =3 the Adriatic and the Central subregions showed maximum average membership grades for the same cluster, while the Western basin and the Eastern characterized the other two clusters. Such division remains also with higher NC

(up to NC = 7), but the average membership grades per subregion are very close to 1/NC, i.e., to being equally distributed among clusters.

Considering the results per GSA subareas (Table 55) at NC = 2 there seem to be a cluster grouping the majority of subareas in the central part of the Mediterranean (CL1), and another grouping those at the westernmost and easternmost part (CL2). At NC = 3 the situation is clearer, with CL2 present from 1b, along the southern Mediterranean coast (14) to the Eastern part, while CL3 is prevalent in the Adriatic and found also in the Tyrrhenian, Ionian, along the French coast and close to Crete. CL1 is prevalent in most of the Western Mediterranean basin, along with subareas 18c and 20a. At NC = 4 the spatial rationale of the clusterization seem more complicated and remains so also with higher NC. The average membership grades are also very similar among clusters at higher NC, possibly indicating a low level of diversification in the genetic data.

Table 54: Spatial interpretation of fuzzy clusters results for the GEN dataset of *Mullus barbatus* considering GFCM subregions at NC = 2:4. Highlighted in green the highest average membership values for each subregion.

GEN	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.506	0.494	0.351	0.324	0.325	0.250	0.227	0.280	0.243
Center	0.517	0.483	0.335	0.331	0.335	0.247	0.256	0.240	0.257
Adriatic	0.529	0.471	0.309	0.313	0.378	0.233	0.280	0.216	0.271
East	0.488	0.512	0.303	0.354	0.343	0.269	0.267	0.206	0.258

Table 55: Spatial interpretation of fuzzy clusters results for the GEN dataset of *Mullus barbatus* considering GSA subareas at NC = 2:4. Highlighted in green the highest average membership values for each subarea.

GEN		2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
1b	West	0.369	0.631	0.282	0.496	0.222	0.433	0.140	0.235	0.192
5a	West	0.496	0.504	0.421	0.293	0.286	0.196	0.191	0.371	0.242
6a	West	0.493	0.507	0.400	0.323	0.277	0.254	0.220	0.332	0.194
6c	West	0.488	0.512	0.293	0.331	0.376	0.189	0.315	0.209	0.287
7a	West	0.512	0.488	0.324	0.326	0.350	0.259	0.179	0.249	0.314
8a	West	0.442	0.558	0.443	0.347	0.210	0.280	0.147	0.353	0.219
8b	West	0.567	0.433	0.437	0.224	0.339	0.168	0.204	0.340	0.288
9a	West	0.498	0.502	0.358	0.306	0.336	0.241	0.244	0.284	0.230
9b	West	0.584	0.416	0.293	0.265	0.442	0.194	0.258	0.193	0.355
10a	West	0.564	0.436	0.336	0.268	0.396	0.204	0.297	0.253	0.246
10c	West	0.514	0.486	0.334	0.334	0.331	0.244	0.285	0.266	0.205
11b	West	0.540	0.460	0.402	0.280	0.318	0.198	0.258	0.321	0.223
11c	West	0.567	0.433	0.274	0.283	0.443	0.215	0.240	0.219	0.326
12	Centre	0.538	0.462	0.374	0.294	0.332	0.225	0.234	0.263	0.278
14	Centre	0.513	0.487	0.307	0.362	0.331	0.291	0.210	0.211	0.288
17a	Adriatic	0.548	0.452	0.302	0.286	0.411	0.212	0.355	0.204	0.229
17b	Adriatic	0.561	0.439	0.282	0.301	0.418	0.226	0.261	0.184	0.329
17c	Adriatic	0.551	0.449	0.339	0.298	0.363	0.227	0.294	0.252	0.228
18a	Adriatic	0.472	0.528	0.358	0.397	0.245	0.287	0.205	0.289	0.219

18b	Adriatic	0.507	0.493	0.265	0.320	0.415	0.225	0.322	0.169	0.285
18c	Adriatic	0.521	0.479	0.355	0.309	0.336	0.248	0.215	0.260	0.277
19b	Centre	0.480	0.520	0.332	0.360	0.307	0.273	0.222	0.276	0.229
19c	Centre	0.535	0.465	0.284	0.335	0.381	0.216	0.383	0.160	0.241
20a	Centre	0.571	0.429	0.422	0.223	0.355	0.147	0.334	0.333	0.186
23	East	0.490	0.510	0.275	0.355	0.370	0.269	0.251	0.189	0.291
25c	East	0.515	0.485	0.388	0.292	0.320	0.194	0.247	0.286	0.272
26b	East	0.486	0.514	0.320	0.353	0.327	0.268	0.249	0.209	0.273
26c	East	0.423	0.577	0.299	0.401	0.300	0.318	0.248	0.215	0.219
27a	East	0.510	0.490	0.265	0.370	0.365	0.288	0.344	0.161	0.208
27b	East	0.461	0.539	0.245	0.391	0.363	0.332	0.247	0.160	0.261

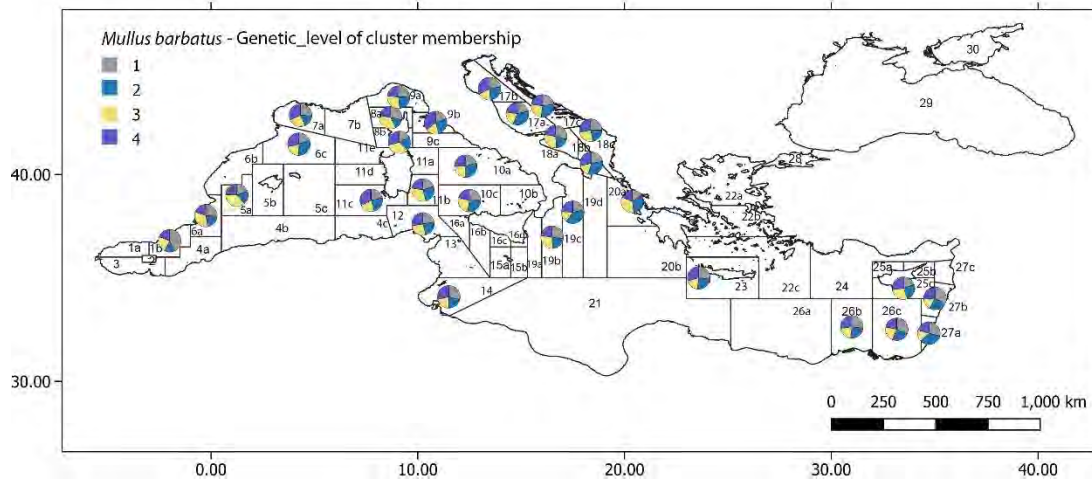
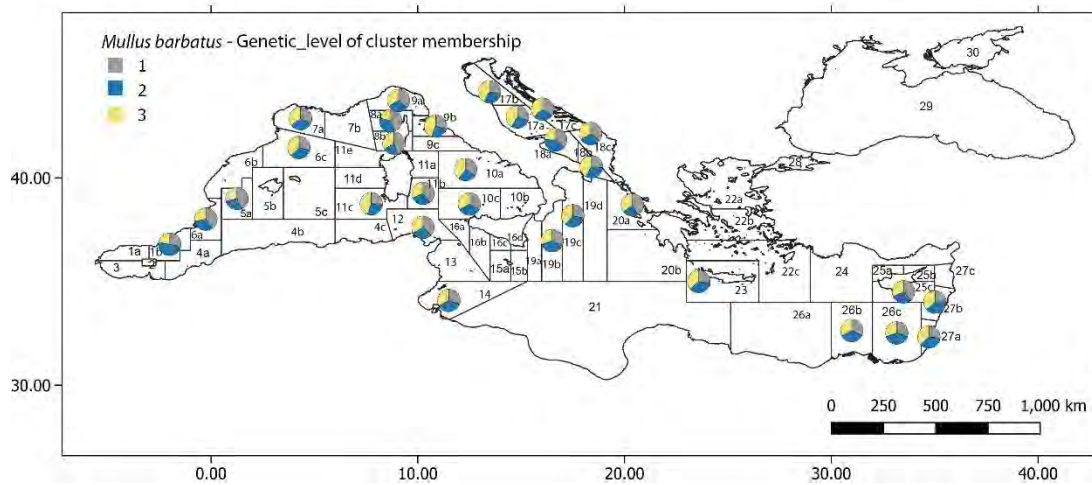


Figure 58: Pie charts with average membership grades per GSA subarea for the GEN dataset of *Mullus barbatus* at NC = 3 (upper panel) and NC = 4 (lower panel).

The results for the OTHO dataset, with the highest number of sampling location and of individuals, at the level of GFCM subregions (Table 56) does not give a clear picture of possible spatial structures. In fact, while at NC = 2 the Adriatic subareas are separated from the others, already at NC = 3 the highest average membership grade for each subregion is

for the same clusters (CL1). With increasing NC, the subregions shift affiliation among them and with clusters in way that seems random. Thus, these results cannot help in identifying the existence of diversification in *Mullus barbatus* populations in the Mediterranean.

At the level of GSA subareas, the picture is similar (Table 57): at NC = 2 there are some high average membership grades (e.g., 24, 26b, 26c, 27b) nevertheless by aggregating the results per subareas no clear spatial distribution seem to emerge. At NC = 3 and higher the spatial interpretation based on GSA remains difficult.

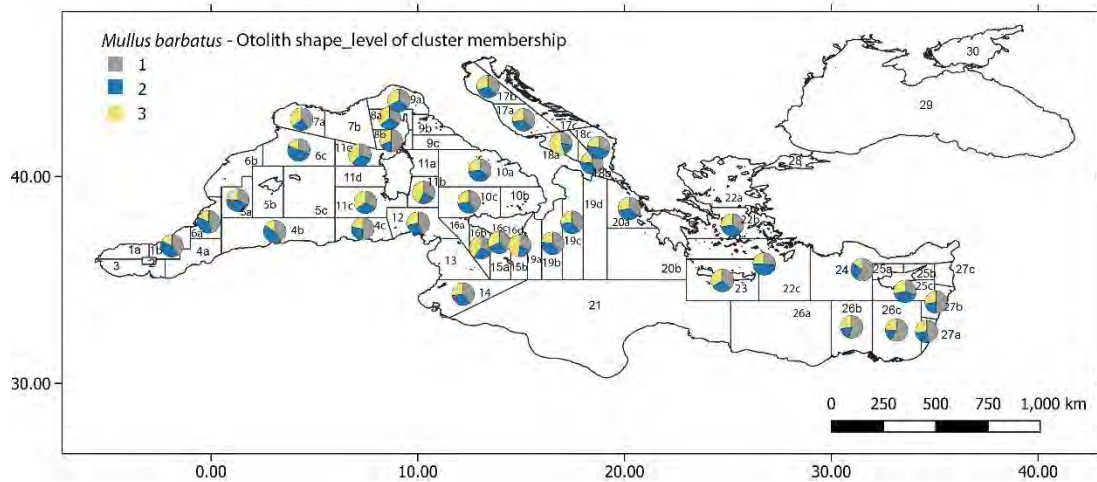
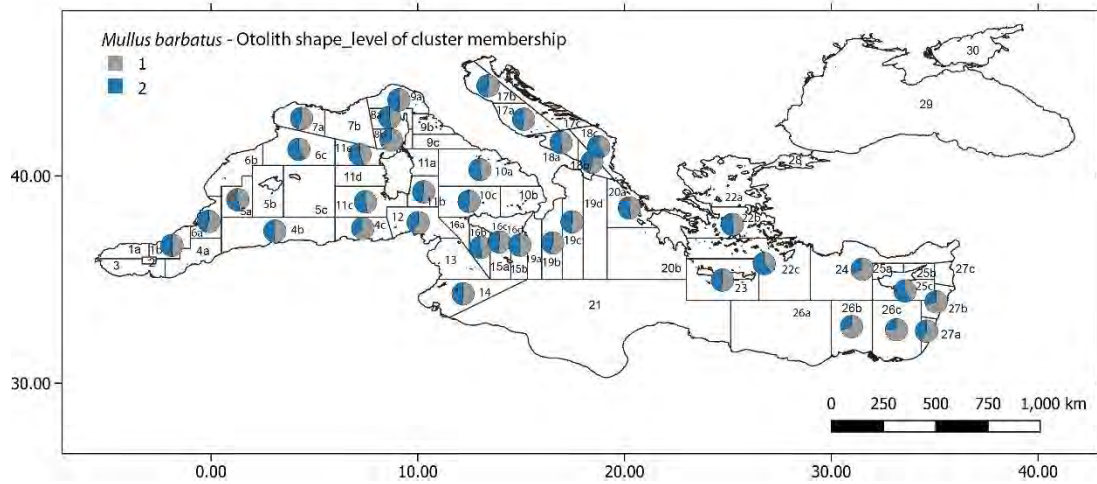
Table 56: Spatial interpretation of fuzzy clusters results for the OTHO dataset of *Mullus barbatus* considering GFCM subregions at NC = 2:4. Highlighted in green the highest average membership values for each subregion.

OTHO	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.511	0.489	0.379	0.341	0.280	0.263	0.203	0.238	0.297
Center	0.527	0.473	0.389	0.304	0.307	0.232	0.220	0.282	0.267
Adriatic	0.492	0.508	0.355	0.326	0.319	0.258	0.255	0.256	0.231
East	0.562	0.438	0.427	0.308	0.264	0.234	0.188	0.317	0.261

Table 57: Spatial interpretation of fuzzy clusters results for the OTHO dataset of *Mullus barbatus* considering GSA subareas at NC = 2:4. Highlighted in green the highest average membership values for each subarea.

OTHO		2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
1b	West	0.498	0.502	0.405	0.418	0.177	0.316	0.103	0.262	0.319
4b	West	0.514	0.486	0.407	0.429	0.165	0.348	0.091	0.257	0.303
4c	West	0.635	0.365	0.510	0.268	0.222	0.186	0.148	0.372	0.294
5a	West	0.471	0.529	0.319	0.445	0.236	0.336	0.138	0.182	0.343
6a	West	0.543	0.457	0.460	0.356	0.184	0.275	0.115	0.306	0.305
6c	West	0.420	0.580	0.312	0.501	0.187	0.400	0.135	0.201	0.264
7a	West	0.556	0.444	0.400	0.240	0.360	0.177	0.283	0.216	0.324
8a	West	0.476	0.524	0.336	0.311	0.353	0.231	0.287	0.208	0.274
8b	West	0.664	0.336	0.500	0.185	0.315	0.130	0.242	0.291	0.337
9a	West	0.493	0.507	0.353	0.319	0.328	0.252	0.224	0.240	0.285
10a	West	0.476	0.524	0.347	0.390	0.263	0.287	0.202	0.226	0.285
10c	West	0.517	0.483	0.404	0.322	0.274	0.261	0.215	0.281	0.244
11b	West	0.544	0.456	0.338	0.233	0.429	0.189	0.335	0.168	0.308
11c	West	0.472	0.528	0.339	0.313	0.348	0.230	0.257	0.206	0.307
11e	West	0.447	0.553	0.301	0.316	0.383	0.256	0.302	0.160	0.281
12	Centre	0.588	0.412	0.444	0.276	0.280	0.191	0.186	0.338	0.286
14	Centre	0.512	0.488	0.393	0.341	0.266	0.260	0.186	0.308	0.246
16b	Centre	0.487	0.513	0.315	0.283	0.402	0.225	0.331	0.276	0.168
16c	Centre	0.518	0.482	0.395	0.284	0.322	0.233	0.231	0.264	0.272
16d	Centre	0.500	0.500	0.338	0.185	0.476	0.134	0.383	0.180	0.302
17a	Adriatic	0.526	0.474	0.373	0.351	0.275	0.279	0.212	0.280	0.230
17b	Adriatic	0.510	0.490	0.378	0.320	0.301	0.247	0.213	0.241	0.300
18a	Adriatic	0.483	0.517	0.276	0.171	0.552	0.137	0.484	0.186	0.193
18b	Adriatic	0.536	0.464	0.436	0.326	0.238	0.240	0.170	0.320	0.269

18c	Adriatic	0.408	0.592	0.311	0.462	0.227	0.388	0.190	0.253	0.168
19b	Centre	0.543	0.457	0.406	0.365	0.230	0.264	0.123	0.316	0.297
19c	Centre	0.514	0.486	0.386	0.346	0.268	0.271	0.183	0.311	0.235
20a	Centre	0.477	0.523	0.358	0.363	0.278	0.309	0.211	0.195	0.285
22b	East	0.490	0.510	0.355	0.368	0.277	0.286	0.192	0.283	0.239
22c	East	0.381	0.619	0.268	0.485	0.247	0.419	0.192	0.225	0.164
23	East	0.514	0.486	0.373	0.291	0.335	0.236	0.278	0.238	0.248
24	East	0.677	0.323	0.601	0.252	0.147	0.176	0.080	0.290	0.454
25c	East	0.414	0.586	0.306	0.412	0.282	0.345	0.227	0.232	0.196
26b	East	0.704	0.296	0.536	0.186	0.278	0.106	0.179	0.446	0.268
26c	East	0.764	0.236	0.587	0.167	0.246	0.098	0.144	0.534	0.224
27a	East	0.589	0.411	0.468	0.261	0.271	0.186	0.187	0.305	0.322
27b	East	0.706	0.294	0.497	0.230	0.273	0.127	0.170	0.441	0.262



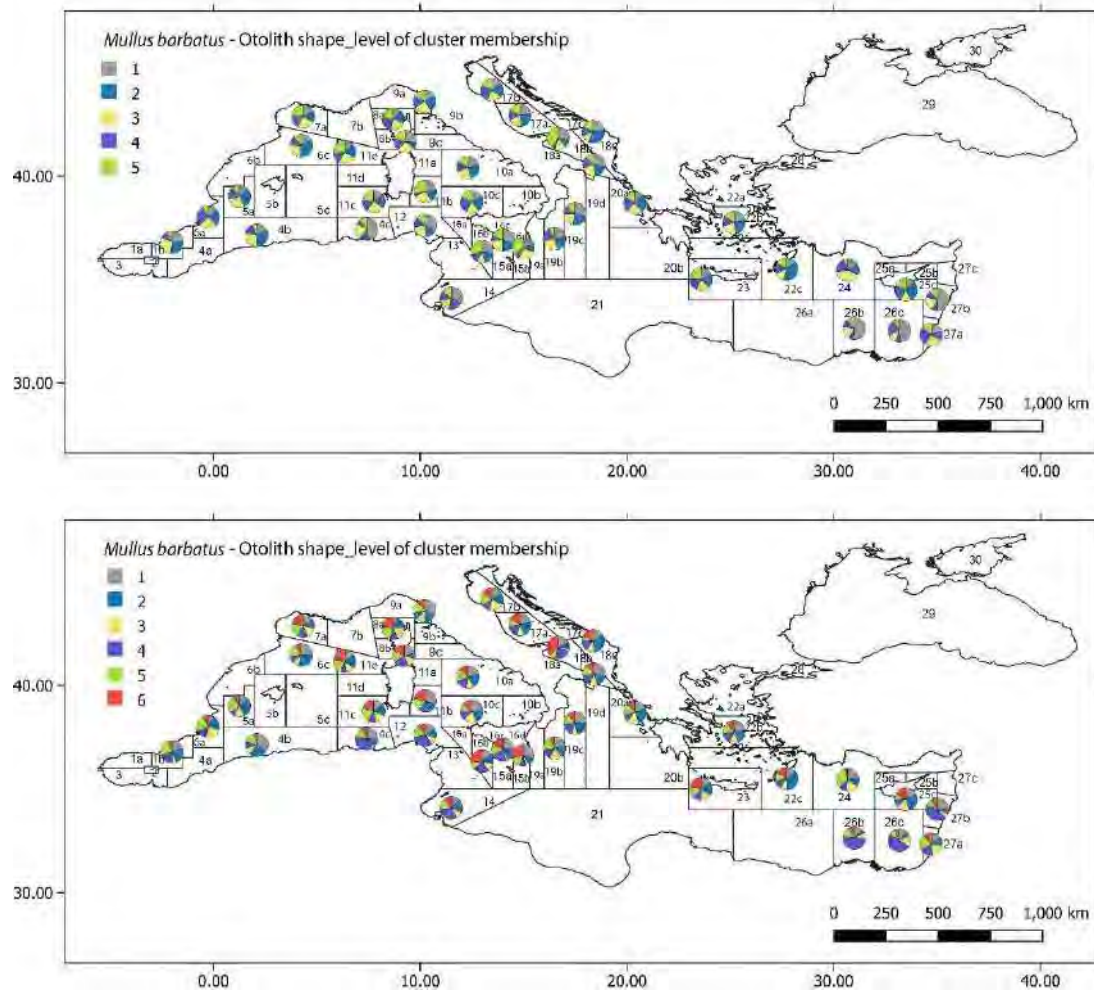


Figure 59: Pie charts with average membership grades per GSA subarea for the OTHO dataset of *Mullus barbatus* at NC = 2, 3, 5, 6 (top to bottom).

The MICRO dataset had individuals sampled in only 10 GSA subareas. At the level of GFCM subregions for NC = 2 there was no diversification among the subregions (Table 58). At NC = 3 the Eastern subregion separates from the rest, and from NC = 4 to NC = 10 the same pattern emerged: the Western and Adriatic subregion associated to the same cluster, while the Central and the Eastern with high average membership grades for other two clusters.

At the level of GSA subareas again the case of NC = 2 is useless: all subareas show the highest membership for the same cluster (Table 59). At NC = 3 subareas 1b, 22c, 27a are associated in one cluster, all the other subareas in another, while CL 2 has no GSA showing the maximum membership grade. From NC = 4 the subarea 22c separates from the rest and remains so till the maximum number of clusters explored. From NC = 5 also subarea 27a shows its own pattern of behavior. On the other hand, there are other two recognizable groups of subareas sharing clusters: subareas 1b, 6a, 9a and 18b; and subareas 11c, 12, 17a and 19b. It is hard to say if this division reflects different sub-populations of *Mullus barbatus*, possibly linked by dispersal, or it is a product of a much higher number of sub-populations that cannot be correctly identified with such low number of sampling locations. It has also to be reminded that these results are individual membership grade averaged per GSA: if the

boundaries of a GSA do not correspond to discontinuities in the species sub-populations, this analysis cannot reveal such discontinuities.

Table 58: Spatial interpretation of fuzzy clusters results for the MICRO dataset of Mullus barbatus considering GFCM subregions at NC = 2:4. Highlighted in green the highest average membership values for each subregion.

MICRO	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.802	0.198	0.389	0.119	0.492	0.391	0.291	0.222	0.096
Center	0.834	0.166	0.295	0.108	0.597	0.280	0.427	0.197	0.096
Adriatic	0.806	0.194	0.344	0.120	0.536	0.391	0.310	0.197	0.102
East	0.816	0.184	0.609	0.090	0.301	0.374	0.168	0.387	0.071

Table 59: Spatial interpretation of fuzzy clusters results for the OTHO dataset of Mullus barbatus considering GSA subareas at NC = 2:4. Highlighted in green the highest average membership values for each subarea.

MICRO		2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
1b	West	0.789	0.211	0.461	0.137	0.403	0.429	0.208	0.254	0.109
6a	West	0.789	0.211	0.438	0.113	0.449	0.493	0.191	0.244	0.071
9a	West	0.871	0.129	0.425	0.088	0.487	0.438	0.243	0.237	0.081
11c	West	0.761	0.239	0.231	0.138	0.630	0.202	0.523	0.153	0.122
12	Center	0.840	0.160	0.344	0.124	0.532	0.351	0.314	0.220	0.114
17a	Adriatic	0.760	0.240	0.301	0.156	0.544	0.305	0.383	0.182	0.130
18b	Adriatic	0.852	0.148	0.388	0.084	0.528	0.477	0.238	0.212	0.073
19b	Center	0.828	0.172	0.246	0.093	0.662	0.208	0.540	0.174	0.078
22c	East	0.811	0.189	0.590	0.126	0.284	0.342	0.156	0.391	0.112
27a	East	0.821	0.179	0.627	0.054	0.319	0.407	0.180	0.384	0.029

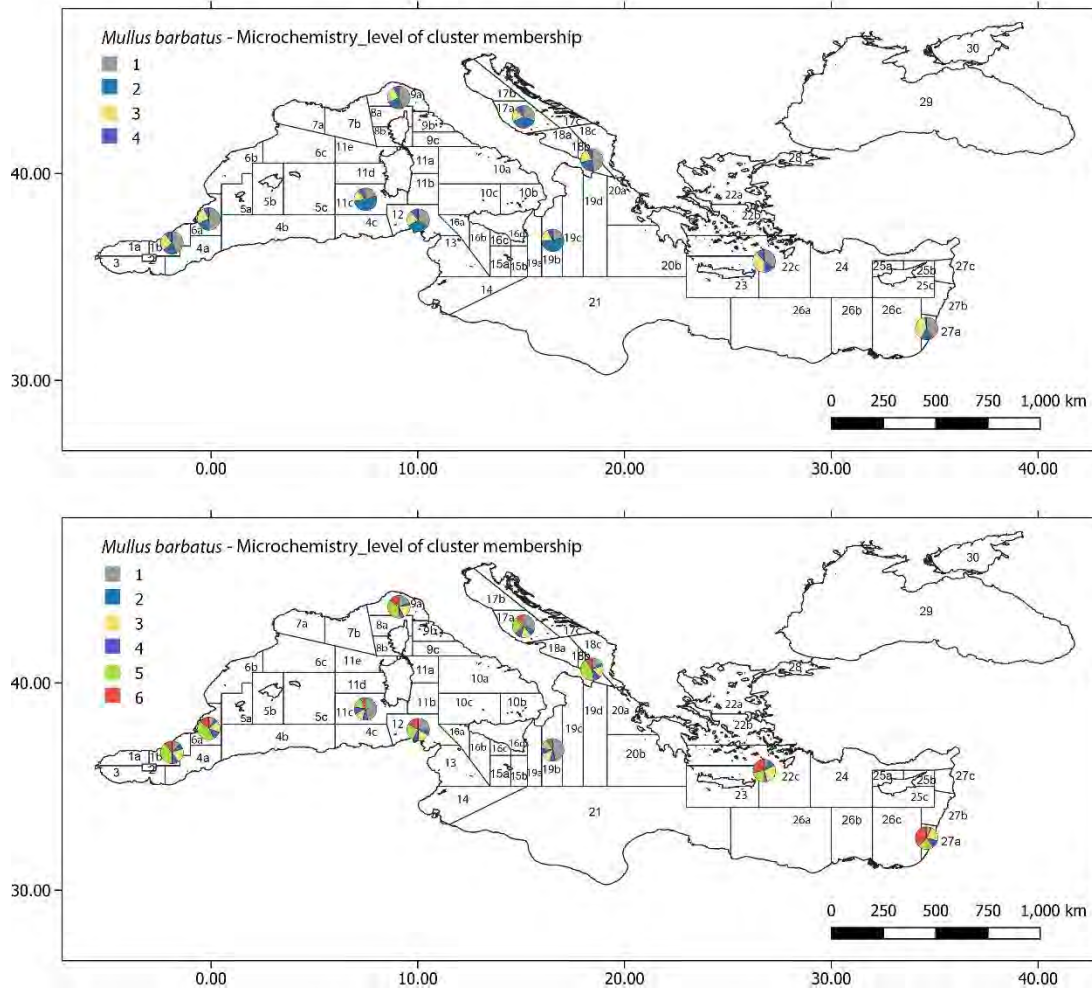


Figure 60: Pie charts with average membership grades per GSA subarea for the MICRO dataset of *Mullus barbatus* at NC = 4 (upper panel) and NC = 6 (lower panel).

The inspection of the results of the combination of genetic and otolith datasets at the GFCM level at NC = 2 shows the West clearly separated from the rest of the subregions. At NC = 3 the West and Center subregions are associated to the same cluster, while the Adriatic and East subregions are also associated with high average membership grade to the same cluster. At NC = 3 the associations among subregions shift again, with West associated to CL1, East to CL4 and Center and Adriatic to CL4 (Table 60). With increasing NC, the subregions shift association among them almost at every NC.

At the level of GSA subareas, no apparent clear spatial distribution of the clusters at various NC can be discerned (Table 61). While some neighboring subareas are associated to the same clusters at certain NC, they appear to be separated at different NC.

Table 60: Spatial interpretation of fuzzy clusters results for the GEN.OTHO dataset of *Mullus barbatus* considering GFCM subregions at NC = 2:4. Highlighted in green the highest average membership values for each subregion.

GEN.OTHO	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.479	0.521	0.358	0.310	0.331	0.277	0.232	0.229	0.261

Center	0.518	0.482	0.358	0.346	0.296	0.264	0.198	0.262	0.276
Adriatic	0.548	0.452	0.352	0.374	0.274	0.247	0.184	0.284	0.285
East	0.526	0.474	0.289	0.382	0.329	0.224	0.240	0.308	0.227

Table 61: Spatial interpretation of fuzzy clusters results for the GEN.OTHO dataset of *Mullus barbatus* considering GSA subareas at NC = 2:4. Highlighted in green the highest average membership values for each subarea.

GEN.OTHO		2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
1b	West	0.526	0.474	0.384	0.334	0.282	0.296	0.183	0.255	0.266
5a	West	0.465	0.535	0.375	0.307	0.318	0.310	0.207	0.237	0.246
6a	West	0.405	0.595	0.373	0.239	0.389	0.286	0.296	0.171	0.246
6c	West	0.592	0.408	0.321	0.442	0.238	0.220	0.112	0.304	0.364
7a	West	0.502	0.498	0.303	0.350	0.347	0.242	0.258	0.269	0.231
8a	West	0.474	0.526	0.316	0.312	0.372	0.251	0.266	0.248	0.235
8b	West	0.484	0.516	0.319	0.333	0.349	0.287	0.262	0.268	0.183
9a	West	0.420	0.580	0.423	0.232	0.344	0.348	0.251	0.136	0.266
10a	West	0.520	0.480	0.354	0.337	0.309	0.226	0.211	0.243	0.320
10c	West	0.459	0.541	0.375	0.281	0.343	0.268	0.236	0.186	0.309
11b	West	0.539	0.461	0.372	0.389	0.239	0.327	0.130	0.309	0.234
11c	West	0.420	0.580	0.329	0.242	0.430	0.239	0.343	0.163	0.255
12	Centre	0.512	0.488	0.362	0.348	0.291	0.268	0.185	0.274	0.273
14	Centre	0.571	0.429	0.344	0.394	0.262	0.240	0.174	0.307	0.279
17a	Adriatic	0.563	0.437	0.370	0.396	0.234	0.279	0.147	0.290	0.284
17b	Adriatic	0.523	0.477	0.354	0.352	0.293	0.263	0.201	0.266	0.270
18a	Adriatic	0.686	0.314	0.325	0.516	0.159	0.187	0.090	0.419	0.304
18b	Adriatic	0.523	0.477	0.341	0.354	0.305	0.257	0.209	0.272	0.262
18c	Adriatic	0.534	0.466	0.361	0.346	0.294	0.222	0.202	0.255	0.321
19b	Centre	0.501	0.499	0.339	0.323	0.337	0.248	0.253	0.234	0.265
19c	Centre	0.514	0.486	0.403	0.336	0.261	0.303	0.151	0.242	0.304
20a	Centre	0.404	0.596	0.256	0.256	0.488	0.222	0.390	0.184	0.204
23	East	0.431	0.569	0.334	0.251	0.416	0.204	0.335	0.162	0.299
25c	East	0.452	0.548	0.333	0.259	0.408	0.207	0.330	0.171	0.291
26b	East	0.645	0.355	0.223	0.553	0.224	0.202	0.135	0.497	0.166
26c	East	0.578	0.422	0.224	0.475	0.301	0.199	0.210	0.417	0.175
27a	East	0.453	0.547	0.335	0.291	0.374	0.275	0.281	0.213	0.231
27b	East	0.590	0.410	0.278	0.462	0.261	0.228	0.160	0.376	0.236

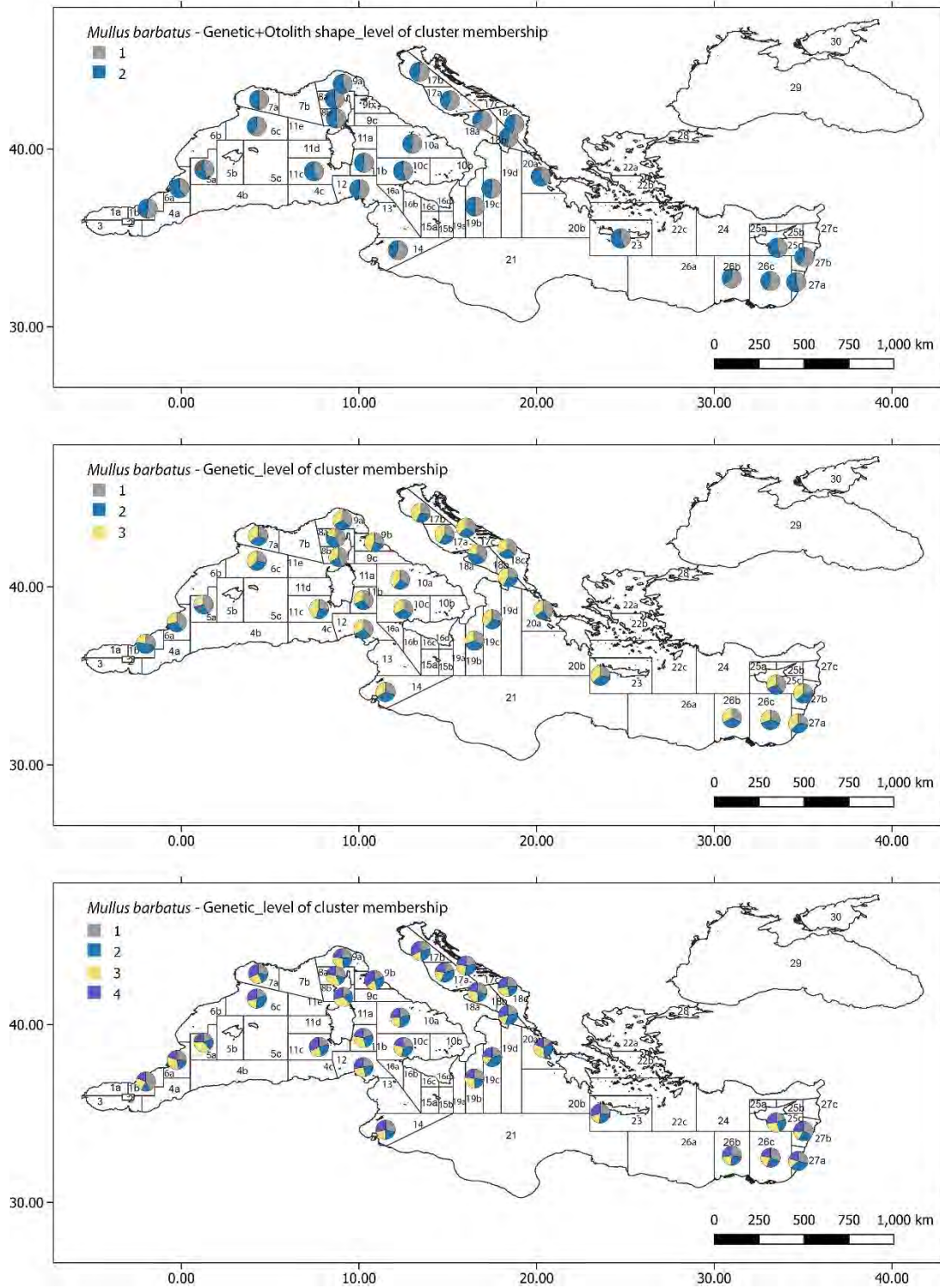


Figure 61: Pie charts with average membership grades per GSA subarea for the GEN.OTHO dataset of *Mullus barbatus* at NC = 2, 3, 4 (top to bottom).

When combining the OHTO and MICRO datasets together and averaging the membership grades per GFCM subregions (Table 62), at NC = 2 all subregions show a high membership grade for the same cluster. At NC = 3 up to NC = 10 the subregions show the same average

membership grade for two or more clusters. Thus, no meaningful information can be inferred from these results.

At the level of GSA subregions at NC = 2 only subarea 11c showed a high average membership for CL1, all other 9 subareas having high average memberships for CL2 (Table 63). From NC = 2 on the same situation as for the GFCM averaging appears, i.e., each GSA subareas showing the same average membership grade for two or more clusters. All in all, the results for the OTHO.MICRO dataset seem those contributing less information to the quest of finding diverging sub-populations in the *Mullus barbatus* Mediterranean populations.

Table 62: Spatial interpretation of fuzzy clusters results for the OTHO.MICRO dataset of *Mullus barbatus* considering GFCM subregions at NC = 2:4. Highlighted in green the highest average membership values for each subregion.

OTHO.MICRO	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.393	0.607	0.426	0.426	0.147	0.301	0.097	0.301	0.301
Center	0.413	0.587	0.432	0.432	0.136	0.302	0.094	0.302	0.302
Adriatic	0.395	0.605	0.427	0.427	0.146	0.300	0.101	0.300	0.300
East	0.401	0.599	0.432	0.432	0.137	0.306	0.081	0.306	0.306

Table 63: Spatial interpretation of fuzzy clusters results for the OTHO.MICRO dataset of *Mullus barbatus* considering GSA subareas at NC = 2:4. Highlighted in green the highest average membership values for each subarea.

OTHO.MICRO		2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
1b	West	0.334	0.666	0.437	0.437	0.127	0.304	0.087	0.304	0.304
6a	West	0.372	0.628	0.424	0.424	0.153	0.306	0.081	0.306	0.306
9a	West	0.354	0.646	0.446	0.446	0.107	0.307	0.078	0.307	0.307
11c	West	0.518	0.482	0.396	0.396	0.207	0.285	0.144	0.285	0.285
12	Center	0.377	0.623	0.431	0.431	0.137	0.299	0.104	0.299	0.299
17a	Adriatic	0.442	0.558	0.410	0.410	0.181	0.291	0.127	0.291	0.291
18b	Adriatic	0.348	0.652	0.444	0.444	0.112	0.308	0.075	0.308	0.308
19b	Center	0.449	0.551	0.433	0.433	0.135	0.305	0.084	0.305	0.305
22c	East	0.394	0.606	0.424	0.424	0.152	0.296	0.112	0.296	0.296
27a	East	0.408	0.592	0.440	0.440	0.120	0.317	0.049	0.317	0.317

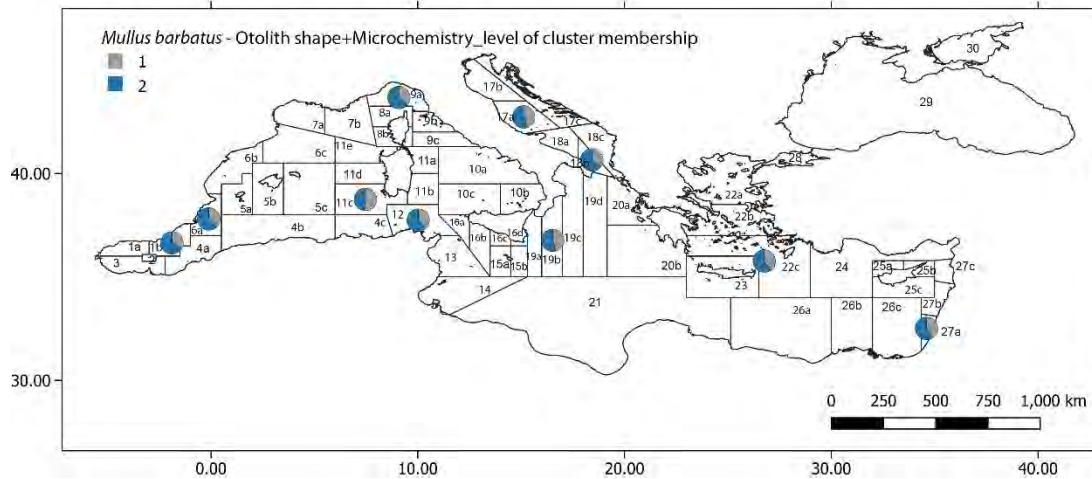


Figure 62: Pie charts with average membership grades per GSA subarea for the OTHO.MICRO dataset of *Mullus barbatus* at NC = 2.

When combining all three datasets together (ALL dataset) again the averaging per GFCM subregion is not much helpful (Table 64). The subregions show the same preference at NC = 2, then the Eastern subregion splits from the others. Nevertheless, at increasing NC subregions shift association among them and with clusters, without a clear spatial pattern to emerge.

At the level of GSA subareas there seem to be no differentiation among most of the explored subareas (Table 65). In fact, at NC = 2 all subareas show the highest averaged membership for the same cluster, and only one subarea (27a) splits from the others at NC = 3 and NC = 4. At NC = 5 subareas 1b, 6a join 27a, but also with higher NC the clusters to which the subareas are clearly associated are only a handful of those computed. This might be either the result of the existence of only one red muller population in the Mediterranean, or on the other hand, of the existence of a higher number of sub-populations that cannot be clearly identified based on data from only 9 subareas of the Mediterranean.

Table 64: Spatial interpretation of fuzzy clusters results for the ALL dataset of *Mullus barbatus* considering GFCM subregions at NC = 2:4. Highlighted in green the highest average membership values for each subregion.

ALL	2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
West	0.736	0.264	0.487	0.366	0.148	0.408	0.348	0.152	0.092
Center	0.714	0.286	0.541	0.274	0.1858	0.507	0.213	0.135	0.145
Adriatic	0.725	0.275	0.540	0.296	0.164	0.468	0.271	0.188	0.073
East	0.800	0.200	0.245	0.705	0.050	0.243	0.624	0.129	0.004

Table 65: Spatial interpretation of fuzzy clusters results for the ALL dataset of *Mullus barbatus* considering GSA subareas at NC = 2:4. Highlighted in green the highest average membership values for each subarea.

ALL		2 CLUSTERS		3 CLUSTERS			4 CLUSTERS			
AREA	GFCM	FKM1	FKM2	FKM1	FKM2	FKM3	FKM1	FKM2	FKM3	FKM4
1b	West	0.815	0.185	0.529	0.387	0.084	0.471	0.352	0.150	0.027
6a	West	0.716	0.284	0.457	0.373	0.170	0.361	0.357	0.183	0.099

9a	West	0.780	0.220	0.469	0.361	0.170	0.468	0.347	0.018	0.168
11c	West	0.621	0.379	0.493	0.317	0.191	0.363	0.324	0.180	0.133
12	Center	0.667	0.333	0.379	0.365	0.255	0.380	0.264	0.185	0.170
17a	Adriatic	0.637	0.363	0.446	0.325	0.230	0.391	0.311	0.201	0.098
18b	Adriatic	0.795	0.205	0.615	0.273	0.111	0.529	0.240	0.178	0.053
19b	Center	0.750	0.250	0.662	0.205	0.133	0.603	0.174	0.096	0.127
22c	East	0.800	0.200	0.245	0.705	0.050	0.243	0.624	0.129	0.004
27a	East	0.815	0.185	0.529	0.387	0.084	0.471	0.352	0.150	0.027

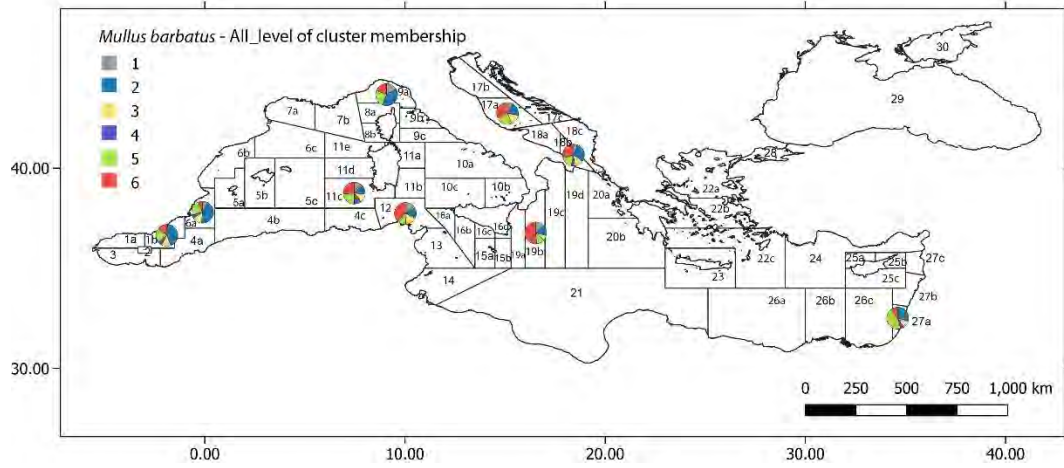
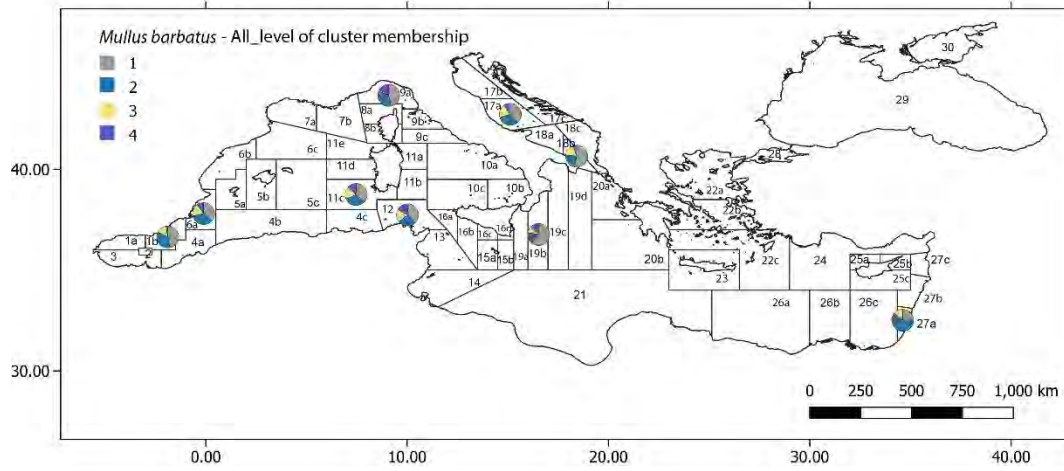
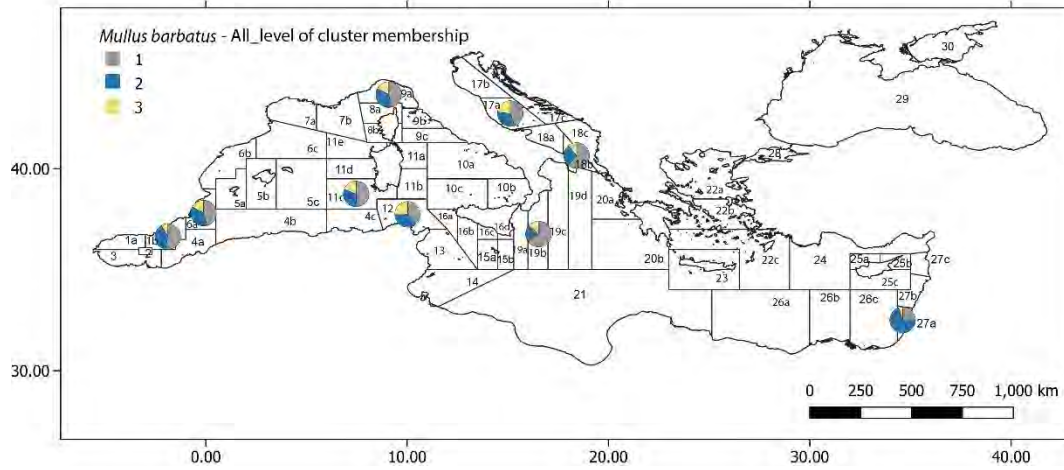


Figure 63: Pie charts with average membership grades per GSA subarea for the ALL dataset of *Mullus barbatus* at NC = 3, 4, 6 (top to bottom).

Spatial analysis

The spatial analysis was performed using all the potential spatial predictors, i.e., depth, longitude and latitude, significant MEMs, to build RDA models explaining the variance in *Mullus barbatus* genetic, otolith shape and otolith microchemistry datasets, as well as in their mutual combinations (Table 66).

The explanatory power of depth was almost always either non-significant or negligible. Only for the MICRO dataset for clusterizations from NC = 4 to 10 and for the OTHO.MICRO at NC = 3 did depth have a significant effect. Thus, we can conclude that depth does not play a role in determining possible *Mullus barbatus* stocks in the Mediterranean.

Contrary to hake, also the longitude and latitude were seldom significant and explaining a relevant fraction of the variance of the analyzed datasets. For the OTHO and GEN.OTHO datasets longitude and latitude were significant but explained low fractions of variance. The same was true for higher number of clusters in the MICRO dataset, while the highest contribution in terms of explained variance was for higher numbers of clusterizations of the ALL dataset. These results show that the stocks of *Mullus barbatus* in the Mediterranean are not defined on a large scale that might be captured by a linear gradient based on the longitude and latitude. This is in contrast with the hake, and it can be explained with the different lifestyle and habits of the red mullet, a more sedentary and less gregarious species.

Table 66: Spatial analysis results for *Mullus barbatus*. For each dataset the explanatory power of each spatial variable or group of variables is reported as adjusted R². All reported models with depth and LongLat were tested for significance with an ANOVA and “ns” indicates those that were not significant. The significance of MEMs was tested with a forward selection procedure and a double-stopping criterion. Only significant MEMs are reported. For MEMs also the number of significant MEMs, their names and the winner spatial weighting matrix are reported.

GEN	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	ns	0.00	3	MEM7, MEM3, MEM14	Gabriel_Down_3
2 CLUSTERS	ns	ns	0.00	1	MEM17	Delaunay_Up_0.9
3 CLUSTERS	0.00	ns	0.02	3	MEM19, MEM7, MEM12	Delaunay_Down_5
4 CLUSTERS	ns	ns	0.02	6	MEM10, MEM15, MEM9, MEM20, MEM17, MEM6	Delaunay_Up_0.2
5 CLUSTERS	0.00	0.01	0.01	4	MEM10, MEM17, MEM3, MEM15	Delaunay_Up_0.4
6 CLUSTERS	ns	ns	0.01	3	MEM10, MEM17, MEM7	Delaunay_Up_0.3
7 CLUSTERS	ns	0.00	0.01	3	MEM15, MEM5, MEM13	Relative_Up_0.9
8 CLUSTERS	ns	0.00	0.01	3	MEM22, MEM19, MEM3	MST_Down_10

9 CLUSTERS	ns	0.00	0.01	4	MEM10, MEM3, MEM11, MEM17	Delaunay_Up_0.8
10 CLUSTERS	ns	0.00	0.01	3	MEM8, MEM10, MEM24	Delaunay_Up_0.3
OTHO	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	ns	0.01	0.04	18	MEM35, MEM27, MEM14, MEM26, MEM18, MEM22, MEM20, MEM5, MEM21, MEM19, MEM24, MEM23, MEM11, MEM32, MEM34, MEM1, MEM3, MEM2	MST_Up_0.3
2 CLUSTERS	ns	0.01	0.05	5	MEM37, MEM26, MEM20, MEM24, MEM25	Relative_Up_0.3
3 CLUSTERS	ns	0.00	0.04	12	MEM37, MEM25, MEM26, MEM18, MEM22, MEM21, MEM35, MEM4, MEM2, MEM3, MEM13, MEM29	Relative_Up_0.2
4 CLUSTERS	ns	0.00	0.03	13	MEM37, MEM25, MEM35, MEM26, MEM18, MEM22, MEM21, MEM36, MEM13, MEM32, MEM4, MEM30, MEM14	Relative_Up_0.2
5 CLUSTERS	ns	0.01	0.04	19	MEM37, MEM35, MEM17, MEM20, MEM31, MEM23, MEM15, MEM9, MEM30, MEM4, MEM3, MEM2, MEM10, MEM12, MEM36, MEM18, MEM24, MEM32, MEM34	Relative_Up_0.1
6 CLUSTERS	ns	0.00	0.04	18	MEM35, MEM26, MEM22, MEM32, MEM20, MEM30, MEM21, MEM5, MEM11, MEM33, MEM24, MEM1, MEM14, MEM19, MEM13, MEM28, MEM18, MEM27	MST_Up_0.3
7 CLUSTERS	ns	0.00	0.03	18	MEM37, MEM25, MEM18, MEM29, MEM35, MEM31, MEM13, MEM36, MEM15, MEM16, MEM14, MEM26, MEM22, MEM10, MEM9, MEM8, MEM5, MEM34	Relative_Up_0.2

8 CLUSTERS	ns	0.00	0.03	15	MEM35, MEM22, MEM26, MEM32, MEM20, MEM30, MEM5, MEM21, MEM27, MEM13, MEM33, MEM11, MEM19, MEM14, MEM9	MST_Up_0.3
9 CLUSTERS	ns	0.00	0.02	16	MEM35, MEM26, MEM23, MEM25, MEM6, MEM28, MEM32, MEM27, MEM10, MEM20, MEM19, MEM22, MEM16, MEM21, MEM2, MEM30	MST_Up_0.5
10 CLUSTERS	ns	0.00	0.02	15	MEM37, MEM25, MEM34, MEM27, MEM33, MEM23, MEM6, MEM24, MEM14, MEM30, MEM36, MEM10, MEM2, MEM3, MEM4	Relative_Up_0.9
MICRO	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	0.01	0.09	10	MEM5, MEM9, MEM6, MEM2, MEM10, MEM1, MEM11, MEM12, MEM4, MEM7	Relative_Up_1
2 CLUSTERS	ns	ns	ns			
3 CLUSTERS	ns	ns	0.07	5	MEM1, MEM6, MEM8, MEM11, MEM2	Relative_Up_1
4 CLUSTERS	0.02	ns	0.10	9	MEM1, MEM2, MEM5, MEM11, MEM12, MEM7, MEM10, MEM6, MEM8	Relative_Up_1
5 CLUSTERS	0.01	ns	0.08	8	MEM12, MEM6, MEM4, MEM7, MEM11, MEM8, MEM2, MEM10	Gabriel_Down_5
6 CLUSTERS	0.02	ns	0.09	8	MEM1, MEM12, MEM4, MEM6, MEM7, MEM11, MEM8, MEM10	Gabriel_Down_4
7 CLUSTERS	0.02	ns	0.09	8	MEM1, MEM12, MEM4, MEM6, MEM8, MEM7, MEM5, MEM11	Gabriel_Down_3
8 CLUSTERS	0.02	0.01	0.10	8	MEM1, MEM7, MEM6, MEM11, MEM5, MEM12, MEM4, MEM8	Gabriel_Down_2

9 CLUSTERS	0.02	0.01	0.09	9	MEM1, MEM7, MEM6, MEM8, MEM11, MEM5, MEM4, MEM12, MEM3	Gabriel_Linear
10 CLUSTERS	0.02	0.01	0.08	9	MEM1, MEM6, MEM7, MEM5, MEM11, MEM10, MEM4, MEM12, MEM8	Gabriel_Down_2
GEN.OTHO	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	0.00	0.00	0.01	7	MEM13, MEM25, MEM6, MEM29, MEM12, MEM21, MEM11	Relative_Down_10
2 CLUSTERS	ns	0.01	0.04	4	MEM19, MEM15, MEM23, MEM24	Delaunay_Up_0.7
3 CLUSTERS	ns	0.02	0.05	7	MEM23, MEM29, MEM12, MEM26, MEM14, MEM22, MEM10	Relative_Up_0.2
4 CLUSTERS	0.00	0.01	0.04	7	MEM23, MEM29, MEM22, MEM14, MEM12, MEM26, MEM27	Relative_Up_0.2
5 CLUSTERS	ns	0.01	0.03	8	MEM15, MEM12, MEM22, MEM14, MEM29, MEM26, MEM27, MEM23	Relative_Up_0.2
6 CLUSTERS	ns	0.01	0.03	8	MEM13, MEM25, MEM11, MEM29, MEM12, MEM3, MEM6, MEM24	Relative_Down_9
7 CLUSTERS	ns	0.01	0.02	8	MEM23, MEM24, MEM21, MEM18, MEM16, MEM1, MEM22, MEM28	MST_Up_0.8
8 CLUSTERS	ns	0.00	0.02	8	MEM25, MEM21, MEM23, MEM16, MEM29, MEM18, MEM1, MEM17	Relative_Up_1
9 CLUSTERS	ns	0.00	0.01	7	MEM14, MEM12, MEM29, MEM27, MEM9, MEM23, MEM11	Relative_Up_0.1
10 CLUSTERS	ns	0.00	0.01	6	MEM13, MEM11, MEM29, MEM24, MEM25, MEM12	Relative_Down_9
OTHO.MICRO	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX

DATA	0.01	0.01	0.08	11	MEM5, MEM10, MEM9, MEM6, MEM2, MEM11, MEM1, MEM12, MEM4, MEM7, MEM8	Relative_Up_1
2 CLUSTERS	0.03	ns	0.07	4	MEM2, MEM11, MEM8, MEM1	Gabriel_Up_0.2
3 CLUSTERS	ns	ns	ns			
4 CLUSTERS	ns	ns	ns			
5 CLUSTERS	ns	ns	ns			
6 CLUSTERS	ns	ns	ns			
7 CLUSTERS	ns	ns	ns			
8 CLUSTERS	ns	ns	ns			
9 CLUSTERS	ns	ns	ns			
10 CLUSTERS	ns	ns	ns			
ALL	VS Depth	VS LongLat	VS MEM	NMEM	MEM	SPATIAL MATRIX
DATA	ns	0.01	0.05	7	MEM9, MEM8, MEM4, MEM6, MEM7, MEM5, MEM2	Relative_Down_2
2 CLUSTERS	ns	ns	0.04	1	MEM6	Relative_Up_0.2
3 CLUSTERS	ns	ns	0.15	5	MEM9, MEM3, MEM8, MEM7, MEM6	Relative_Down_9
4 CLUSTERS	ns	ns	0.12	6	MEM9, MEM7, MEM5, MEM4, MEM2, MEM8	Relative_Down_2
5 CLUSTERS	ns	ns	0.09	6	MEM9, MEM3, MEM7, MEM6, MEM4, MEM8	Relative_Down_10
6 CLUSTERS	ns	0.02	0.09	6	MEM1, MEM9, MEM5, MEM2, MEM7, MEM4	Relative_Down_3
7 CLUSTERS	ns	0.02	0.08	6	MEM9, MEM2, MEM4, MEM5, MEM3, MEM7	Relative_Linear
8 CLUSTERS	ns	0.03	0.12	6	MEM9, MEM6, MEM2, MEM4, MEM5, MEM8	Relative_Down_2

9 CLUSTERS	ns	0.02	0.11	6	MEM9, MEM6, MEM2, MEM4, MEM5, MEM8	Relative_Down_2
10 CLUSTERS	ns	0.02	0.09	6	MEM9, MEM6, MEM2, MEM4, MEM8, MEM5	Relative_Down_2

The significant MEMs also did not always explain much of the variance of each dataset. For the GEN dataset MEMs explained a maximum adjusted R^2 of 0.02 for NC = 3 (with 3 significant MEMs) and NC = 4 (with 6 significant MEMs). For the OTHO dataset also the contribution of selected MEMs was always significant but did not exceed an adjusted R^2 of 0.05 with 5 up to 19 significant MEMs, mostly medium to small-scale (Fig. 63). In fact, the best result was obtained for 2 clusters with the lower number of predictors. For the MICRO dataset the significant MEMs explained up to an adjusted R^2 of 0.10 at NC = 4 and NC = 8. Furthermore, the selected MEMs were similar for all NC (e.g., MEM6, MEM12) as were the selected spatial weighting matrixes (mostly the relative neighborhood and Gabriel's connectivity scheme matrix). Nevertheless, it should be reminded that the spatial coverage of the MICRO dataset is very low (25 sampling locations) compared to the GEN and OTHO datasets. For the GEN.OTHO dataset the selected MEMs were always significant, explaining up to an adjusted R^2 of 0.05 at NC = 3 with 9 mostly medium scale MEMs (Figs. 64-65). The explanatory power of MEMs for GEN.OTHO clusters decreased with increasing NC. For OTHO-MICRO only NC = 2 had a significant explanatory power of four selected MEMs with an adjusted $R^2 = 0.07$. No other clusterization of this dataset could be explained by any combination of spatial components. Finally, for the combination of all three datasets, selected MEMs could explain relevant adjusted R^2 at NC = 3 (0.15), 4 (0.12), 8 (0.12) and 9 (0.11) (Fig. 66). Again, it should be reminded that also this dataset had very low spatial coverage with only 20 sampling locations in 10 GSA subareas.

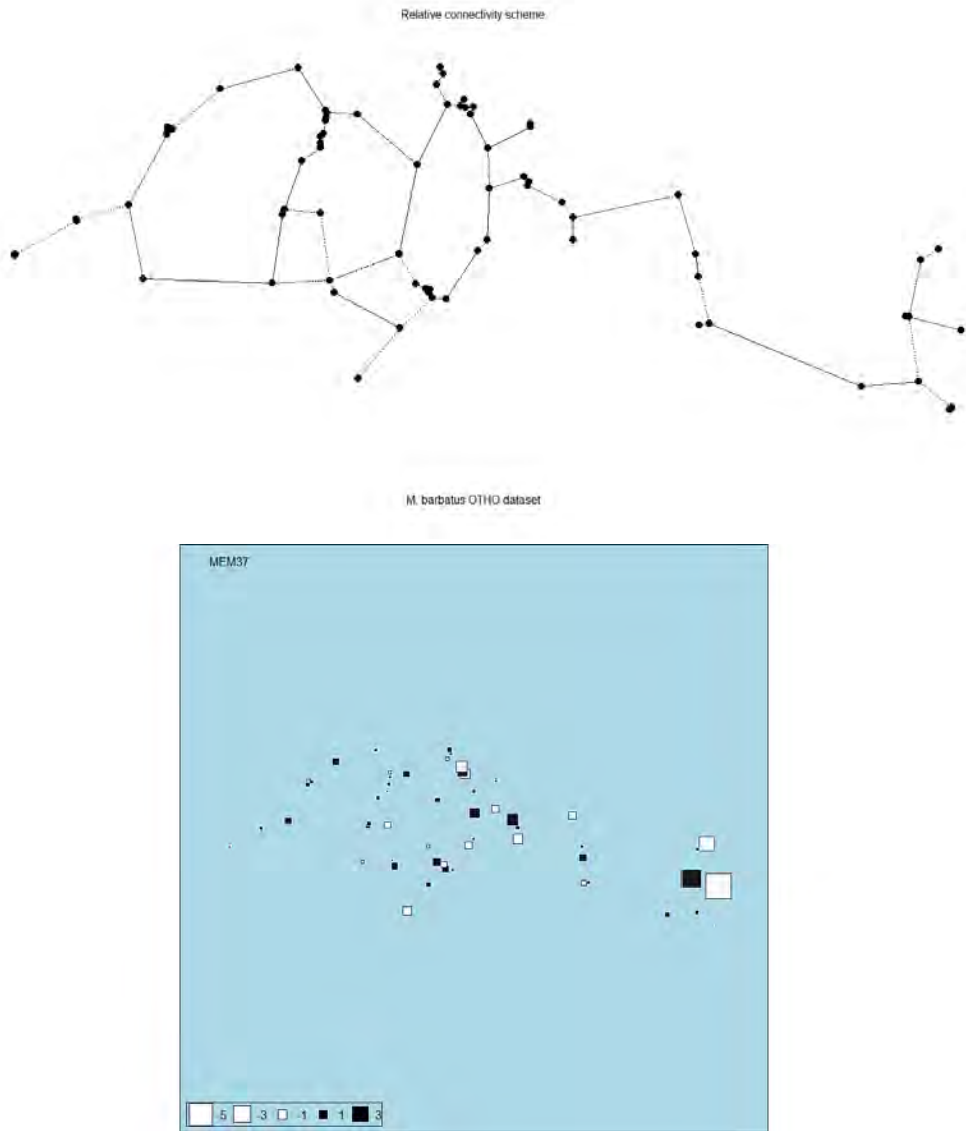
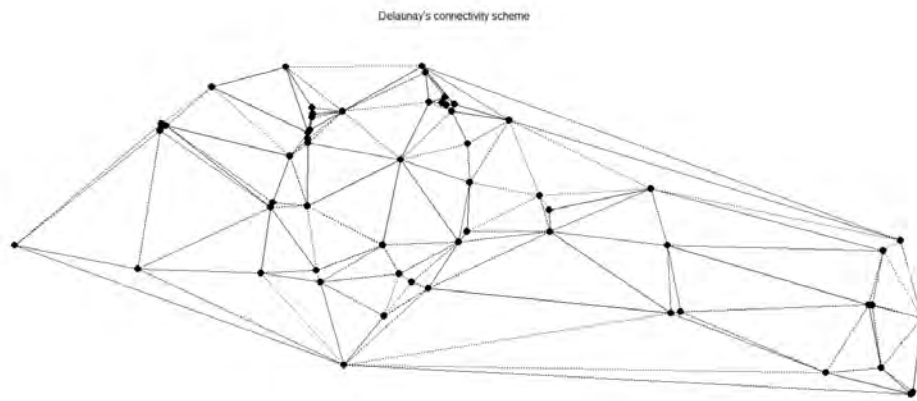


Figure 63: The relative neighborhood connectivity scheme for the OTHO dataset of *Mullus barbatus* (upper panel) and the MEM37 built on top of it (lower panel).



M barbatus GEN.OTHO dataset

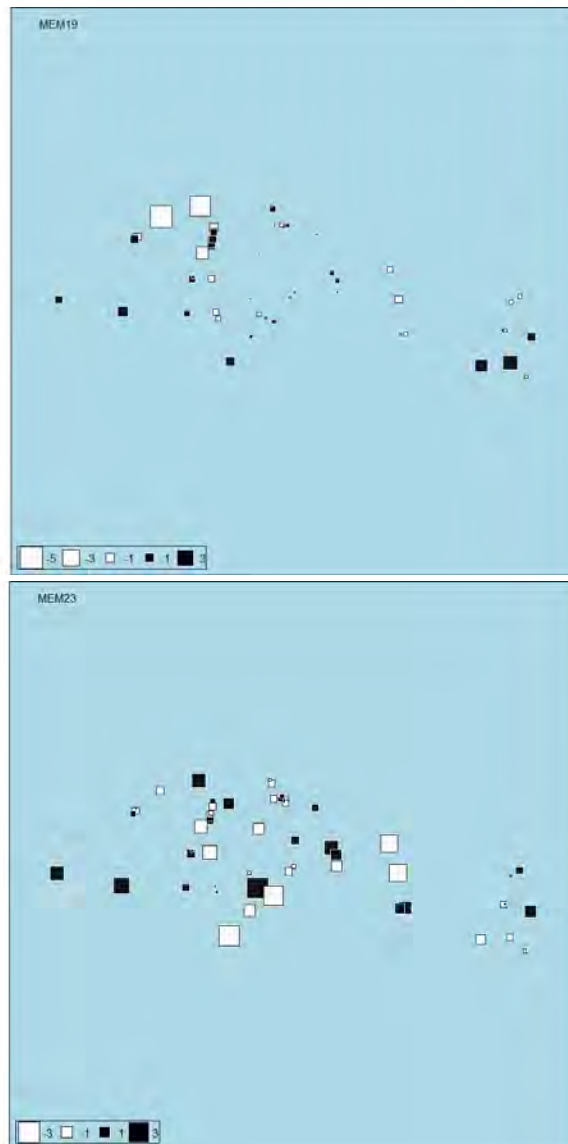


Figure 64: The Delaunay's connectivity scheme for the GEN.OTHO dataset of *Mullus barbatus* and the MEM19 and MEM23 built on top of it (top to bottom).

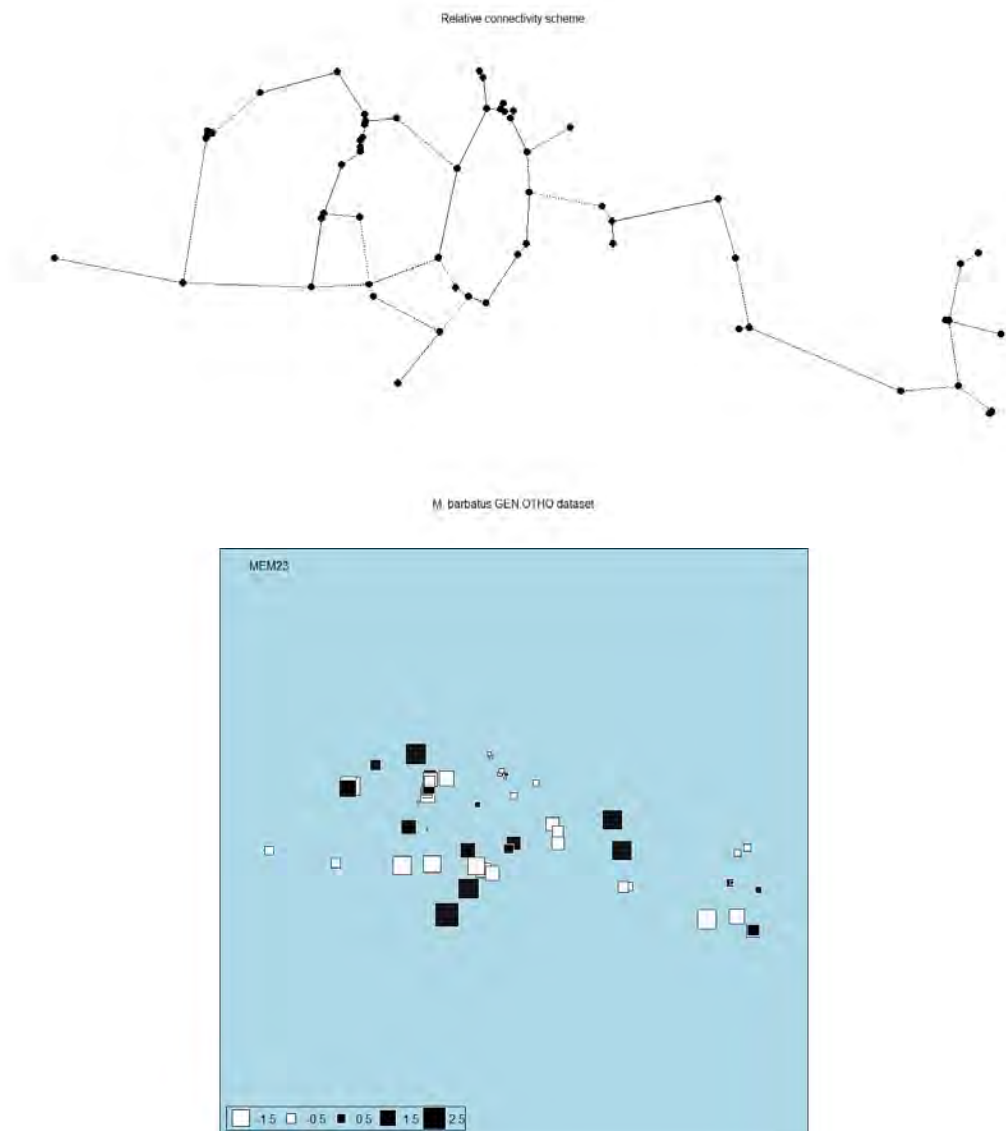
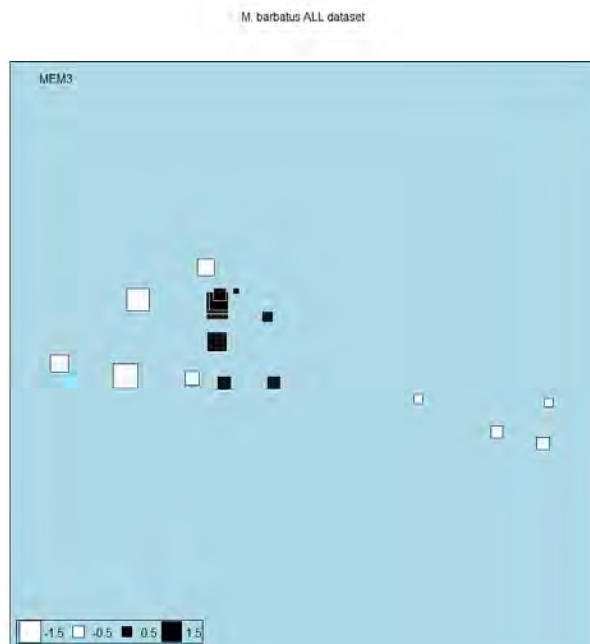
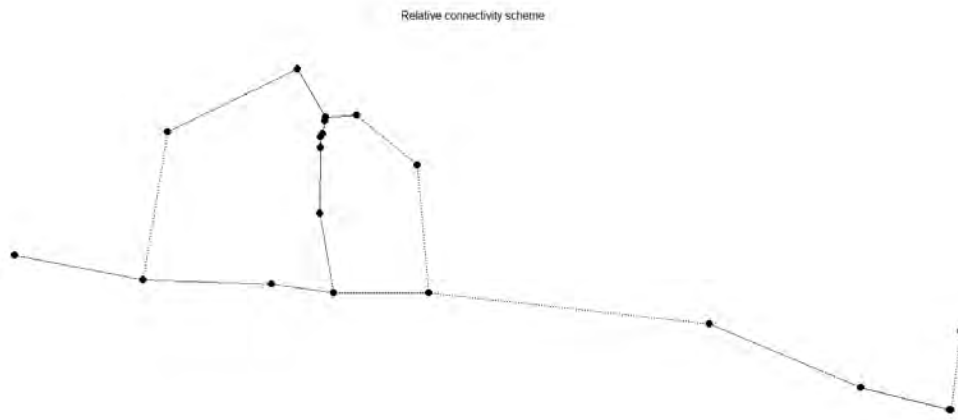


Figure 65: The relative neighborhood connectivity scheme for the GEN.OTHO dataset of *Mullus barbatus* (upper panel) and the MEM23 built on top of it (lower panel).

The general conclusion of the spatial analysis is that there are no large-scale spatial structures in the *Mullus barbatus* population, and that depth does not play a role in a possible diversification of it in different stocks. The MEMs explain low fractions of variance in the most abundant datasets (GEN, OTHO, GEN.OTHO), mostly with medium and small-scale components, while only for the less abundant datasets (MICRO, ALL) the amount of explained variance is higher, never exceeding an adjusted R^2 of 0.15. Nevertheless, that the results on different datasets diverge also in the indication as to which are the more spatially structured clusterizations cast serious doubts to the possibility of using these data for the identification of stocks. In fact, based on spatial analysis only, the GEN dataset would suggest 3-4 different stocks, the OTHO dataset 2-3 or 5-6, the GEN.OTHO from 2 to 5. For the less abundant datasets, the MICRO would suggest 4 clusters or 6 to 9 clusters, the OTHO.MICRO 2 clusters and the ALL dataset 3-4 or 8-9 clusters. The usefulness of exploring 8

or 9 clusters on only 20-25 sampling location remains doubtful, since it would mean an average of 2-3 sampling locations per cluster.



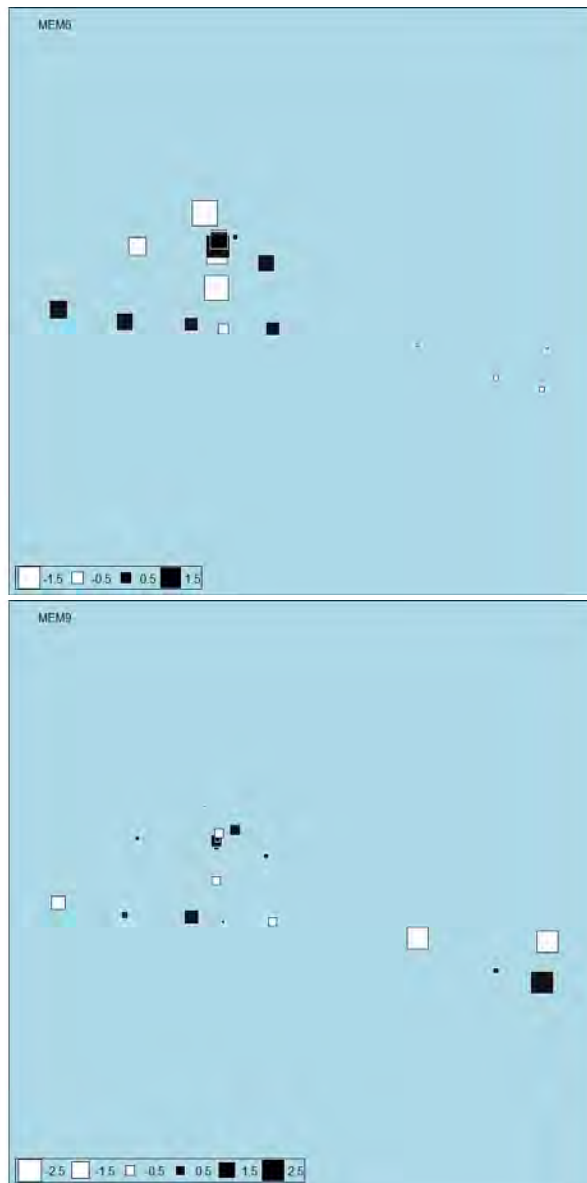


Figure 66: The relative neighborhood connectivity scheme for the ALL dataset of *Mullus barbatus* and the MEM3, MEM6 and MEM9 built on top of it (top to bottom).

With the significant spatial components, we built the complete spatial models for the most interesting cases: the OTHO dataset at NC = 2-6 clusters, the GEN.OTHO dataset at NC = 2-3, and the ALL dataset at NC = 3. The complete spatial models were subjected to a forward selection procedure with double-stopping criterion in order to select the best subset of explanatory variables.

The forward selection for the OTHO clusters at NC = 2 selected 5 MEMs as the best predictors, explaining together an adjusted R^2 of 0.05 (Table 67). For NC = 3, 5 and 6 the number of selected spatial components was 11, 18 and 17 variables, respectively, explaining an adjusted R^2 of 0.04, 0.04 and 0.03. Only at NC = 5 one of these variables was not a MEM, but the longitude. From a pure statistical point of view, it is doubtful the utility of building models which explain so low amount of variance with so high numbers of predictors. Only the model at NC = 2 would seem justifiable from this point of view. The selected MEMs were

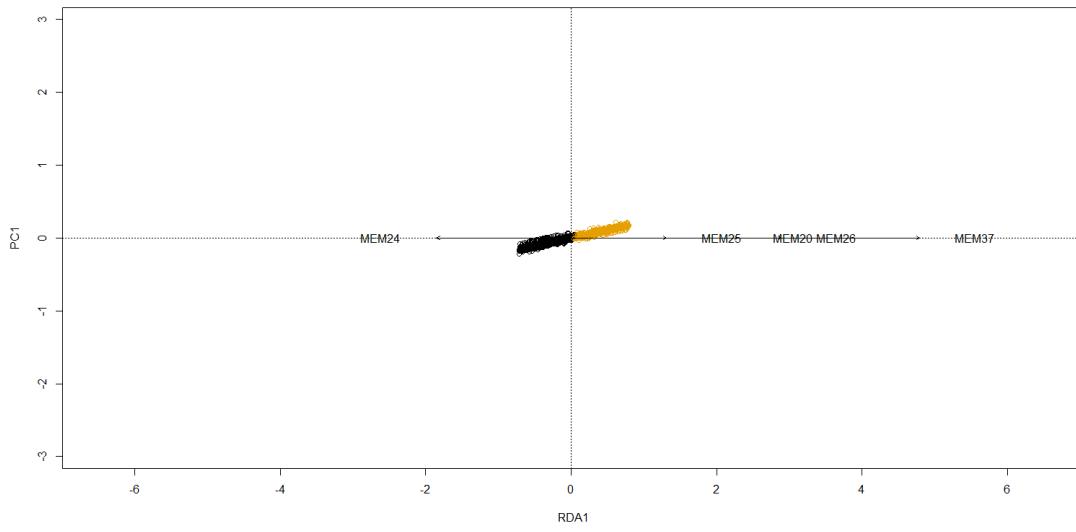
mostly medium and small-scale components at NC = 2, but there were also big-scale components (MEM1, MEM2, MEM3, MEM4, MEM5) for the models of the other three clusterizations (Fig. 67).

Table 67: Forward selection on spatial components for M. barbatus otolith shape PCAs at NC = 2, 3, 5, 6. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

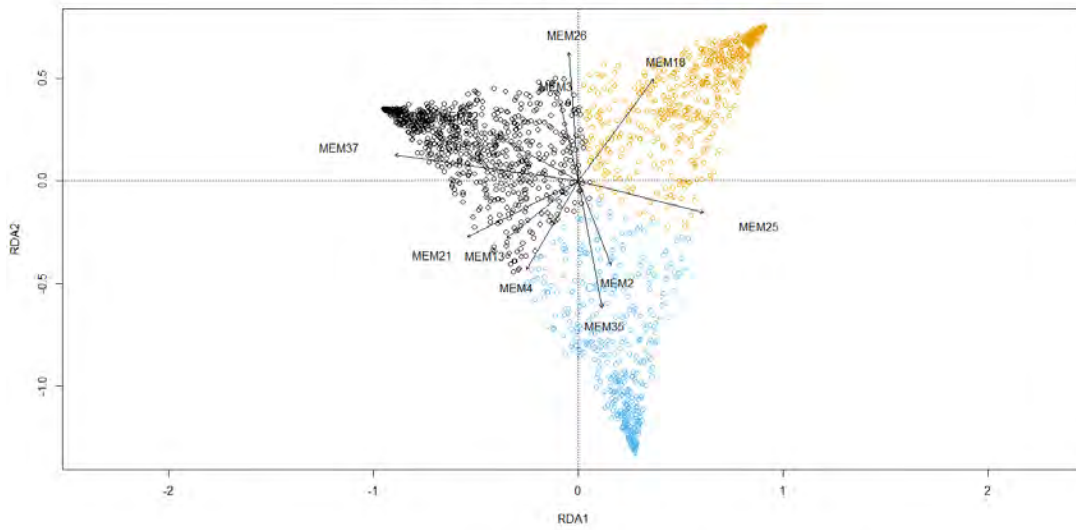
MUT OTHO 2 CLUSTERS			MUT OTHO 3 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
MEM37	0.03	0.001	MEM37	0.01	0.001
MEM26	0.04	0.001	MEM25	0.02	0.001
MEM20	0.04	0.002	MEM26	0.02	0.001
MEM24	0.04	0.004	MEM18	0.02	0.001
MEM25	0.05	0.046	MEM22	0.03	0.005
			MEM21	0.03	0.002
			MEM35	0.03	0.006
			MEM4	0.03	0.012
			MEM2	0.03	0.023
			MEM3	0.04	0.03
			MEM13	0.04	0.042
MUT OTHO 5 CLUSTERS			MUT OTHO 6 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
MEM37	0.01	0.001	MEM35	0.01	0.001
MEM35	0.01	0.001	MEM26	0.02	0.001
MEM17	0.01	0.001	MEM22	0.02	0.001
MEM20	0.02	0.001	MEM32	0.02	0.002
MEM31	0.02	0.001	MEM20	0.02	0.001
MEM23	0.02	0.002	MEM30	0.02	0.001
MEM15	0.02	0.001	MEM21	0.03	0.005
MEM9	0.03	0.002	MEM5	0.03	0.009

MEM30	0.03	0.006	MEM11	0.03	0.012
MEM4	0.03	0.004	MEM33	0.03	0.026
MEM3	0.03	0.011	MEM24	0.03	0.025
longUTM	0.03	0.015	MEM1	0.03	0.04
MEM10	0.03	0.004	MEM14	0.03	0.047
MEM24	0.03	0.013	MEM19	0.03	0.032
MEM18	0.03	0.033	MEM13	0.03	0.032
MEM36	0.04	0.036	MEM28	0.03	0.047
MEM32	0.04	0.045	MEM18	0.03	0.043
MEM12	0.04	0.034			

SELECTED SPATIAL RDA M. BARBATUS OTHO 2 CLUSTERS



SELECTED SPATIAL RDA M. BARBATUS OTHO 3 CLUSTERS



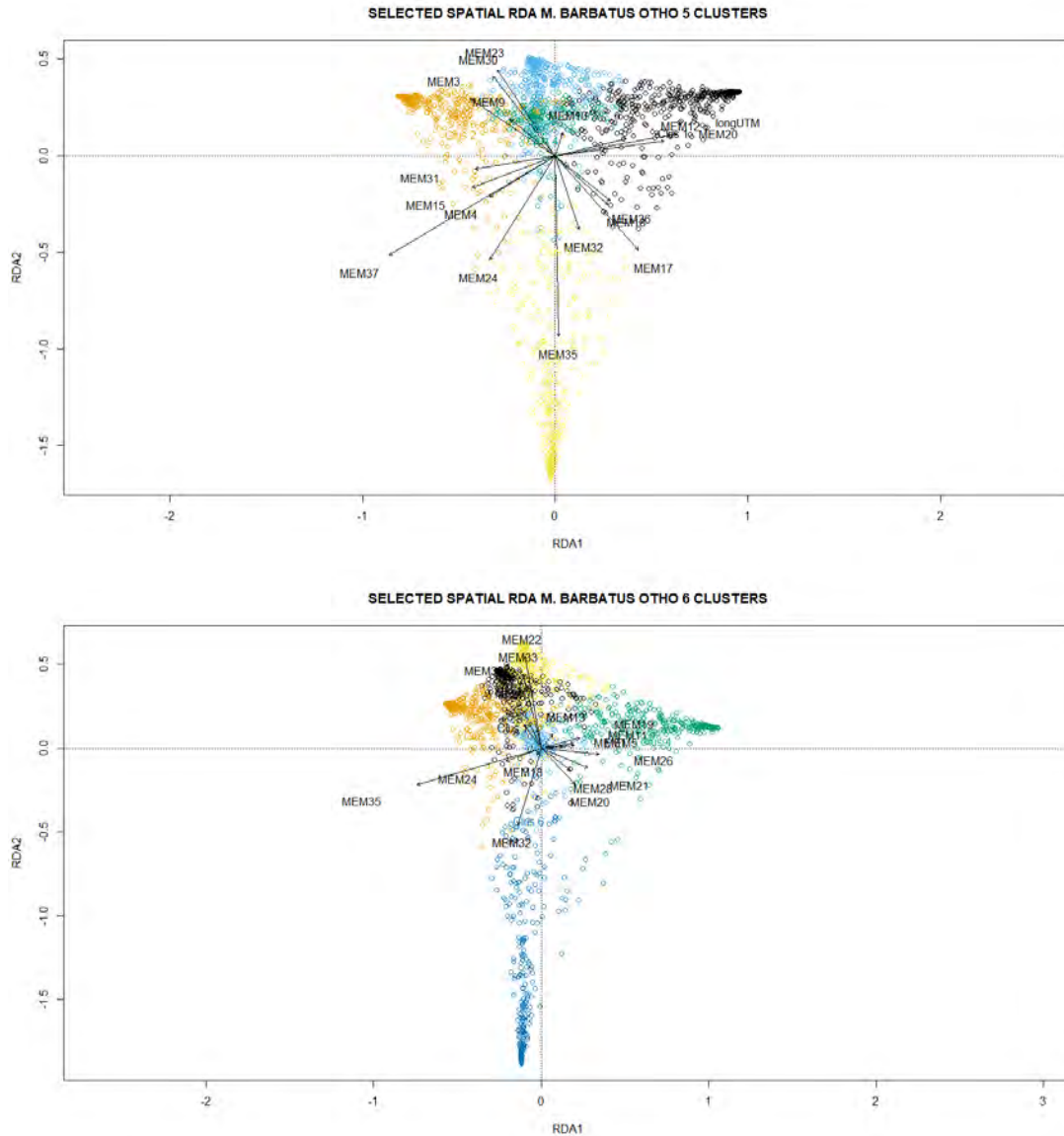


Figure 67: Best spatial parsimonious RDA models for *M. barbatus* OTHO dataset at NC = 2, 3, 5, 6 (from top to bottom). Explained variance on the first and only RDA axis was 0.05 for NC = 2; 0.02 and 0.02 on the first and second RDA axis at NC = 3; 0.02 and 0.01 on the first and second RDA axis at NC = 5; 0.02 and 0.01 on the first and second RDA axis at NC = 6.

For the GEN.OTHO dataset at NC = 2 only three MEMs were selected as significant (Table 68), one (MEM15) related to cluster 1, the other two (MEM23, MEM19) to cluster 2 (Fig. 68). At NC = 3 six medium and small-scale MEMs were selected as significant, showing different associations with the three clusters (Fig. 68).

Table 68: Forward selection on spatial components for *M. barbatus* GEN.OTHO dataset at NC = 2 and NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT GEN.OTHO 2 CLUSTERS			MUT 3 GEN.OTHO CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
MEM19	0.01	0.007	MEM23	0.01	0.001

MEM15	0.02	0.005	MEM29	0.02	0.001
MEM23	0.03	0.024	MEM12	0.03	0.006
			MEM26	0.04	0.007
			MEM14	0.04	0.012
			MEM22	0.05	0.024

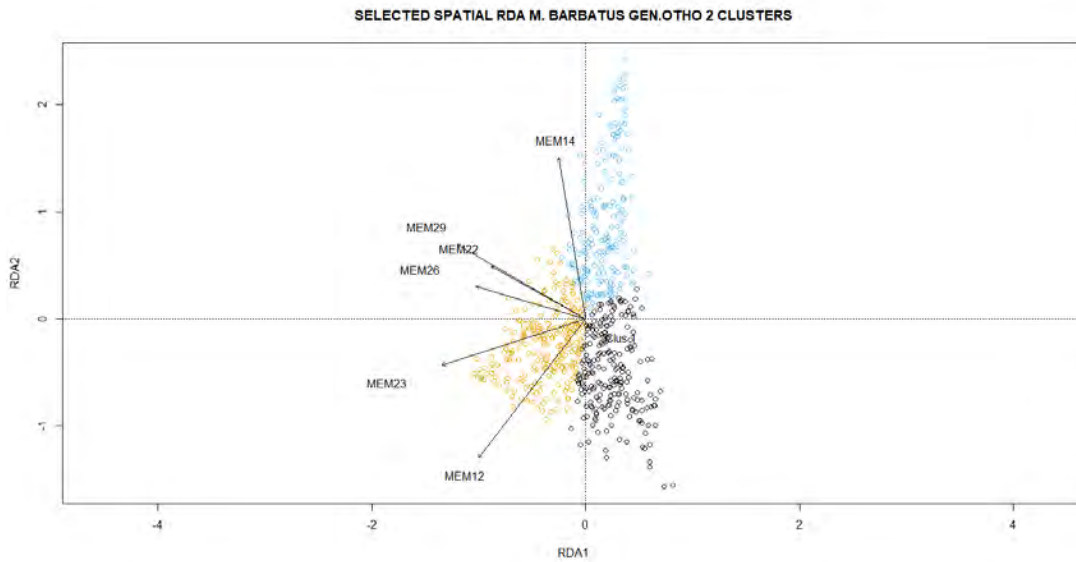
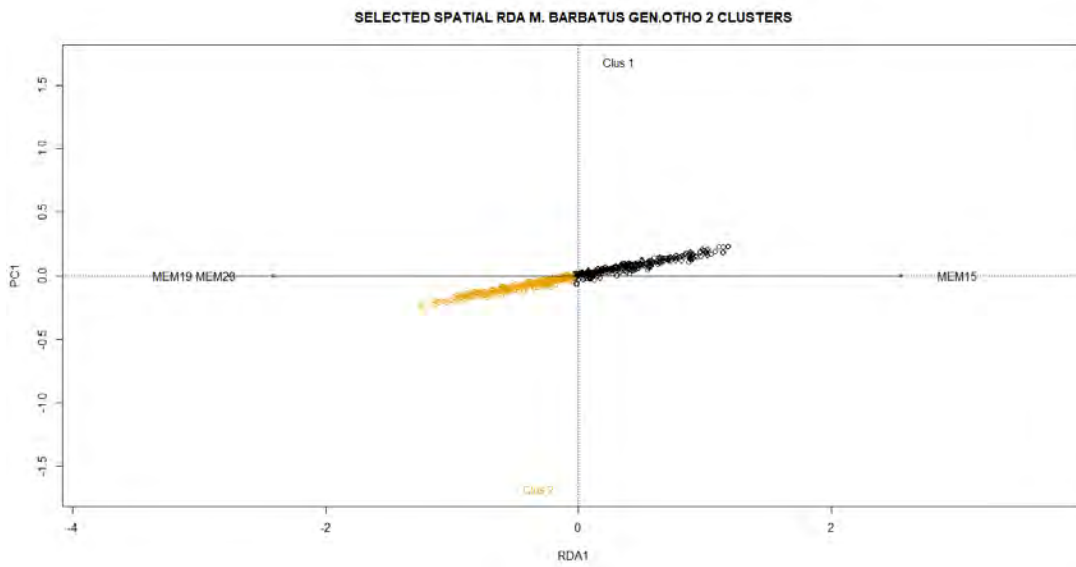


Figure 68: Best spatial parsimonious RDA models for *M. barbatus* GEN.OTHO dataset at NC = 2 (upper panel) and NC = 3 (lower panel). Explained variance on the first and only RDA axis was 0.03 for NC = 2; 0.04 and 0.01 on the first and second RDA axis at NC = 3.

The forward selection on the ALL dataset at NC = 3 did select 4 big and medium-scale MEMs (MEM3, MEM7, MEM8, MEM9) as the most significant predictors. Positive values of MEM8 and MEM9 (Fig. 69) are associated to high membership grades of cluster 1 and low membership grades of cluster 2. The second axis separates cluster 3 membership grades from the other two clusters and is mostly defined by MEM7 and MEM3.

Table 69: Forward selection on spatial components for *M. barbatus* ALL dataset at NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT ALL 3 CLUSTERS		
variables	AdjR2Cum	p-value
MEM9	0.08	0.001
MEM3	0.10	0.022
MEM8	0.12	0.032
MEM7	0.13	0.039

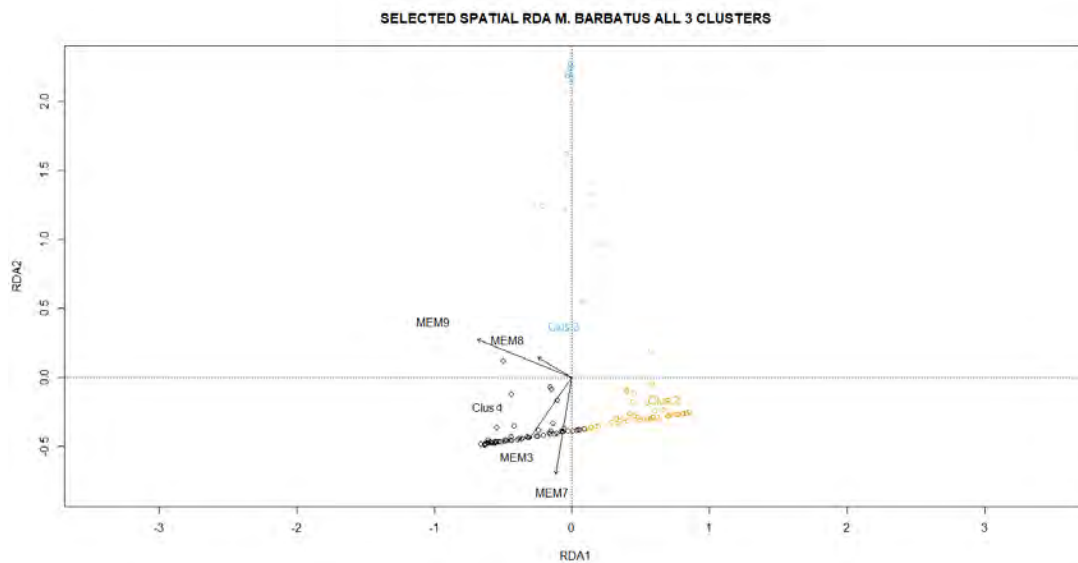


Figure 69: Best spatial parsimonious RDA model for *M. barbatus* ALL dataset at NC = 3. Explained variance was 0.13 and 0.03 on the first and second RDA axis.

Environmental analysis

The environmental analysis was performed by computing separate RDA using the bottom and euphotic zone medians of the environmental variables over three different time intervals (1 year, 3 years, 5 years), to explain each dataset of the original *Mullus barbatus* data (GEN, OTHO, MICRO), including their combinations (Table 70).

Since the depth range of this species is 20-300 m and the euphotic zone was considered as 0-200 m, there was no big difference in the explanatory power of the bottom and the

euphotic layer variables. Furthermore, in many cases, the same patterns could be observed as in the spatial analysis: e.g., for the OTHO.MICRO but for NC = 2 no significant result was obtained in either case; for the MICRO dataset NC = 4 and NC = 8 are the two best explained clusterizations both with spatial and with environmental variables; etc. This is because also the environmental variables in the Mediterranean are spatially structured and thus their contribution to the explanation of the fuzzy clusters is overlapping with those of the spatial components. All these aspects will be thoroughly analyzed with the variation partitioning procedures below.

For the GEN dataset the RDA models with environmental variables were not significant but for NC = 3 and NC = 4, explaining at most an adjusted R^2 of 0.02. As for the spatial analysis it seems that there are no significant differences in the environmental characteristics of the fuzzy clusters obtained from the genetic PCA.

The membership grades for the clusters of the OTHO dataset were always significantly related to the environmental parameters, but still the adjusted R^2 did not exceed 0.03. Thus, the diversification based on otolith shape PCAs for the red mullet is stronger than the one based on genetic PCAs, but still not very strong.

For the MICRO dataset fuzzy clusters, because of a reduced number of sampling locations, the relationships with environmental variables were stronger, reaching an adjusted $R^2 = 0.14-0.15$ for NC = 8 and $0.13-0.14$ for NC = 4. Nevertheless, with 8 clusters the number of sampling locations and individuals per cluster would be very low (respectively an average of ca. 3 sampling locations and 30 individuals per clusters), thus for the less abundant datasets does not seem justifiable to consider such high number of clusters.

The GEN.OTHO dataset clusters were much better explained by the environmental parameters than those based on either the genetic or the otolith shape datasets alone. It has to be reminded though, that in this case the genetic dataset was much less abundant, and the combined dataset had only 624 individuals, compared to the 1756 of the otolith shape dataset and the 771 of the genetic dataset. Thus, at least part of this increase in explanatory power of environmental variables is due to a reduction of complexity in the data. Nevertheless, as for the hake, it seems that the combination of datasets of different origin might help in a better identification of different fish stocks.

The OTHO.MICRO dataset, as in the spatial analysis, was completely unrelated to gradients in environmental factors, but for NC = 2. The causes of such low relationships with either the spatial components and the environmental factors of the combined otolith shape and otolith microchemistry datasets are unknown and should be investigated better.

Finally, clusters obtained on the combination of all three original datasets of *Mullus barbatus* were again well related to the environmental factors, especially at NC = 4 and NC = 8. As for all cases with low numbers of individuals and sampling locations, considering many clusters, with few individuals/sampling locations per cluster, might not give much information than the obvious conclusion that sampling locations and individuals sampled far away from each other tend to be more different than those sampled close.

Table 70: Adjusted R² for each dataset and their combinations of *Mullus barbatus*, calculated for RDA models with environmental variables as explanatory variables. Different models were built for each set of explanatory variables: bottom and surface, and the three time intervals considered (5 years, 3 years, 1 year). Results are reported for each original dataset, and the clusterizations from 2 to 10 clusters.

GENETIC															
INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.00	0.00	0.00	ns	ns	ns	0.02	0.01	0.02	0.02	0.02	0.02	ns	ns	ns
surface	0.00	0.00	0.00	ns	ns	ns	0.01	0.01	0.01	0.01	0.01	0.01	ns	ns	ns
6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.00	ns	ns	0.00	ns	ns
surface	ns	ns	ns	ns	ns	ns	ns	ns	0.00	0.00	ns	0.00	ns	ns	ns
OTOLITH SHAPE															
INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03
surface	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.03
6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
surface	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
OTOLITH MICROCHEMISTRY															
INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.11	0.11	0.12	ns	ns	ns	0.10	0.11	0.11	0.12	0.12	0.14	0.10	0.09	0.11
surface	0.12	0.12	0.12	ns	ns	ns	0.11	0.11	0.12	0.13	0.13	0.13	0.10	0.10	0.10
6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.11	0.11	0.12	0.12	0.12	0.12	0.14	0.14	0.14	0.12	0.12	0.12	0.12	0.11	0.12
surface	0.12	0.11	0.12	0.13	0.12	0.12	0.15	0.14	0.14	0.12	0.12	0.12	0.12	0.12	0.12
GENETIC + OTOLITH SHAPE															
INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.02	0.02	0.01	0.08	0.09	0.07	0.08	0.08	0.07	0.06	0.06	0.06	0.04	0.05	0.05
surface	0.01	0.01	0.01	0.06	0.04	0.07	0.06	0.05	0.07	0.04	0.04	0.05	0.03	0.03	0.04
6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
surface	0.03	0.02	0.03	0.02	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.02
OTOLITH SHAPE + OTOLITH MICROCHEMISTRY															
INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS			
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.11	0.11	0.11	0.04	ns	0.04	ns	ns	ns	ns	ns	ns	ns	ns	ns
surface	0.11	0.11	0.11	ns	0.04	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS			

	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
surface	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
GENETIC + OTOLITH SHAPE + OTOLITH MICROCHEMISTRY															
	INDIVIDUALS			2 CLUSTERS			3 CLUSTERS			4 CLUSTERS			5 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.06	0.07	0.07	ns	ns	ns	0.13	0.10	0.13	0.06	ns	0.06	ns	ns	ns
surface	0.06	0.06	0.06	ns	ns	ns	0.13	0.12	0.12	0.05	ns	0.05	ns	ns	ns
	6 CLUSTERS			7 CLUSTERS			8 CLUSTERS			9 CLUSTERS			10 CLUSTERS		
	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1	y5	y3	y1
bottom	0.08	0.07	0.08	0.09	0.09	0.10	0.14	0.14	0.15	0.13	0.12	0.14	0.11	0.11	0.12
surface	0.08	0.07	0.07	0.09	0.09	0.09	0.15	0.14	0.15	0.13	0.12	0.12	0.11	0.11	0.11

Based on the environmental analysis alone, the suggested best choices of clusterizations would be: 3-4 for the GEN dataset, 2 or 5-6 for the OTHO dataset, 2-3 for the GEN.OTHO; as for the less abundant datasets: 4 or anything between 6 and 10 for the MICRO dataset, 2 for the OTHO.MICRO dataset and 3 or 8-9 for the ALL dataset. As already said above, clusterizations of datasets with few sampling points and individuals, sampled far away from each other, tend to be explained better than clusterizations of dense sampling design simply because points far away from each other tend to have different environmental characteristics. Nevertheless, care should be taken when extrapolating such results to a much wider area where the environmental characteristics might be much different from those of the sampled areas.

For the most interesting results (i.e., the OTHO dataset clusterizations because of the highest number of individuals and sampling locations; the GEN.OTHO dataset because of a relatively high number of individuals and sampling locations, together with a large number of variables; the ALL dataset because of the combination of the variables of all three original datasets) the environmental models were subjected to a forward selection procedure to select the best environmental predictors. It was done for the OTHO clusterizations at NC = 2, 5 and 6, for the GEN.OTHO clusterizations at NC = 2 and 3, and for the ALL clusterizations at NC = 3. In all three cases the bottom medians of environmental variables over 5 years were used as predictors.

The forward selection procedure applied to the environmental model of the OTHO NC = 2 results did select only the median temperature and the median concentration of NO₃ as significant variables (Table 71). In Fig. 70 both are positively related to cluster 1 membership grades and negatively to cluster 2 membership grades. At NC = 5 and NC = 6 both variables were again selected as the most important but joined with other 8 and 10 variables respectively. In any case, the two clusterizations at NC = 5 and 6 are related to the same patterns since in both cases the first ten selected variables are the same. Nevertheless, when interpreting the relations found it should be reminded the low explanatory power of these models.

Table 70: Forward selection on the environmental variables of the bottom over 5 years for *M. barbatus* OTHO dataset at NC = 2, 5 and 6. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT OTHO 2 CLUSTERS					
variables	AdjR2Cum	p-value			
P50_b_T	0.01	0.001			
P50_b_NO3	0.02	0.001			
MUT OTHO 5 CLUSTERS			MUT OTHO 6 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_b_T	0.01	0.001	P50_b_T	0.01	0.001
P50_b_NO3	0.01	0.001	P50_b_NO3	0.01	0.001
P50_b_chlf	0.01	0.001	P50_b_chlf	0.02	0.002
P50_b_NH4	0.02	0.003	P50_b_NH4	0.02	0.001
P50_b_sal	0.02	0.01	P50_b_vel	0.02	0.016
P50_b_PO4	0.02	0.011	P50_b_sal	0.02	0.035
P50_b_pH	0.02	0.002	P50_b_phy	0.02	0.003
P50_b_O2o	0.02	0.002	P50_b_PO4	0.02	0.021
P50_b_vel	0.02	0.002	P50_b_O2o	0.03	0.001
P50_b_phy	0.03	0.024	P50_b_pH	0.03	0.001
			P50_b_ALK	0.03	0.013
			P50_b_DIC	0.03	0.006

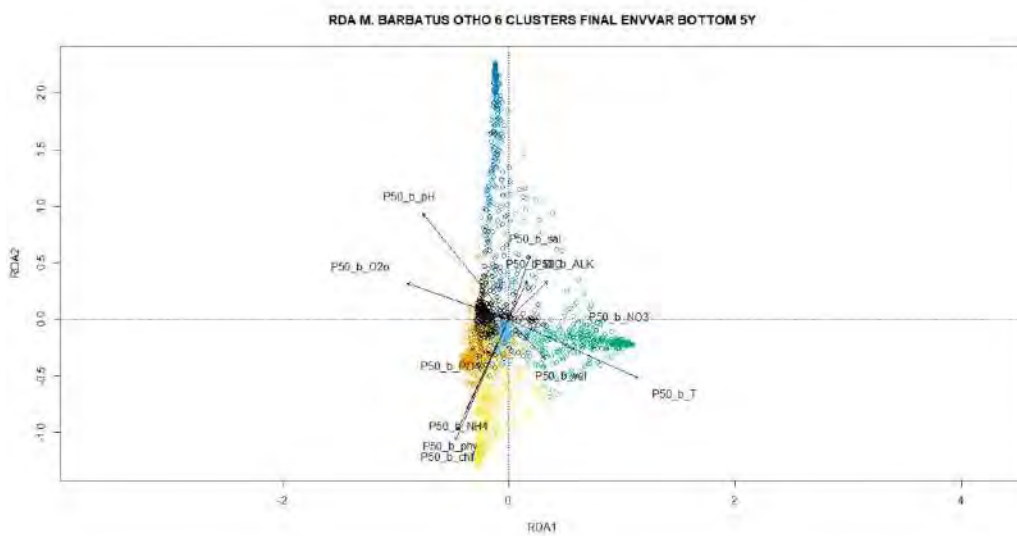
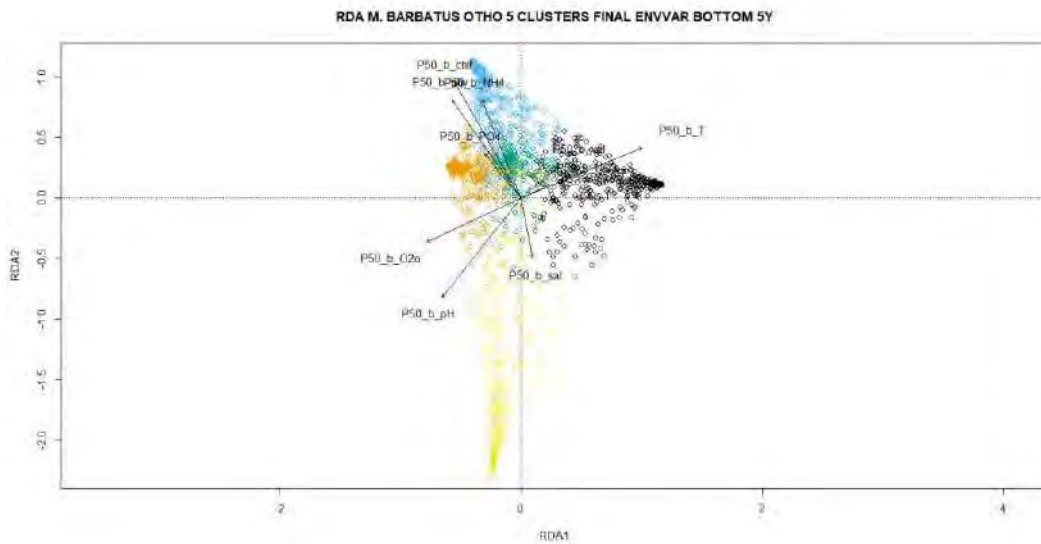
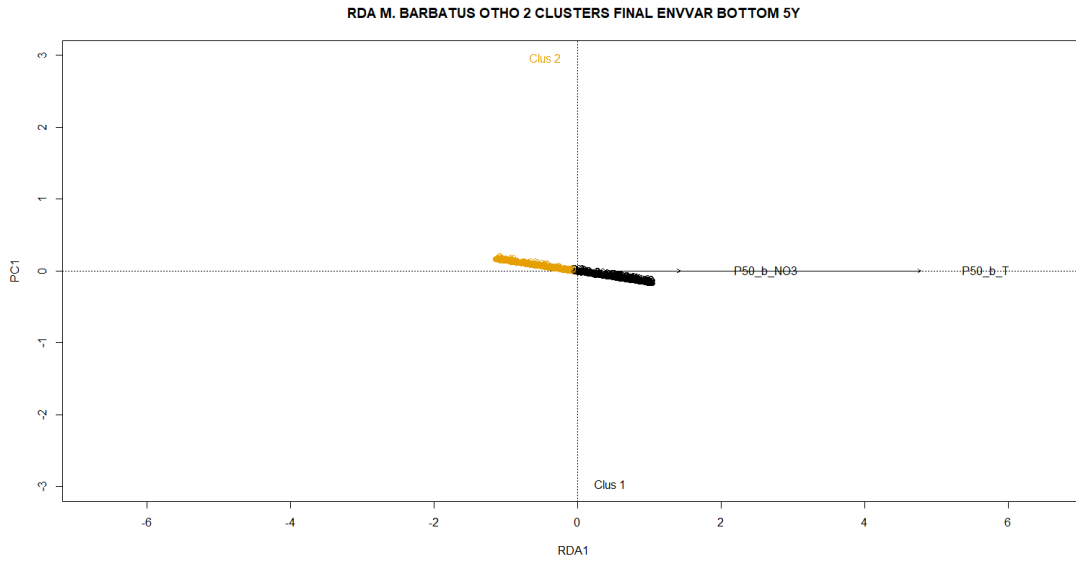


Figure 70: Best environmental parsimonious RDA models for *M. barbatus* OTHO dataset at NC = 2, 5 and 6 (from top to bottom). Explained variance on the first and only RDA axis was 0.02 for NC = 2; 0.02 and 0.01 on the first and second RDA axis at NC = 5; 0.02 and 0.01 on the first and second RDA axis at NC = 6.

The forward selection applied to the environmental models of the GEN.OTHO dataset clusterizations (Table 72) selected two variables (concentrations of PO4 and NO3) at NC = 2, and 6 variables at NC = 3 (temperature, dissolved oxygen, concentrations of NO3 and DIC, primary production and pH). While at NC = 2 both variables contribute to the first gradient and are positively related to cluster 2 membership grades, at NC = 3 only temperature and DIC seem to have an important role in defining the first gradient (Fig. 71): concentration of NO3 and dissolved oxygen are related to the second gradient, while ppn and DIC have marginal roles in defining both these two gradients. The explained variance is in any case higher than for the models of the OTHO dataset, in particular at NC = 3, thus these results should be considered more interesting than the former.

Table 72: Forward selection on the environmental variables of the bottom over 5 years for M. barbatus GEN.OTHO dataset at NC = 2 and 3. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT GEN.OTHO 2 CLUSTERS			MUT GEN.OTHO 3 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_b_PO4	0.02	0.001	P50_b_T	0.02	0.001
P50_b_NO3	0.04	0.001	P50_b_O2o	0.04	0.001
			P50_b_NO3	0.05	0.001
			P50_b_ppn	0.06	0.001
			P50_b_DIC	0.07	0.01
			P50_b_pH	0.07	0.003

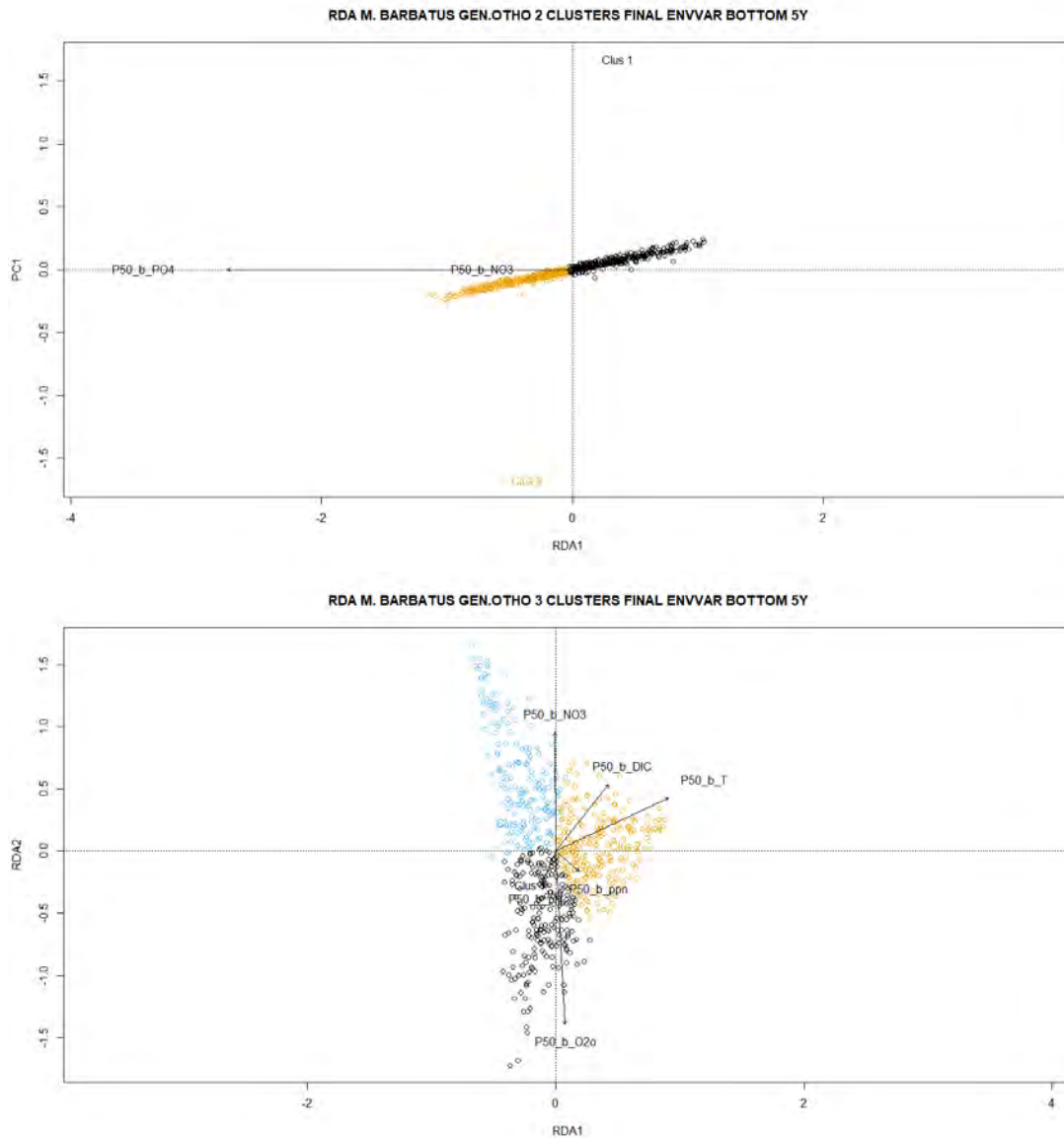


Figure 71: Best environmental parsimonious RDA models for *M. barbatus* GEN.OTHO dataset at NC = 2 (upper panel) and 3 (lower panel). Explained variance on the first and only RDA axis was 0.04 for NC = 2; 0.07 and 0.02 on the first and second RDA axis at NC = 3.

When applying the forward selection procedure to the (Table 73) only two variables were selected as significant: temperature and primary production. Temperature defines mostly the first gradient, which separates cluster 1 from cluster 2, and together with ppn the second gradient, which separates between cluster 3 and the other two clusters (Fig. 72). Nevertheless, the variance explained by this second axis was almost negligible. Thus, on a sparse sampling design covering the whole Mediterranean, only temperature was important in explaining the distribution of three fuzzy membership grades for the ALL dataset.

Table 73: Forward selection on the environmental variables of the bottom over 5 years for *M. barbatus* ALL dataset at NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT ALL 3 CLUSTERS		
variables	AdjR2Cum	p-value
P50_b_T	0.03	0.008
P50_b_ppn	0.05	0.032

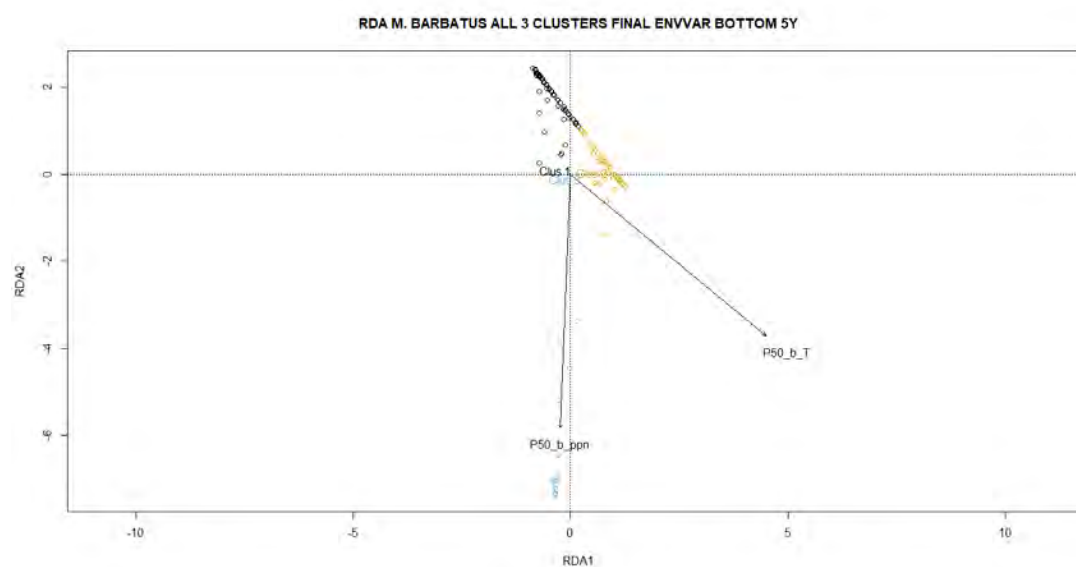


Figure 72: Best environmental parsimonious RDA models for *M. barbatus* ALL dataset at NC = 3. Explained variance on the first and second RDA axis was respectively 0.06 and 0.00.

Variation partitioning and best model selection

The variation partitioning procedure was applied on the most interesting models, i.e., models for the GEN.OTHO clusterizations at NC = 2 and 3, and models for the ALL clusterization at NC = 3. For the OTHO clusterizations this procedure was not applied since the explained variance of the best models did not exceed 0.04-0.05 and generally with a very large number of significant predictors.

From Table 74 is clear that for the GEN.OTHO clusterizations the environmental variables were those with the highest explanatory power on the combination of genetic and otolith shape PCAs at NC =2 and NC = 3. Around half of it was shared with the significant MEMs, much less with the geographic coordinates. In fact, environmental variables alone could explain all variance that can be explained. The unexplained fraction was very high, reaching 0.92 in both cases.

Table 74: Variation partitioning for GEN.OTHO dataset of *M. barbatus* at NC = 2 and NC = 3. + indicates the total variation explained by the groups of variables; | indicates the variation explained by the first group of variables

conditioned to the second; U indicates the fraction of variance explained jointly by the two groups of variables. Compare to Fig. 73 for a graphical representation.

M.BARBATUS GEN.OTHO	2 CLUSTERS	3 CLUSTERS
Explanatory group of variables	Adjusted R²	Adjusted R²
<i>LongLat</i>	0.01	0.02
<i>MEM</i>	0.04	0.05
<i>EnvVar Bot 5y</i>	0.08	0.08
<i>LongLat + MEM</i>	0.03	0.05
<i>LongLat + EnvVar Bot 5y</i>	0.07	0.08
<i>MEM + EnvVar Bot 5y</i>	0.08	0.08
<i>LongLat + MEM + EnvVar Bot 5y</i>	0.08	0.08
<i>LongLat EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat MEM</i>	0.00	0.00
<i>MEM EnvVar Bot 5y</i>	0.00	0.01
<i>MEM LongLat</i>	0.02	0.03
<i>EnvVar Bot 5y LongLat</i>	0.06	0.06
<i>EnvVar Bot 5y MEM</i>	0.04	0.03
<i>LongLat MEM + EnvVar Bot 5y</i>	0.00	0.00
<i>MEM LongLat + EnvVar Bot 5y</i>	0.00	0.00
<i>EnvVar Bot 5y LongLat + MEM</i>	0.04	0.03
<i>LongLat \cup MEM</i>	0.00	0.00
<i>MEM \cup EnvVar Bot 5y</i>	0.02	0.03
<i>LongLat \cup EnvVar Bot 5y</i>	0.00	0.00
<i>LongLat \cup MEM \cup EnvVar Bot 5y</i>	0.01	0.01
<i>Residuals</i>	0.92	0.92

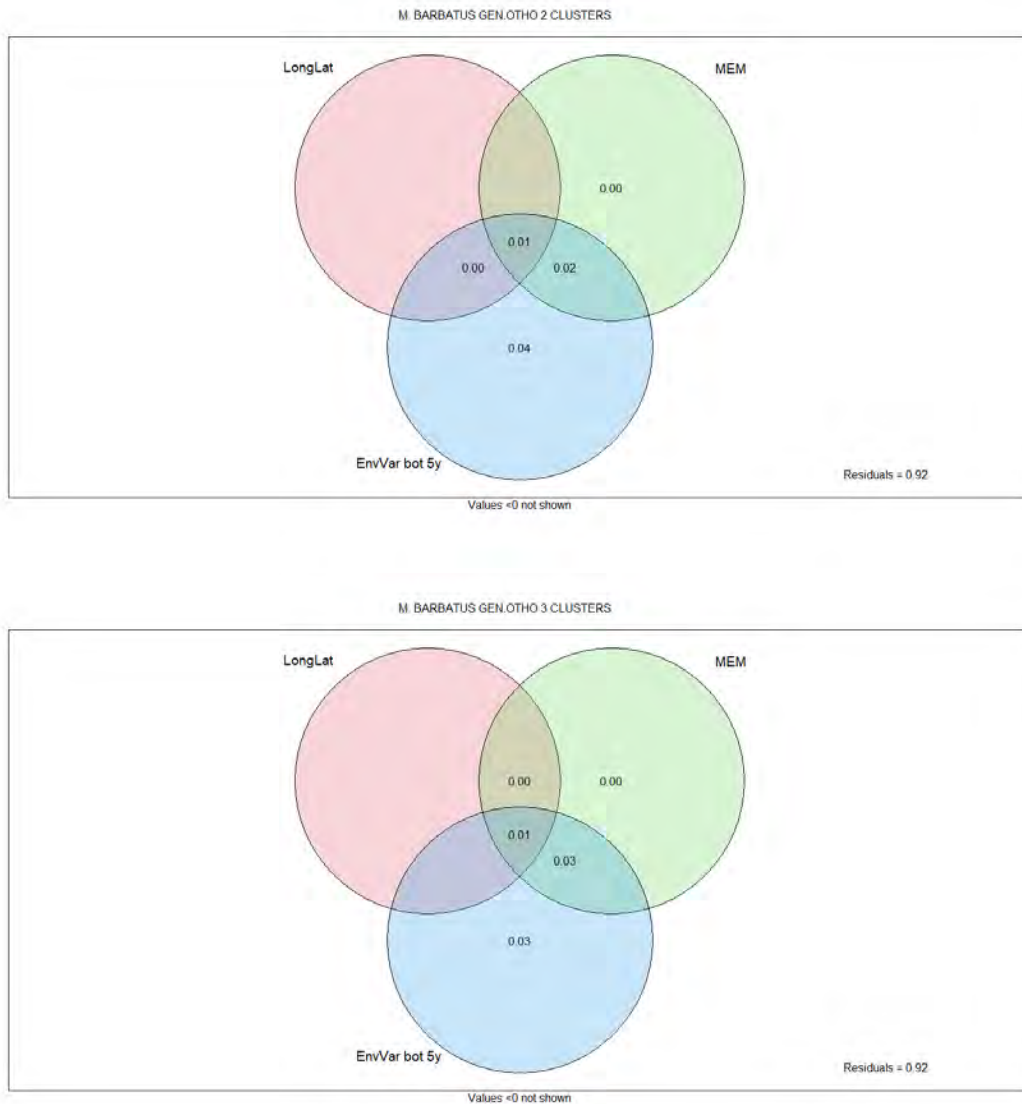


Figure 73: Graphical representation of the variation partitioning of GEN.OTHO clusterizations for *M. barbatus* at 2 and 3 clusters among three groups of variables: LongLat, MEMs and EnvVar bot 5y. Negative or zero fractions are not displayed. See Table 74 for more details.

In the case of the combination of all three original dataset of *Mullus barbatus* (ALL dataset), the MEMs had a higher explanatory power than the environmental variables, but the two set of descriptors shared much of their explanatory power (Table 75). The amount of unexplained variance was lower than for the GEN.OTHO clusterizations, but it has to be reminded that the number of individuals and the spatial coverage was much lower in the ALL dataset.

Table 75: Variation partitioning for ALL dataset of *M. barbatus* at NC = 3. + indicates the total variation explained by the groups of variables; | indicates the variation explained by the first group of variables conditioned to the second; U indicates the fraction of variance explained jointly by the two groups of variables. Compare to Fig. 74 for a graphical representation.

M.BARBATUS ALL	3 CLUSTERS
----------------	------------

Explanatory group of variables	Adjusted R ²
MEM	0.15
EnvVar Bot 5y	0.13
MEM + EnvVar Bot 5y	0.19
MEM EnvVar Bot 5y	0.06
EnvVar Bot 5y MEM	0.04
MEM \cup EnvVar Bot 5y	0.09
Residuals	0.81



Figure 74: Graphical representation of the variation partitioning of ALL clusterization for *M. barbatus* at 2 clusters among two groups of variables: MEMs and EnvVar bot 5y. Negative or zero fractions are not displayed. See Table 75 for more details.

Considering the results of the variation partitioning procedure, it is no surprise that for the GEN.OTHO final models at NC = 2 and NC = 3 many environmental parameters were selected. At NC = 3 in particular the concentration of PO₄, MEM23 and MEM19 are associated to cluster 2 membership grades, while velocity to cluster 1 membership grades. NO₃ concentration and velocity were selected also in the NC = 3 case, together with temperature, dissolved oxygen, ppn, DIC and pH. Temperature, velocity and DIC were mostly related to the gradient of the first RDA axis (Fig. 75), which tried to separate cluster 2 from cluster 1 and cluster 3 samples. The second axis was mostly determined by O₃ concentration, increasing with increasing cluster 3 memberships, and dissolved oxygen,

increasing with increasing cluster 1 memberships. Significantly, no spatial component was selected for the best model at NC = 3.

Table 76: Forward selection of the final models for M. barbatus GEN.OTHO dataset at NC = 2 and 3. The adjusted R² is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT GEN.OTHO 2 CLUSTERS			MUT GEN.OTHO 3 CLUSTERS		
variables	AdjR2Cum	p-value	variables	AdjR2Cum	p-value
P50_b_PO4	0.02	0.001	P50_b_T	0.02	0.001
P50_b_NO3	0.04	0.002	P50_b_O2o	0.04	0.001
MEM23	0.05	0.01	P50_b_NO3	0.05	0.001
MEM19	0.05	0.028	P50_b_ppn	0.06	0.003
P50_b_vel	0.06	0.048	P50_b_DIC	0.07	0.015
			P50_b_pH	0.07	0.006
			P50_b_vel	0.08	0.019

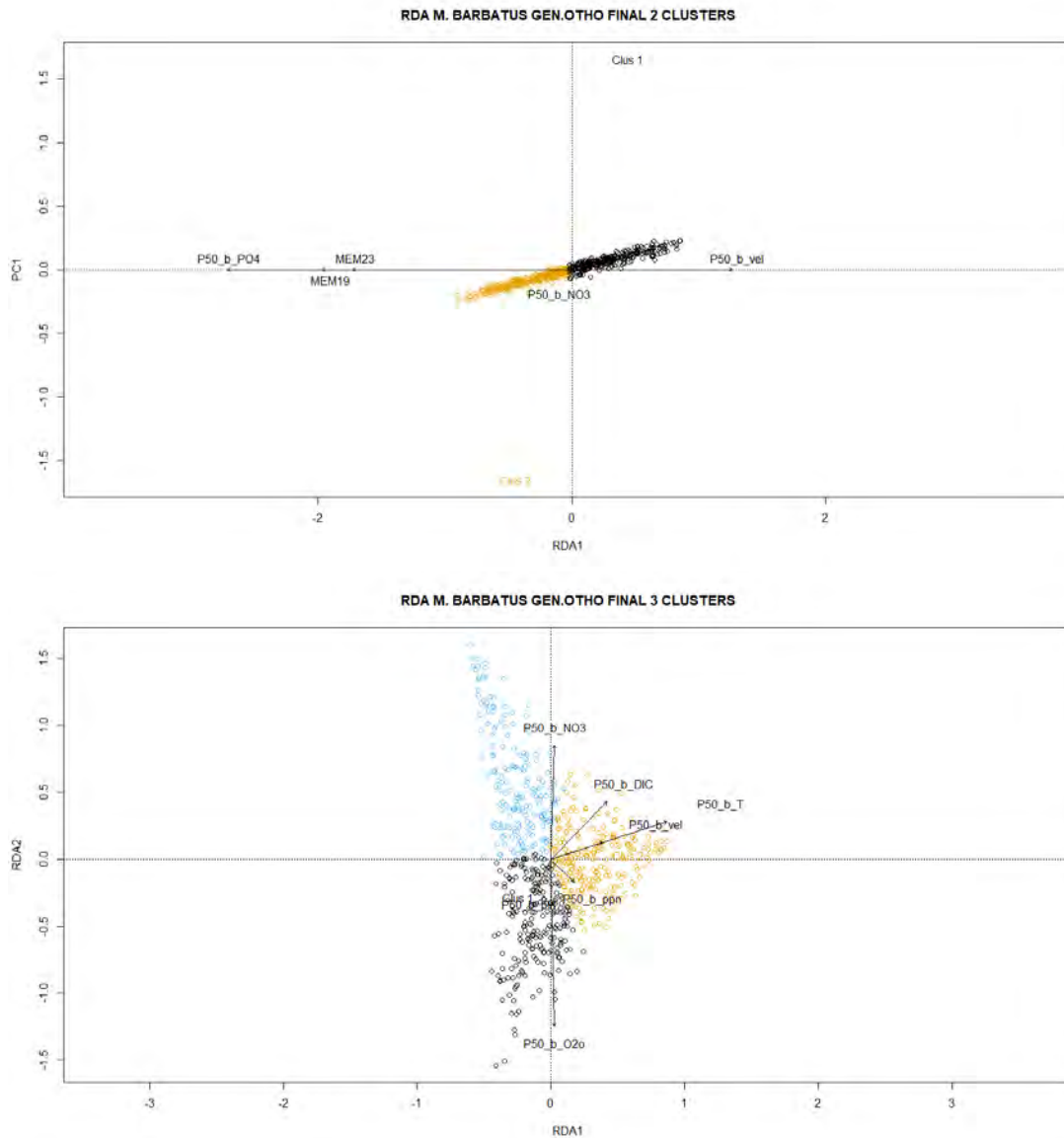


Figure 75: Best final RDA models for *M. barbatus* GEN.OTHO dataset at NC = 2 (upper panel) and 3 (lower panel). Explained variance on the first and only RDA axis was 0.07 for NC = 2; 0.07 and 0.02 on the first and second RDA axis at NC = 3.

The forward selection for the ALL dataset at NC = 3 did select only one environmental parameter (velocity), along with five large to medium-scale MEMs (MEM3, MEM6, MEM7, MEM8, MEM9) (Table 77). The main gradient (associated mostly to MEM9, MEM3 and MEM6 positive values) tried to separate cluster 1 from cluster 2, while the second axis (associated mostly to a combination of MEMs, with MEM7 positive values oriented inversely to high cluster 3 memberships) separated cluster 3 membership grades from the rest (Fig. 75).

Table 77: Forward selection of the final models for *M. barbatus* ALL dataset at NC = 3. The adjusted R^2 is reported as cumulative values and the variables are in the order in which they were added to the model. P-values were computed after 999 permutations.

MUT ALL 3 CLUSTERS		
variables	AdjR2Cum	p-value
MEM9	0.08	0.001
MEM3	0.10	0.013
MEM8	0.12	0.025
MEM7	0.13	0.032
MEM6	0.15	0.029
P50_b_vel	0.18	0.009

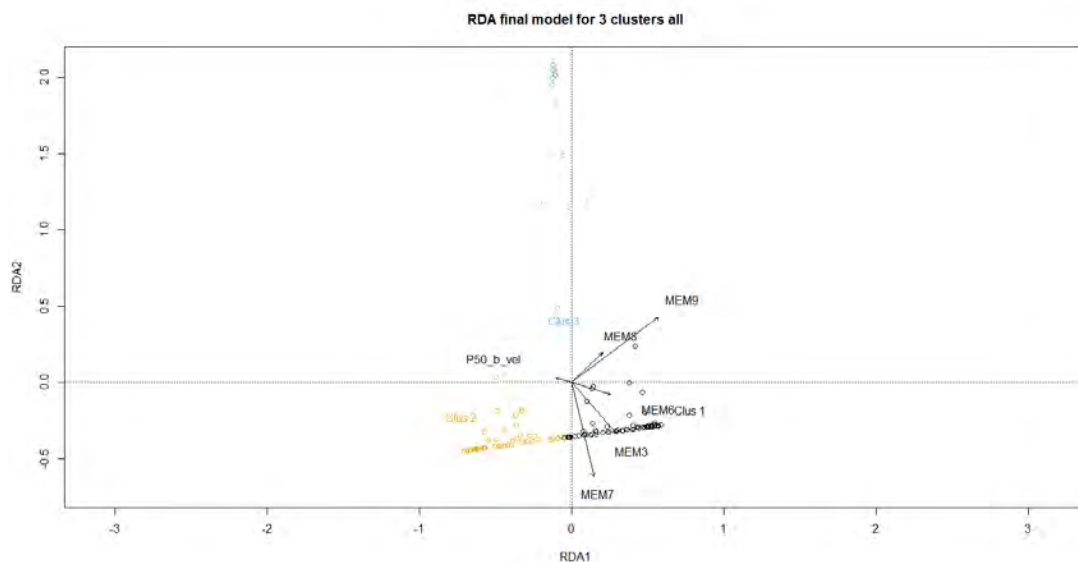


Figure 76: Best final RDA models for *M. barbatus* ALL dataset at NC = 3. Explained variance on the first and second RDA axis was 0.18 and 0.04.

Conclusions

The analysis of the *Mullus barbatus* datasets did not give clear results, neither considering the internal quality indexes, nor the distribution of membership grades averaged per GSA subarea, nor even with the external validation through spatial and environmental analysis. While, as for the hake, the combination of different datasets generally increased the relationships with the spatial and environmental data, the variance explained in the RDA models remained very low. Some of the datasets did not give useful information at all, such as the OTHO.MICRO dataset, which showed only few significant relations with any predictor.

The best results were obtained for the clusterizations of the GEN.OTHO dataset at NC = 2 and NC = 3, and for the clusterization of the ALL dataset at NC = 3. Nevertheless, the spatial

distribution in the first case was not clear: if at NC = 2 there seem to be a cluster grouping together most of the Adriatic, Ionian and Central Mediterranean subareas, this group is broken at NC = 3. Furthermore, other situations are more dubious, for instance subarea 1b is always grouped together with the Adriatic 17b and 18c, while the two neighboring subareas 27a and 27b in both cases are associated to different clusters. Areas 20a, 23 and 25c are instead always grouped with most of the Western Mediterranean subareas. Thus, no clear spatial rationale could be inferred even from the clusterizations that were best explained by the spatial and environmental variables. The same was even more true for the ALL dataset at NC = 3, where all GSA subareas did show a neat preference for just one cluster.

In both cases, the explained variance was low reaching less than 10% for the GEN.OTHO dataset and around 20% for the ALL dataset (with much reduced sampling coverage and individuals' number). Any conclusion drawn from these results should take into account this low explained variance. Depth was never an important factor, and the geographic coordinates did not show significant relationships with the distribution of fuzzy membership grades.

Thus, from the data used in this study and with the methods applied here the conclusion is that there are no diverging populations of the red mullet in the Mediterranean. Most of the variability in the datasets analyzed is either random or at an individual level and no spatially defined grouping could be identified.

There are some indications that might indicate that *M. barbatus* populations in the Mediterranean diverge on scales smaller than those of the sampling design in the present study, such as: the fact that mostly medium and small-scale MEMs were selected as significant in the various RDA models (the only exception being in the datasets with less spatial coverage, the MICRO and ALL datasets); the increase of the total variance explained with either spatial or environmental variables for clusterizations with more clusters that was observed for several datasets (e.g., in the OTHO, MICRO, and ALL datasets); the indications emerging from the internal validity indexes that in several cases pointed to higher numbers of clusters as the optimal choice for clusterization (in the GEN, OTHO and GEN.OTHO datasets). Nevertheless such indications are not conclusive and only a detailed and spatially dense sampling design (perhaps focused on smaller areas) might confirm or reject this hypothesis.

General conclusions

In this Task the interpretation of the quality of the clustering results was made only on the basis of internal cluster quality indexes, and then on the basis of the relationships between the observed fuzzy clusters and the spatial and environmental variables. Further insights might be gained from the comparison of the results of this Task with the results coming from the genetic WP1 (D.1.5.1 and D.1.5.2), and otolith shape (D.2.4) and otolith microchemistry analysis (D.2.5) performed in WP2.

The combination of datasets as attempted in this Task is a promising way. In fact, for the hake and to a lesser degree for the red mullet, by combining GEN and OTHO data together the relations with both the spatial and the environmental variables increased, as judged by their respective explanatory power on the GEN, the OTHO and the GEN.OTHO dataset. The same happened when adding GEN and OTHO to the MICRO dataset: the best explanatory power was reached for the ALL dataset. This is not an obvious result. Combining datasets means also increasing the total variance due to biases, errors and randomness in each dataset: the combined datasets have more variables and are thus more complex than each single dataset is. On the contrary, from our results emerged that by combining different datasets their common patterns are reinforced by each other. Thus, the observed increase in relations with spatial and/or environmental variables when combining dataset, indicates that together these different variables might make the patterns in the fish populations of the Mediterranean more easily identifiable and interpretable.

The fuzzy k-means applied to the PCA derived on genetic data allowed to detect the discontinuities in the data and to inspect their inherent fuzziness. Hard edge discontinuities are rare in the real world, where usually gradual changes along a gradient are more common. This is all the truer for genetic data of potential sub-population of the same species living in different areas of the same basin and linked by dispersal through currents advection. FKM on one side permits to detect differences in the data, if these are strong enough; on the other hand, does not impose a hard-edge discontinuity where much probably does not exist, allowing to relate this fuzziness to the drivers that might have caused it. Thus, the FKM is flexible enough to offer both: discontinuity identification and continuity assessment. The advantage over other hard clustering methods is also that FKM membership grades are continuous variables, thus continuous gradient analysis methods, such as RDA, can be used to relate their distribution to other factors: with hard clustering method, DA (Discriminant Analysis) should be used.

The application of internal cluster validity indexes to assess the quality of the clustering is useful, but if the goal is the identification of possible stocks, not only of genetically or morphologically diverging sub-populations, this is not enough. External validation procedures should be applied too, such as those used in the present study, i.e., the spatial and environmental analysis. The use of RDA models with spatial and environmental variables as external validation method is justified by the fact that if a sub-population does exist, much probably is spatially isolated from other sub-populations (otherwise specific traits

would not have emerged, and if they did, they would not have persisted in time), and perhaps experiencing different environmental conditions.

External validation with spatial and environmental variables is useful especially when there is no a priori knowledge on the number of different sub-population (clusters). Obviously, if the number of clusters is known a priori, it can be explained by applying a DA procedure using either internal information (i.e., related to the genetic/morphological/chemical data in which a diversification among clusters has been observed), or external information (i.e., related to factors that may have caused the observed genetic/morphological/chemical structuring). If such a priori knowledge does not exist, besides applying internal cluster validity indexes, only external validation procedure may be applied by using the available variables (spatial variables, environmental variables, variables on competition, predation, etc.).

Nevertheless, some problems do exist in applying the proposed procedures. The number of PCAs to use in the first step of the proposed methods (clusterization) should be addressed on a quantitative basis considering the variance explained by each PCA. Furthermore, the interplay between the number of PCAs and the parameter of fuzziness should be examined and assessed quantitatively. Cluster validity indexes to apply should also be carefully assessed as to their computation and expected behavior. In fact, in many cases in this study, different cluster validity indexes gave different indications, and thus made problematic the choice of the best number of clusters based on their results.

The explicit spatial method applied here, the MEMs, proved useful. They can be calculated on any sampling design; produce independent spatial components; and can model all-scale spatial patterns, from local autocorrelation to large-scale climate zone patterns. The environmental characteristics of the Mediterranean are also strongly spatially organized due to the spatial structuring of the processes affecting them. Thus, the spatial variables and the environmental ones shared much of their respective explanatory power. The environmental variables used in the present study were medians over long time intervals (1, 3, and 5 years), and they did not show significant differences in the strength of their relationships with the datasets to explain. There was also not much difference in the explanatory power of the bottom and the euphotic zone environmental datasets. This was due to the approach used in the computation of the environmental variables as medians over different depth levels and timeframes of the models. In particular, in many coastal areas the values computed for the bottom and the euphotic zone were overlapping, especially for species with a low depth range (e.g., *M. barbatus*). The choice of considering averaged values over long period of time seem reasonable when analyzing combined datasets of different variables, and the environmental variables should be considered as the “average” description of the habitat in which the individuals lived, rather than the “real” conditions encountered by the individuals in their life. Nevertheless, different time intervals than those used here might prove better related to the genetic, morphological or microchemistry characteristics of the species under study, e.g., for genetic characteristics much probably longer periods of time should be considered, for microchemistry variables shorter time intervals (in the order of months) could be more relevant. In any case, it should be reminded that the existence of significant statistical relationships does not imply the existence of any particular mechanism by which

one group of variables is supposed to be shaped by the other, nor any such attempt has been made in the present study. A refinement of the analysis performed in this task could be to explore also the other values that were extracted from Copernicus products (5th and 95th percentiles of environmental variables), and additional time intervals. Notwithstanding the obvious inaccuracies of a whole Mediterranean 3D biogeochemical and physical model, the significant relationships found between the environmental datasets and the (genetic, otolith shape, otolith microchemistry) data for different species, demonstrate the validity of this approach. Model simulations are the only product by which the salient characteristics of the waters at a basin-scale can be described, summarized, and put into relations with data sampled over a discrete sampling grid.

The results did show that different numbers of potential management units (stocks) can be identified in the Mediterranean for the different species under study. In particular, for *Merluccius merluccius* three stocks were identified; for *Nephrops norvegicus* four stocks; and for *Parapenaeus longirostris* two stocks. No stock was clearly identified for either *Aristeomorpha foliacea*, *Aristeus antennatus* and *Mullus barbatus*. The three stocks of *M. merluccius* were identified on the combination of the genetic and otolith shape data, and were distributed one in the Eastern Mediterranean, one in the Adriatic-Ionian-Tyrrhenian basins, and one in the Western Mediterranean. For *N. norvegicus* one stock was characterizing the Adriatic Sea and the sampled subareas of the Ionian and Aegean Sea, while the other three were distributed in the Western Mediterranean. For *P. longirostris*, the discontinuity between the eastern and the western stocks was positioned in the Ionian Sea south of Italy. Obviously, these results depend critically on the sampling design, which was based on the GSA subareas. Even if in some GSA subareas there were more than one sampling locations, a much denser sampling design would be needed to correctly and consistently reconstruct possible sub-populations at a scale smaller than that of the GSA subareas.

These same considerations hold true also for the interpretation of the results for *A. foliacea*, *A. antennatus* and *M. barbatus*. For the two shrimps no clusterization did show a recognizable spatial distribution, and the relation with the spatial and environmental variables was low, non-significant and did not show any pattern related to the number of clusters. The only possible conclusion is that for these two species, based on the data produced in this study and with the methods applied here, only one population could be identified in the sampled areas of the Mediterranean. For *M. barbatus* the conclusion is much the same, even though there were some rather inconclusive evidences for the existence of 2 to 3 sub-population in the Mediterranean. Other evidences pointed to the possibility of a divergence of *M. barbatus* populations at scales smaller than those sampled here, but also these were inconclusive since limited by the scale of the sampling design. A more dense sampling design, perhaps restricted to smaller areas, might help in confirming or rejecting this hypothesis.

References

- Bauman, D., Drouet, T., Dray, S., Vleminckx, J., 2018a. Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. *Ecography* 41, 1638–1649.
<https://doi.org/10.1111/ecog.03380>
- Bauman, D., Drouet, T., Fortin, M.-J., Dray, S., 2018b. Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. *Ecology* 99, 2159–2166.
<https://doi.org/10.1002/ecy.2469>
- Bezdek J.C., 1974. Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3, 58-73.
- Bezdek, J.C., 1981. *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York, NY, USA.
- Bezdek, J.C., Ehrlich, R., Full, W.E., 1984. FCM: the fuzzy c-means clustering algorithm. *Computers and Geosciences* 10, 192–203.
- Blanchet, F.G., Legendre, P., Borcard, D., 2008. Forward selection of explanatory variables. *Ecology* 89, 2623–2632.
- Blanchet, F.G., Legendre, P., Maranger, R., Monti, D., Pepin, P., 2010. Modelling the effect of directional spatial ecological processes at different scales. *Oecologia* 166, 357–368.
<https://doi.org/10.1007/s00442-010-1867-y>
- Borcard, D., Gillet, F., Legendre, P., 2011. *Numerical Ecology with R, Use R!* Springer New York, New York, NY.
- Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model* 153, 51–68.
- Borcard, D., Legendre, P., Avois-Jacquet, C., Tuomisto, H., 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85, 1826–1832.
- Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the Spatial Component of Ecological Variation. *Ecology* 73, 1045–1055.
- Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the Spatial Component of Ecological Variation. *Ecology* 73, 1045–1055.
- Campello R.J.G.B., Hruschka E.R., 2006. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157, 2858-2875.
- Davè R.N., 1996. Validating fuzzy partitions obtained through c-shells clustering. *Pattern Recognition Letters*, 17, 613-623.

- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2021. adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-14. <https://CRAN.R-project.org/package=adespatial>
- Dray, S., Legendre, P., Peres-Neto, P.R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Escudier, R., Clementi, E., Omar, M., Cipollone, A., Pistoia, J., Aydogdu, A., Drudi, M., Grandi, A., Lyubartsev, V., Lecci, R., Cretí, S., Masina, S., Coppini, G., & Pinardi, N. (2020). Mediterranean Sea Physical Reanalysis (CMEMS MED-Currents) (Version 1) [Data set]. Copernicus Monitoring Environment Marine Service (CMEMS). https://doi.org/10.25423/CMCC/MEDSEA_MULTIYEAR_PHY_006_004_E3R1
- Ezekiel, M., 1930, *Methods of correlation analysis*. Wiley, New York, NY, USA.
- Ferraro, M.B., Giordani, P., Serafini, A., 2019. fclust: An R Package for Fuzzy Clustering. *The R Journal* 11, 198–210.
- ICES. 2020. Stock Identification Methods Working Group (SIMWG). *ICES Scientific Reports*. 2:94. 32 pp. <http://doi.org/10.17895/ices.pub.7485>
- Kaufman L., Rousseeuw P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Legendre, P., Legendre, L., 2012. *Numerical ecology*, 3rd English Edition. ed, *Developments in Environmental Modelling*. Elsevier, Amsterdam; Boston.
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2020. vegan: Community Ecology Package. R package version 2.5-7. <https://CRAN.R-project.org/package=vegan>
- Teruzzi, A., Di Biagio, V., Feudale, L., Bolzon, G., Lazzari, P., Salon, S., Di Biagio, V., Coidessa, G., & Cossarini, G. (2021). Mediterranean Sea Biogeochemical Reanalysis (CMEMS MED-Biogeochemistry, MedBFM3 system) (Version 1) [Data set]. Copernicus Monitoring Environment Marine Service (CMEMS). https://doi.org/10.25423/CMCC/MEDSEA_MULTIYEAR_BGC_006_008_MEDBFM3
- van den Wollenberg, A.L., 1977. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* 42, 207–219. <https://doi.org/10.1007/BF02294050>
- Xie X.L., Beni G. (1991). A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 841-847.

Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-
Project MED_UNITS

*“Study on Advancing fisheries assessment and management advice
in the Mediterranean by aligning biological and management units
of priority species”*

WP4 - SYNTHESIS AND PROPOSALS

*Task 4.2 – Integrating results by different WPs and proposals of
new management units*

Deliverable 3.2 – Report with the maps of population units and discontinuities

Annexes

Responsible: Vinko Bandelj (OGS)

Compiled by: Vinko Bandelj (OGS), Fabrizio Gianni (OGS)

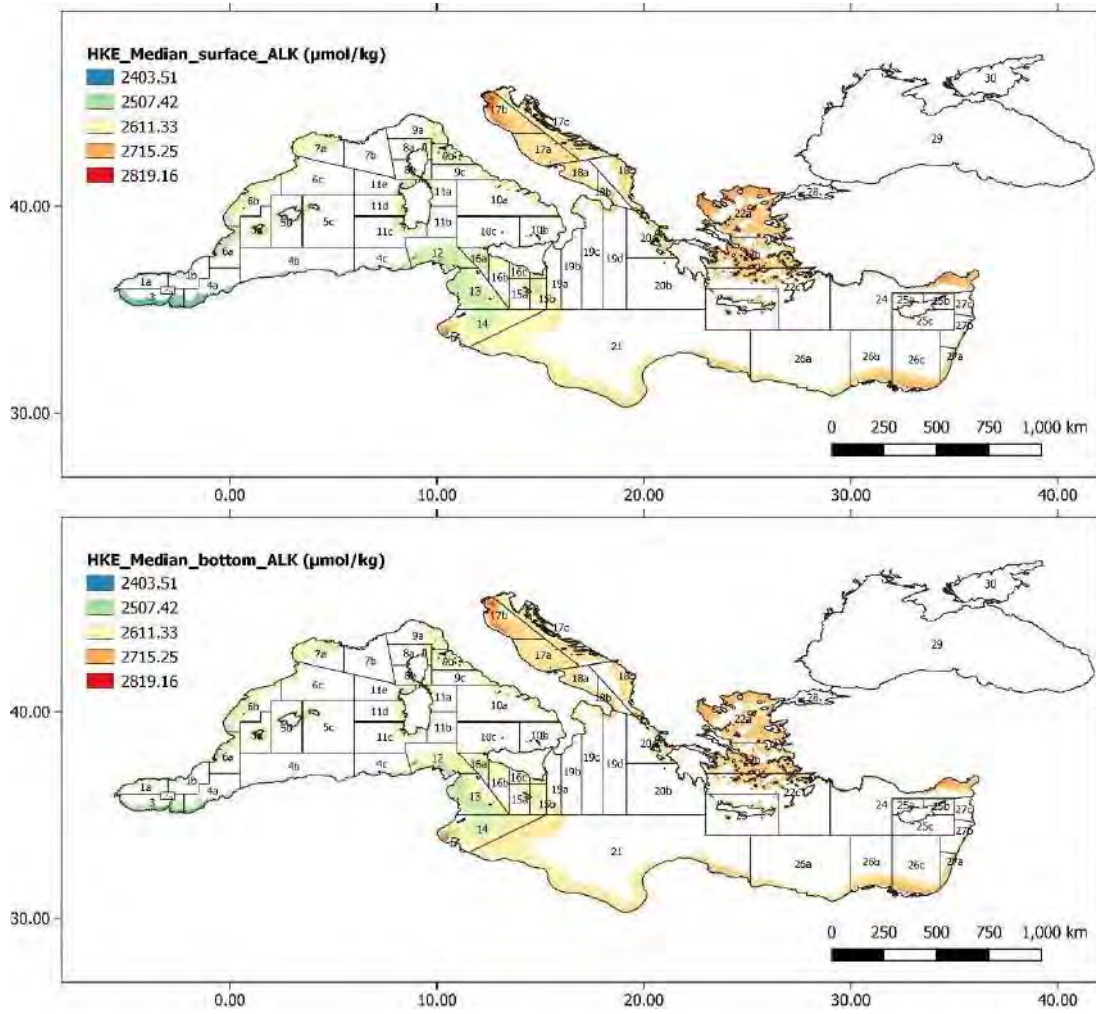
23. 6. 2021

Version: Final

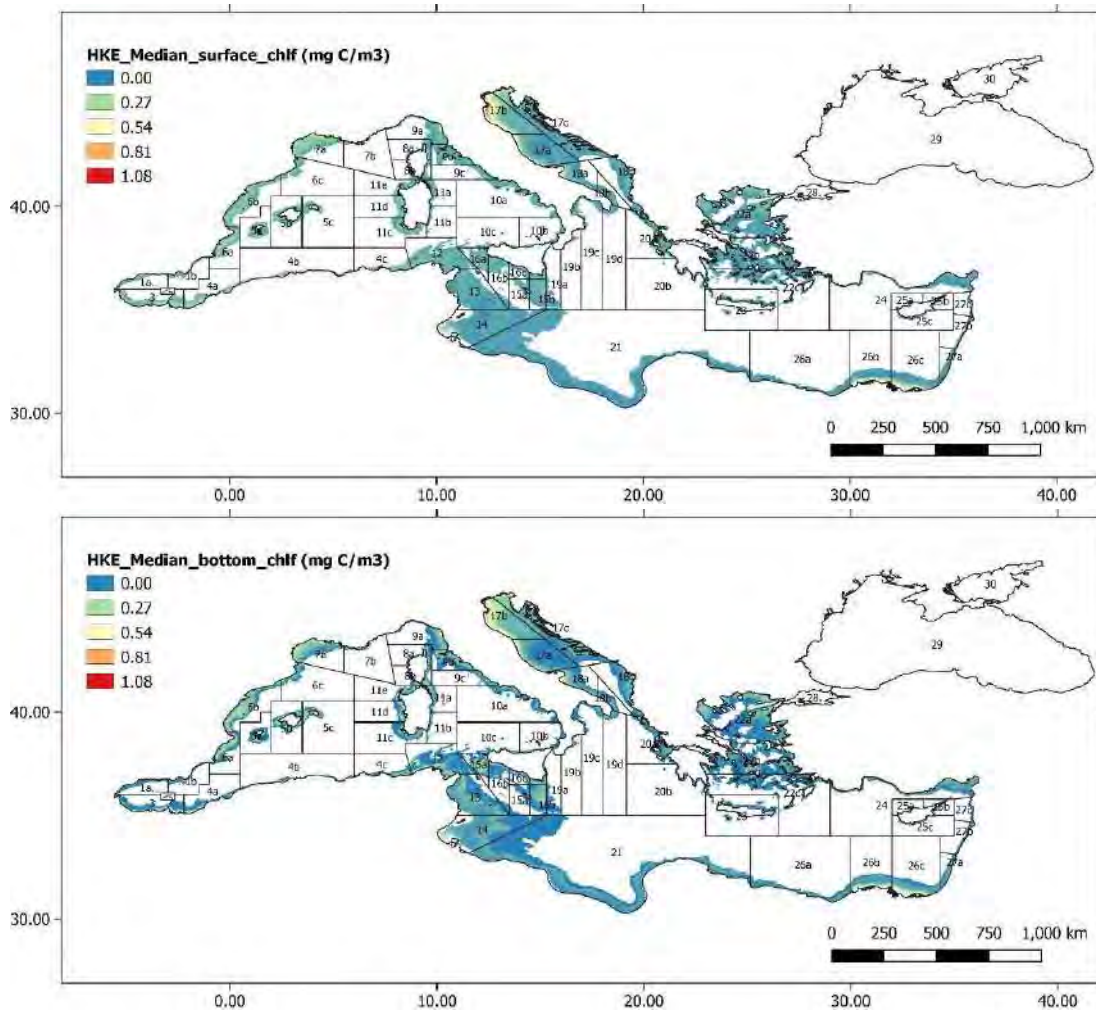
Index

Index	2
Annex 1	3
Annex 2	16
Annex 3	29
Annex 4	42
Annex 5	55
Annex 6	68

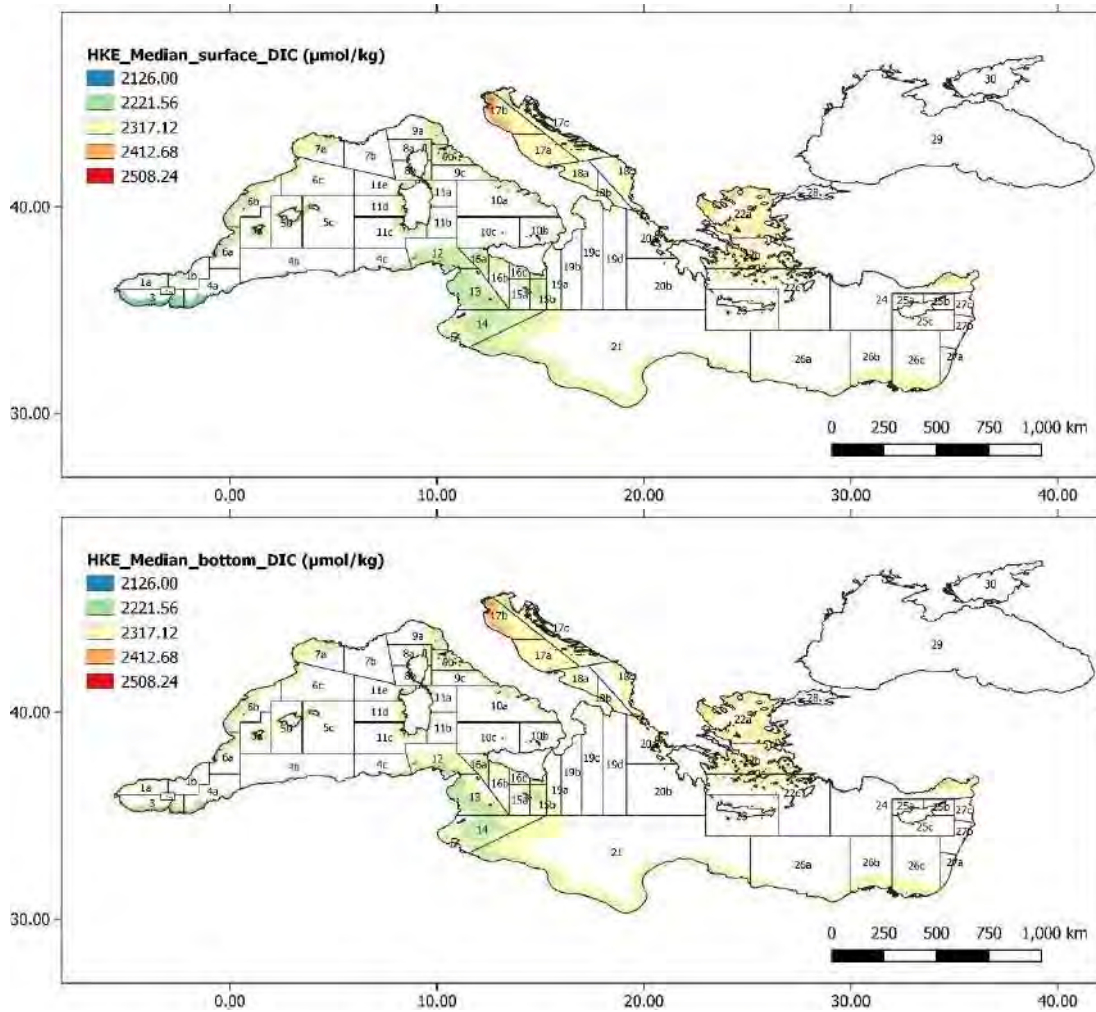
Annex 1



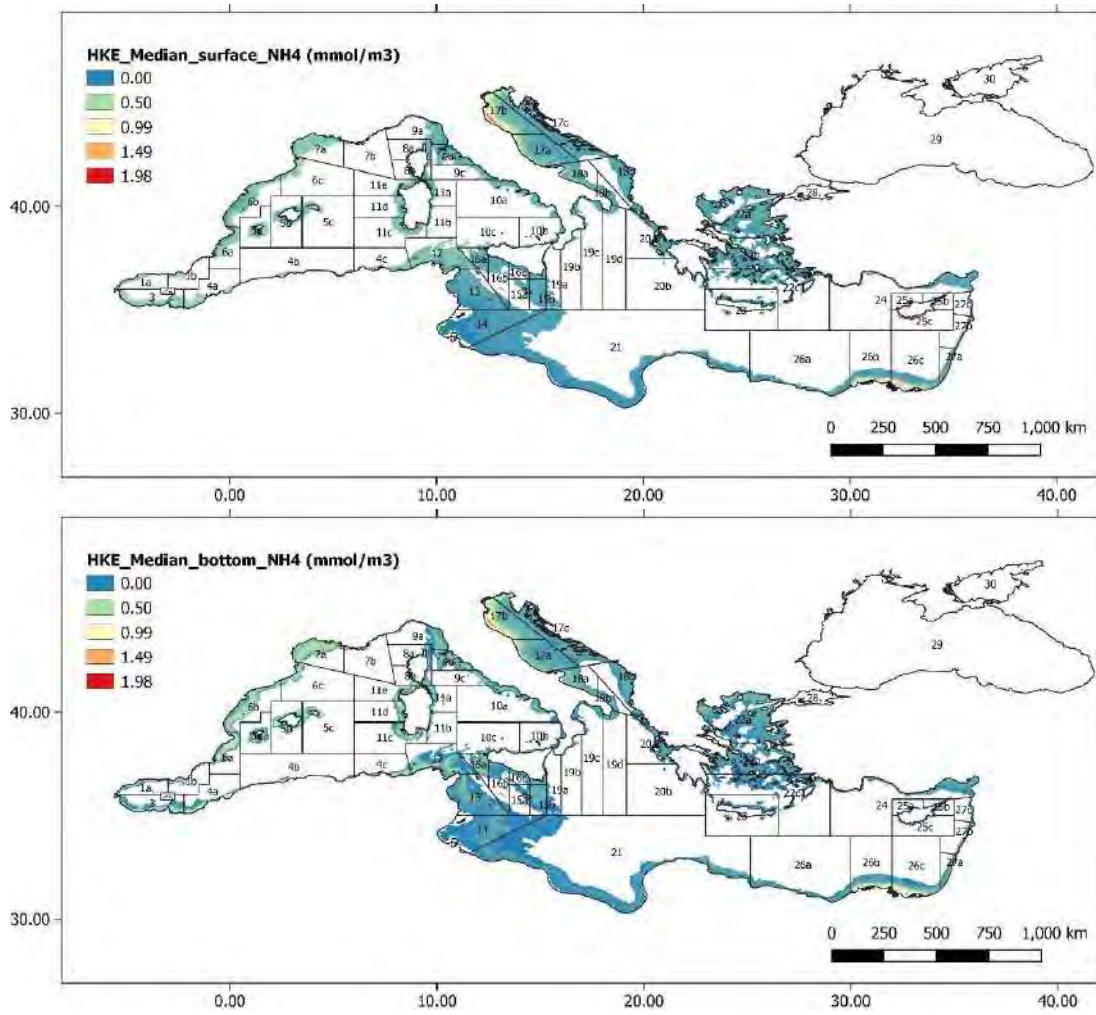
Annex 1 Figure 1: Median euphotic zone (upper panel) and bottom layer (bottom panel) alkalinity over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



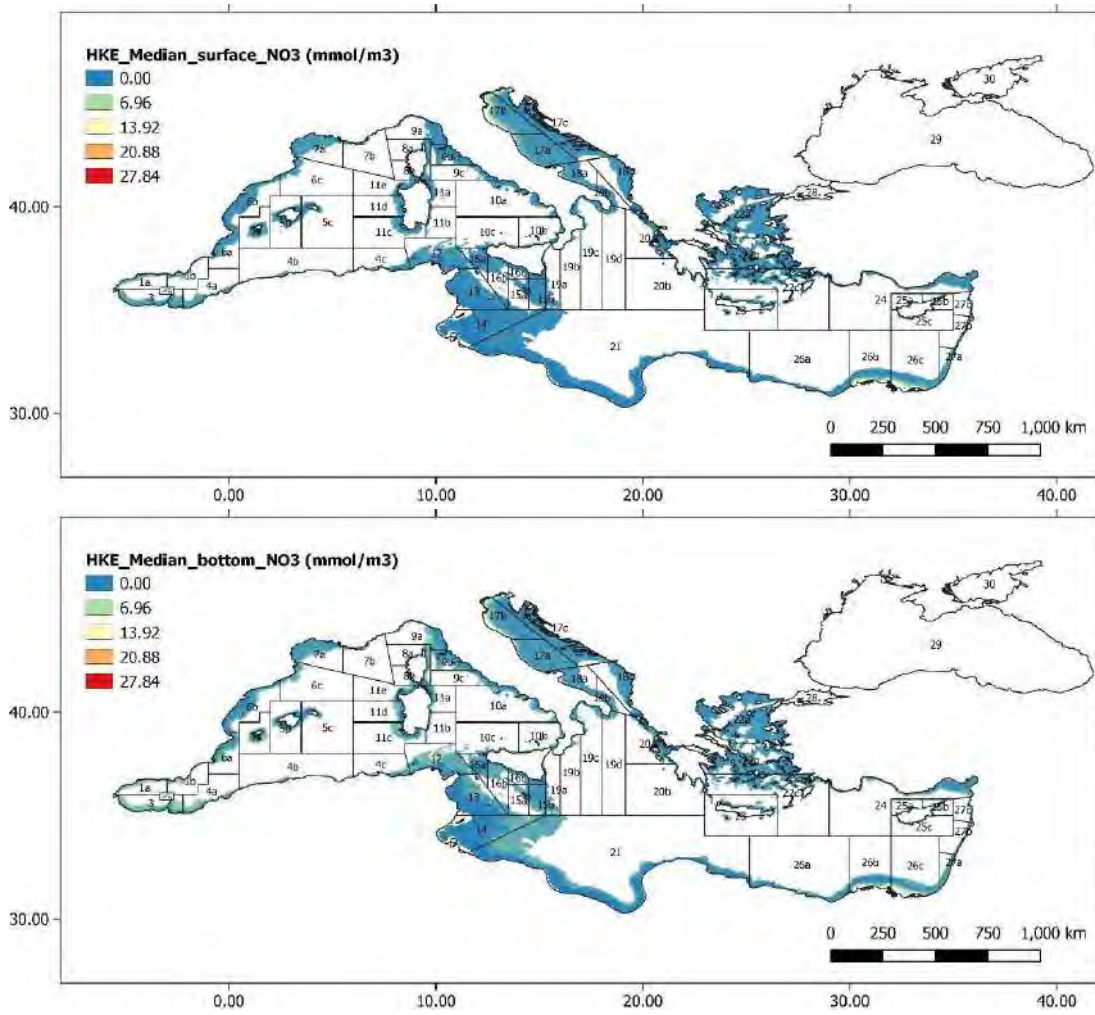
Annex 1 Figure 2: Median euphotic zone (upper panel) and bottom layer (lower panel) chlorophyll a concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



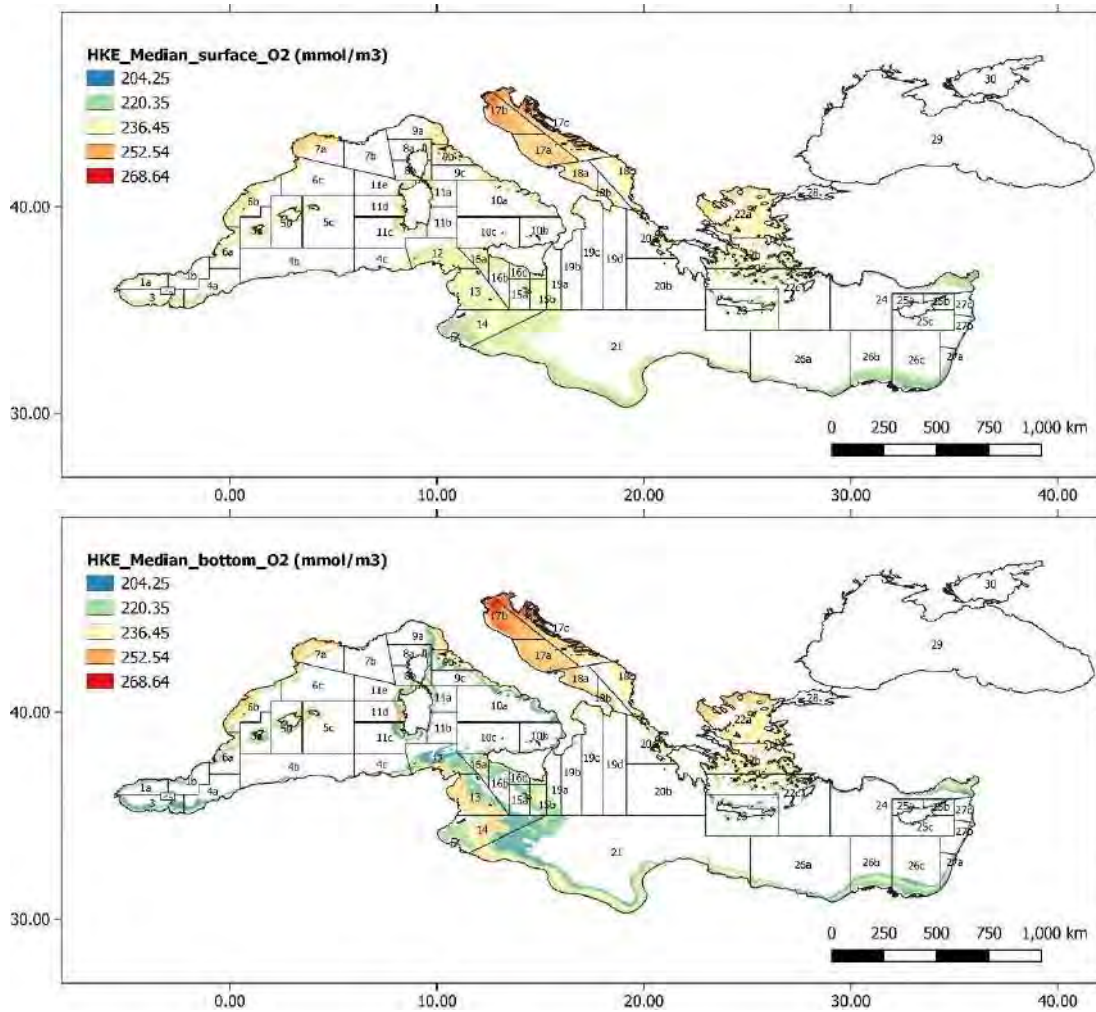
Annex 1 Figure 3: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved inorganic carbon concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



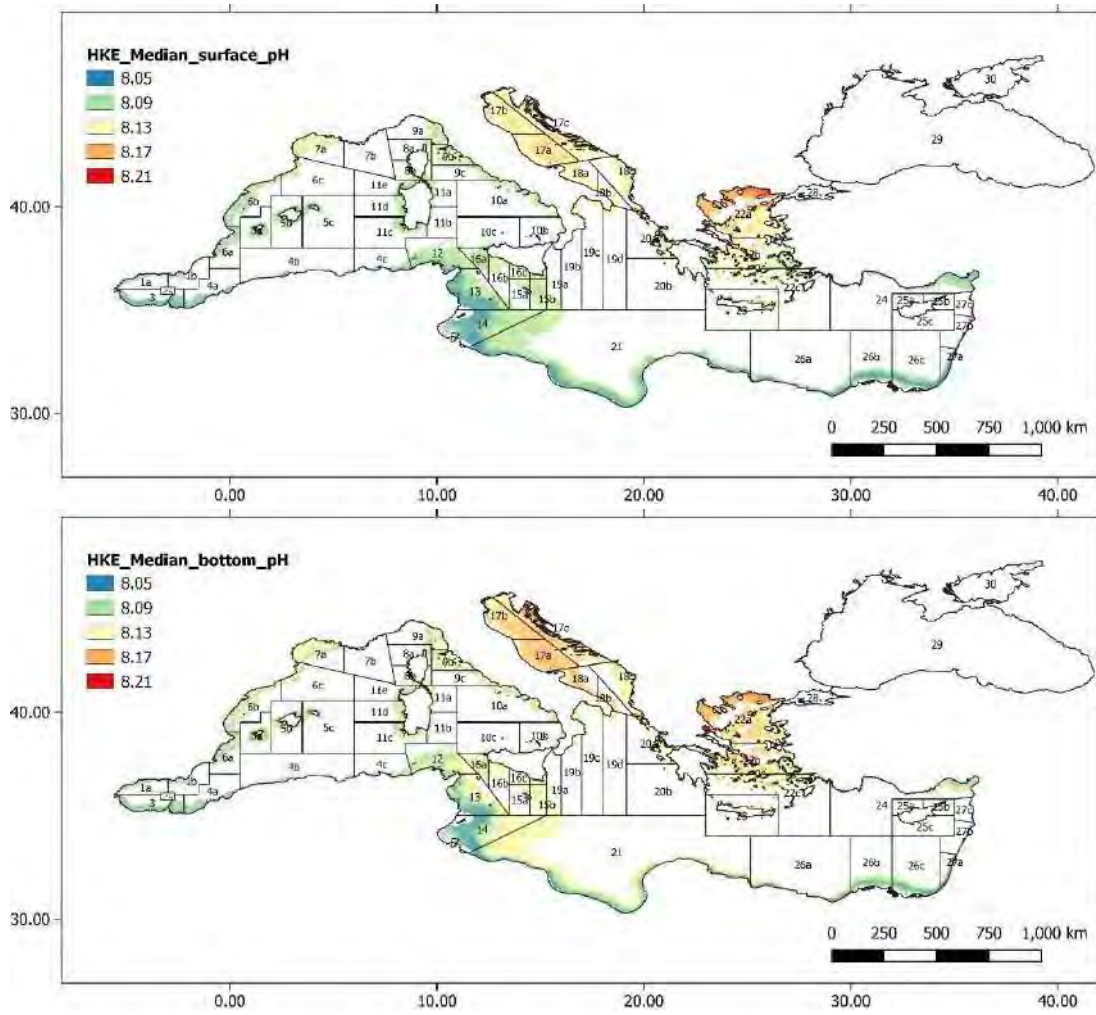
Annex 1 Figure 4: Median euphotic zone (upper panel) and bottom layer (lower panel) ammonium concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



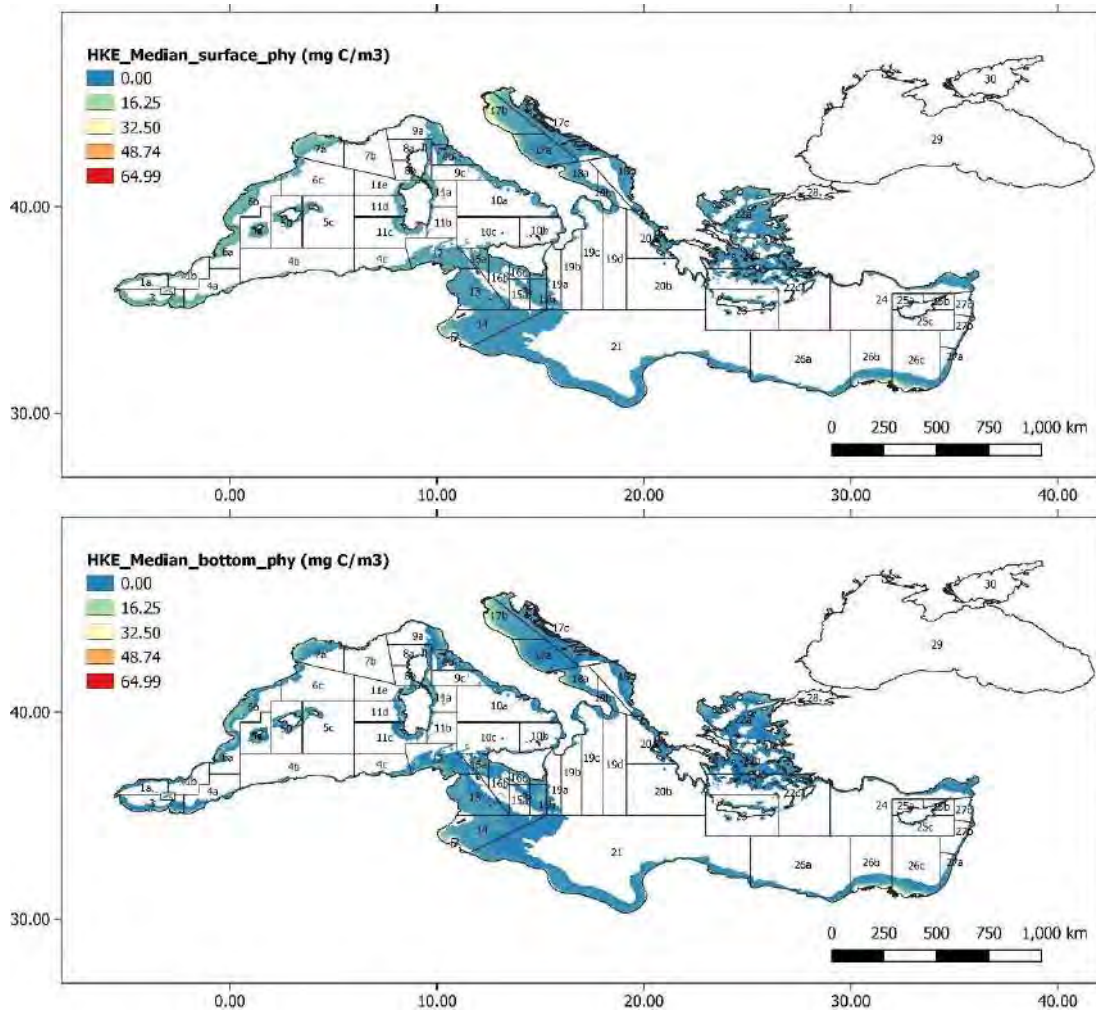
Annex 1 Figure 5: Median euphotic zone (upper panel) and bottom layer (lower panel) nitrates concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



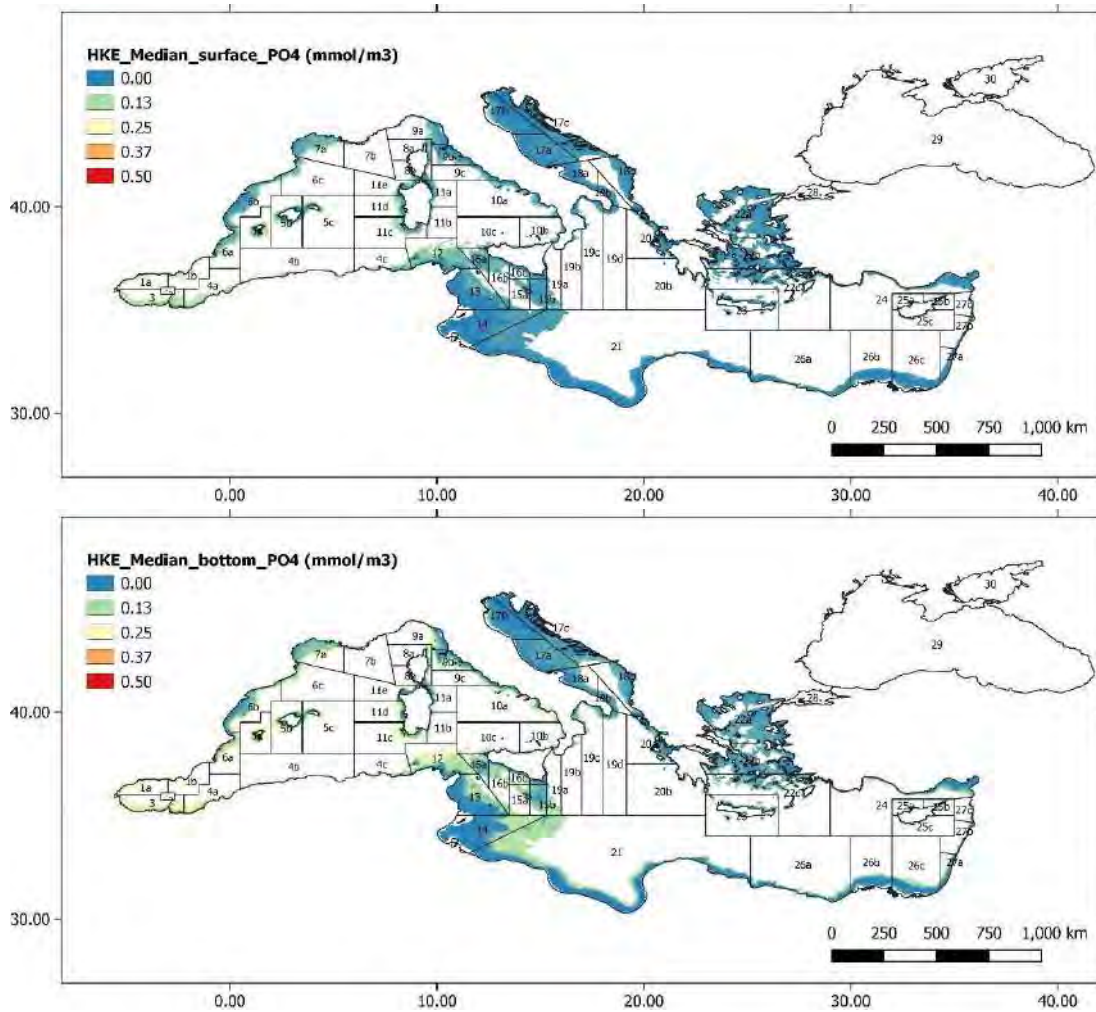
Annex 1 Figure 6: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved oxygen concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



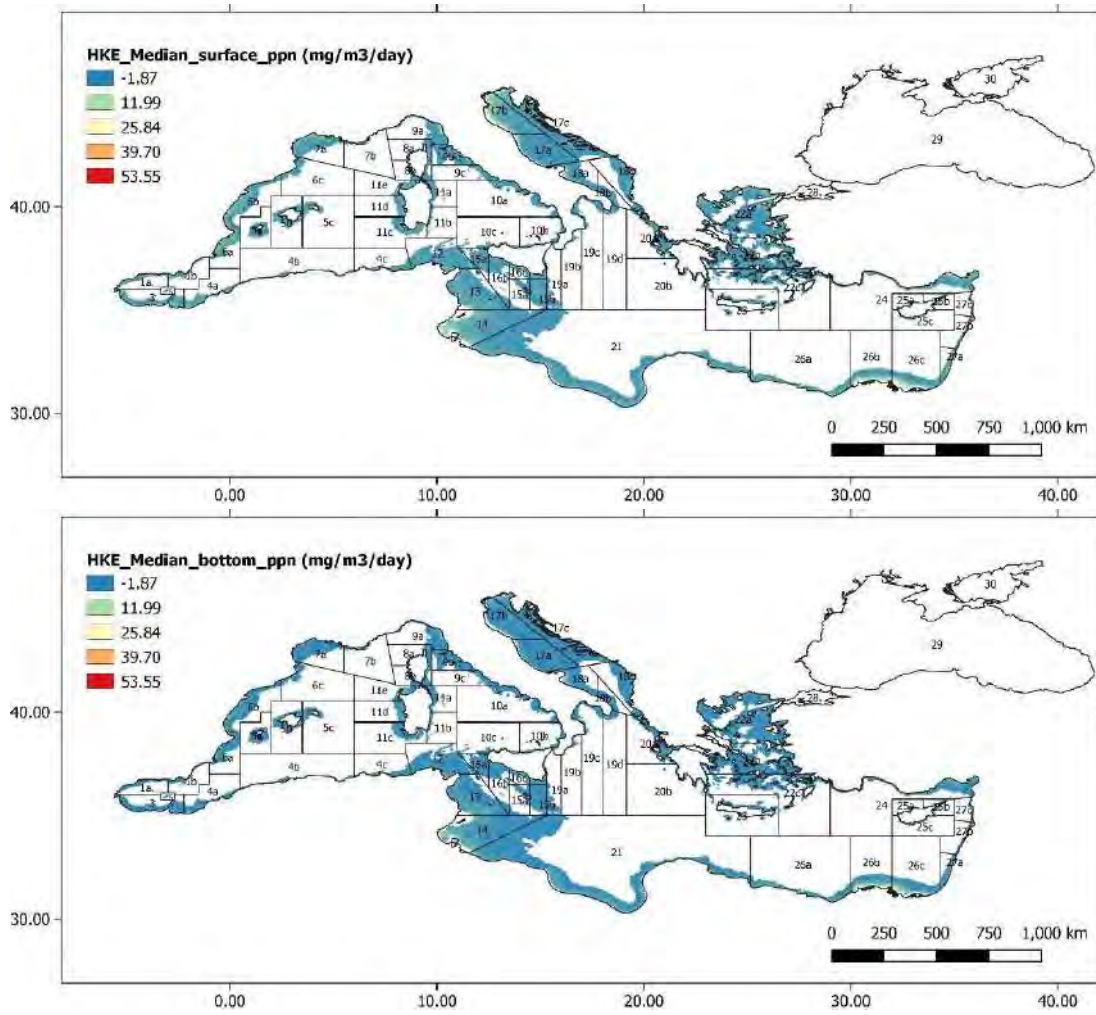
Annex 1 Figure 7: Median euphotic zone (upper panel) and bottom layer (lower panel) pH values over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



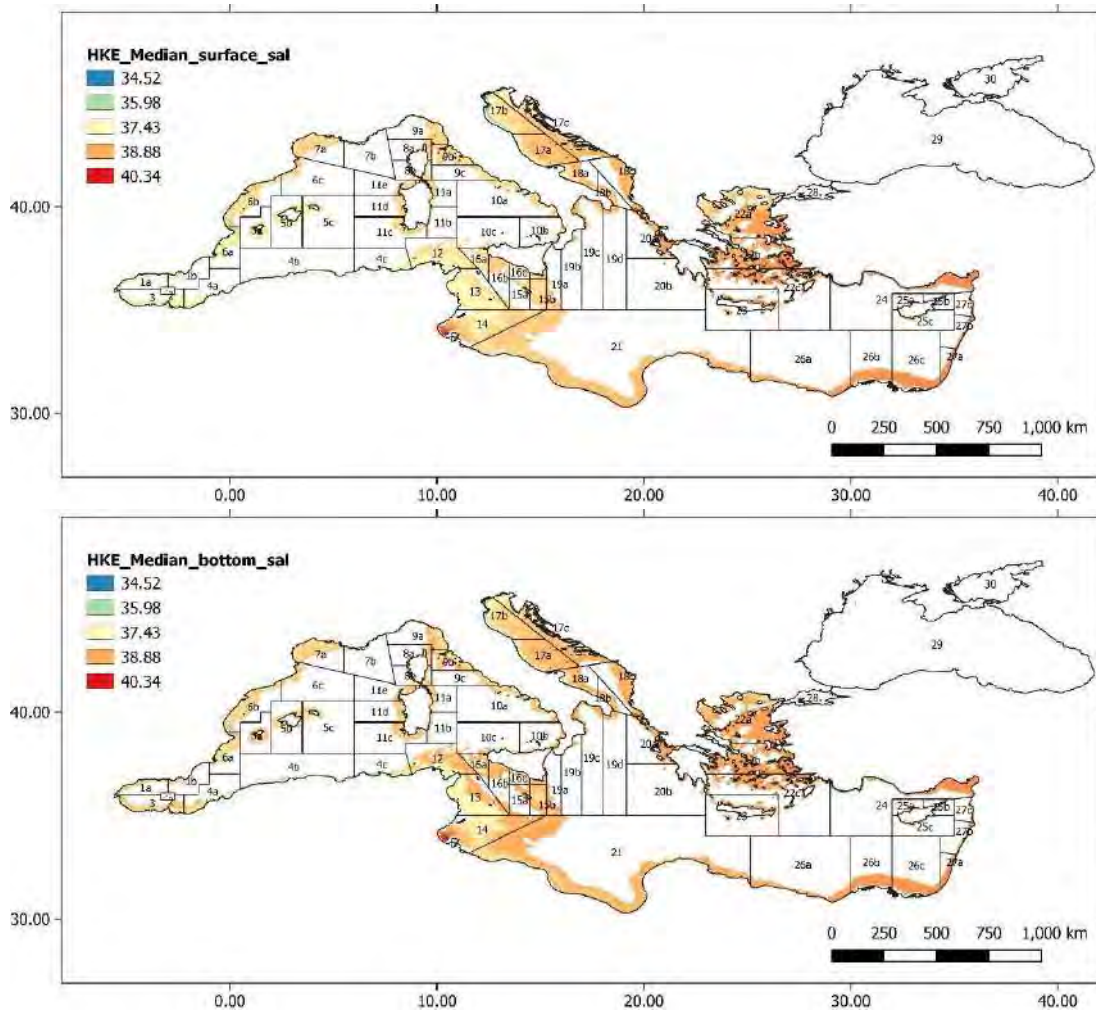
Annex 1 Figure 8: Median euphotic zone (upper panel) and bottom layer (lower panel) phytoplankton biomass concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



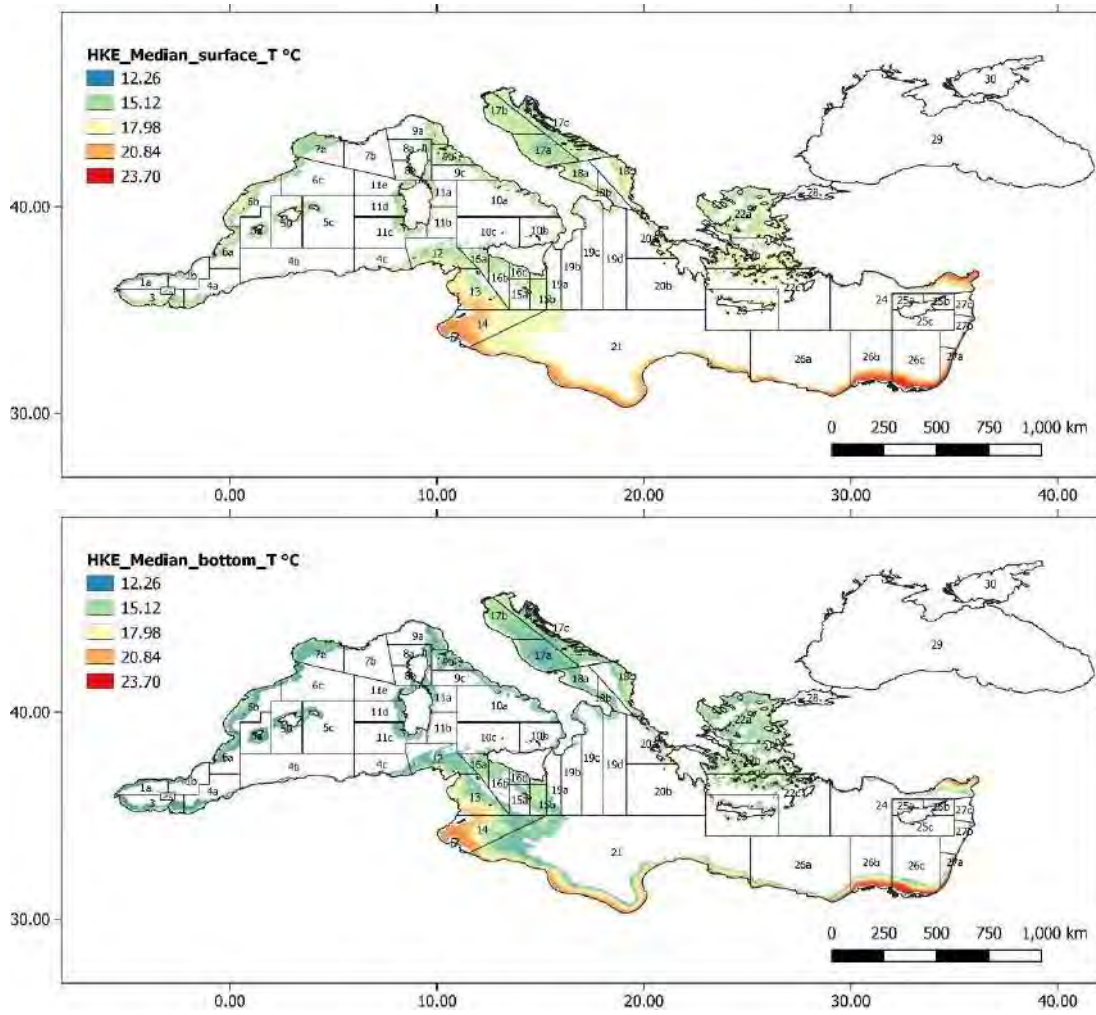
Annex 1 Figure 9: Median euphotic zone (upper panel) and bottom layer (lower panel) phosphates concentration over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



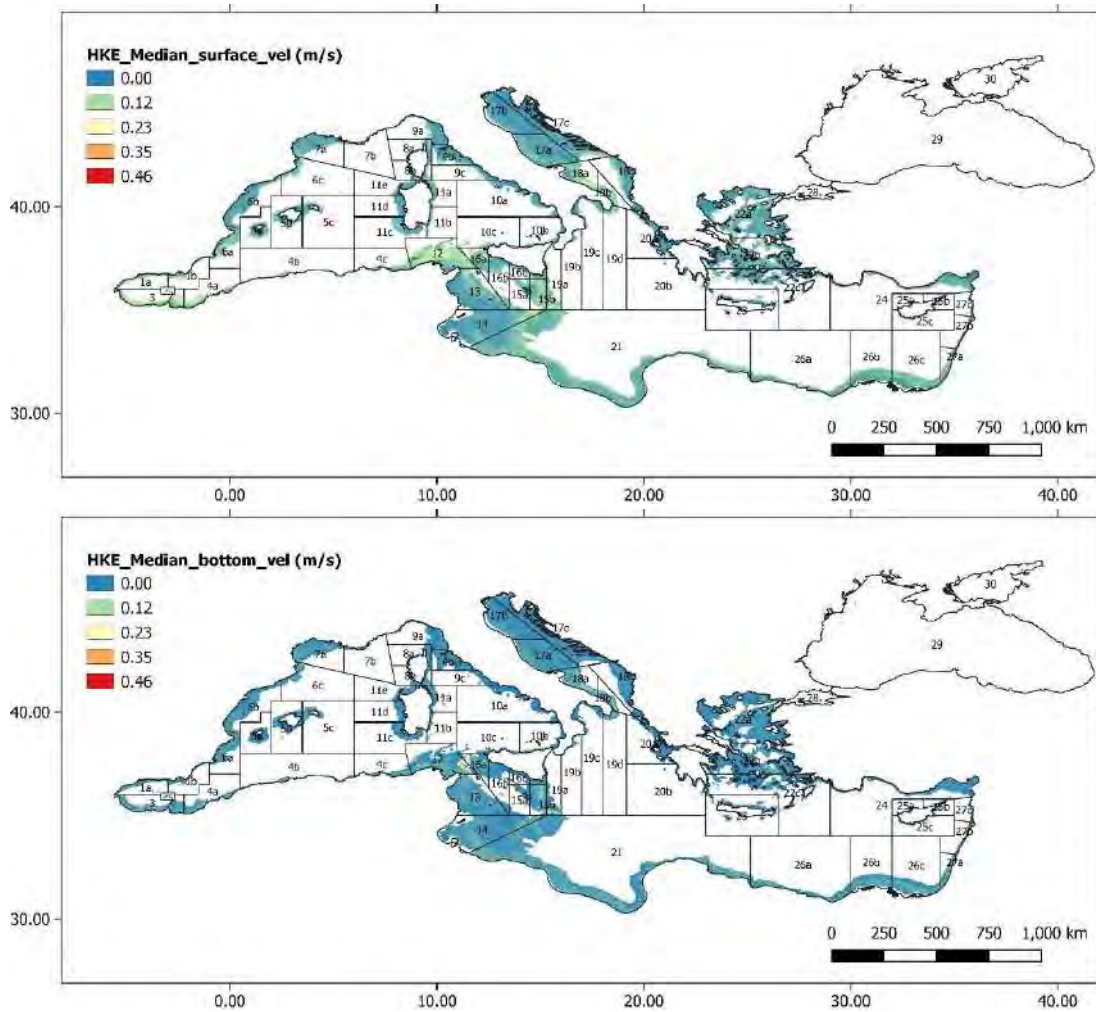
Annex 1 Figure 10: Median euphotic zone (upper panel) and bottom layer (lower panel) primary production over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.



Annex 1 Figure 11: Median euphotic zone (upper panel) and bottom layer (lower panel) salinity over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.

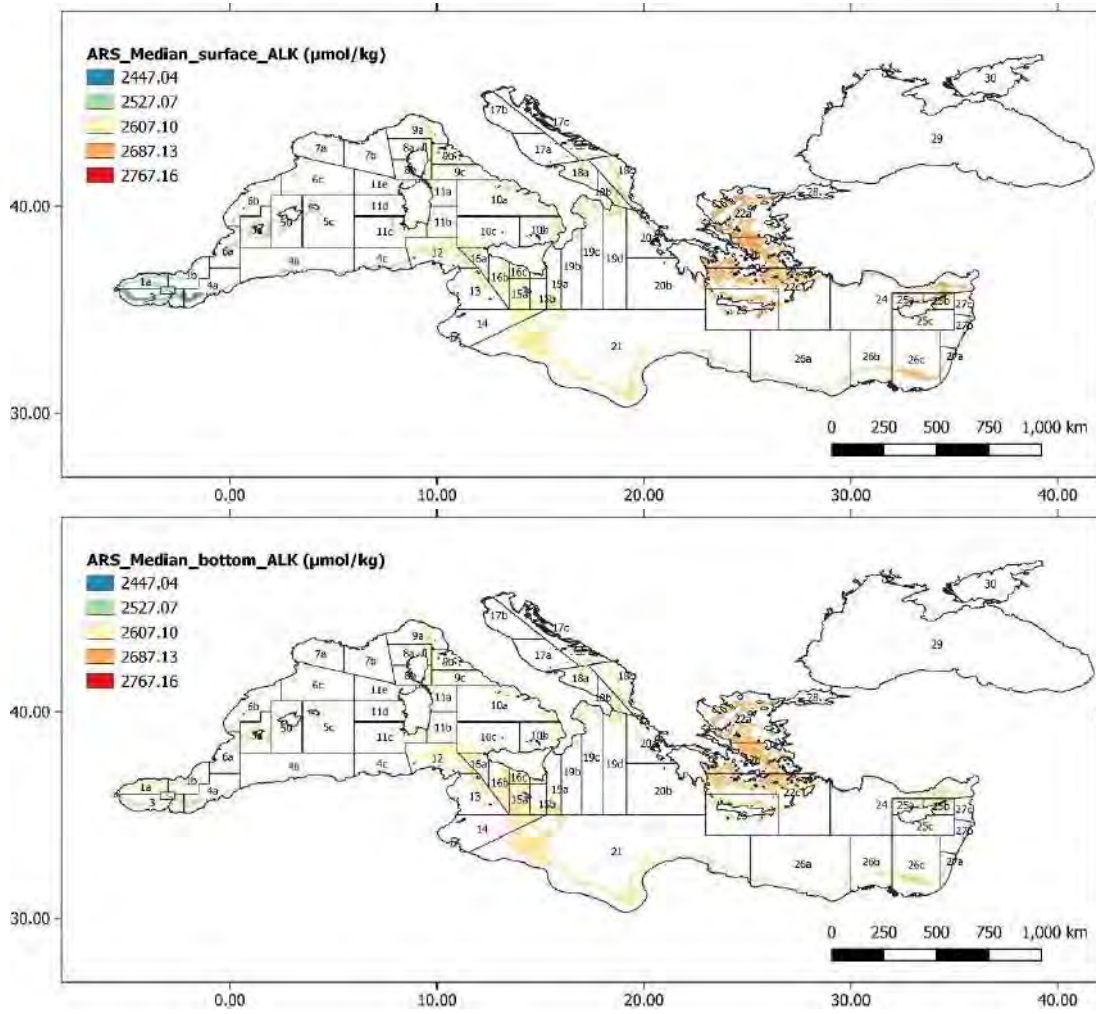


Annex 1 Figure 12: Median euphotic zone (upper panel) and bottom layer (lower panel) temperature over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.

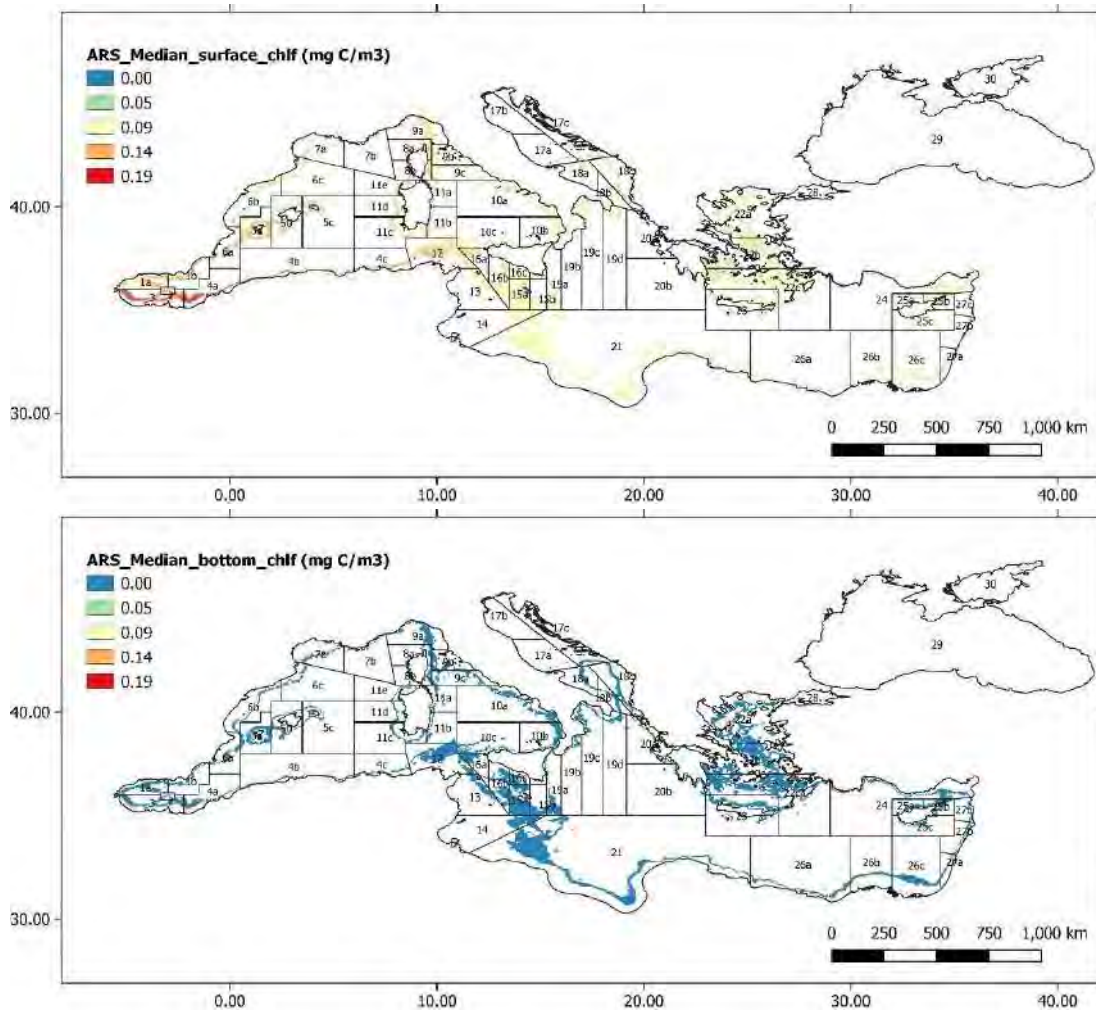


Annex 1 Figure 13: Median euphotic zone (upper panel) and bottom layer (lower panel) current velocity over 5 years in the *Merluccius merluccius* depth range of the Mediterranean.

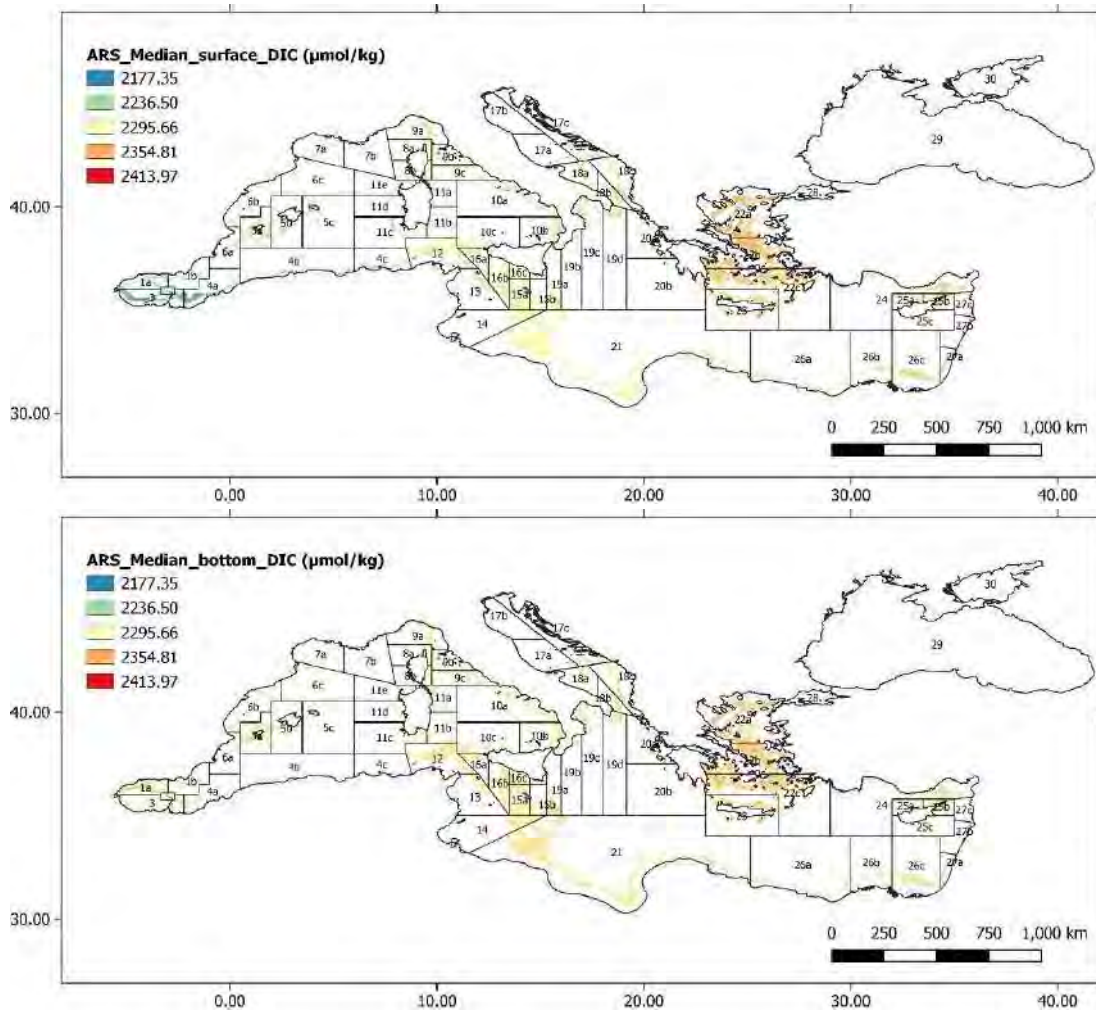
Annex 2



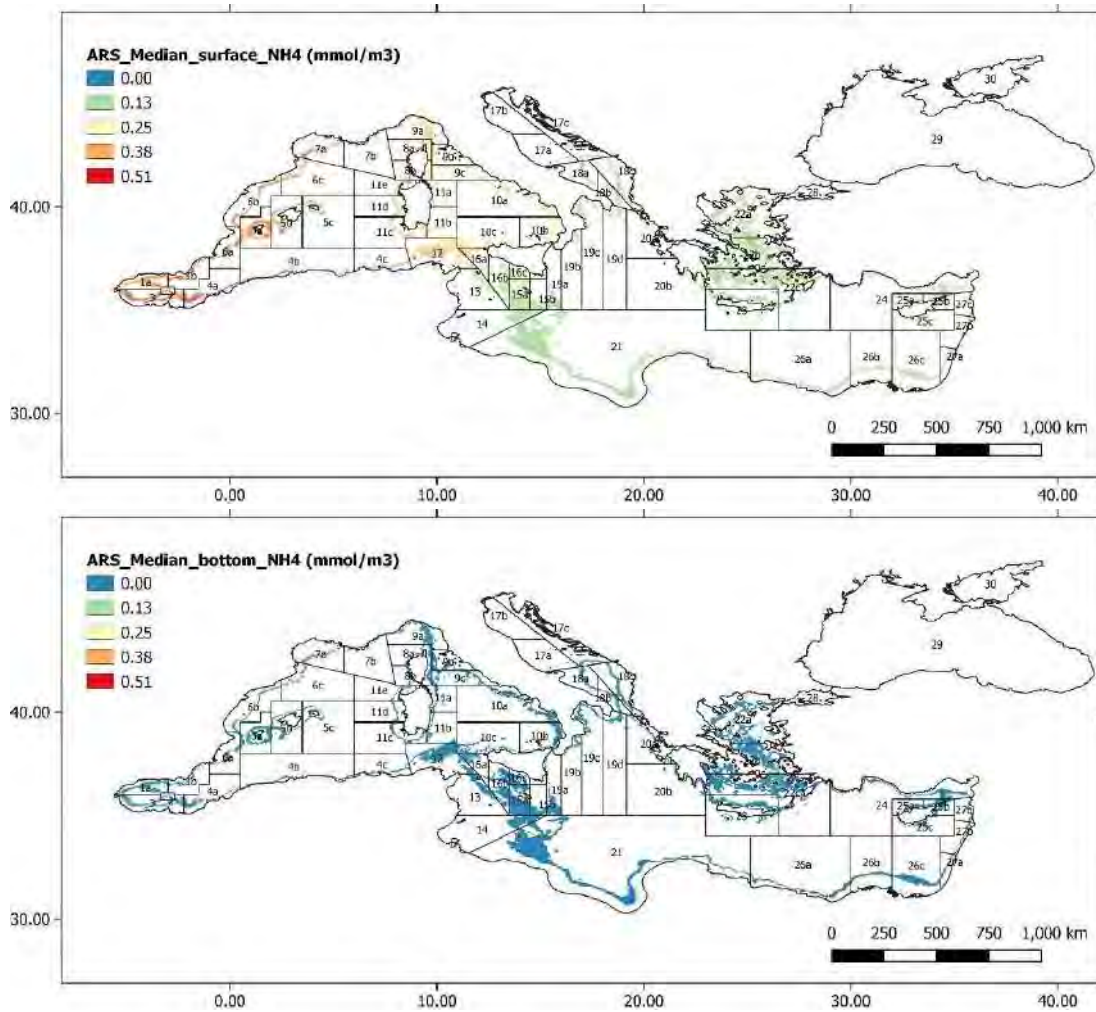
Annex 2 Figure 1: Median euphotic zone (upper panel) and bottom layer (bottom panel) alkalinity over 5 years in the *Aristeomorpha foliacea* depth range of the Mediterranean.



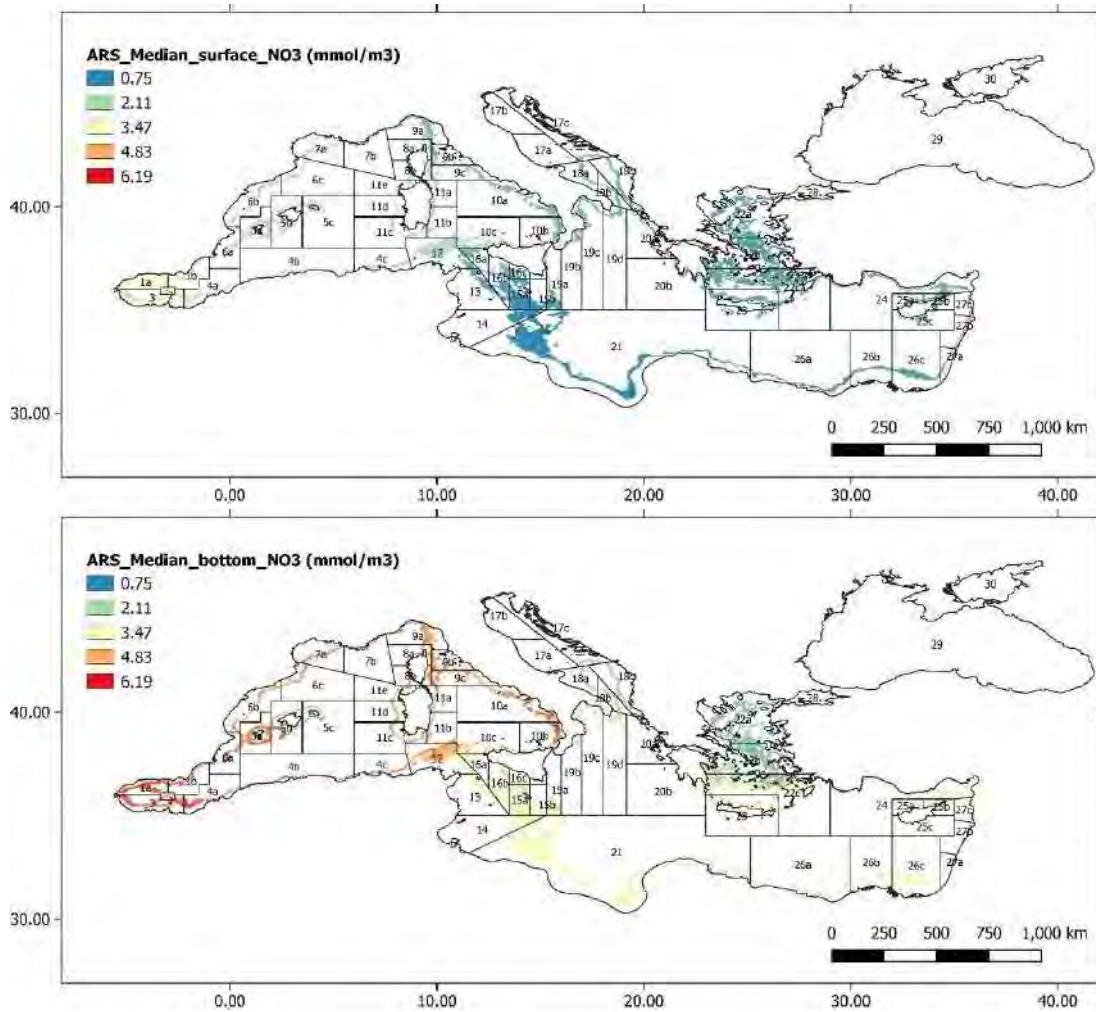
Annex 2 Figure 2: Median euphotic zone (upper panel) and bottom layer (lower panel) chlorophyll a concentration over 5 years in the *Aristeomorpha foliacea* depth range of the Mediterranean.



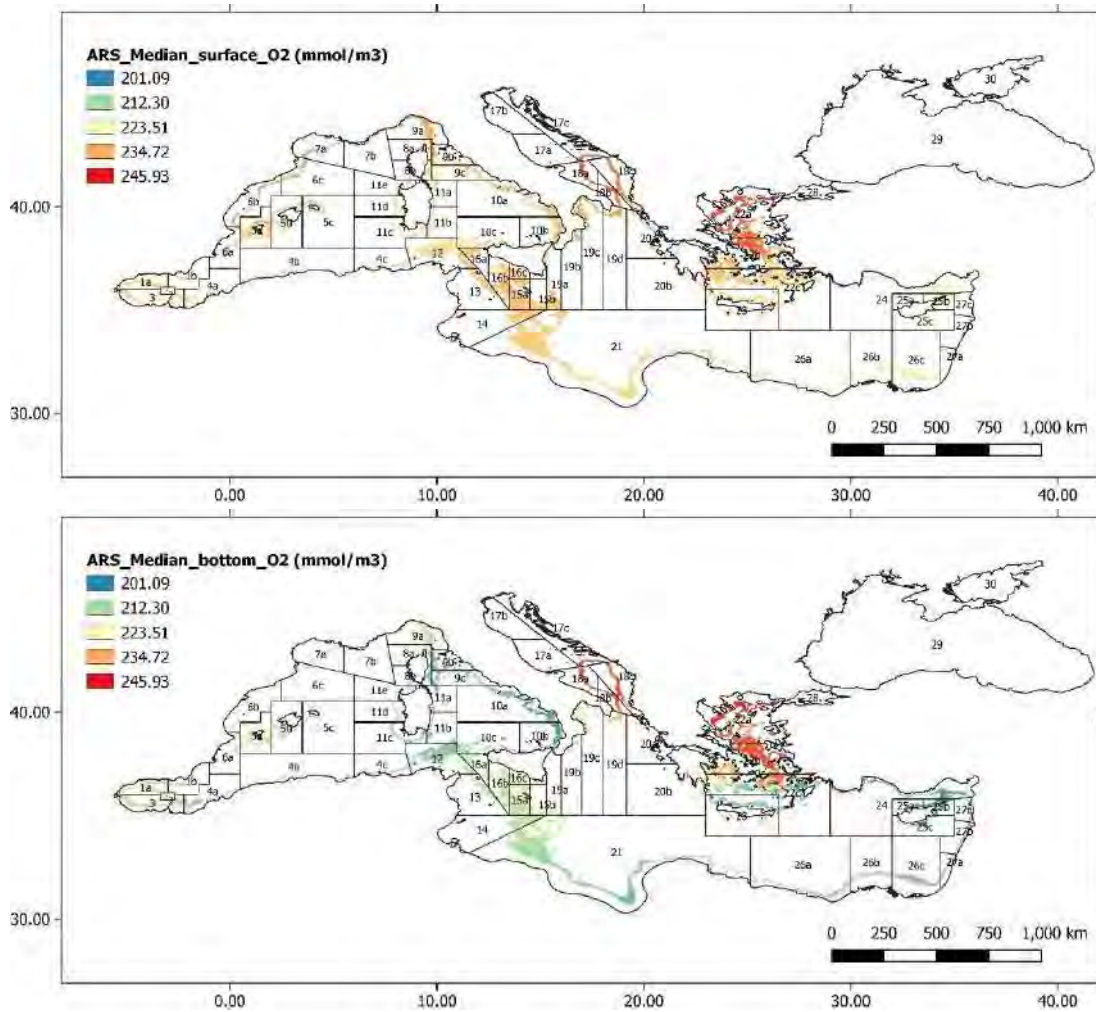
Annex 2 Figure 3: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved inorganic carbon concentration over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.



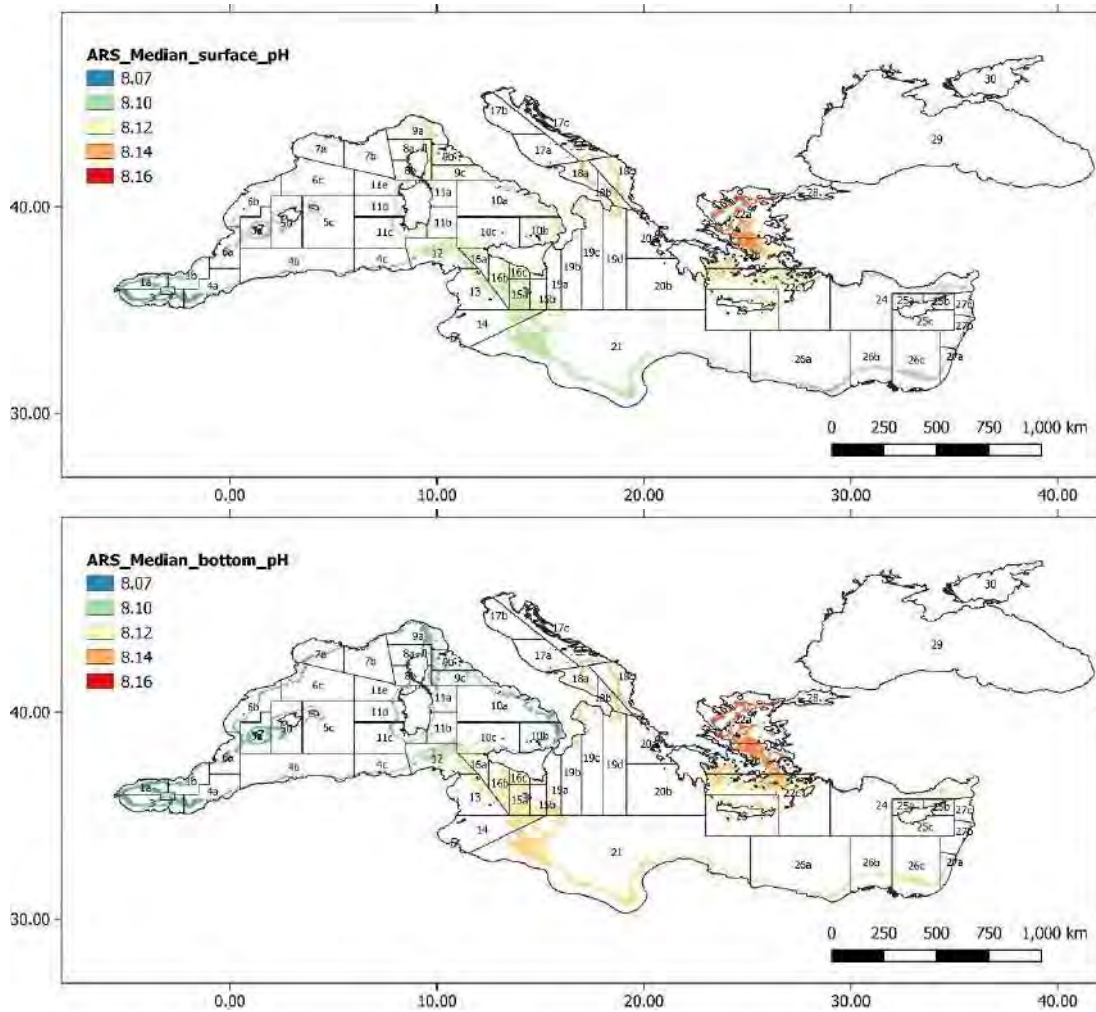
Annex 2 Figure 4: Median euphotic zone (upper panel) and bottom layer (lower panel) ammonium concentration over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.



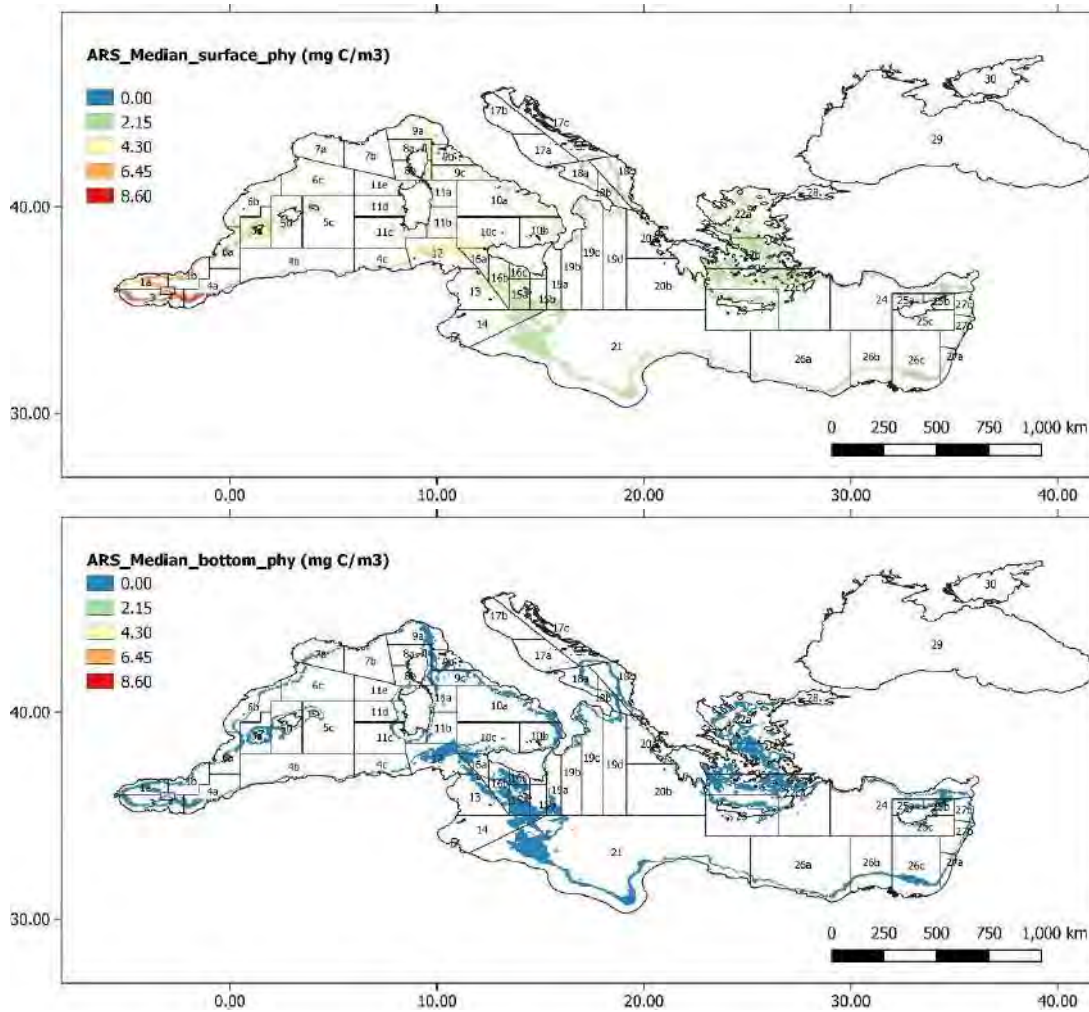
Annex 2 Figure 5: Median euphotic zone (upper panel) and bottom layer (lower panel) nitrates concentration over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.



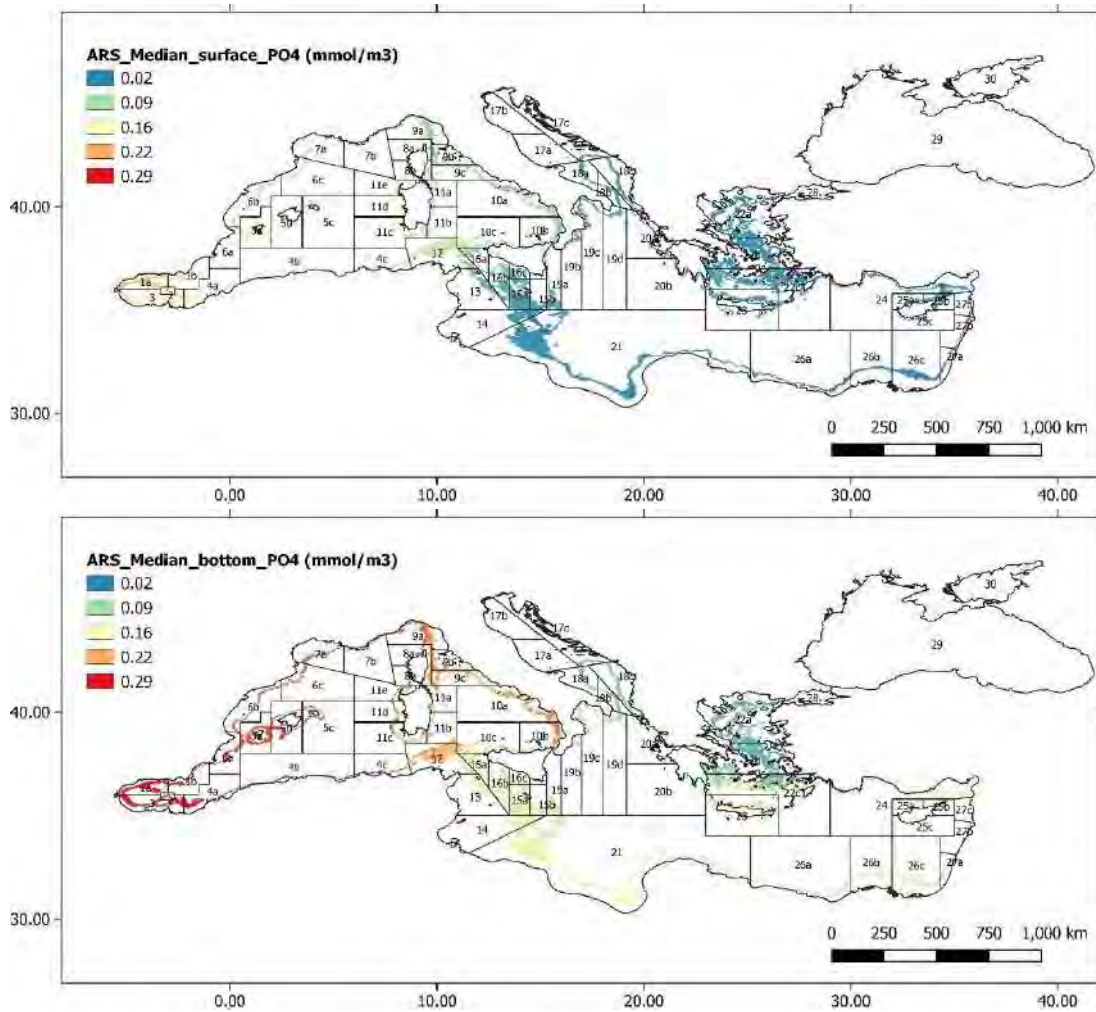
Annex 2 Figure 6: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved oxygen concentration over 5 years in the *Aristeomorpha foliacea* depth range of the Mediterranean.



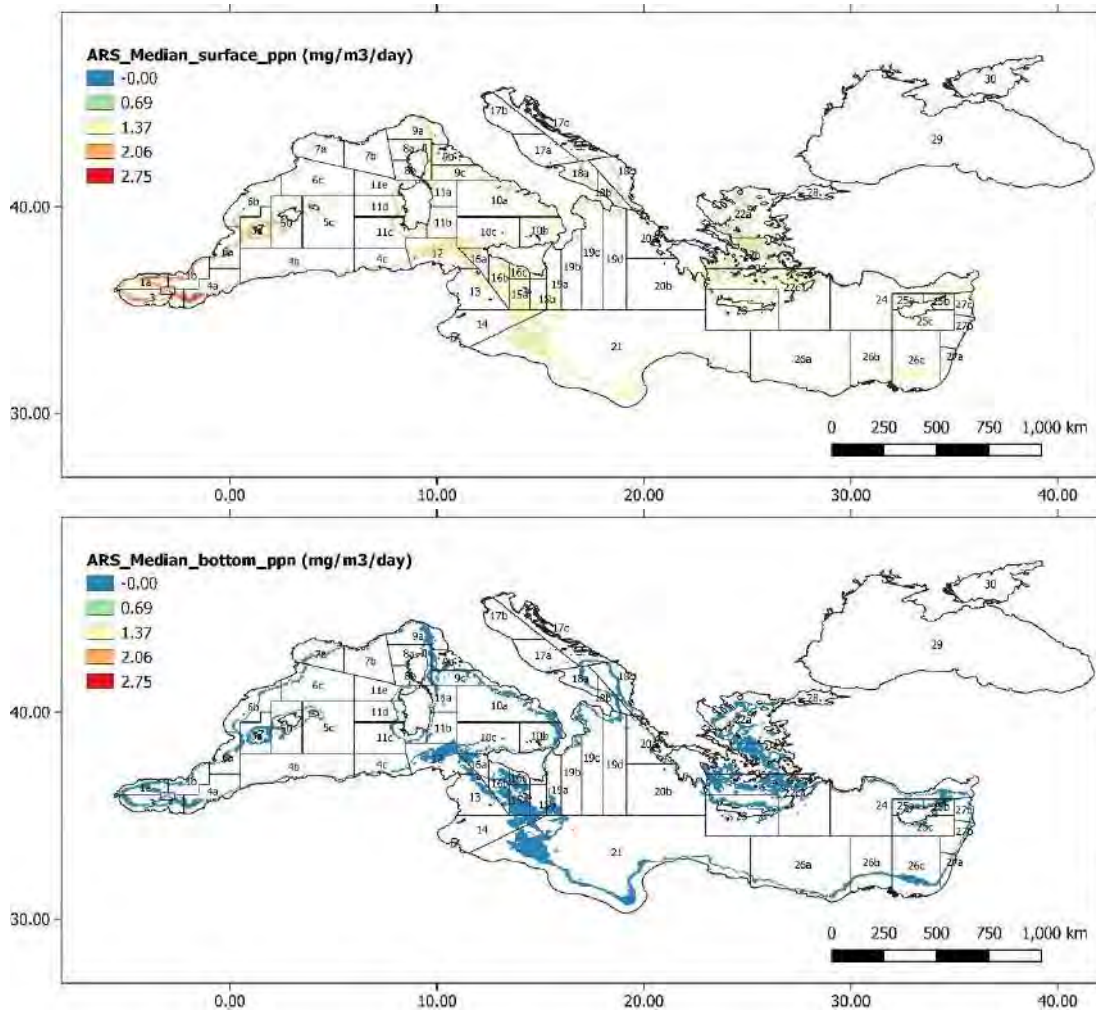
Annex 2 Figure 7: Median euphotic zone (upper panel) and bottom layer (lower panel) pH values over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.



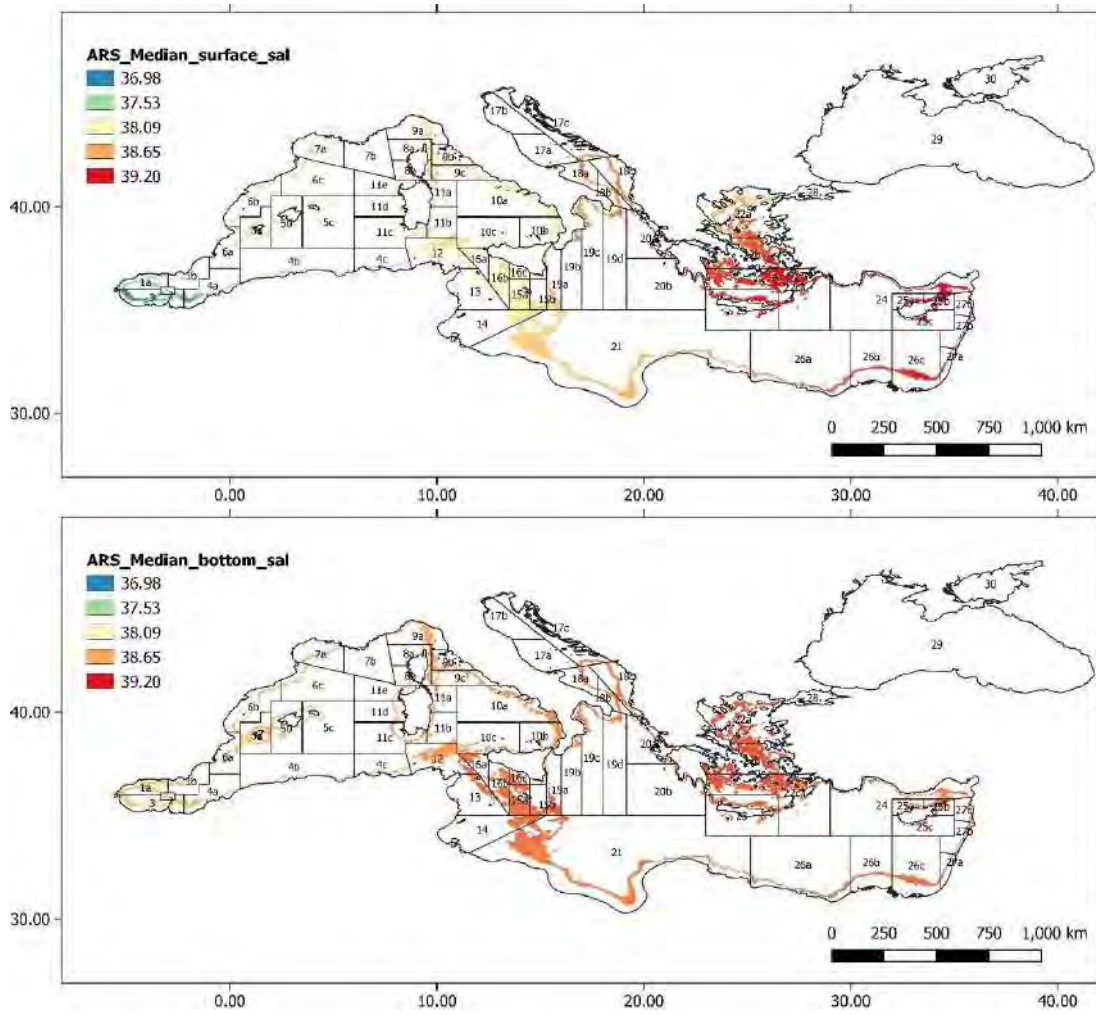
Annex 2 Figure 8: Figure 14: Median euphotic zone (upper panel) and bottom layer (lower panel) phytoplankton biomass concentration over 5 years in the *Aristemomorpha foliacea* depth range of the Mediterranean.



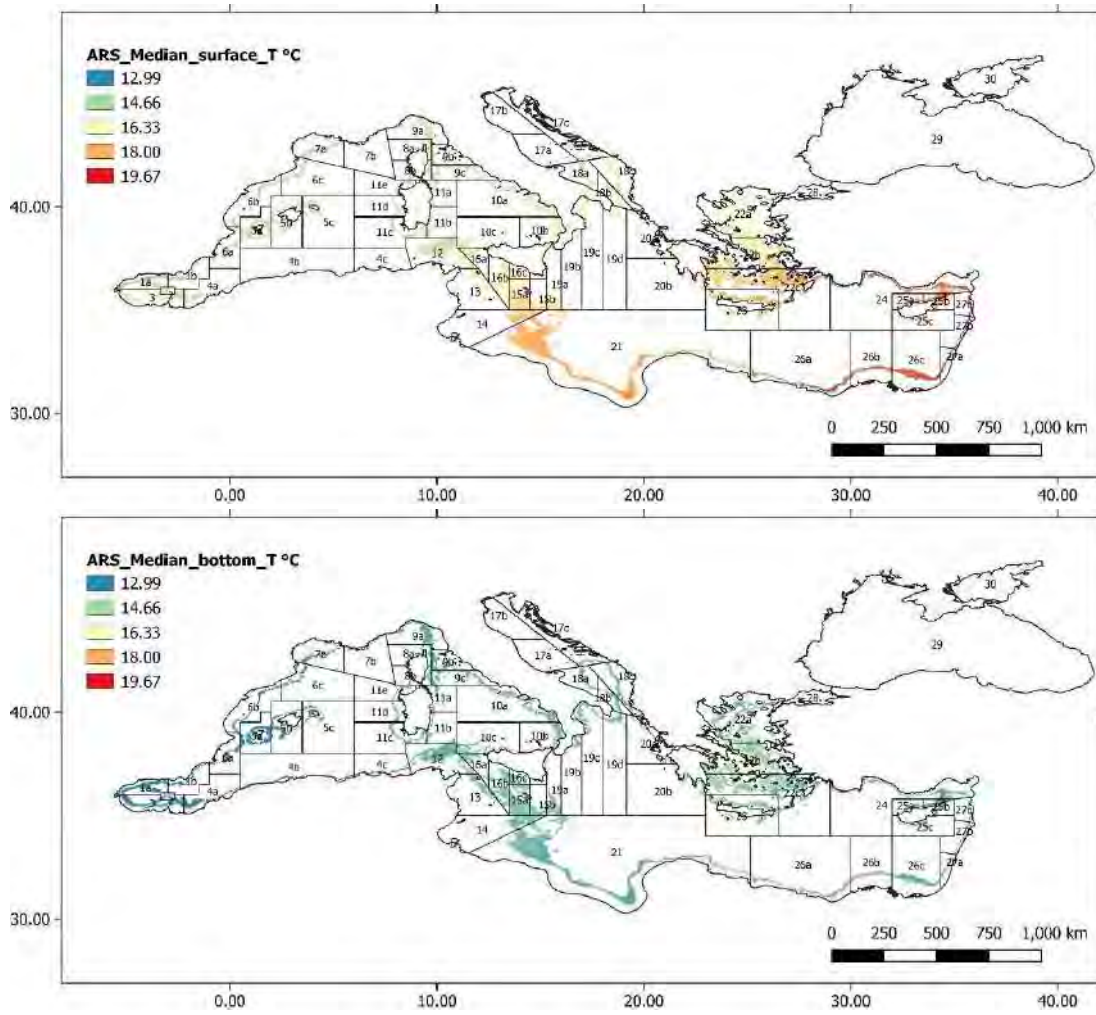
Annex 2 Figure 9: Median euphotic zone (upper panel) and bottom layer (lower panel) phosphates concentration over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.



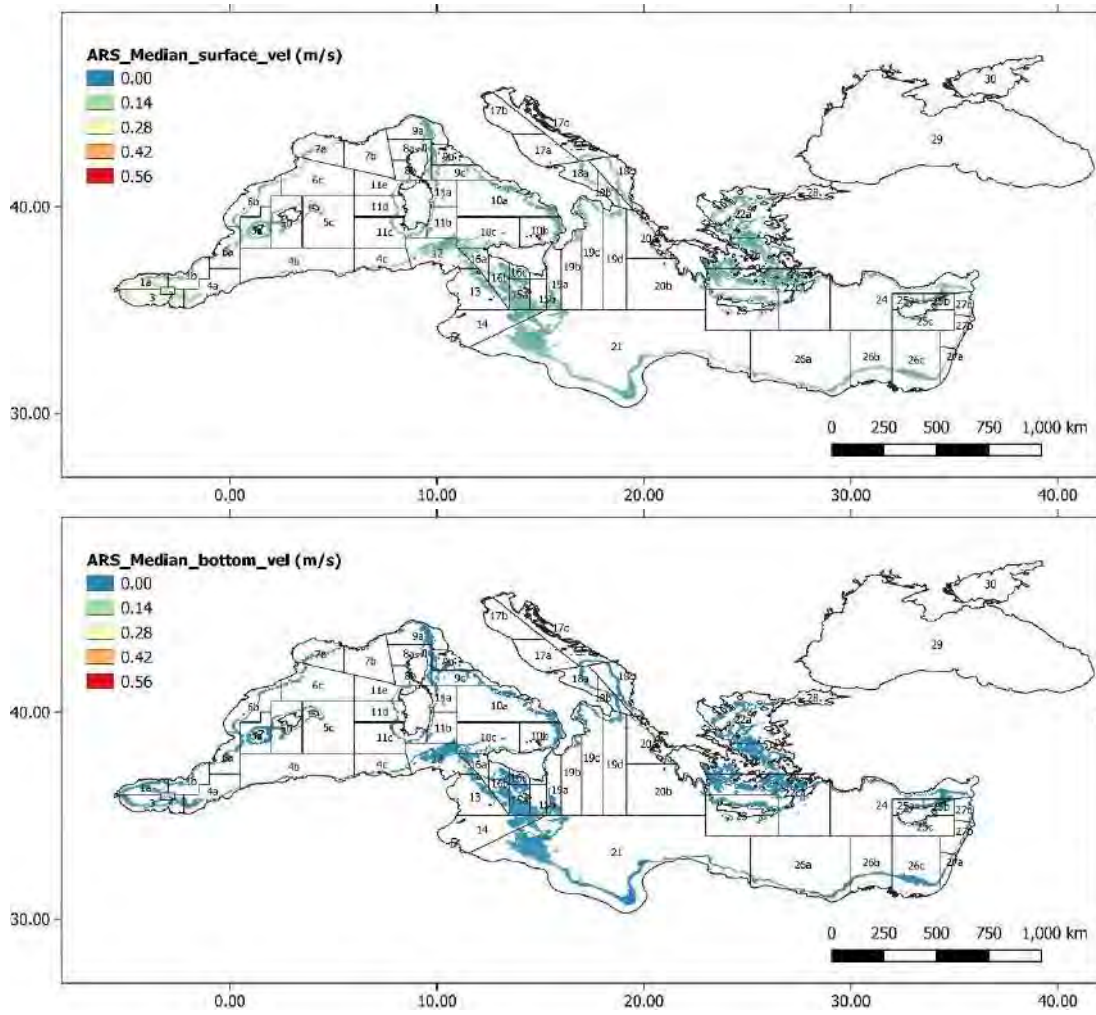
Annex 2 Figure 10: Median euphotic zone (upper panel) and bottom layer (lower panel) primary production over 5 years in the *Aristeomorpha foliacea* depth range of the Mediterranean.



Annex 2 Figure 11: Median euphotic zone (upper panel) and bottom layer (lower panel) salinity over 5 years in the *Aristeomorpha foliacea* depth range of the Mediterranean.

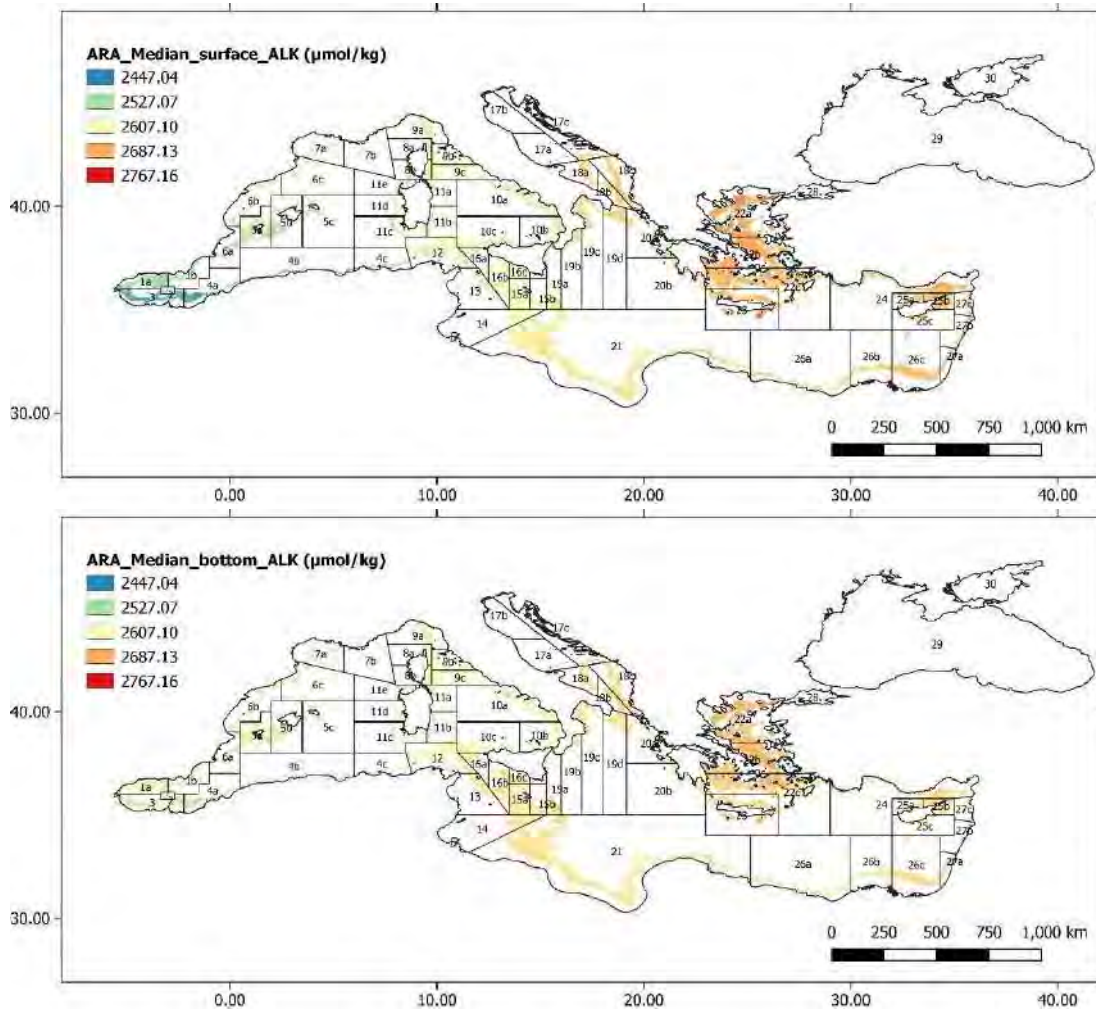


Annex 2 Figure 12: Median euphotic zone (upper panel) and bottom layer (lower panel) temperature over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.

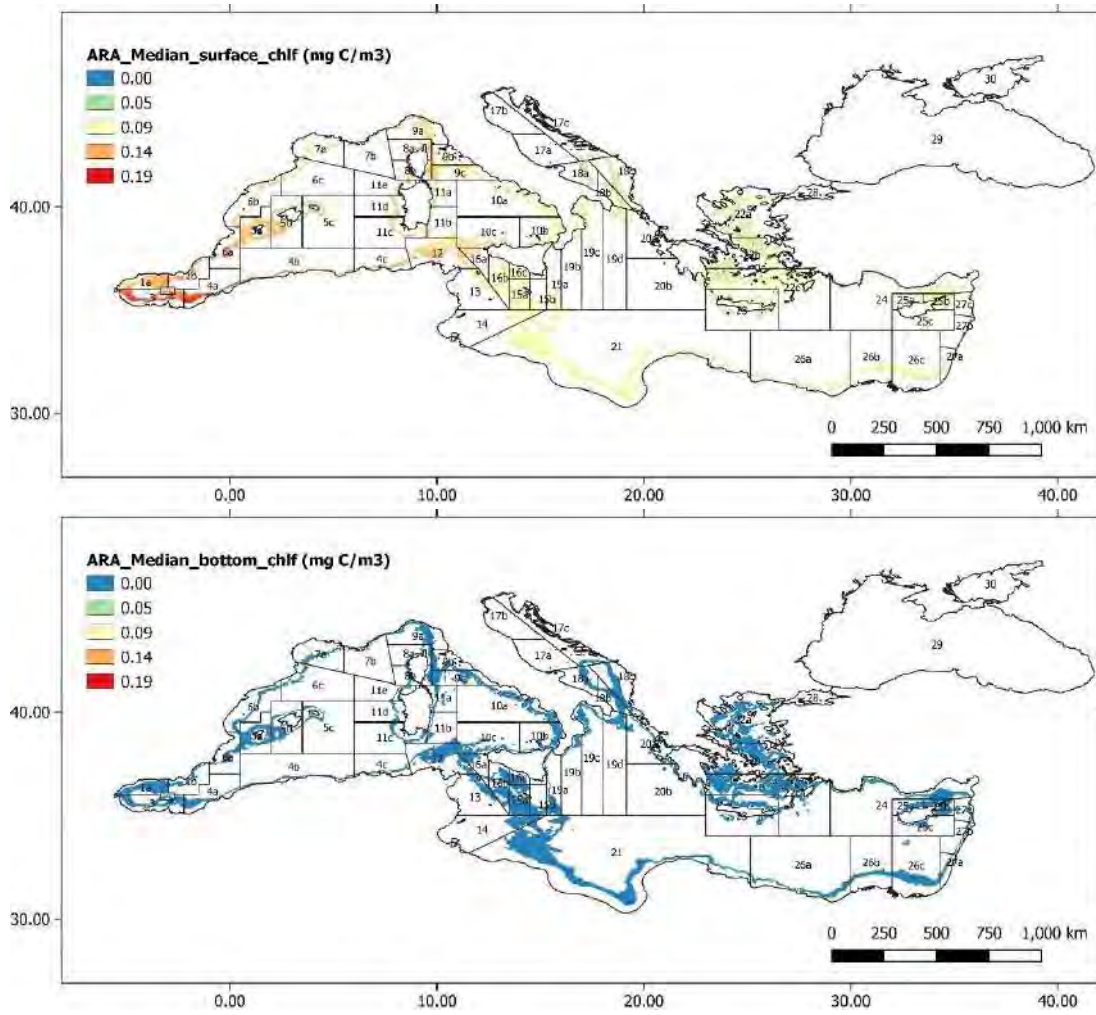


Annex 2 Figure 13: Median euphotic zone (upper panel) and bottom layer (lower panel) current velocity over 5 years in the *Aristomorpha foliacea* depth range of the Mediterranean.

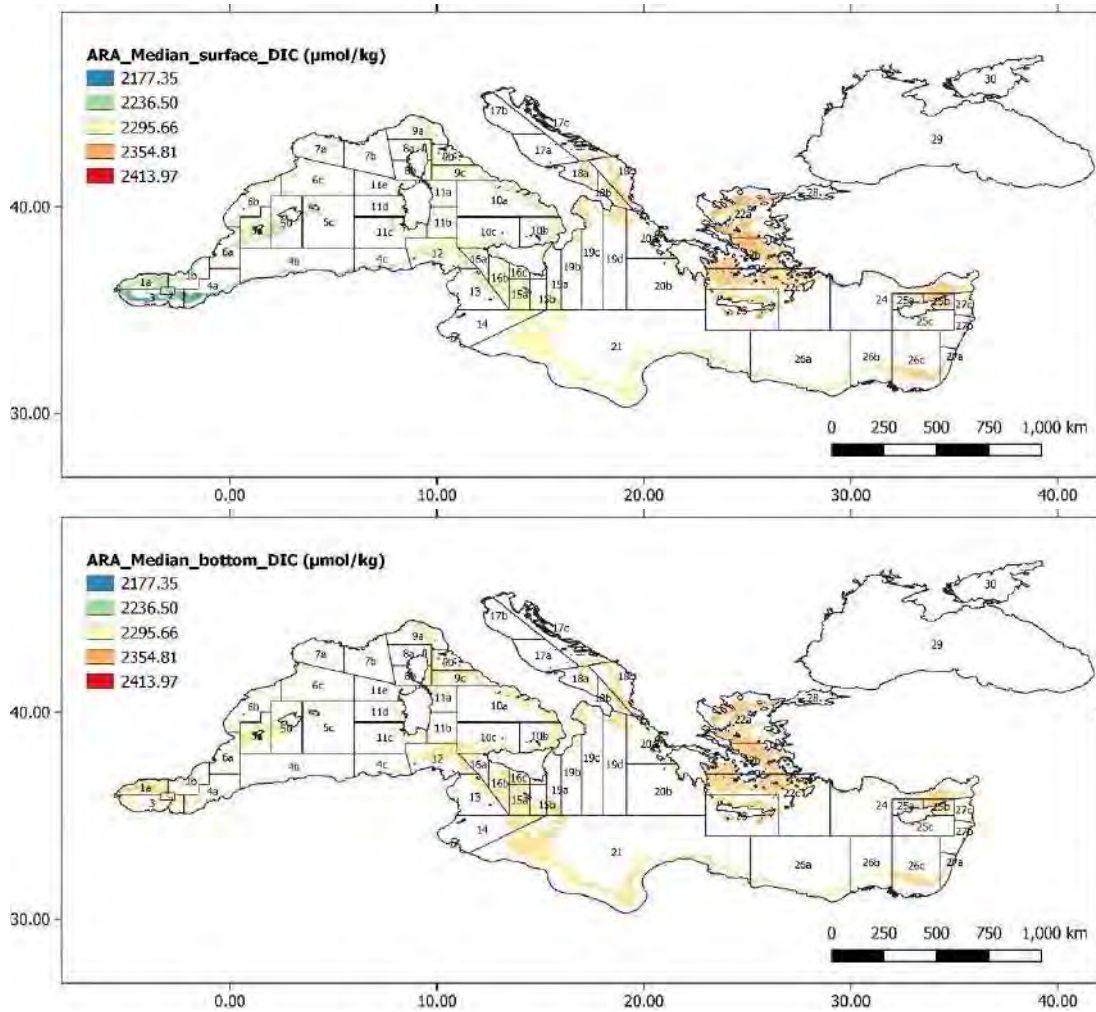
Annex 3



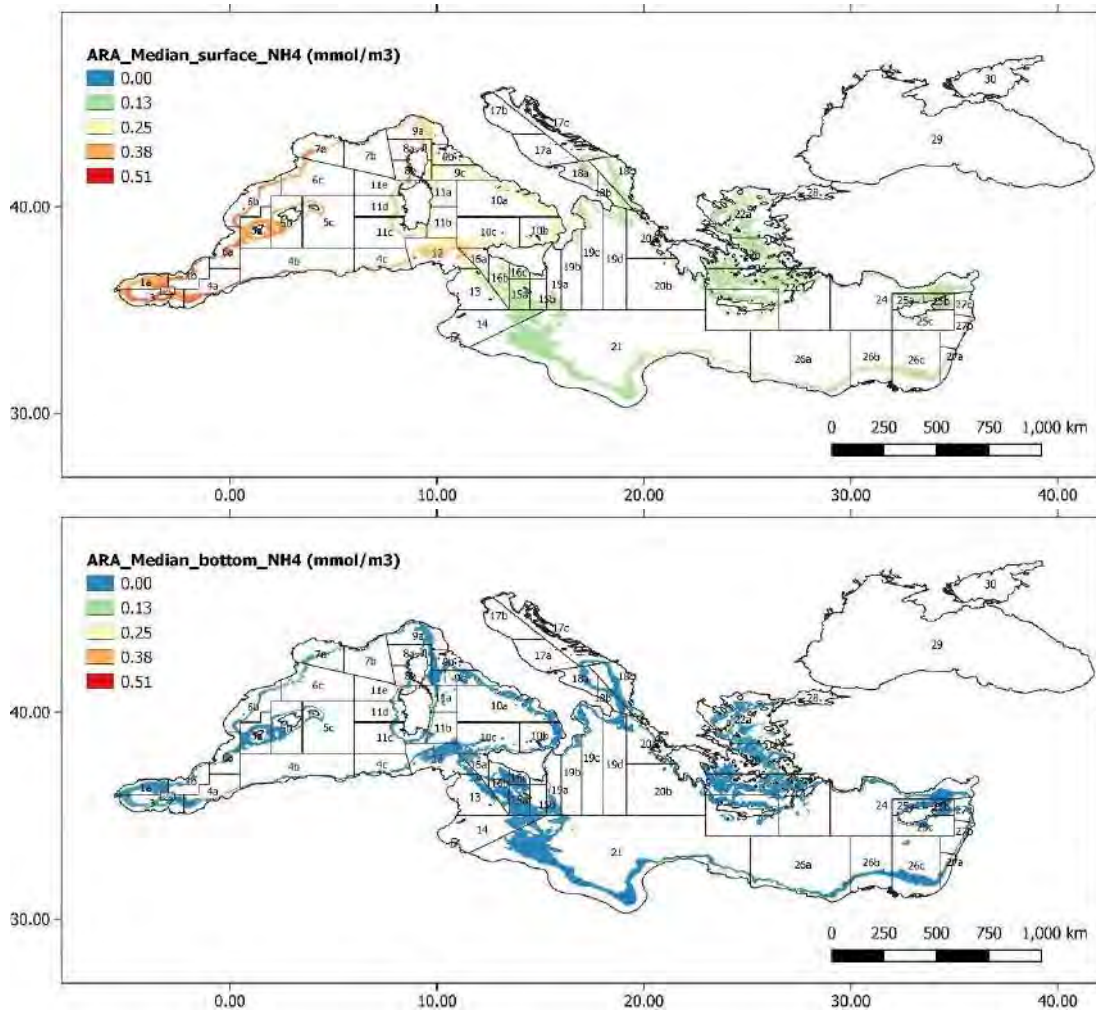
Annex 3 Figure 1: Median euphotic zone (upper panel) and bottom layer (bottom panel) alkalinity over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



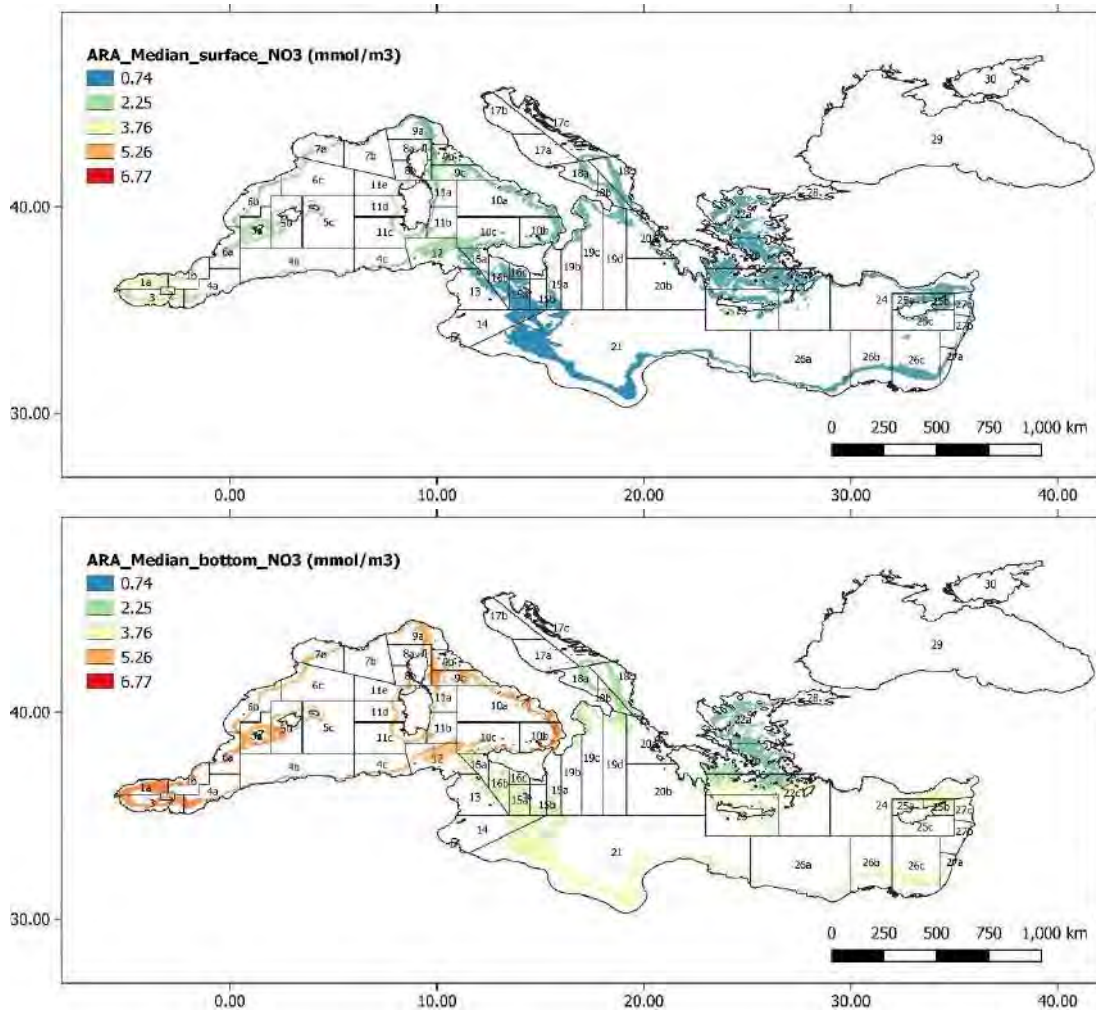
Annex 3 Figure 2: Median euphotic zone (upper panel) and bottom layer (lower panel) chlorophyll a concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



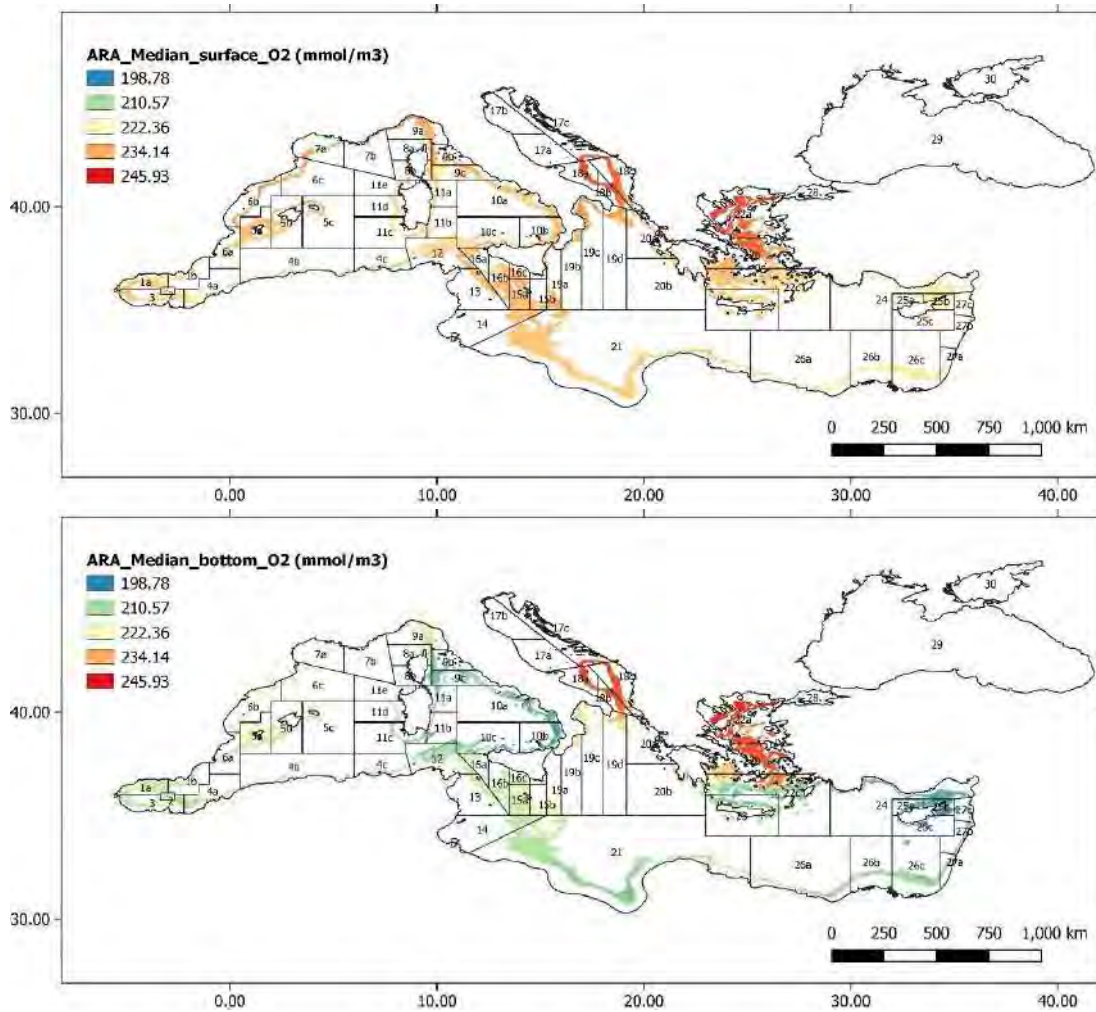
Annex 3 Figure 3: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved inorganic carbon concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



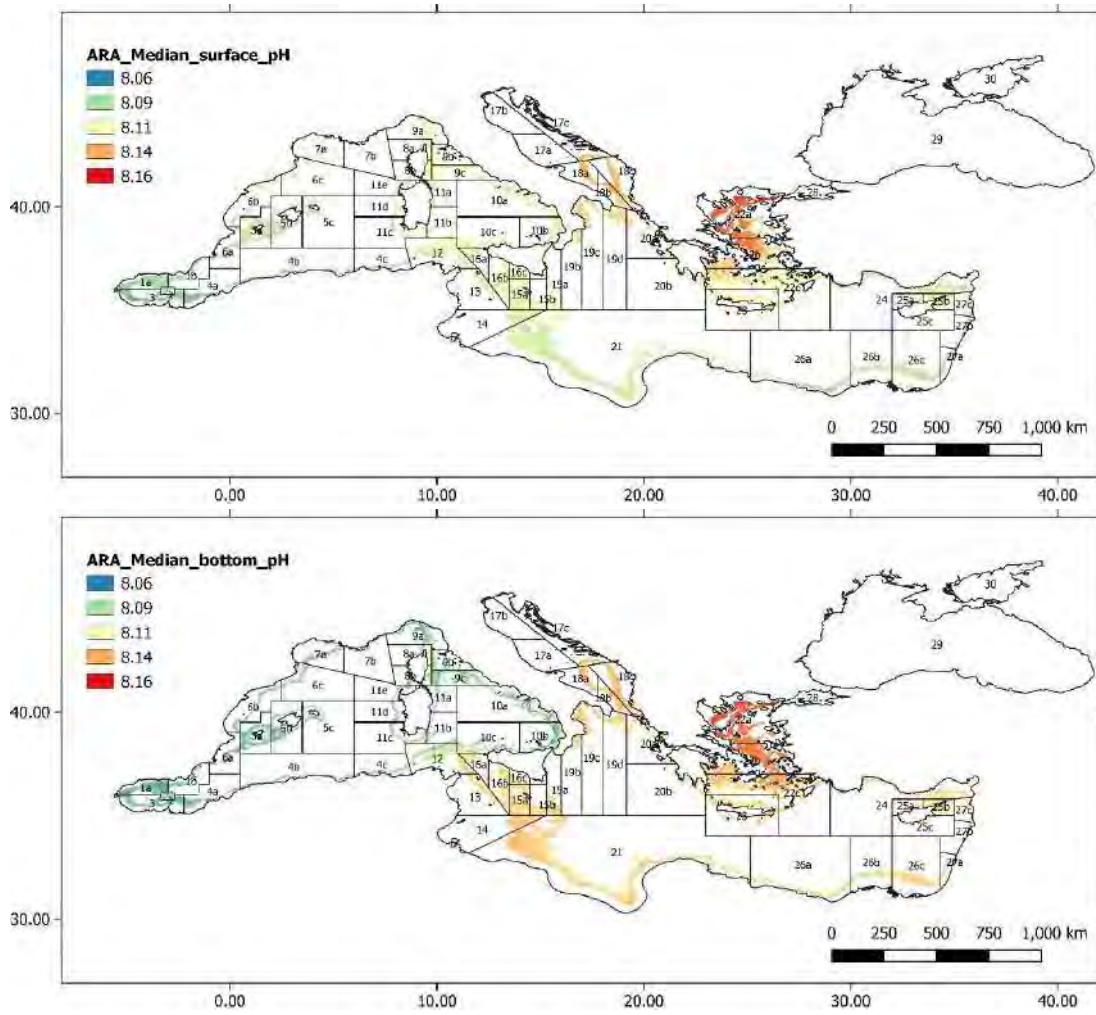
Annex 3 Figure 4: Median euphotic zone (upper panel) and bottom layer (lower panel) ammonium concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



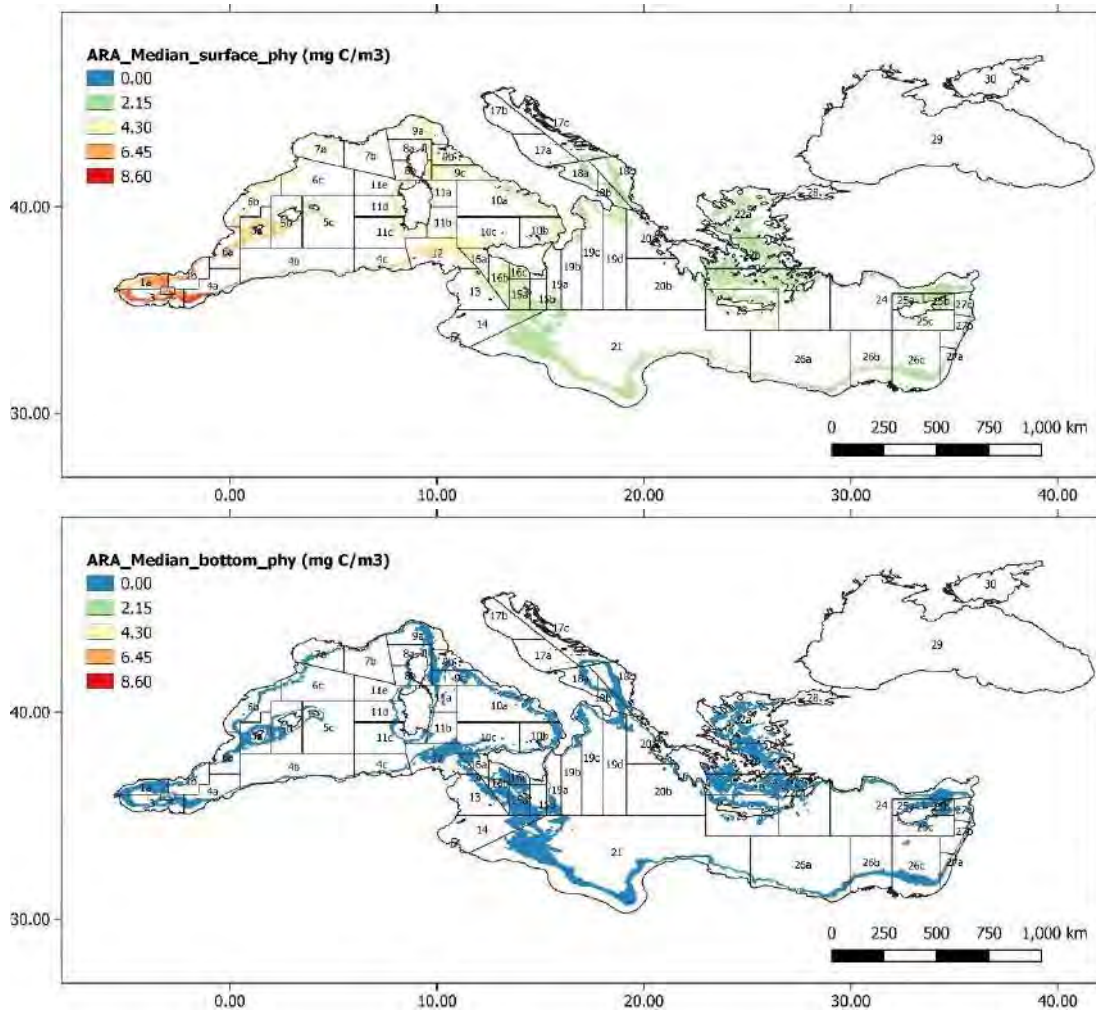
Annex 3 Figure 5: Median euphotic zone (upper panel) and bottom layer (lower panel) nitrates concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



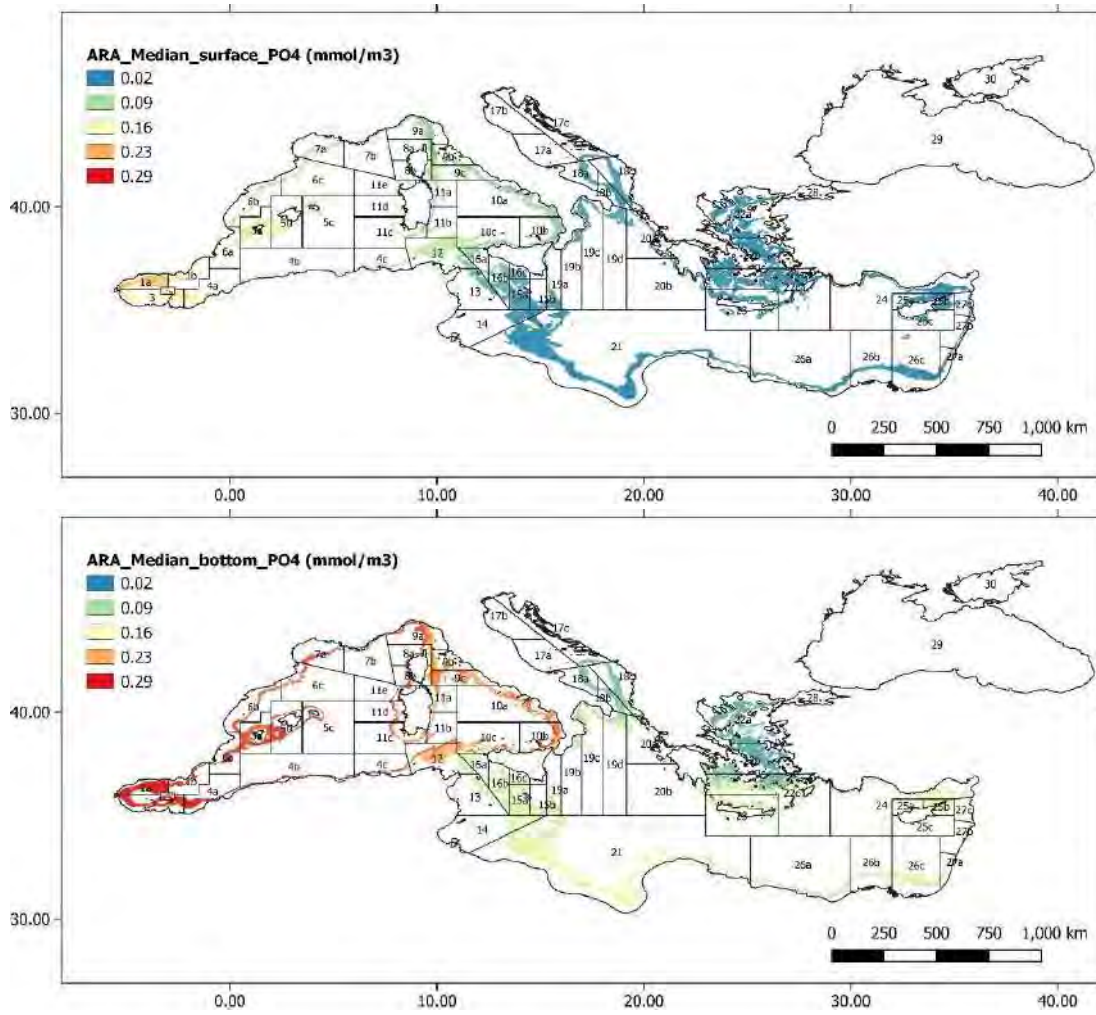
Annex 3 Figure 6: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved oxygen concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



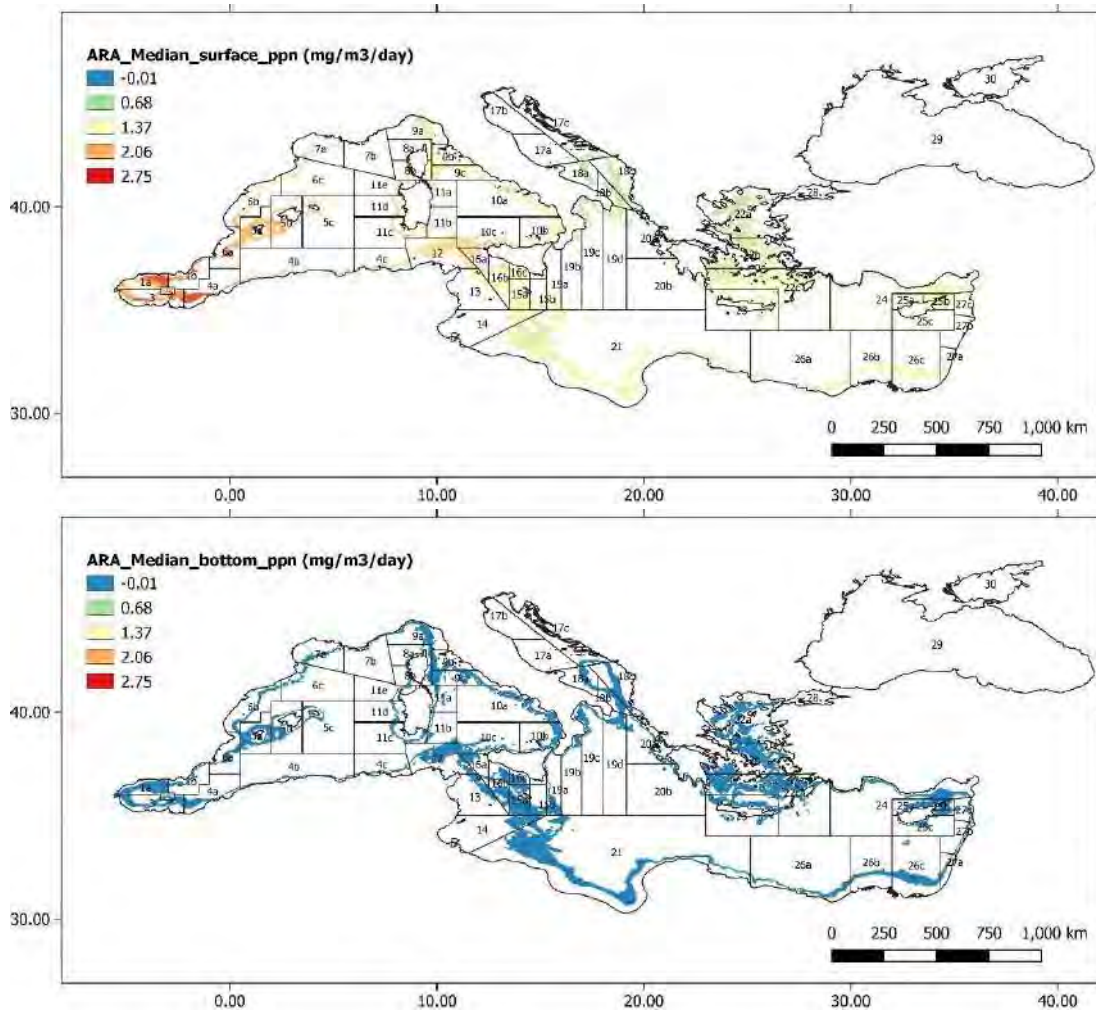
Annex 3 Figure 7: Median euphotic zone (upper panel) and bottom layer (lower panel) pH values over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



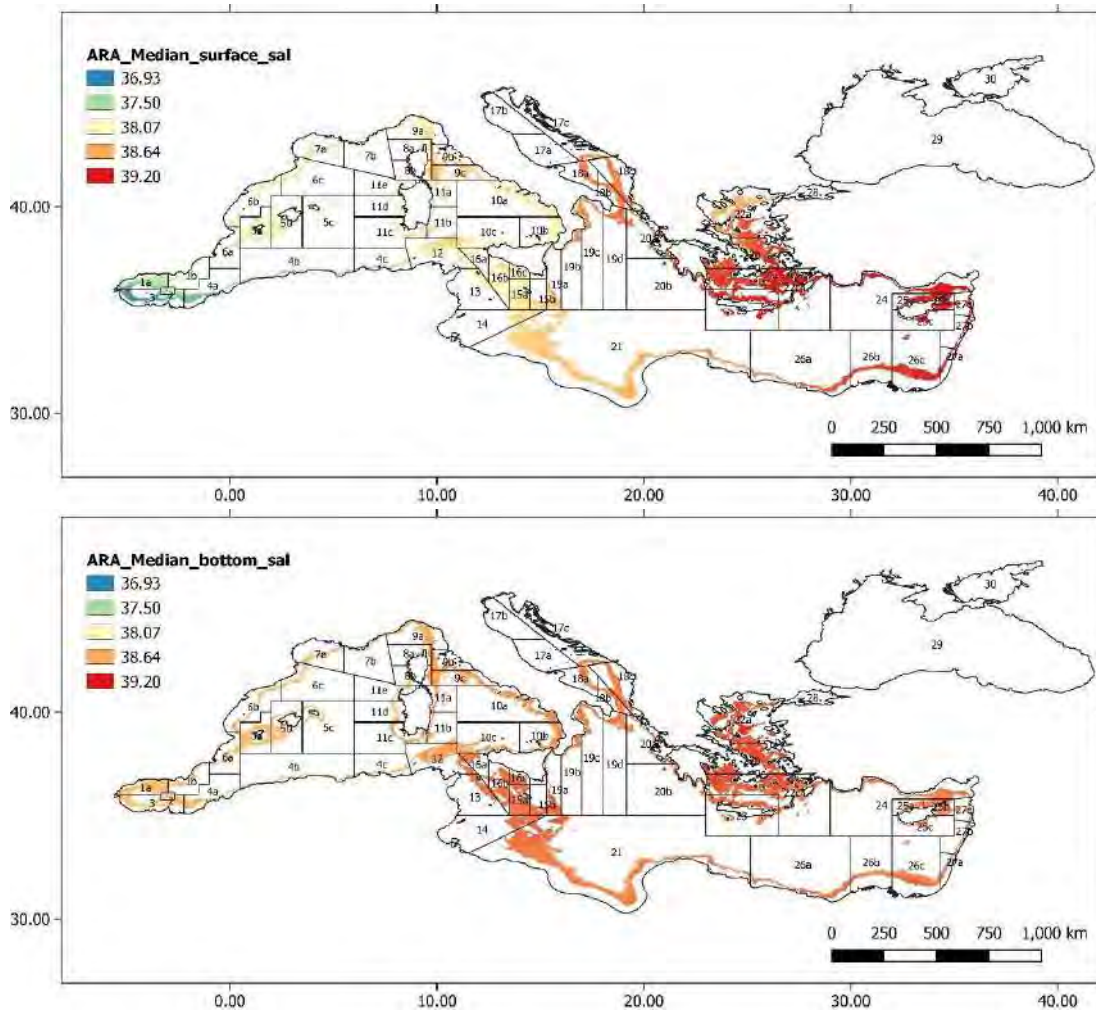
Annex 3 Figure 8: Figure 15: Median euphotic zone (upper panel) and bottom layer (lower panel) phytoplankton biomass concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



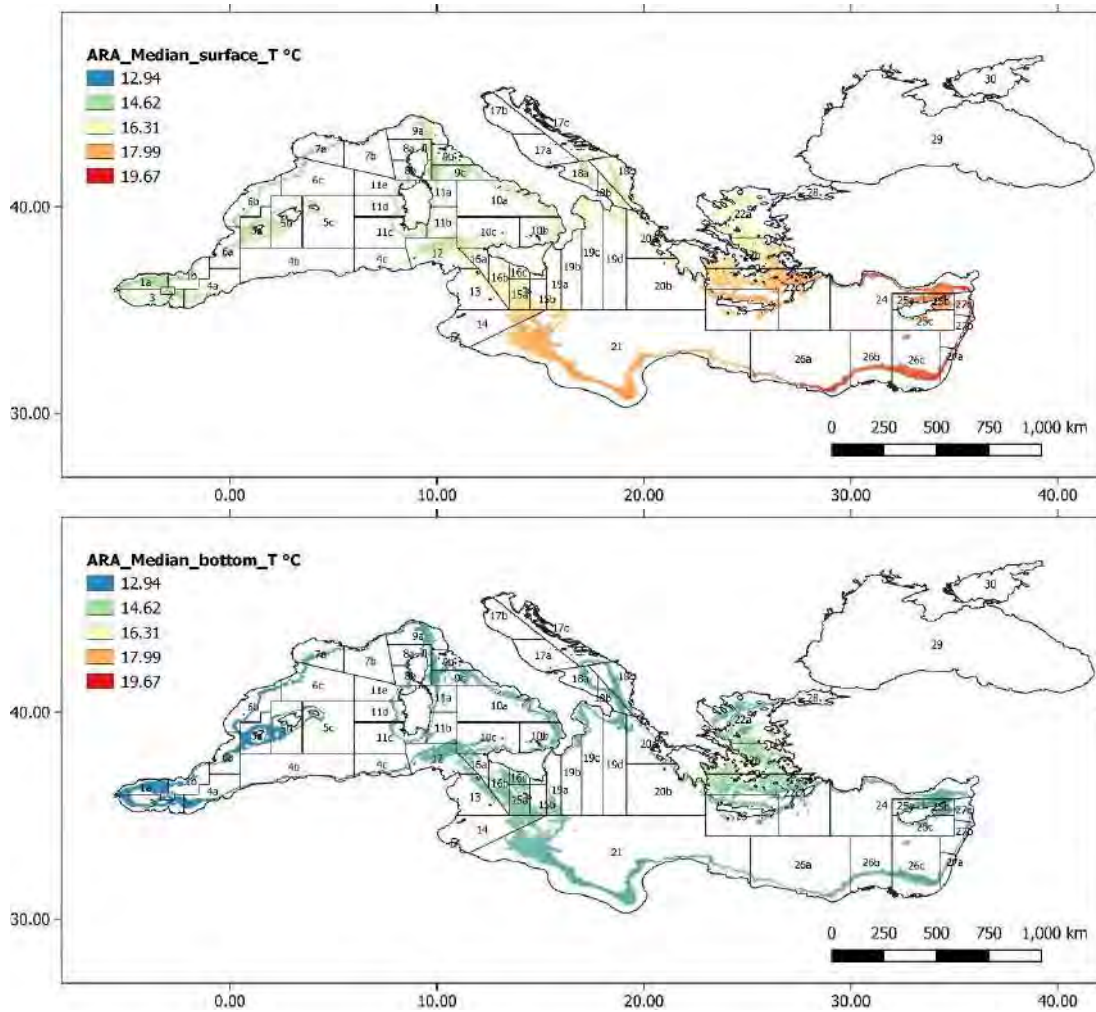
Annex 3 Figure 9: Median euphotic zone (upper panel) and bottom layer (lower panel) phosphates concentration over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



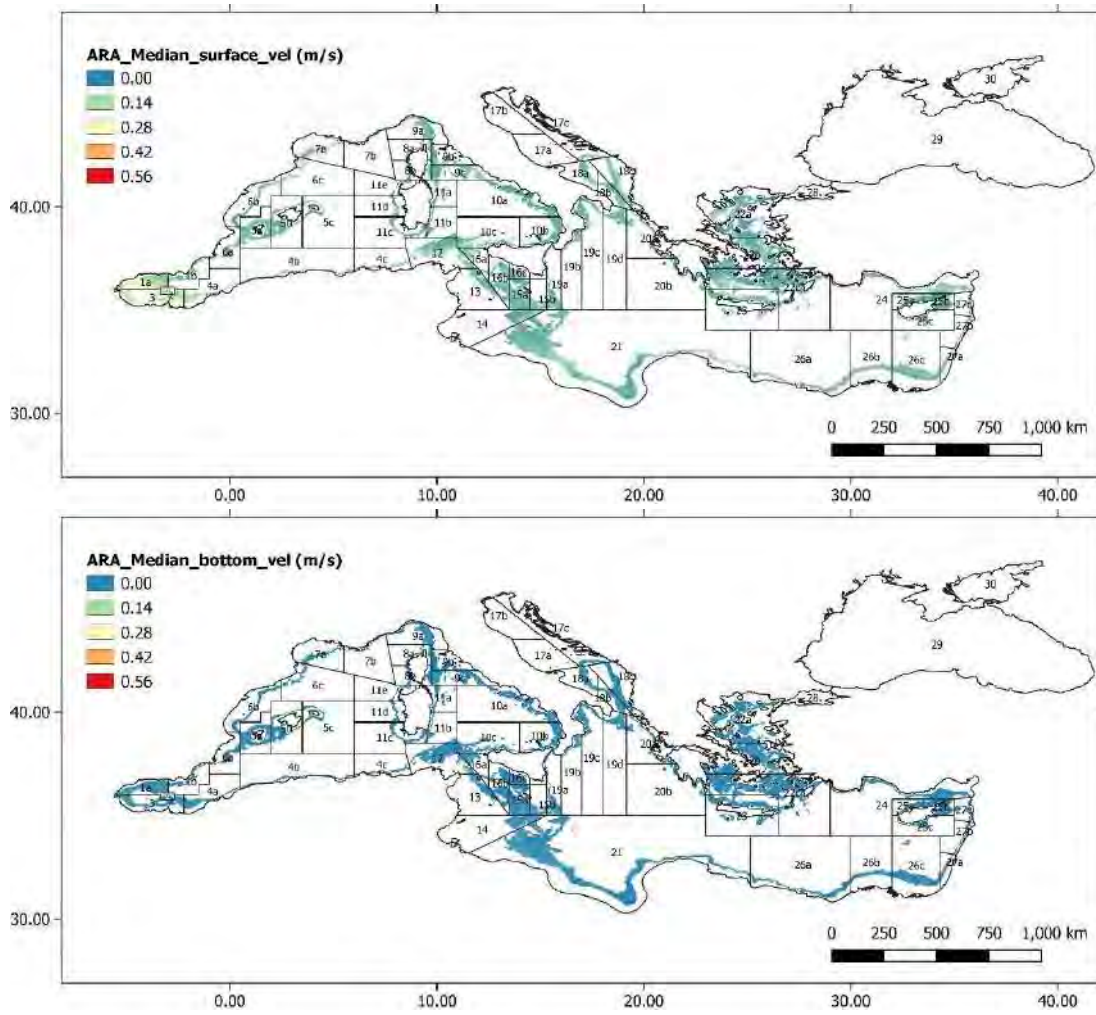
Annex 3 Figure 10: Median euphotic zone (upper panel) and bottom layer (lower panel) primary production over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.



Annex 3 Figure 11: Median euphotic zone (upper panel) and bottom layer (lower panel) salinity over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.

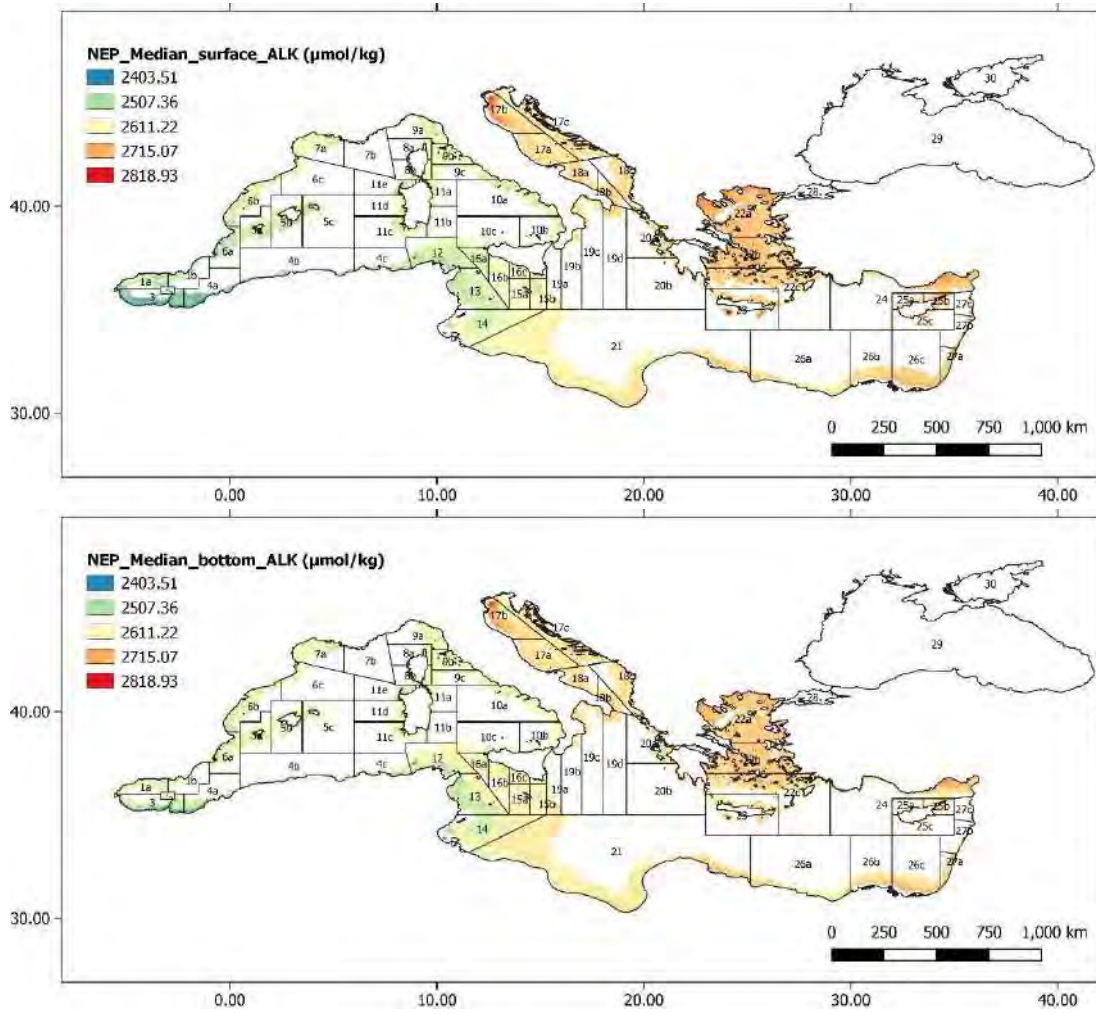


Annex 3 Figure 12: Median euphotic zone (upper panel) and bottom layer (lower panel) temperature over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.

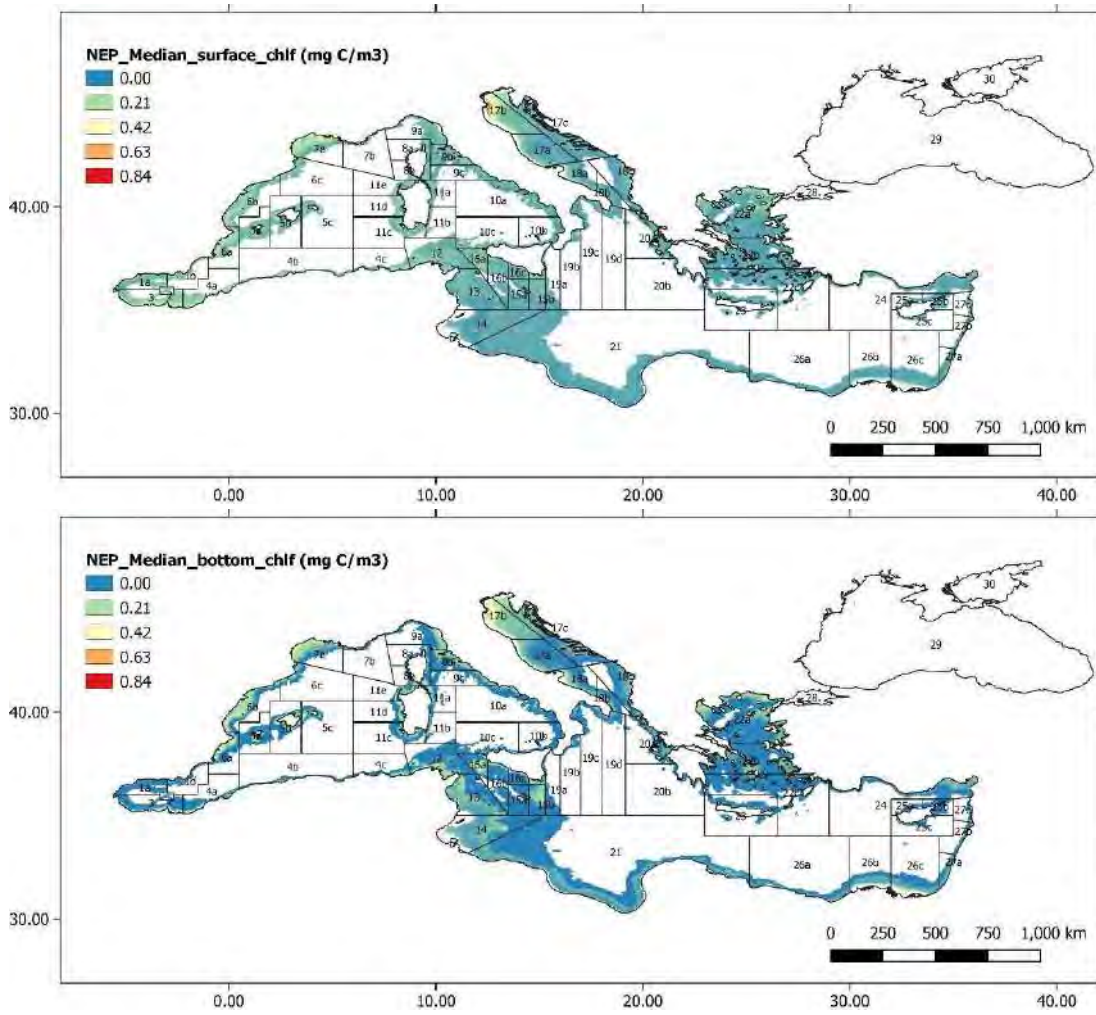


Annex 3 Figure 13: Median euphotic zone (upper panel) and bottom layer (lower panel) current velocity over 5 years in the *Aristeus antennatus* depth range of the Mediterranean.

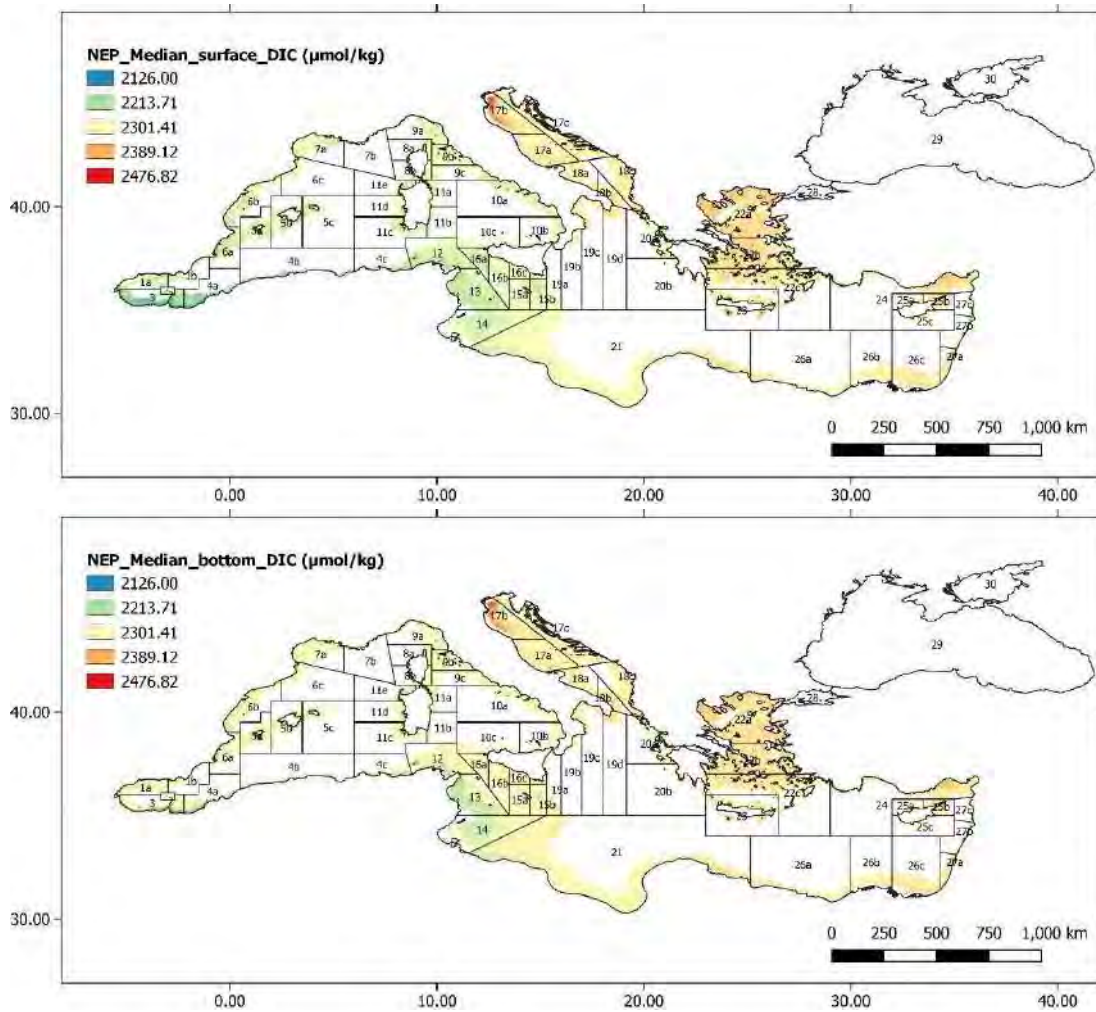
Annex 4



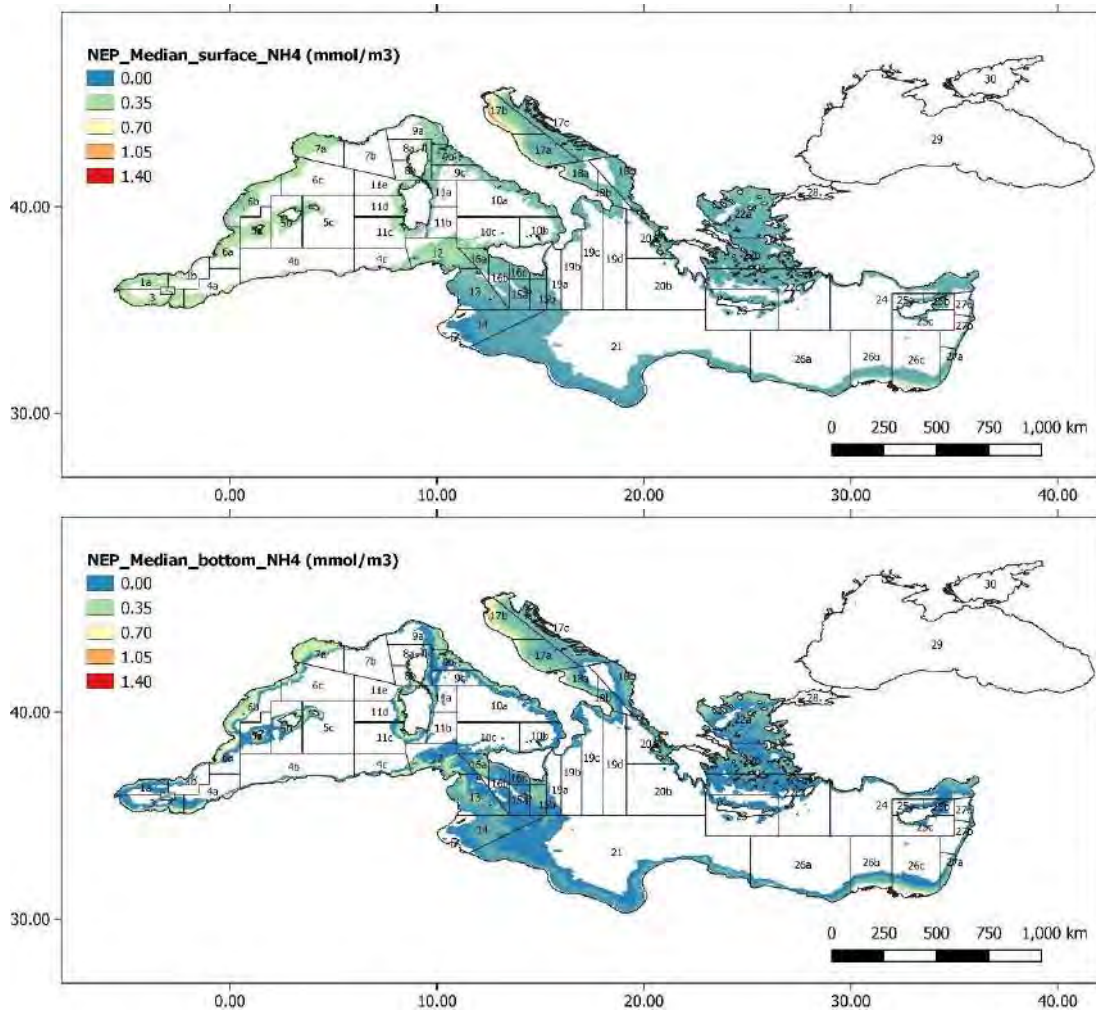
Annex 4 Figure 1: Median euphotic zone (upper panel) and bottom layer (bottom panel) alkalinity over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



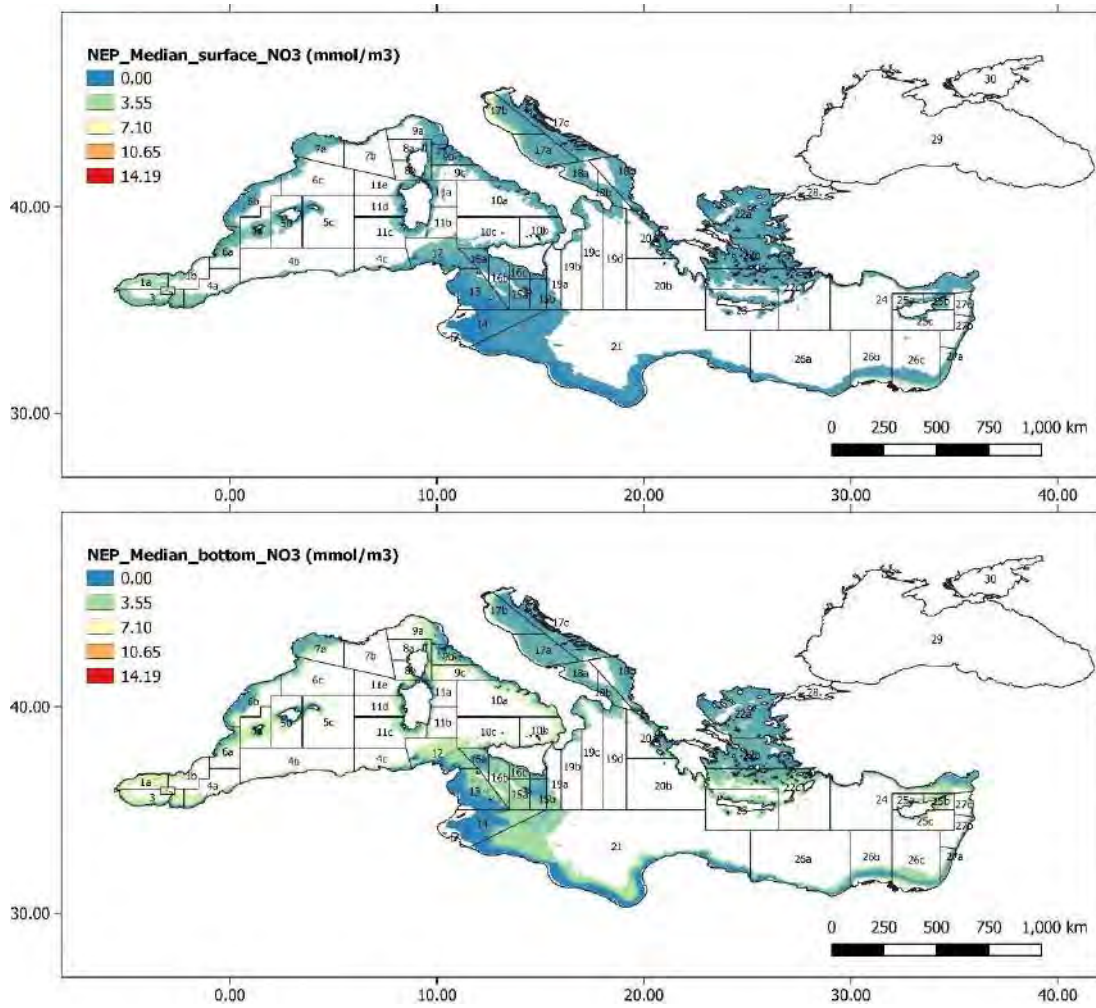
Annex 4 Figure 2: Median euphotic zone (upper panel) and bottom layer (lower panel) chlorophyll a concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



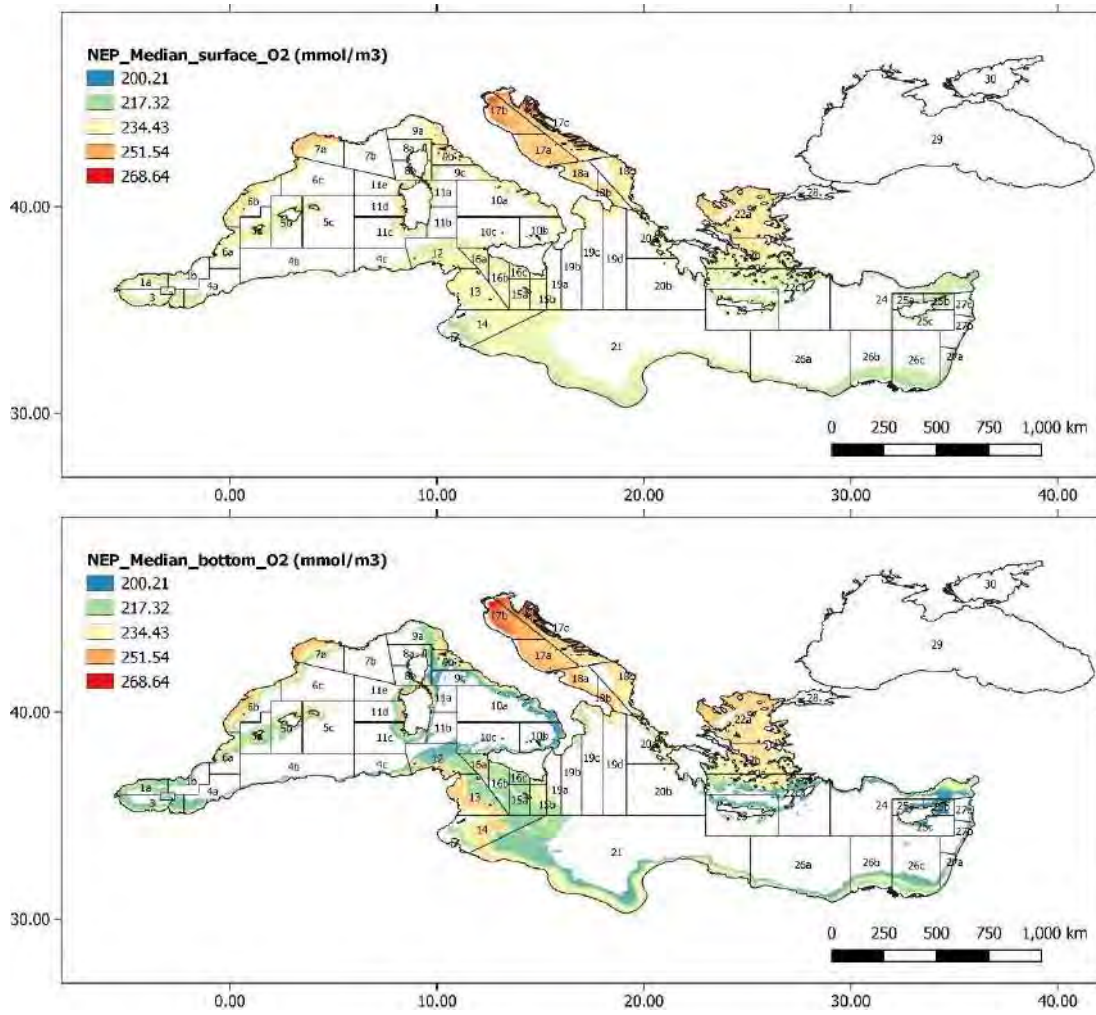
Annex 4 Figure 3: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved inorganic carbon concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



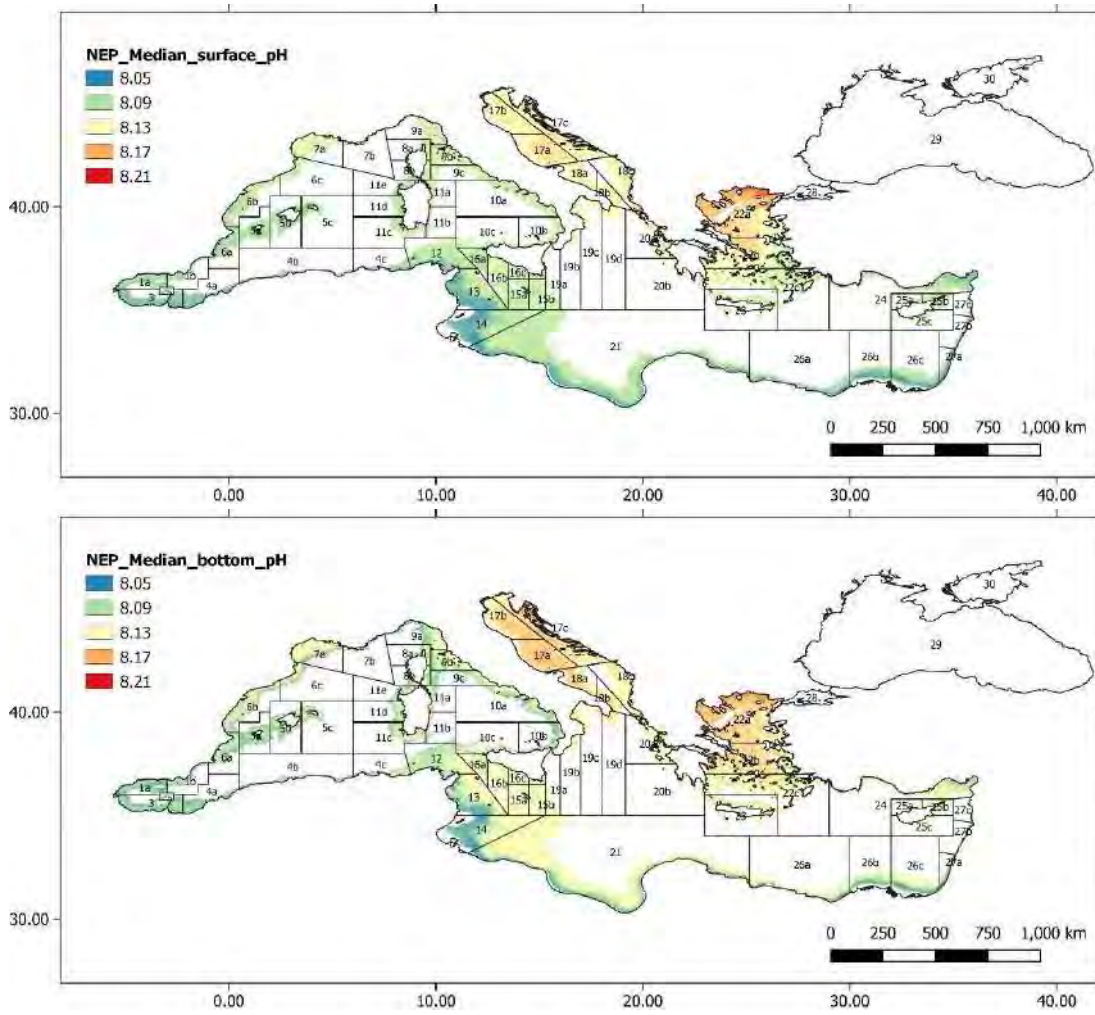
Annex 4 Figure 4: Median euphotic zone (upper panel) and bottom layer (lower panel) ammonium concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



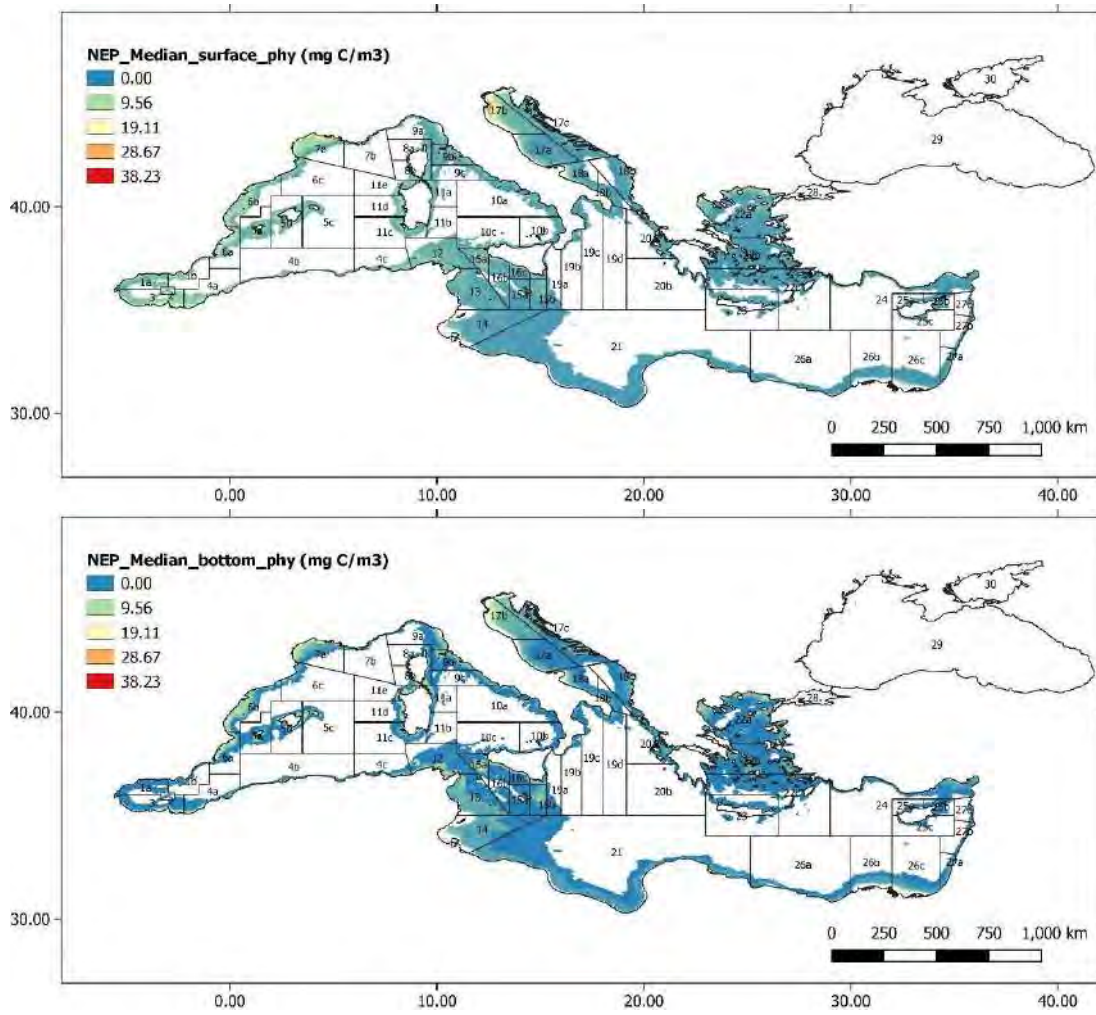
Annex 4 Figure 5: Median euphotic zone (upper panel) and bottom layer (lower panel) nitrates concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



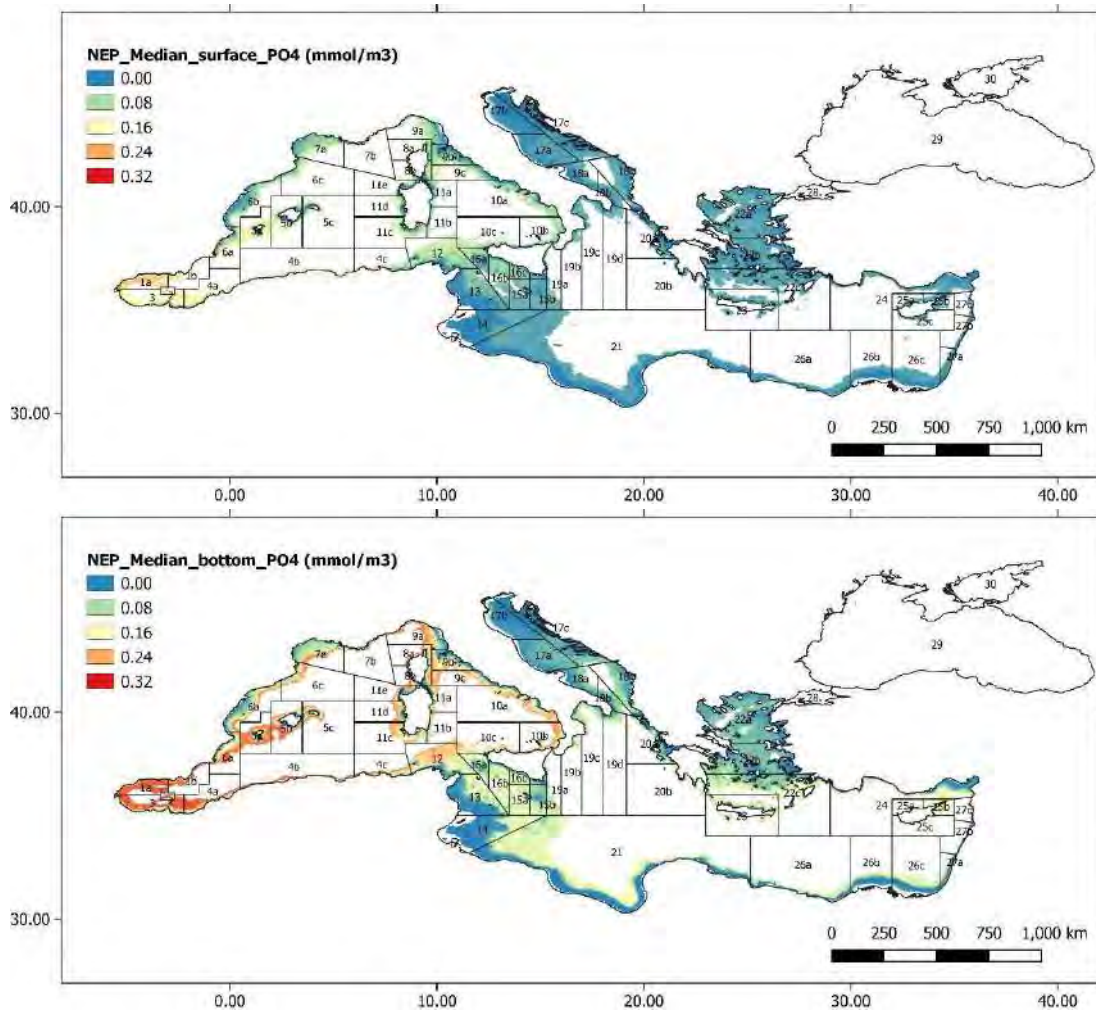
Annex 4 Figure 6: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved oxygen concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



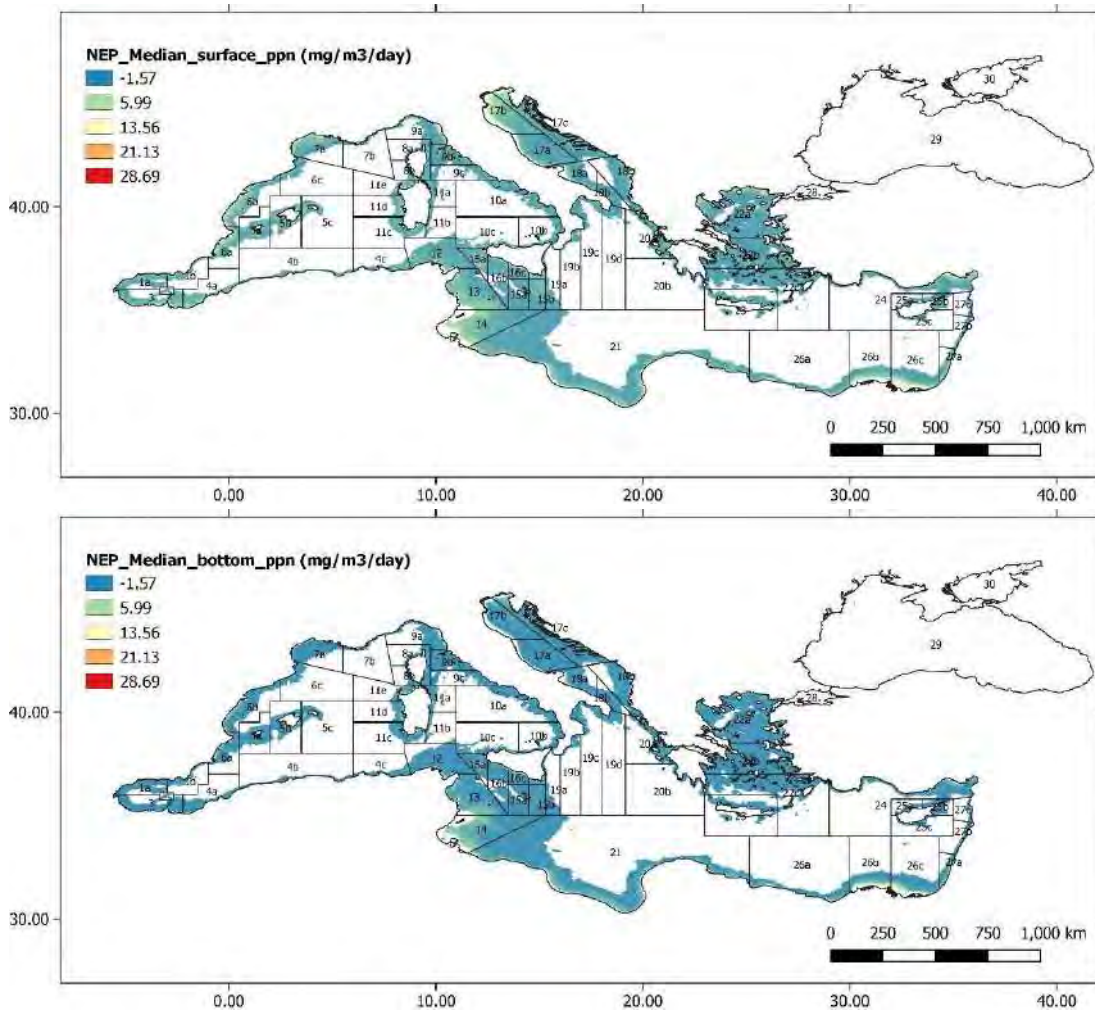
Annex 4 Figure 7: Median euphotic zone (upper panel) and bottom layer (lower panel) pH values over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



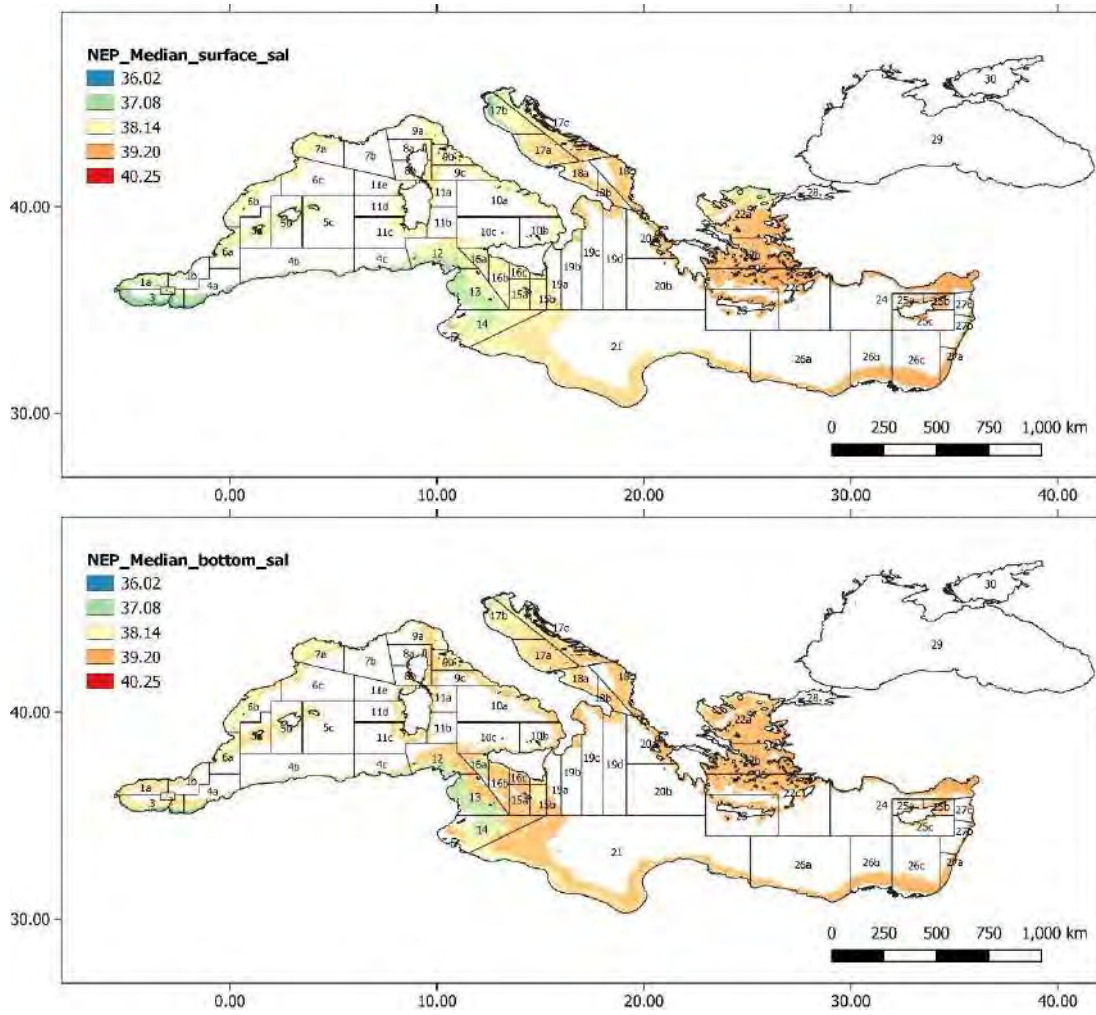
Annex 4 Figure 8: Median euphotic zone (upper panel) and bottom layer (lower panel) phytoplankton biomass concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



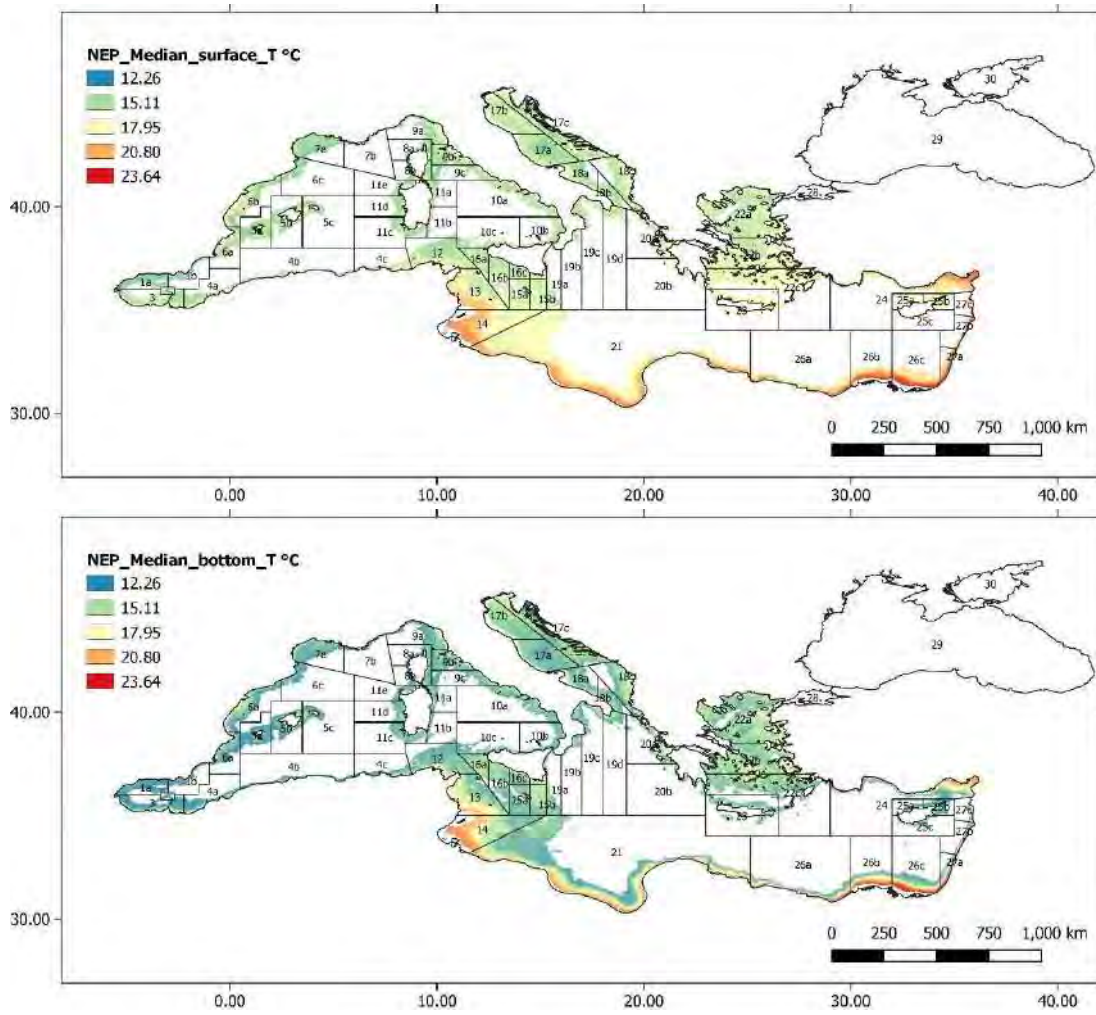
Annex 4 Figure 9: Median euphotic zone (upper panel) and bottom layer (lower panel) phosphates concentration over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



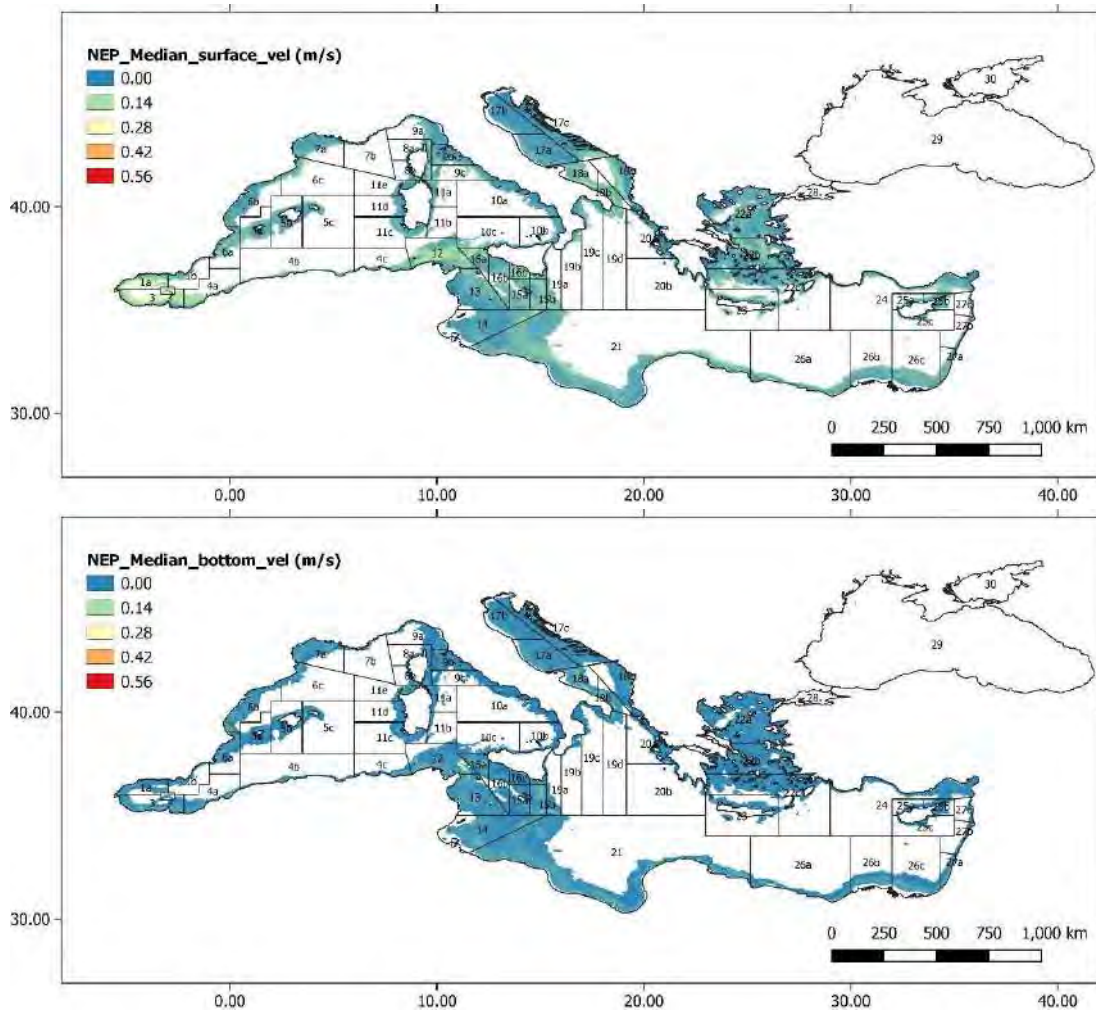
Annex 4 Figure 10: Median euphotic zone (upper panel) and bottom layer (lower panel) primary production over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.



Annex 4 Figure 11: Median euphotic zone (upper panel) and bottom layer (lower panel) salinity over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.

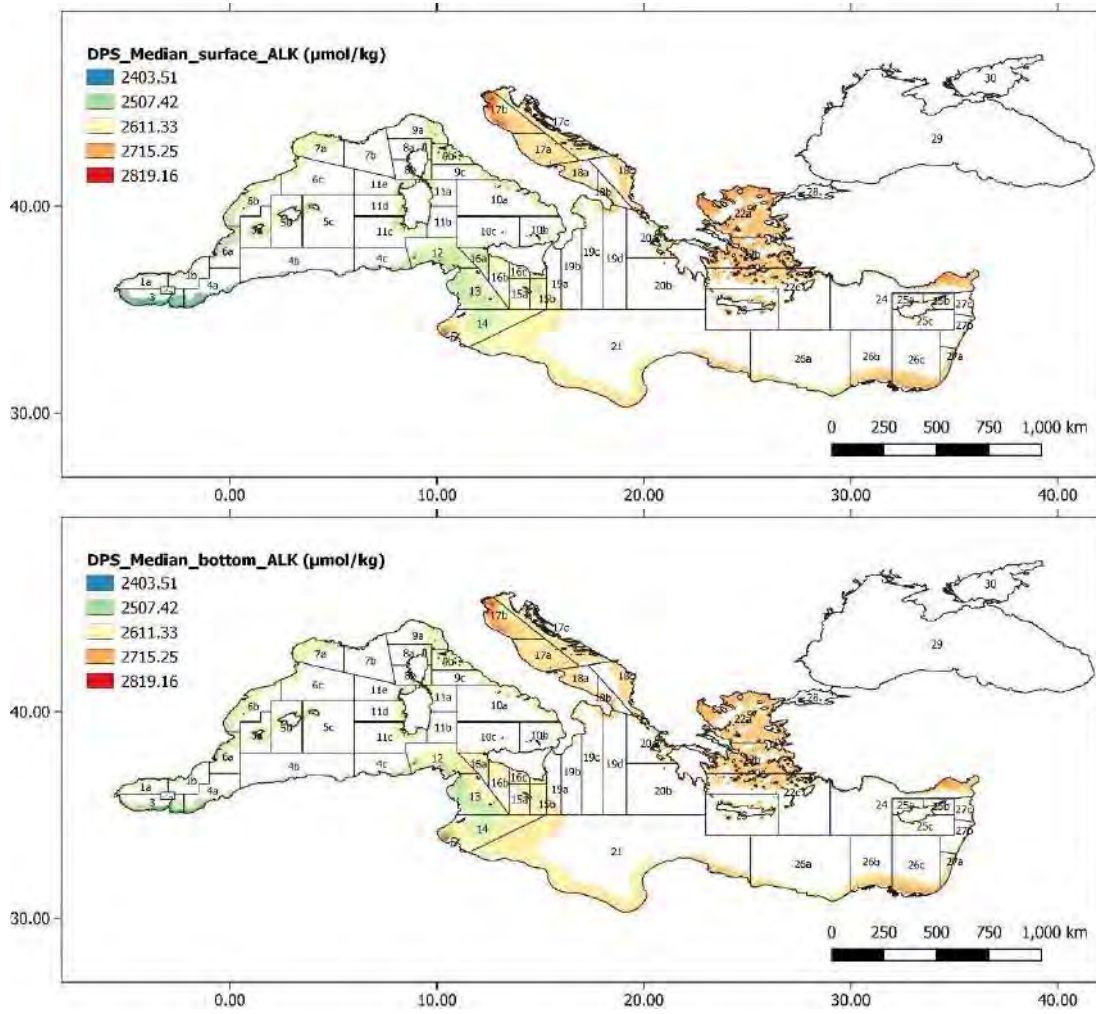


Annex 4 Figure 12: Median euphotic zone (upper panel) and bottom layer (lower panel) temperature over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.

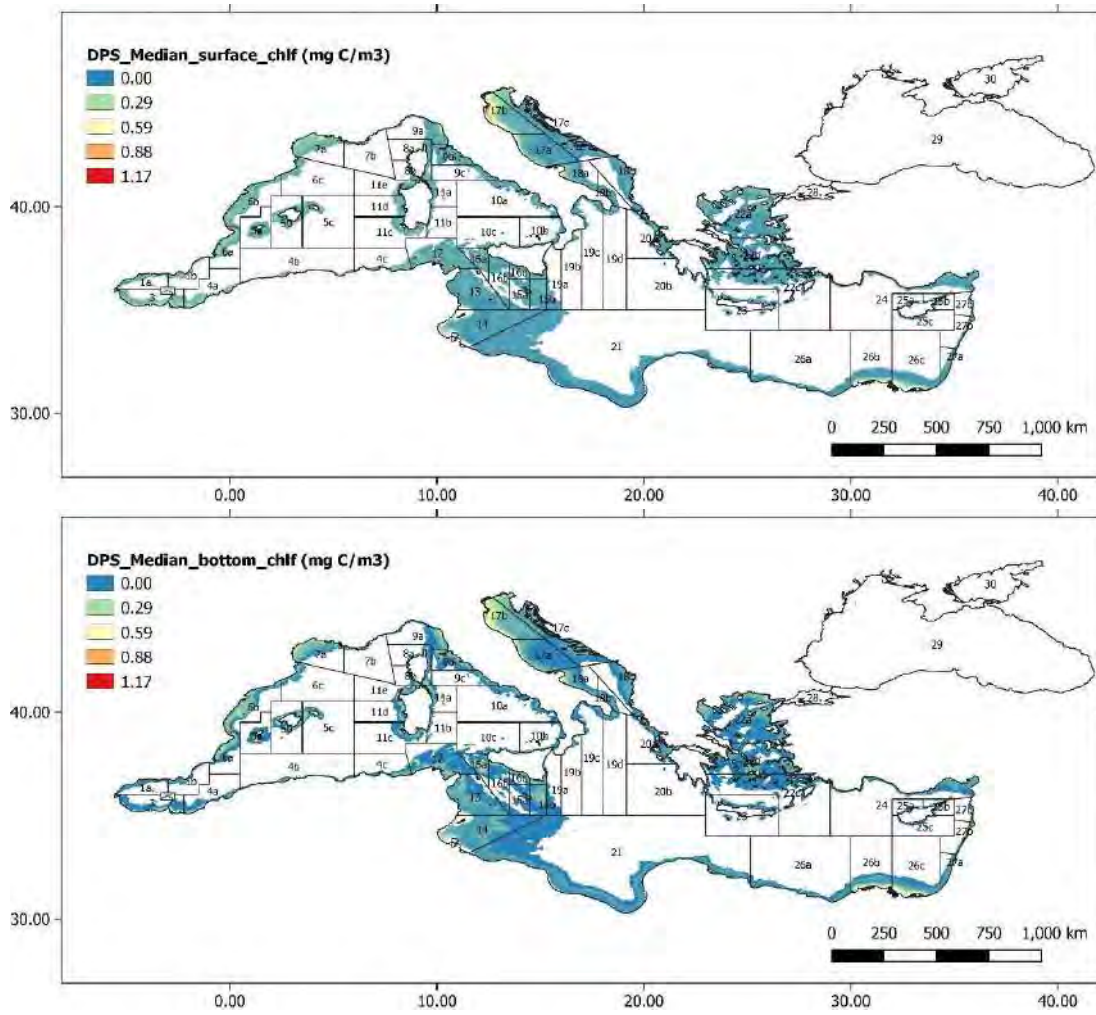


Annex 4 Figure 13: Median euphotic zone (upper panel) and bottom layer (lower panel) current velocity over 5 years in the *Nephrops norvegicus* depth range of the Mediterranean.

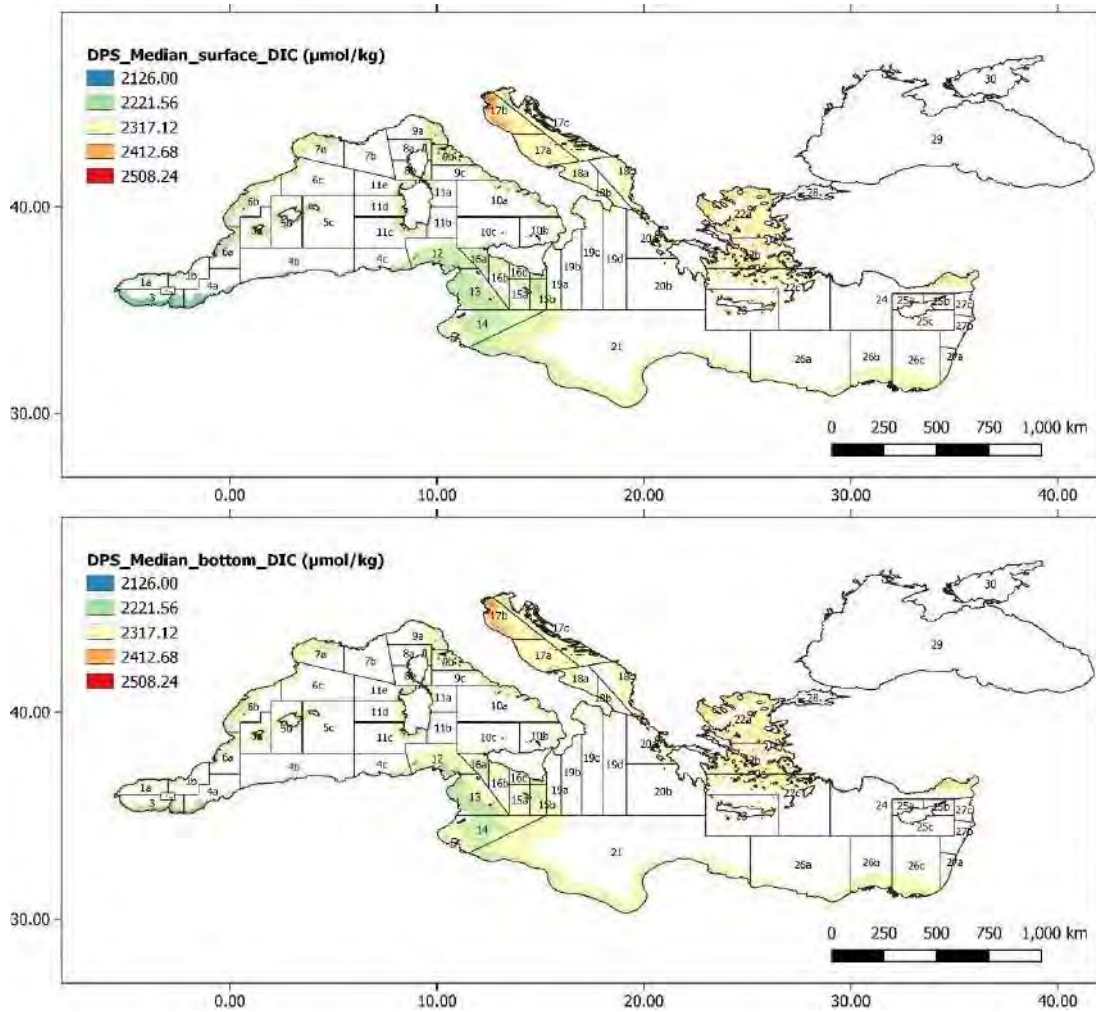
Annex 5



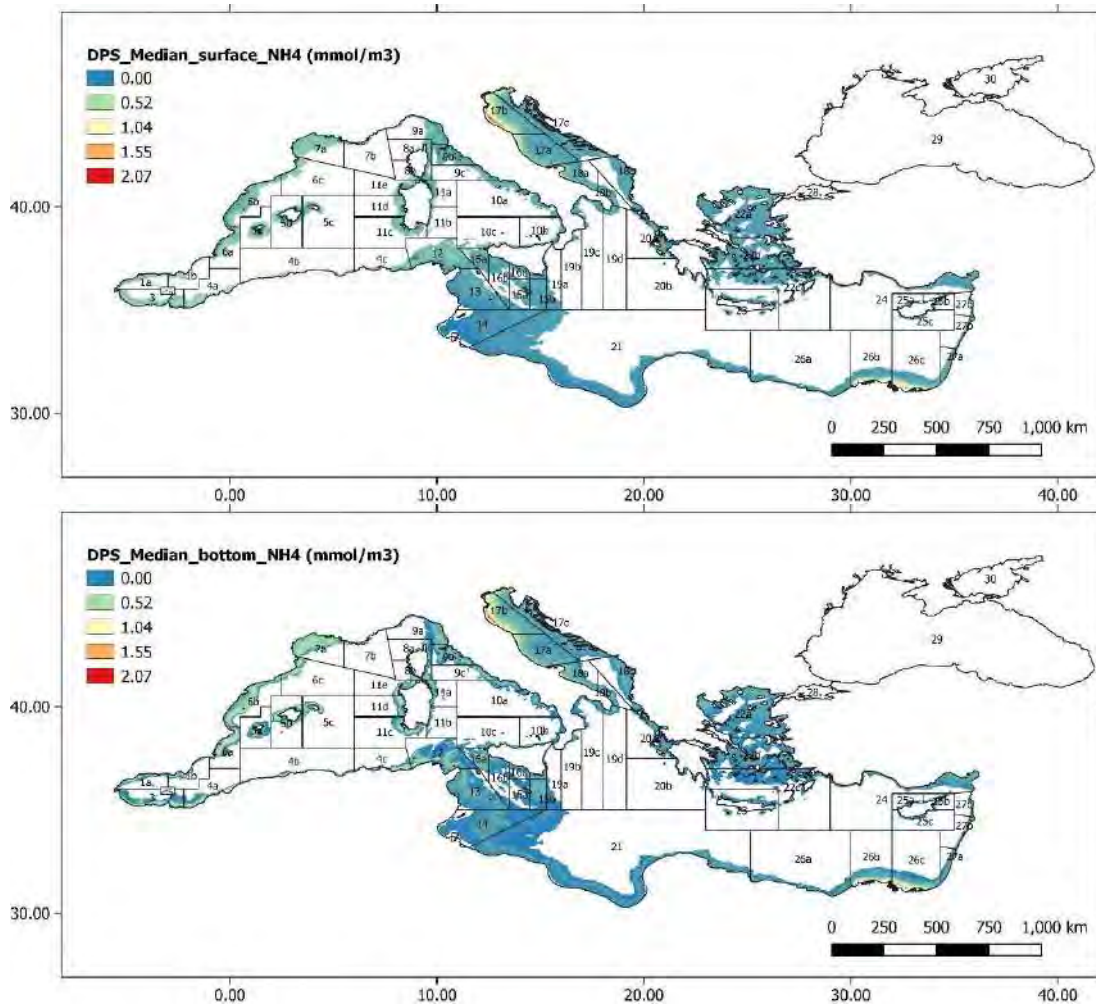
Annex 5 Figure 14: Median euphotic zone (upper panel) and bottom layer (bottom panel) alkalinity over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.



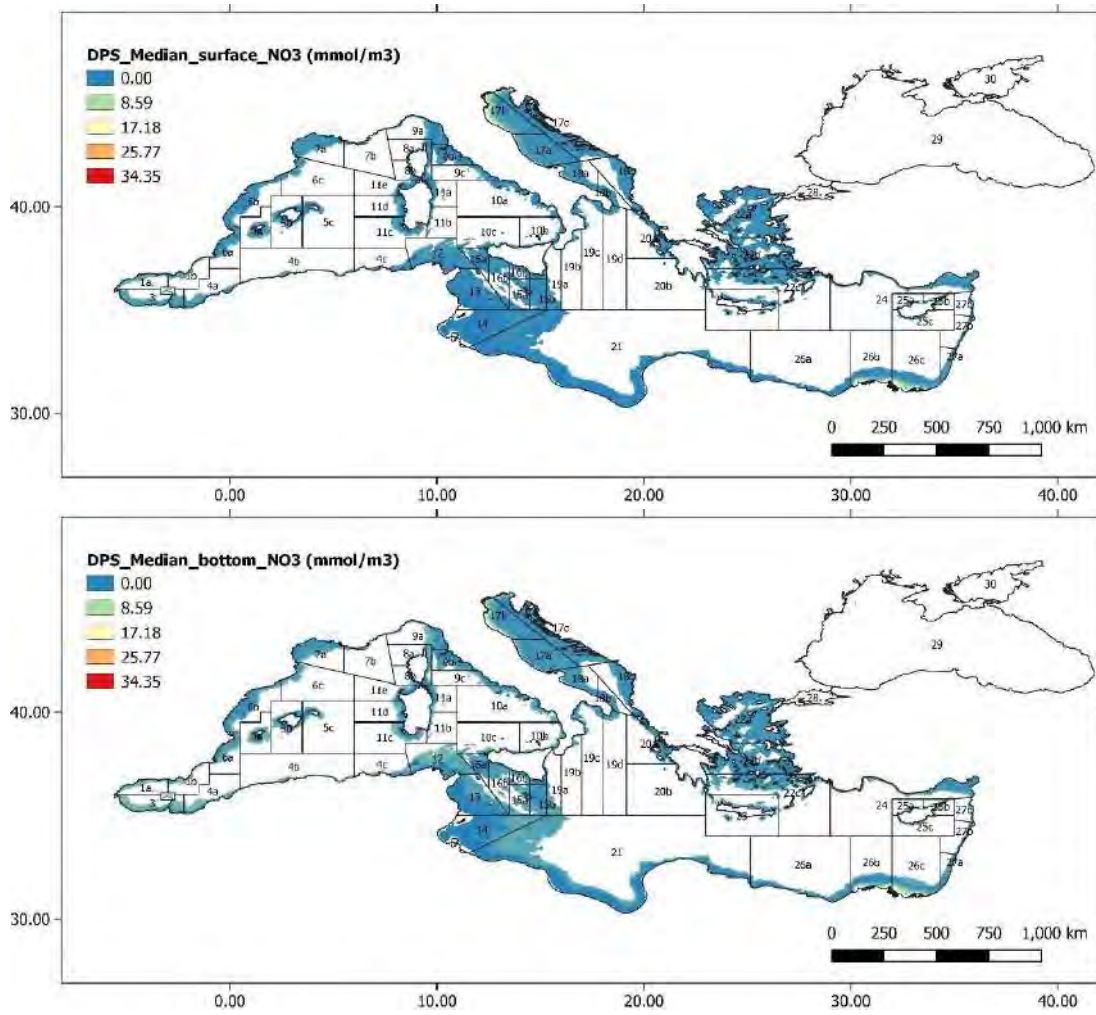
Annex 5 Figure 14: Median euphotic zone (upper panel) and bottom layer (lower panel) chlorophyll a concentration over 5 years in the *Parapneaus longirostris* depth range of the Mediterranean.



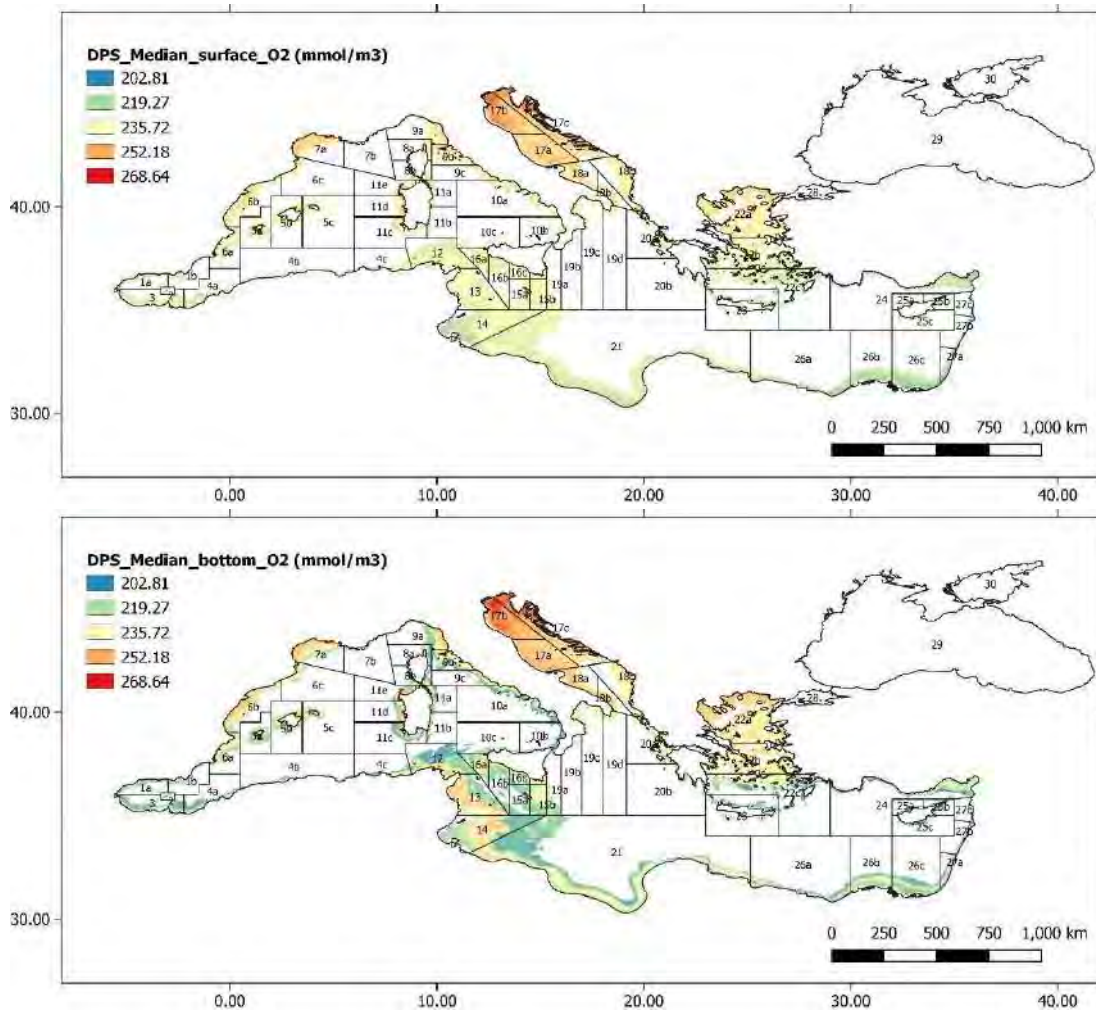
Annex 5 Figure 15: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved inorganic carbon concentration over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.



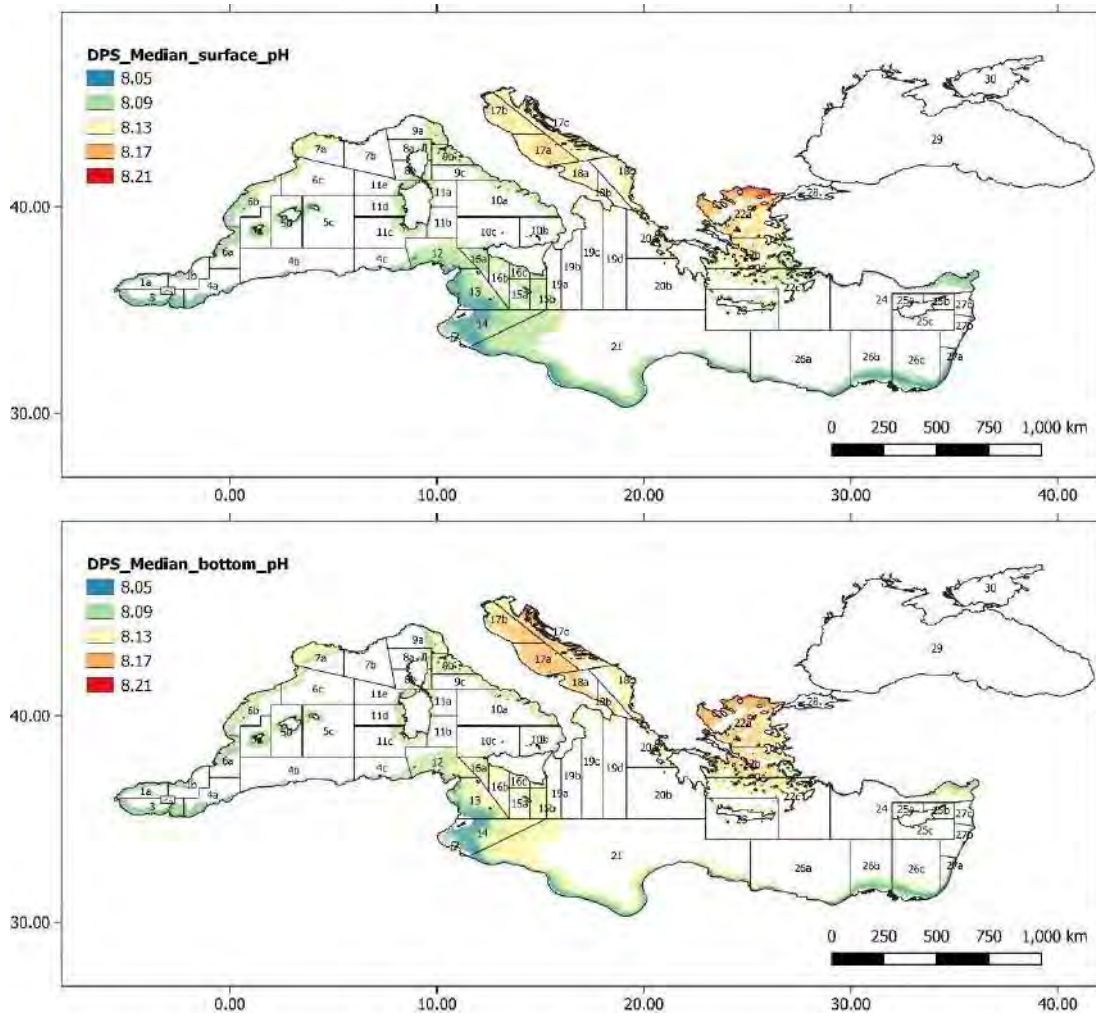
Annex 5 Figure 16: Median euphotic zone (upper panel) and bottom layer (lower panel) ammonium concentration over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.



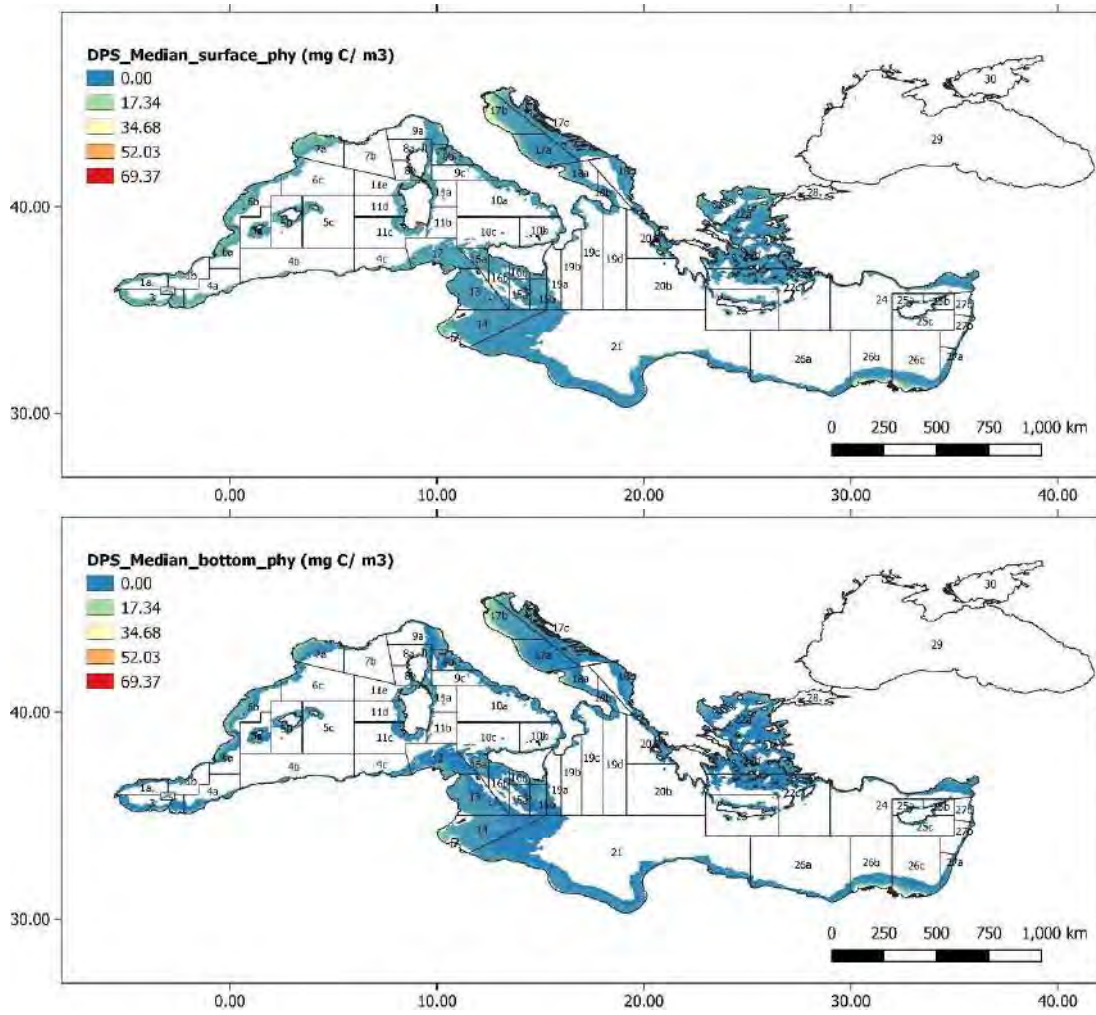
Annex 5 Figure 17: Median euphotic zone (upper panel) and bottom layer (lower panel) nitrates concentration over 5 years in the *Parapneustes longirostris* depth range of the Mediterranean.



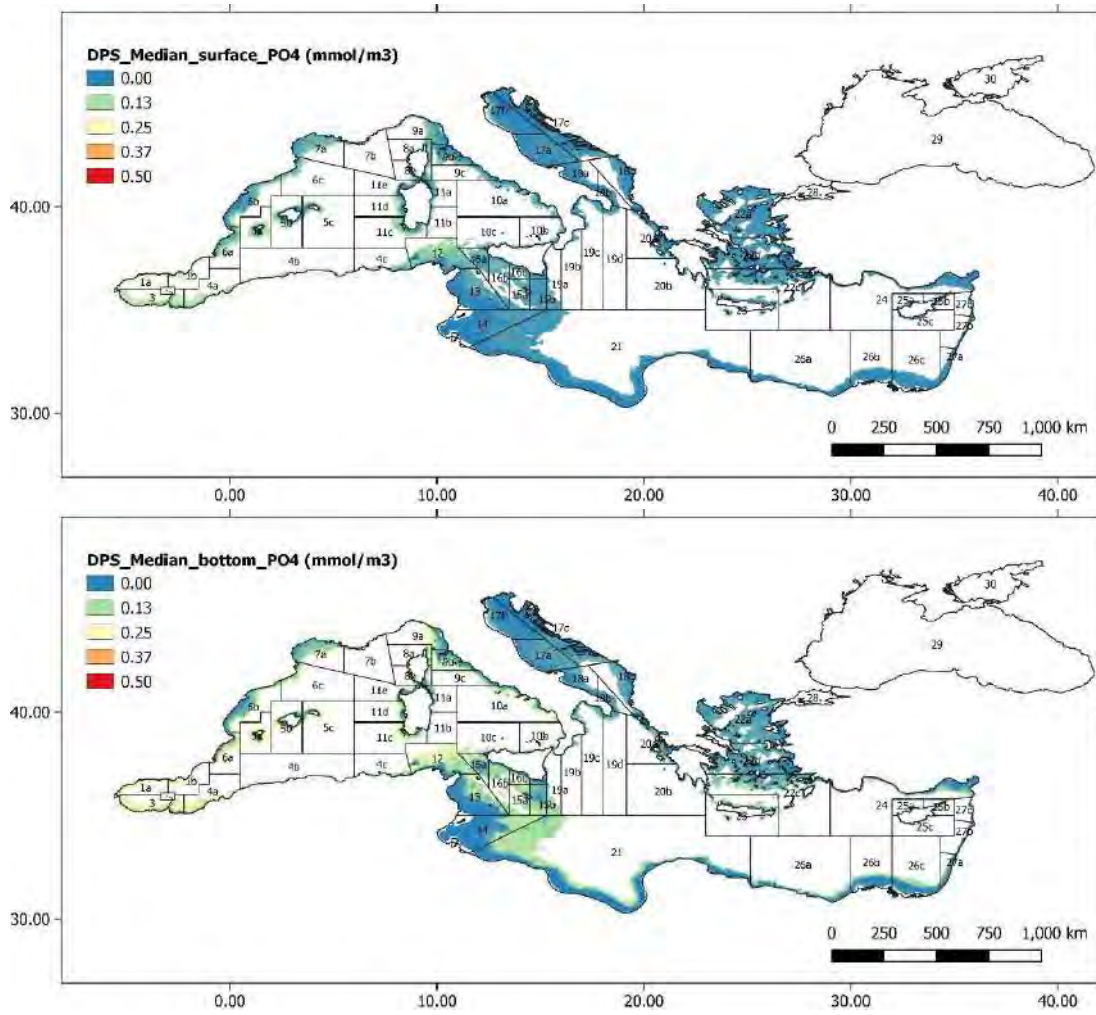
Annex 5 Figure 18: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved oxygen concentration over 5 years in the *Parapneaus longirostris* depth range of the Mediterranean.



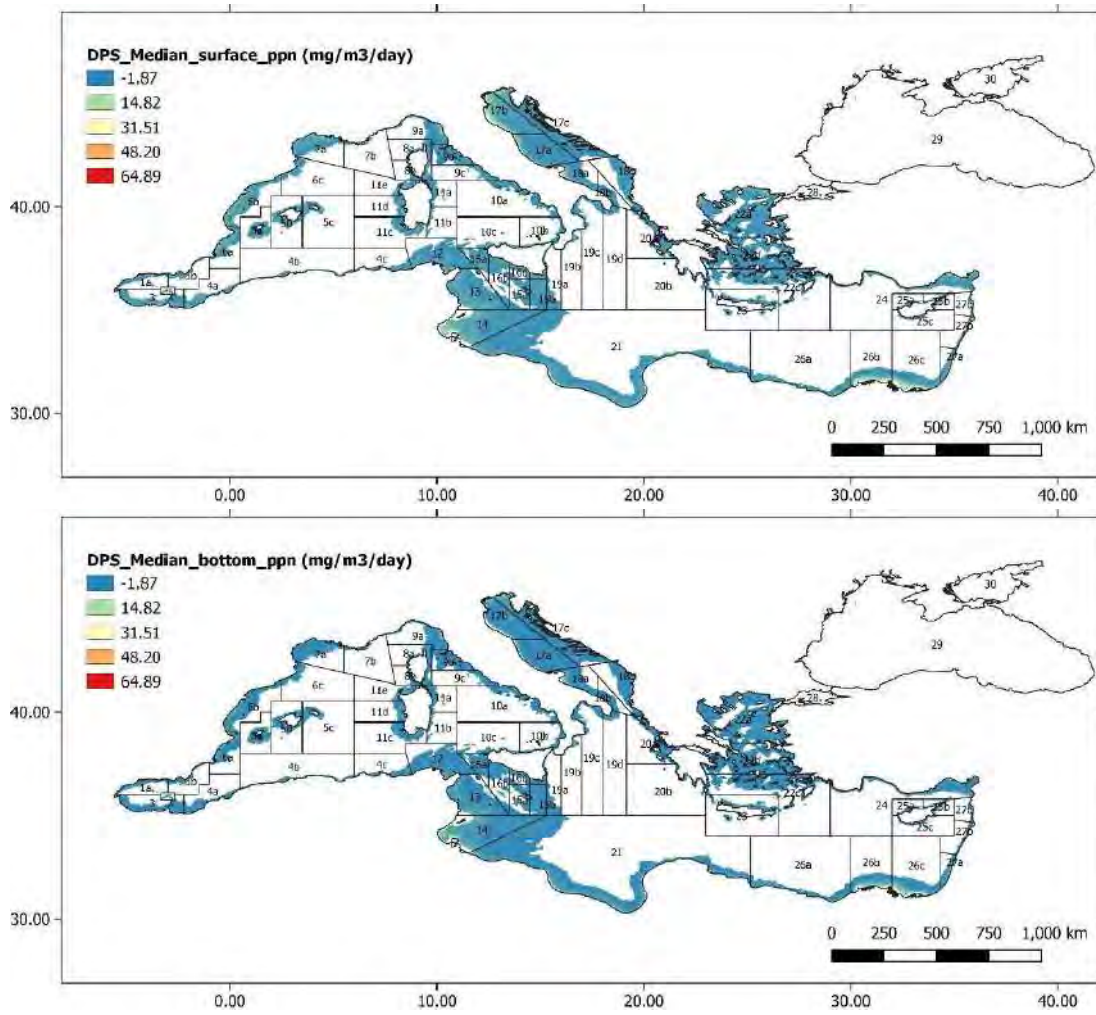
Annex 5 Figure 19: Median euphotic zone (upper panel) and bottom layer (lower panel) pH values over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.



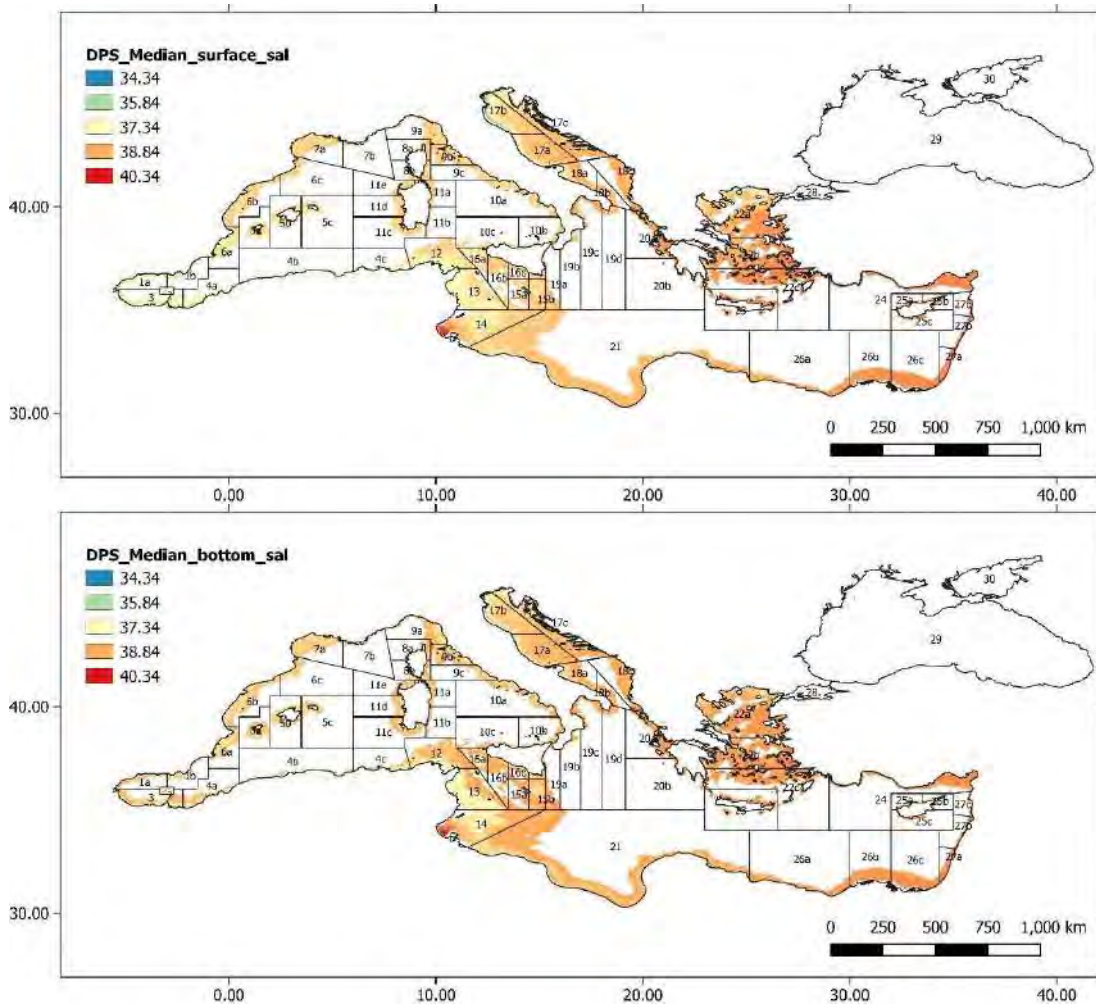
Annex 5 Figure 20: Median euphotic zone (upper panel) and bottom layer (lower panel) phytoplankton biomass concentration over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.



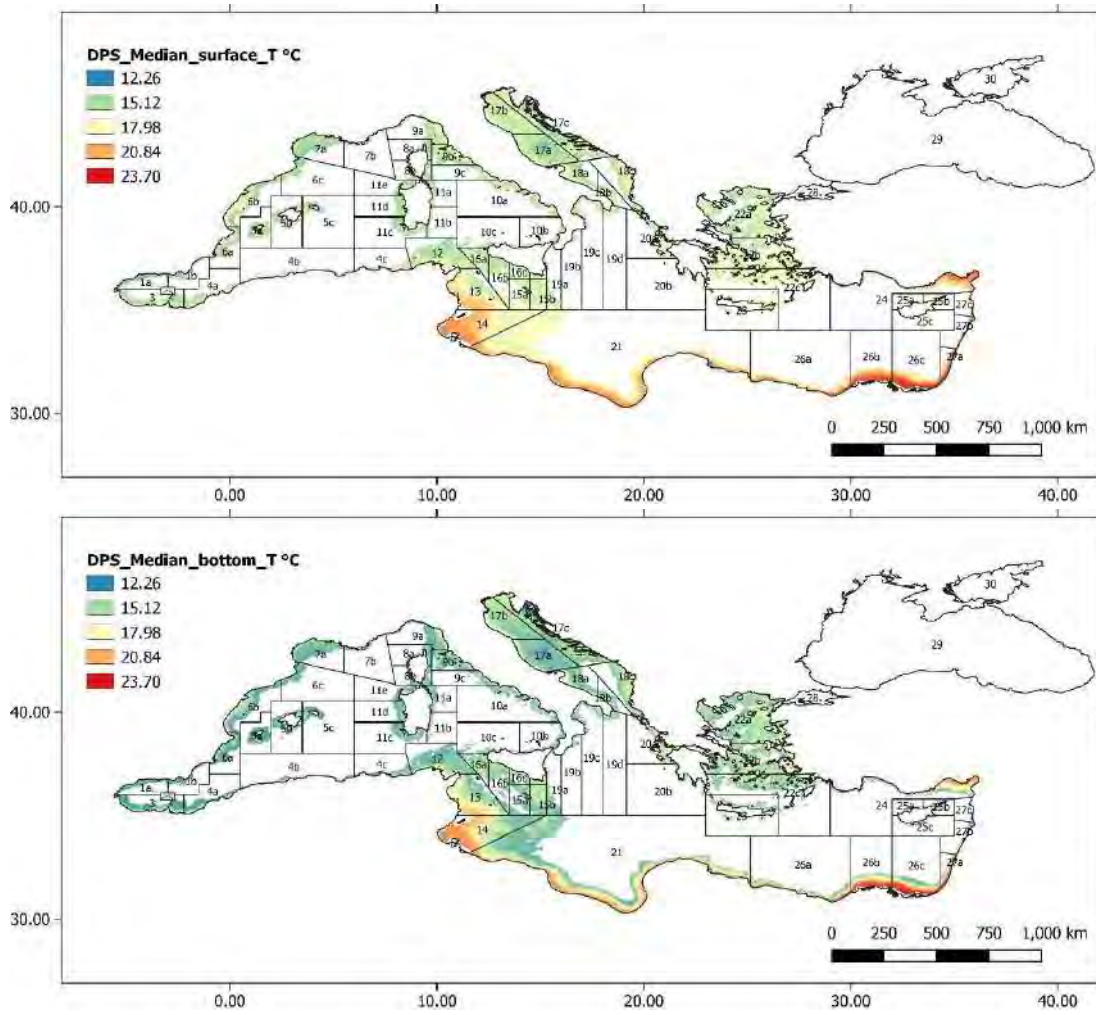
Annex 5 Figure 21: Median euphotic zone (upper panel) and bottom layer (lower panel) phosphates concentration over 5 years in the *Parapneaus longirostris* depth range of the Mediterranean.



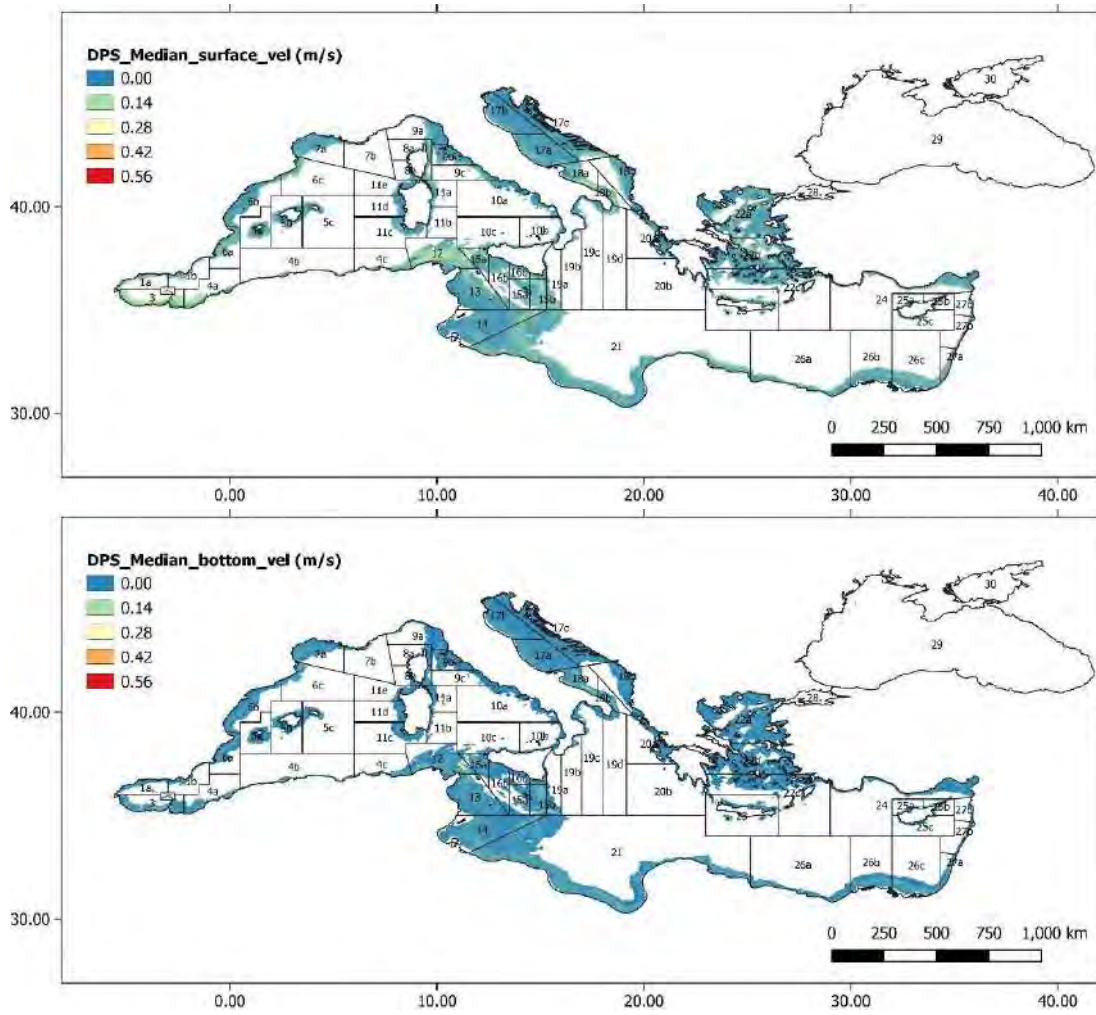
Annex 5 Figure 22: Median euphotic zone (upper panel) and bottom layer (lower panel) primary production over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.



Annex 5 Figure 23: Median euphotic zone (upper panel) and bottom layer (lower panel) salinity over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.

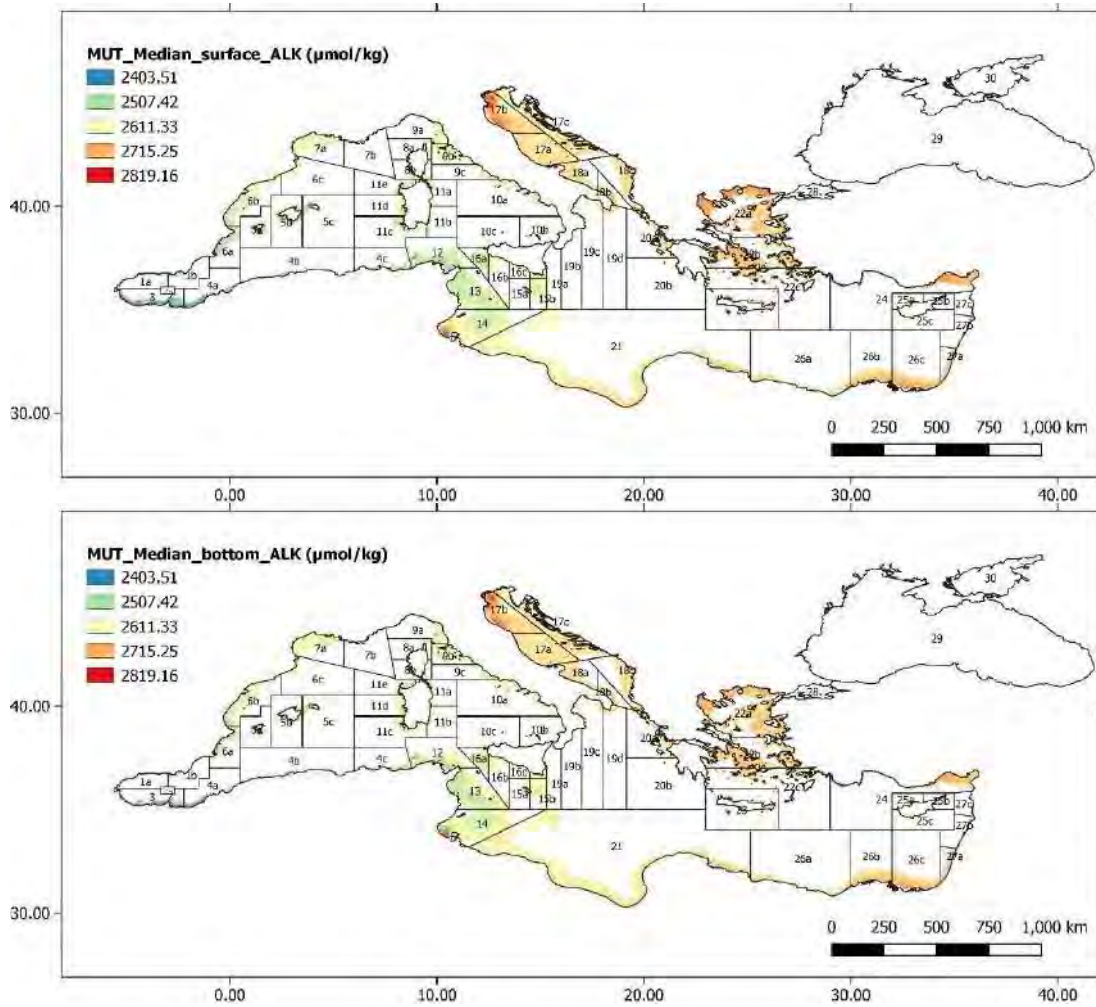


Annex 5 Figure 24: Median euphotic zone (upper panel) and bottom layer (lower panel) temperature over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.

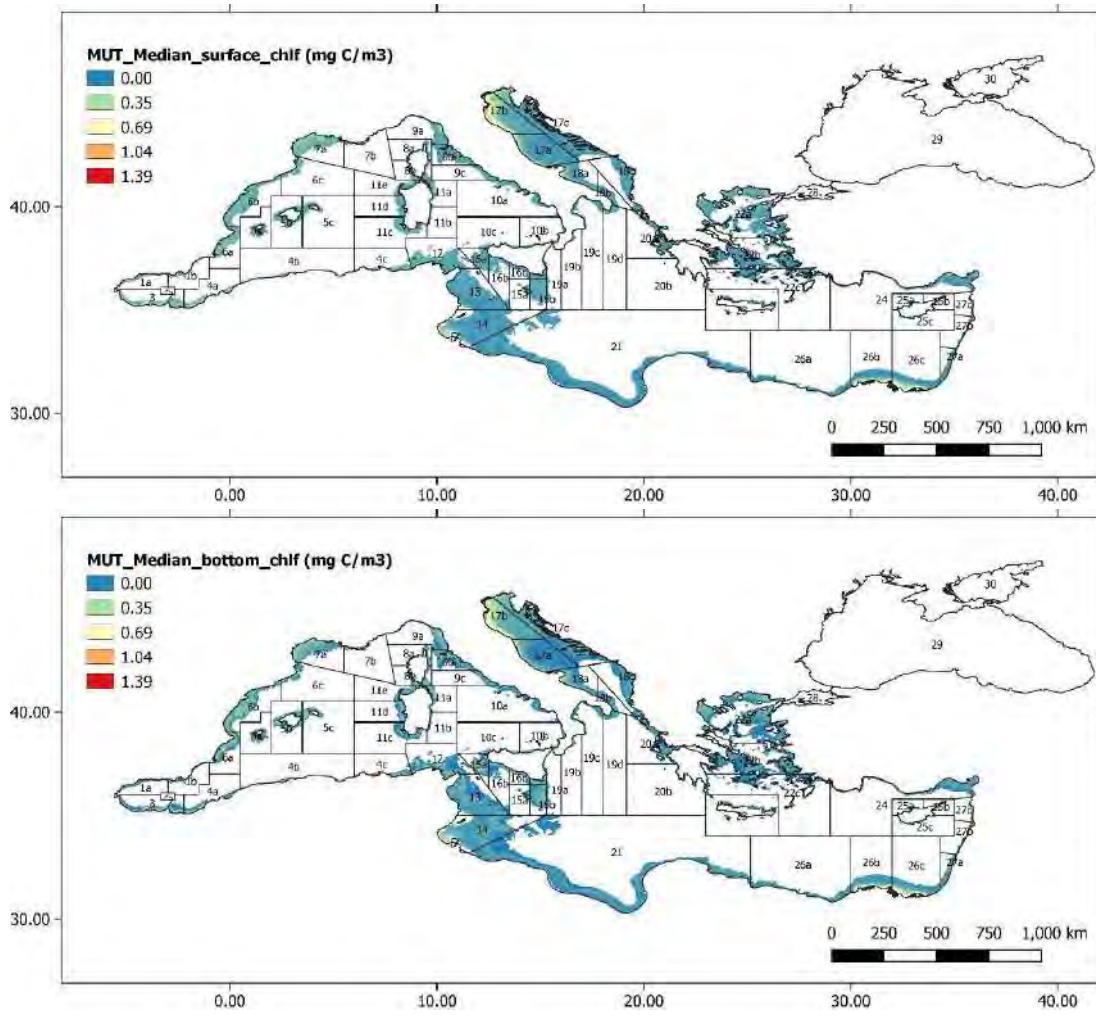


Annex 5 Figure 25: Median euphotic zone (upper panel) and bottom layer (lower panel) current velocity over 5 years in the *Parapenaeus longirostris* depth range of the Mediterranean.

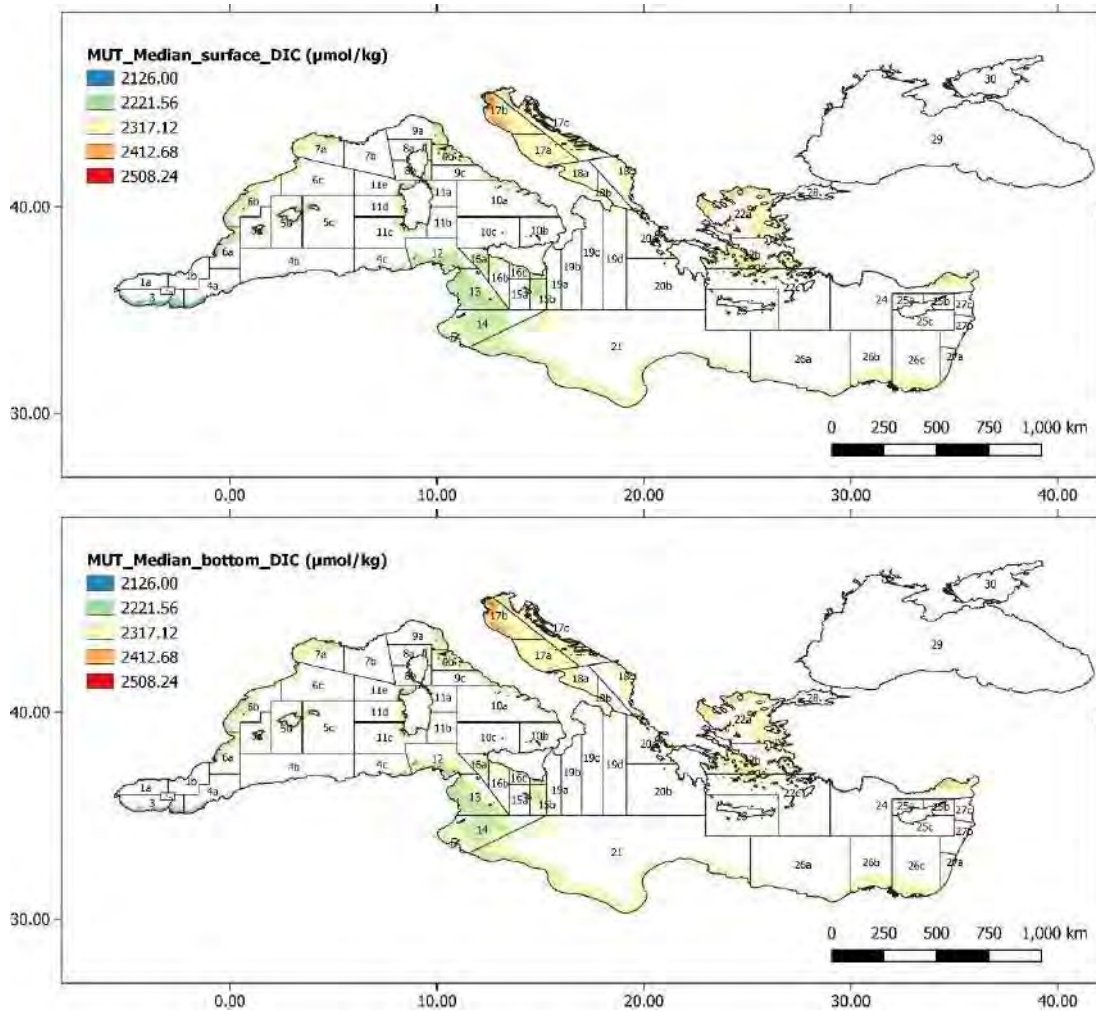
Annex 6



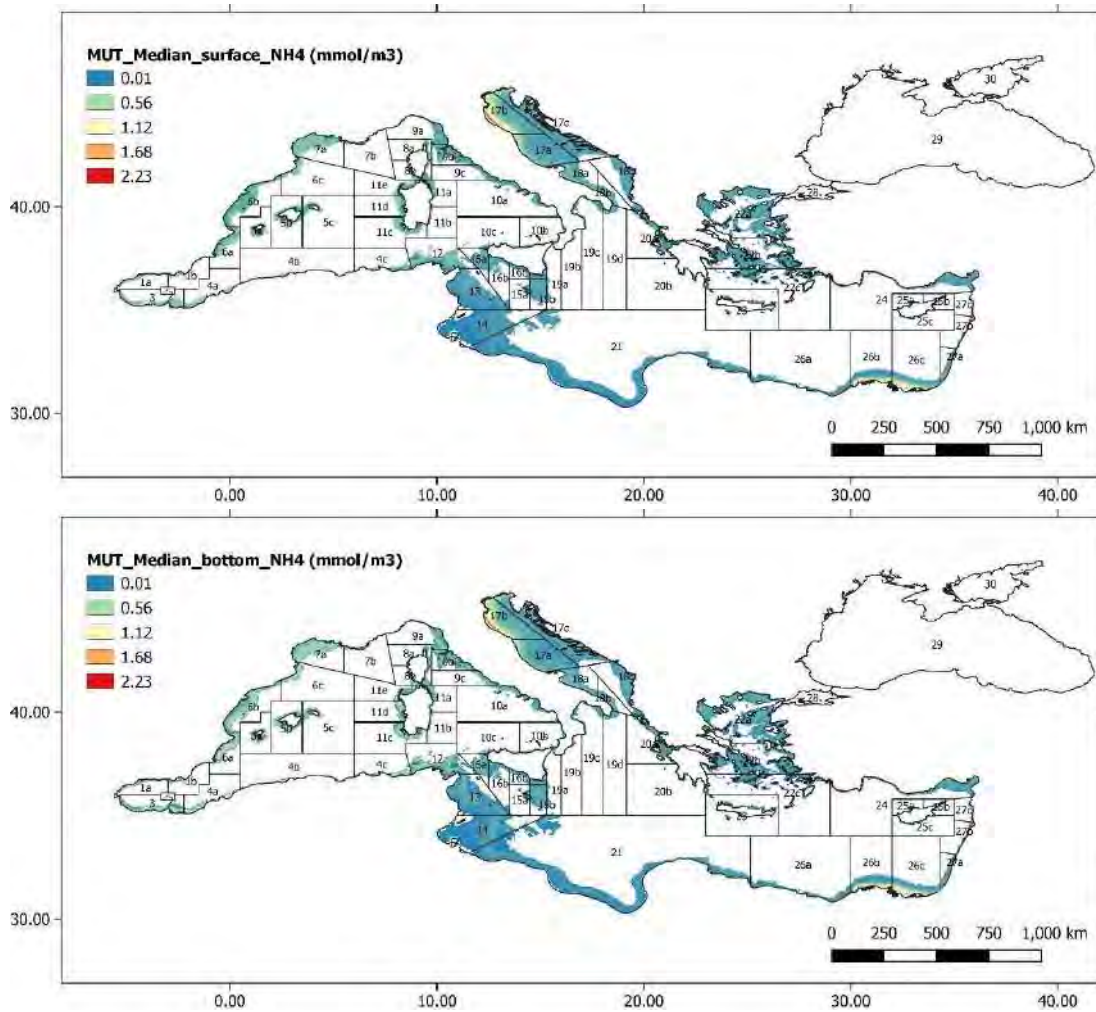
Annex 6 Figure 15: Median euphotic zone (upper panel) and bottom layer (bottom panel) alkalinity over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



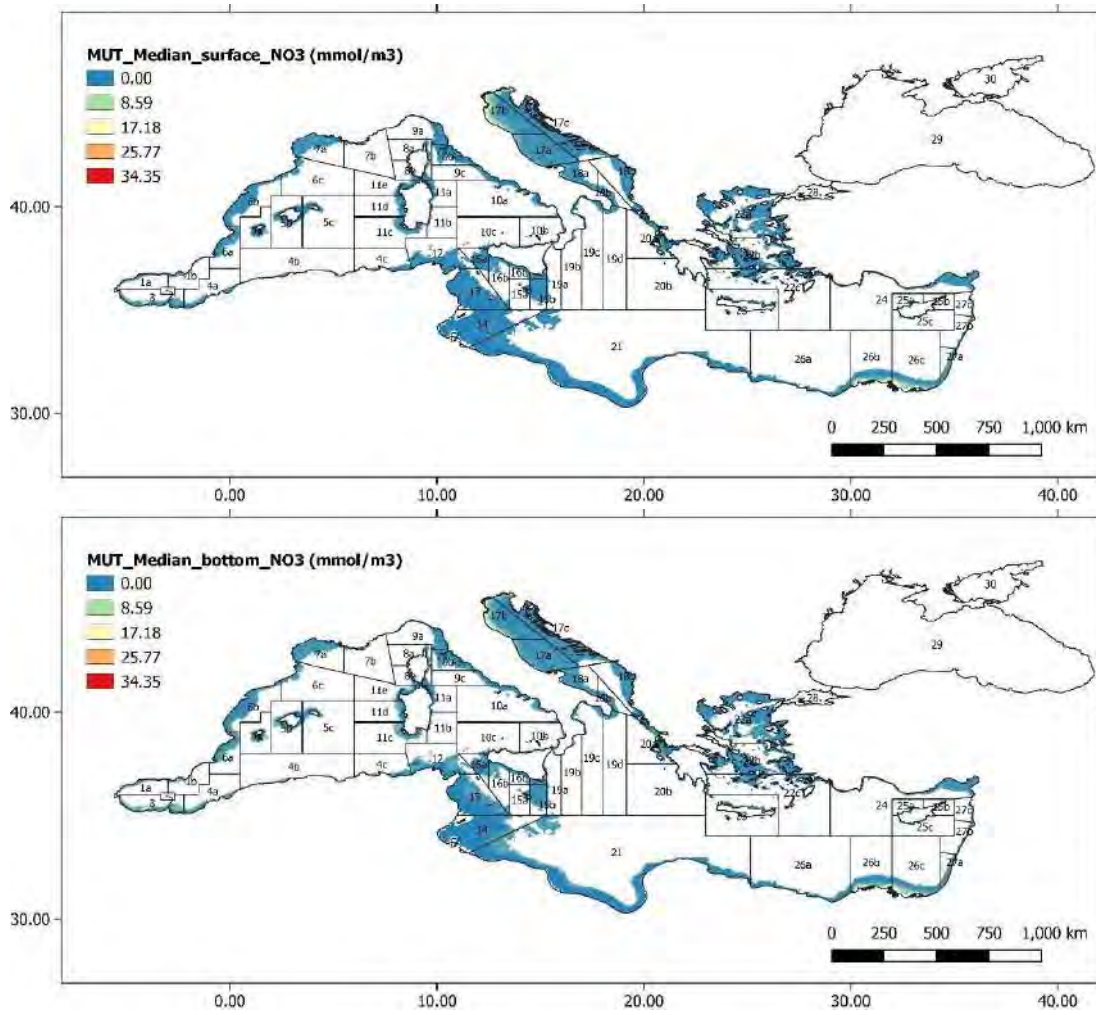
Annex 6 Figure 26: Median euphotic zone (upper panel) and bottom layer (lower panel) chlorophyll a concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



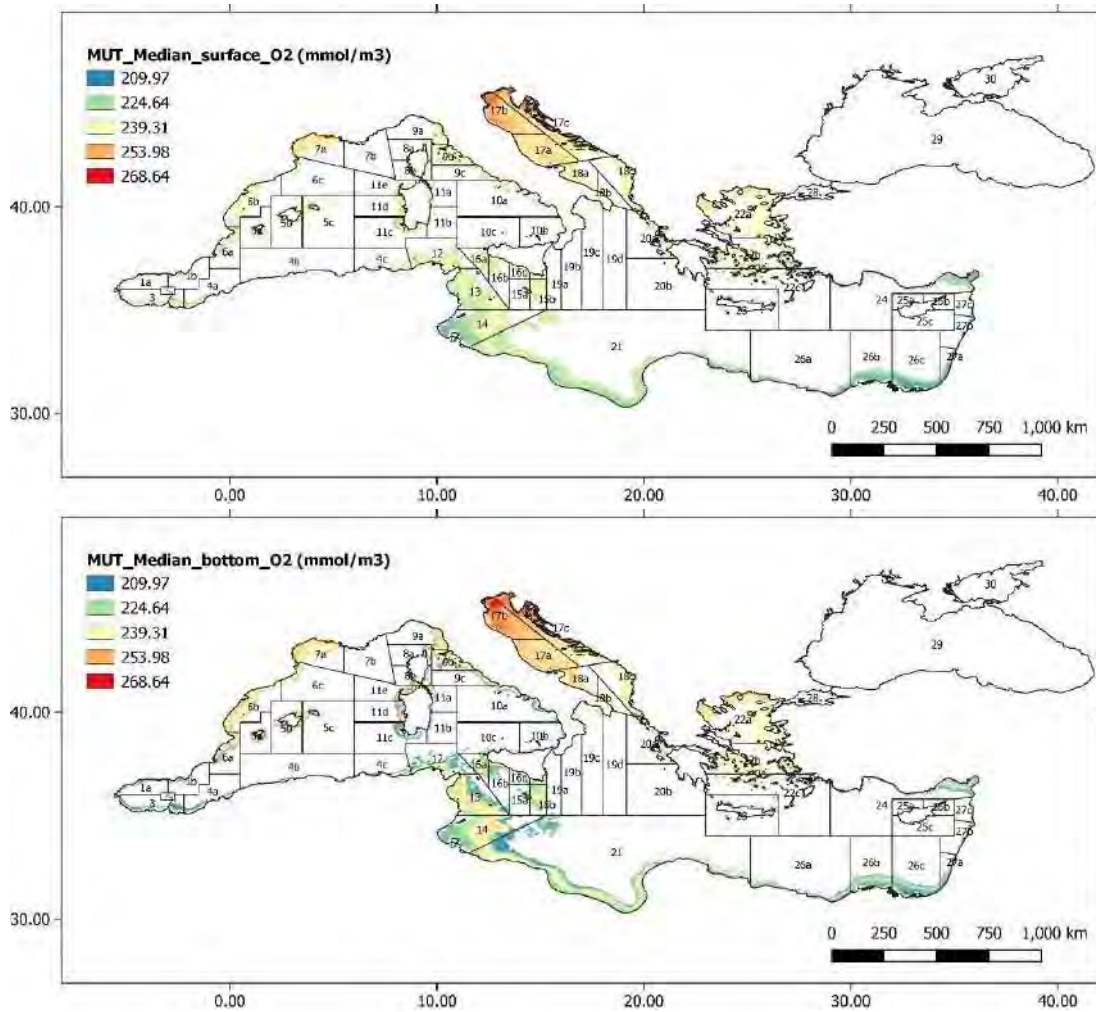
Annex 6 Figure 27: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved inorganic carbon concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



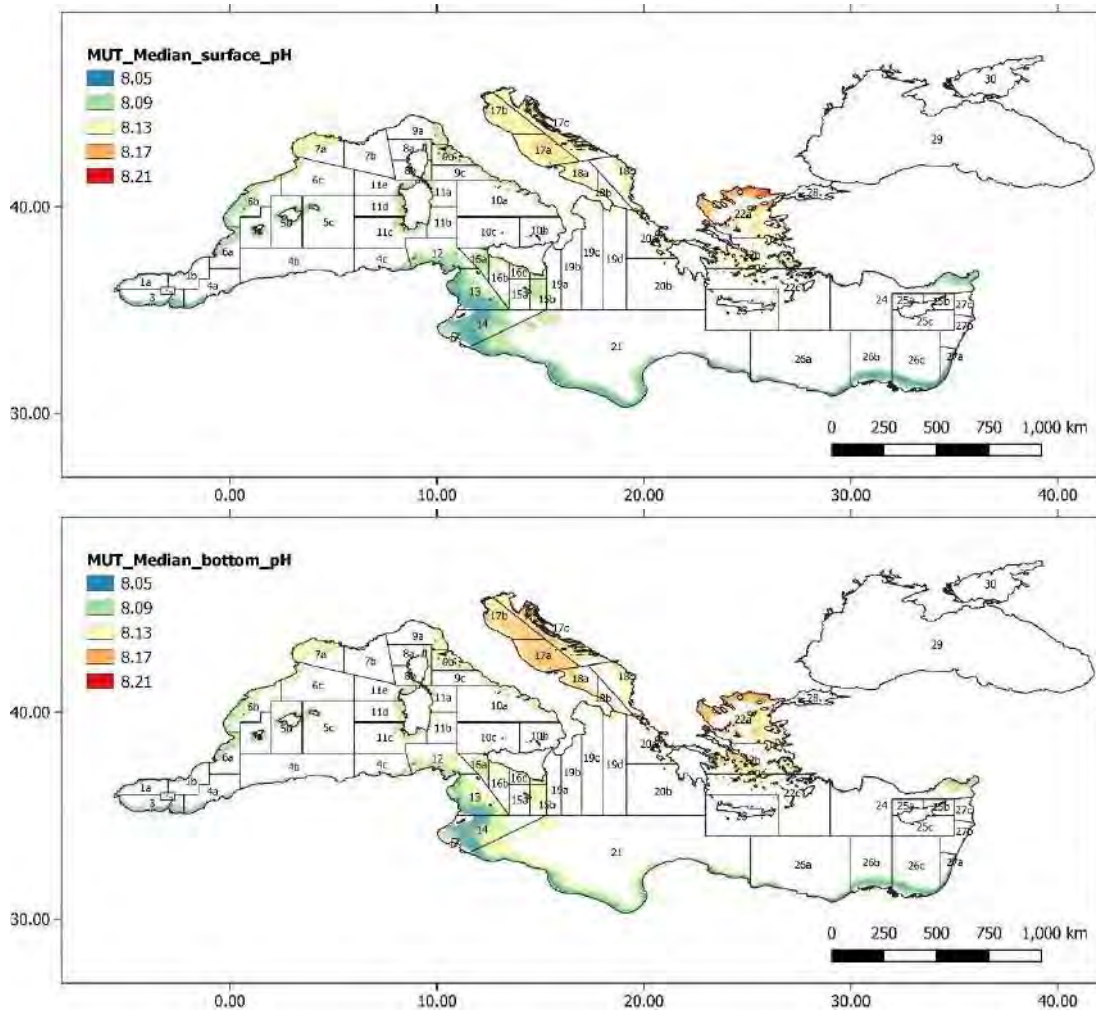
Annex 6 Figure 28: Median euphotic zone (upper panel) and bottom layer (lower panel) ammonium concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



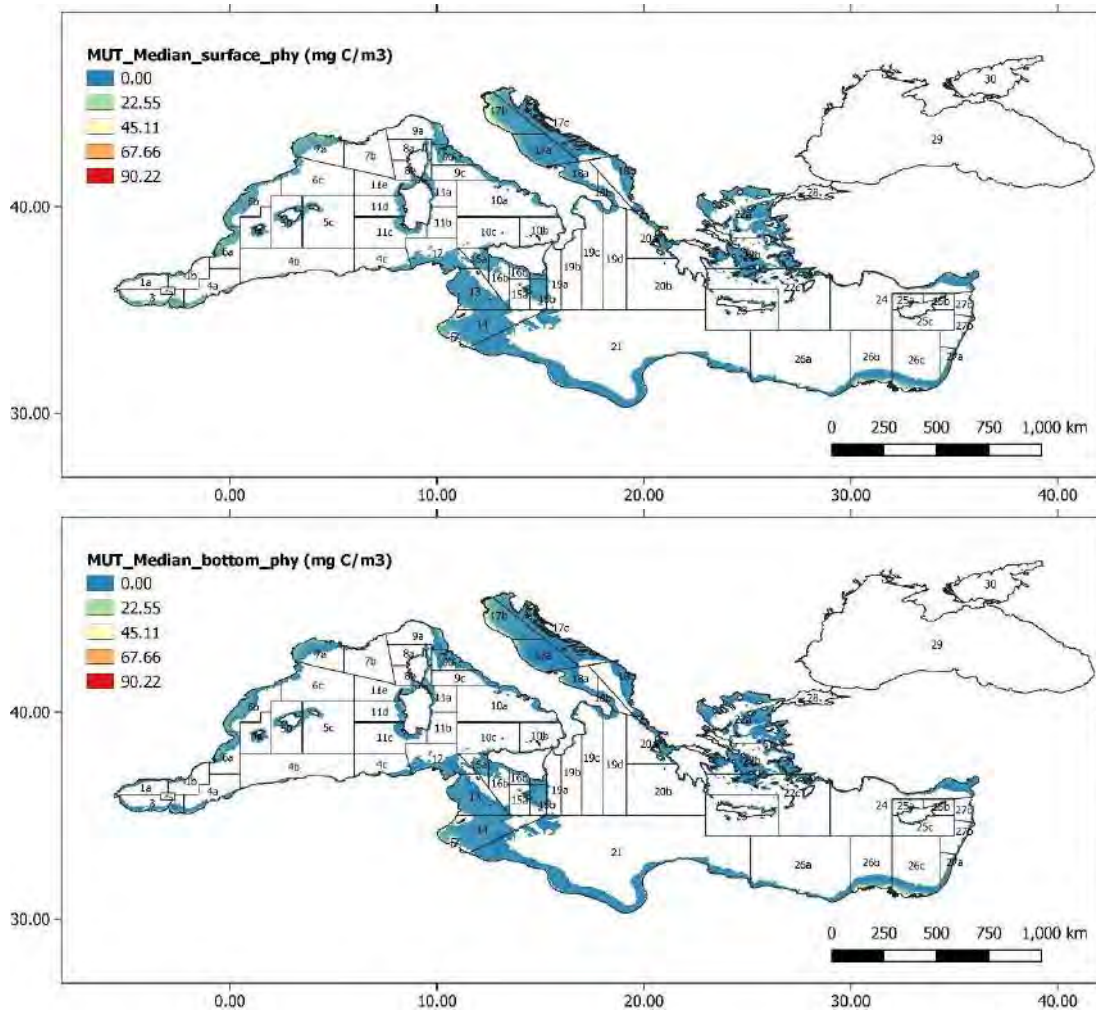
Annex 6 Figure 29: Median euphotic zone (upper panel) and bottom layer (lower panel) nitrates concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



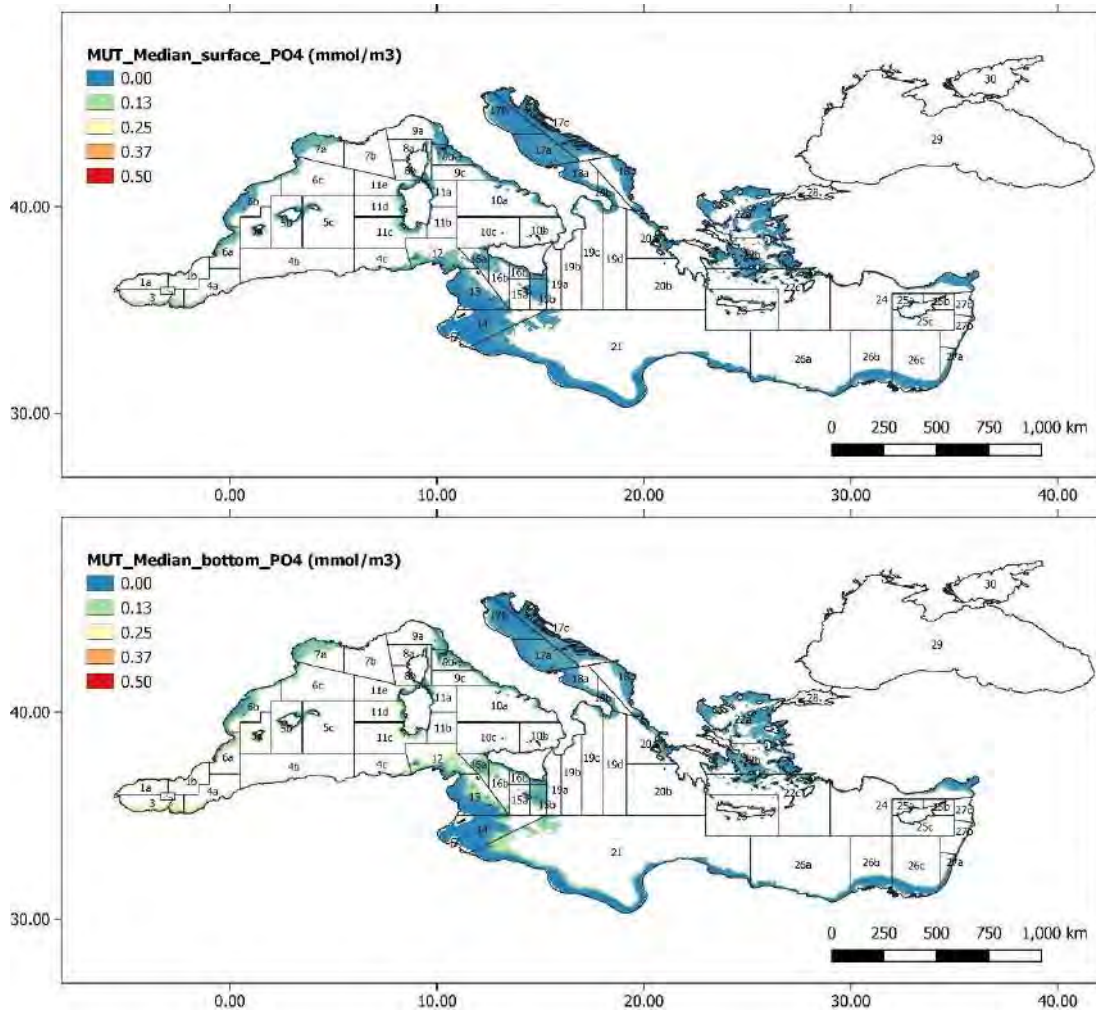
Annex 6 Figure 30: Median euphotic zone (upper panel) and bottom layer (lower panel) dissolved oxygen concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



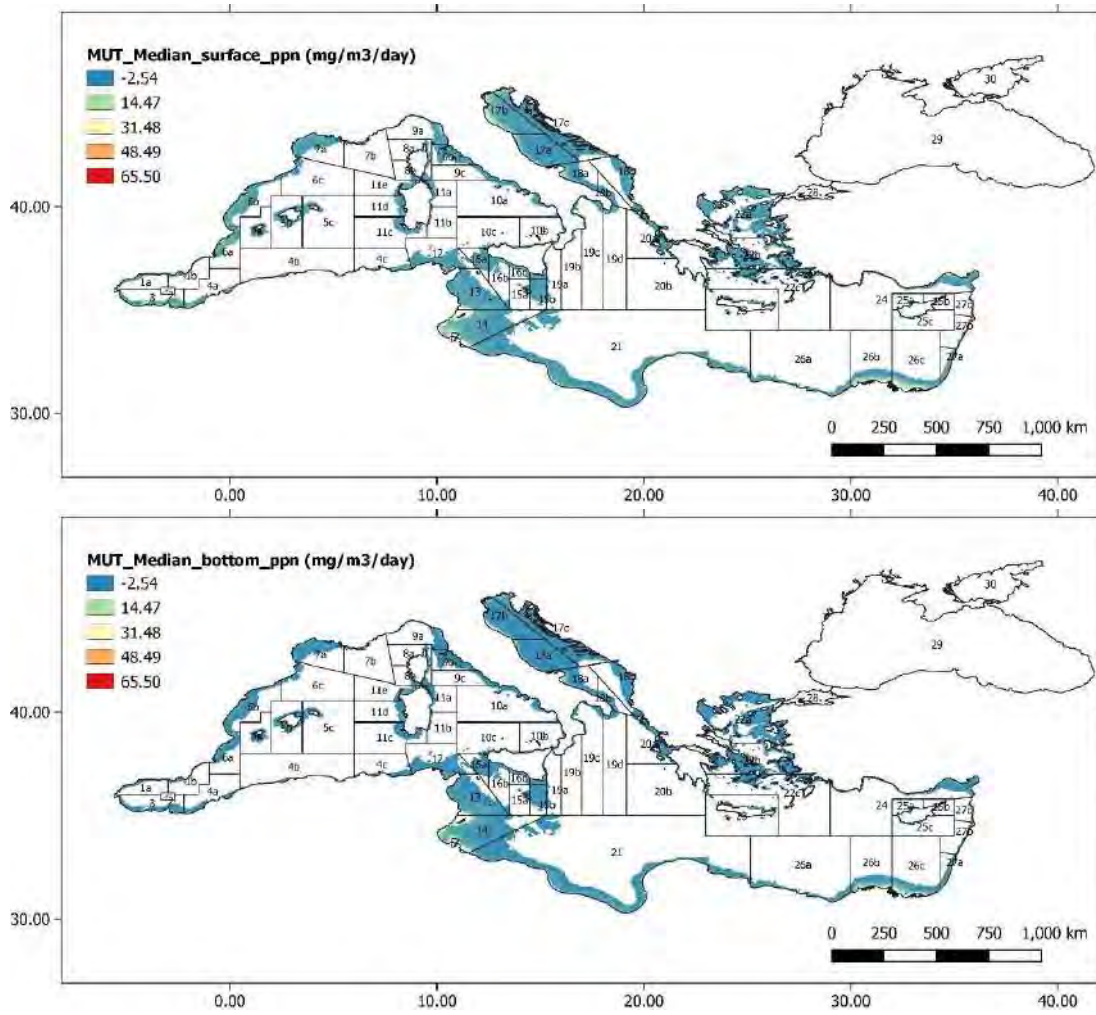
Annex 6 Figure 31: Median euphotic zone (upper panel) and bottom layer (lower panel) pH values over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



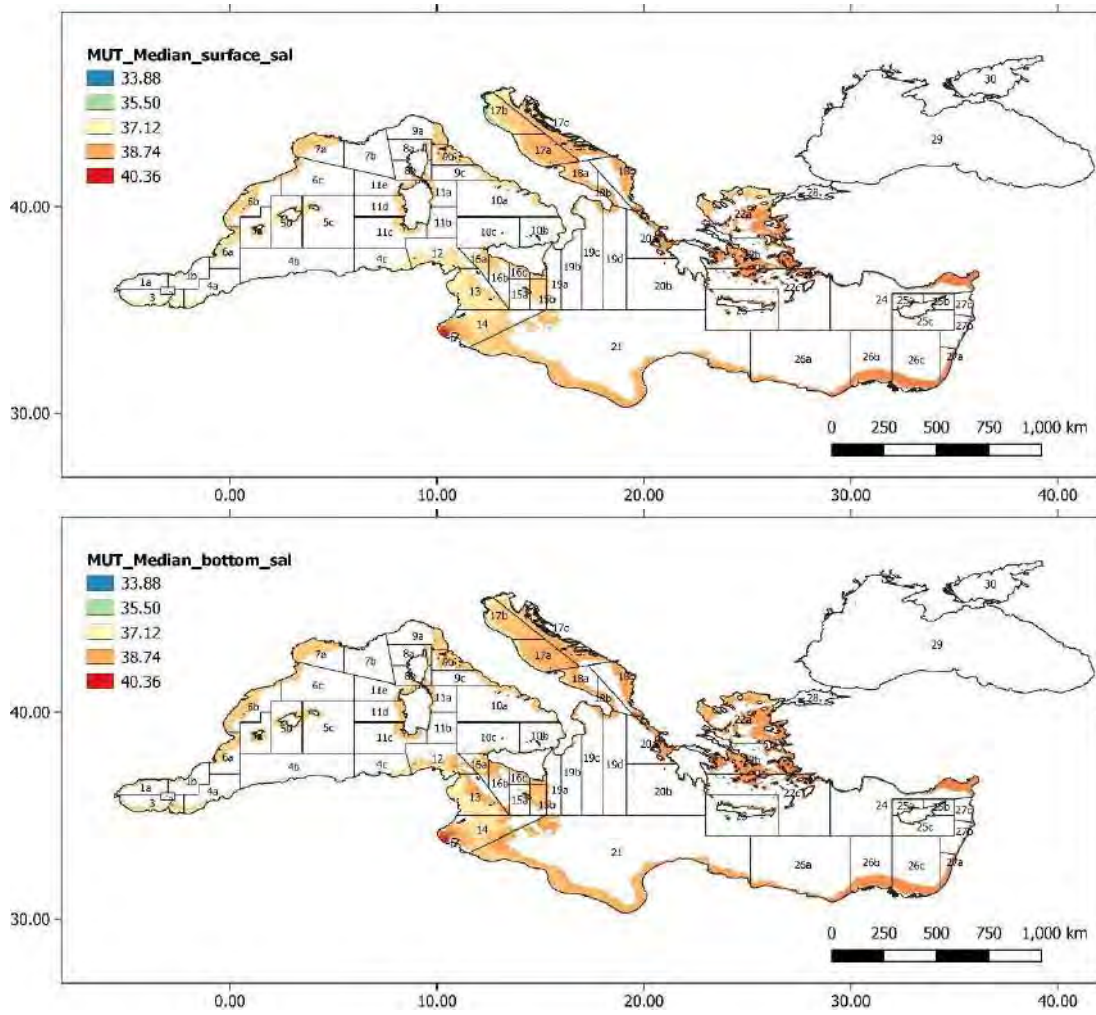
Annex 6 Figure 32: Median euphotic zone (upper panel) and bottom layer (lower panel) phytoplankton biomass concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



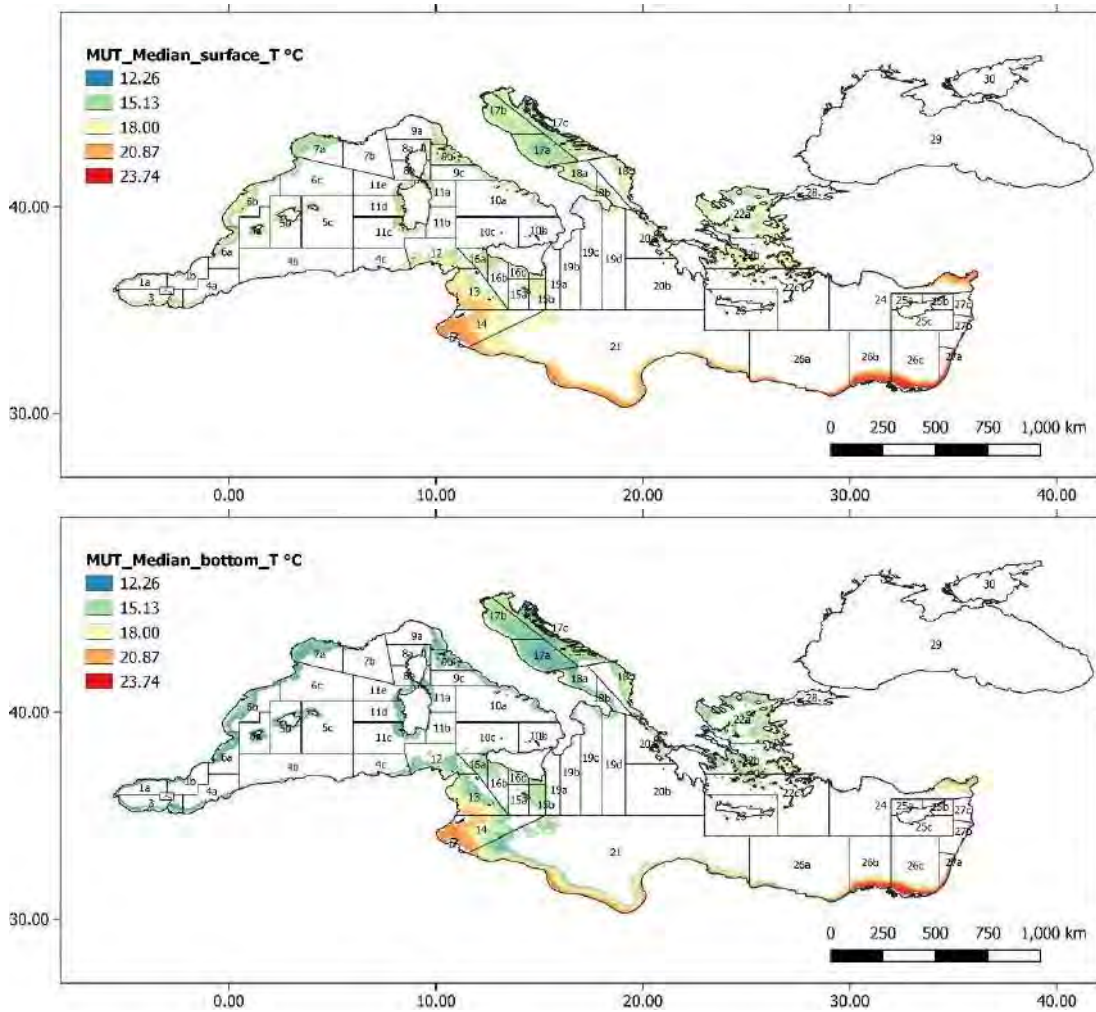
Annex 6 Figure 33: Median euphotic zone (upper panel) and bottom layer (lower panel) phosphates concentration over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



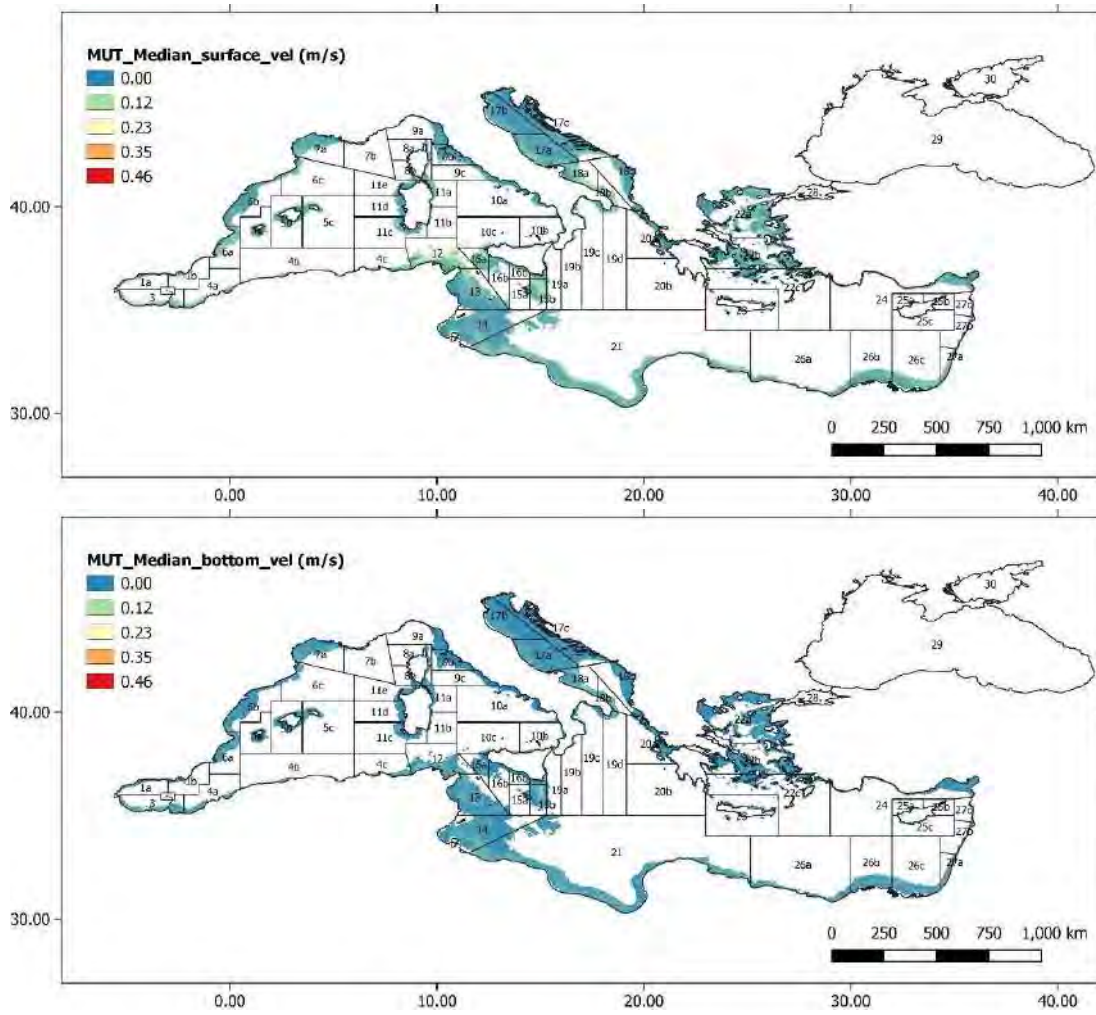
Annex 6 Figure 34: Median euphotic zone (upper panel) and bottom layer (lower panel) primary production over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



Annex 6 Figure 35: Median euphotic zone (upper panel) and bottom layer (lower panel) salinity over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



Annex 6 Figure 36: Median euphotic zone (upper panel) and bottom layer (lower panel) temperature over 5 years in the *Mullus barbatus* depth range of the Mediterranean.



Annex 6 Figure 37: Median euphotic zone (upper panel) and bottom layer (lower panel) current velocity over 5 years in the *Mullus barbatus* depth range of the Mediterranean.

Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201
– SC03- Project MED_UNITS

“Study on Advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species”

WP4 - SYNTHESIS AND PROPOSALS

Task 4.2 – Integrating results by different WPs and proposals of new management units

Deliverable 4.3 – Report with the maps combining population and management units

Responsible: Tommaso Russo (CoNISMa)

Compiled by: Tommaso Russo (CoNISMa)

27. 08. 2021

Version: Final

Index

Index	2
Executive summary	3
Introduction and Objectives	4
Workplan	5
<i>Merluccius merluccius</i>	7
<i>Nephrops norvegicus</i>	10
<i>Parapenaeus longirostris</i>	13

Executive summary

In this project, the quest of identifying fish management units has been addressed analysing multiple sources (e.g., genetic data or other data, such as fish otolith shape and microchemistry composition) of information. Individuals of the different species were collected, characterized and classified according to different stocks structure. The spatial origin of each individual was considered to reconstruct the spatial structure of the stocks configuration and to investigate the relationship of these stock structures with respect to the main environmental drivers. The obtained stock structures were then projected on the management partitioning of the Mediterranean Sea (i.e., FAO Geographical Sub Areas – GSA -or GSA subunits) in order to obtain a series of maps combining population, management units and the fishing footprint (i.e., the gear-specific patterns of the fishing effort).

For some species, it was not possible to identify a clear and reliable stock structure (e.g., when differences do not appear since the populations resulted homogenous for the investigated variables), whereas, for other species, some relevant differences arose and suggested the partitioning of the Mediterranean Sea in some sub-areas occupied by different stocks.

Here, the most reliable and comprehensive stock configurations by species, as identified in the Deliverable 4.2, were selected and a series of interrelated maps were prepared to show, in simple and possibly clear way, what we can reasonably say about spatial correspondence between stocks, management areas, and fishing grounds.

Introduction and Objectives

In this project, the quest of identifying fish management units has been based on the integrated analyses of information from multiple sources (e.g., genetic data or other data, such as fish otolith shape and microchemistry composition), for which divergent characteristics in the sub-populations can be observed and described. In particular, individuals of the different species were collected, characterized and classified according to different stocks structure. Finally, the spatial origin of each individual was considered to reconstruct the spatial structure of the stocks configuration and to investigate the relationship of these stock structures with respect to the main environmental drivers.

According to the aims of MEDUNITS project, the obtained stock structures were projected on the management partitioning of the Mediterranean Sea (i.e., FAO Geographical Sub Areas – GSA -or GSA subunits) in order to obtain a series of maps combining population, management units and the fishing footprint (i.e., the gear-specific patterns of the fishing effort) described in the Deliverable 3.1. – Maps on: (a) the fishing grounds in temporal and spatial scale, (b) hot and cold spots. Hence, the purpose of this Deliverable is to represent, in a synthetic and graphical way, the main results of the analyses presented and extensively discussed in the Deliverable 4.2 – Report with the maps of population units and discontinuities. As stated in the Inception Report, maps obtained for the population units of the target species should be analyzed in combination with the maps on the distribution of the fishing ground from WP3 in order to identify areas of potential joint management at sub-regional and regional level.

However, it is obvious that a clear and coherent pattern can be obtained only if genetic, morphological or chemical composition divergence is observed in a population and is associated with a spatial pattern. Thus, the lack of clear and reliable stock structure (e.g., when differences do not appear since the populations resulted homogenous for the investigated variables) prevents the possibility to corresponding spatial structure.

This document is not aimed at re-discussing or further develop the analyses and the results presented in the Deliverable 4.2 – Report with the maps of population units and discontinuities. The most reliable and comprehensive stock configurations by species, as identified in the Deliverable 4.2, were selected instead and a series of interrelated maps were prepared to show, in simple and possibly clear way, what we can reasonably say about spatial correspondence between stocks, management areas, and fishing grounds. For some species, it was not possible to recognize a clear stock configuration from the Deliverable 4.2. This was the case of *Mullus barbatus* for which either no diversification on the scale of the Mediterranean do exist, or more probably, the diversification is on a local scale that could not be captured by the sampling resolution of the present study, or *A. antennatus* that cannot be divided in different stocks in the Mediterranean Sea. Therefore, maps were prepared only for *Merluccius merluccius* (HKE), *Nephrops norvegicus* (NEP), and *Parapenaeus longirostris* (DPS) for which different stock units were identified in D4.2.

Workplan

The Task 4.2 delivered maps of the delineated population units by species on the basis of biological information (genetic and otolith results) from WPs 1-2. In addition, the results from biological units (WPs 1-2) and those from management units (WP3) were combined in the Task 4.2. Results of biological analysis (Genetic data produced in WP1, Otolith Shape and Otolith microchemistry in WP2) were analyzed with multivariate analysis techniques also by including environmental variables in order to obtain a division in separate, yet possibly overlapping fish population units. Fuzzy clustering methods (Bezdek, 1981) was used since they provide a separation of data based on their similarity, yet with also a different membership linking each sample to each identified cluster/population unit.

For each species, each individual was assigned with the fuzzy k-means algorithm with different memberships to all clusters, and the mean membership grade for each cluster over all the individuals belonging to each GSA Subarea and to each GFCM subregions was calculated. Thus, each GSA subarea and GFCM subregion was assigned with a different membership grade, similarly to the fuzzy assignments of each individual, to the clusters. In this way it was possible to assess the spatial characterization of the clusters, as well as the strength of the association of each subarea/subregion to the clusters.

As a result, the Task 4.2 delivered a series of tables such as the one represented in the Fig. 1. Each of these tables contains, for each row, the probabilities of membership of a given subarea to each of the “candidate stocks” identified through the integrated analysis of different data sources (e.g., genetic plus otolith shape). Although it is always possible to identify, for each subarea, the winner stock/cluster (that is the one with the highest value of probability), it is worth noting that almost never happens that a subarea is associated to a given stock/cluster with more than 50% of the probability.

SUBAREA	WINNER STOCK CLUSTER	STOCK/CLUSTER #1	STOCK/CLUSTER #2	STOCK/CLUSTER #3
6B	1	0.40	0.22	0.37
9B	1	0.41	0.26	0.32
11C	1	0.43	0.21	0.35
11E	1	0.43	0.23	0.33
16B	1	0.37	0.26	0.35
18A	1	0.38	0.27	0.34
18C	1	0.39	0.28	0.31
19C	1	0.35	0.29	0.34
19D	1	0.43	0.20	0.36
19B	3	0.30	0.30	0.39
17A	1	0.37	0.26	0.35
17C	1	0.34	0.32	0.32

Figure 1 – The first rows of the table representing the probabilities of membership of the different subareas to the three stocks/clusters identified for *Merluccius merluccius* through the combined analysis of genetic and otolith shape analysis.

This implies that forcing the classification of subareas with a score lower than 50% could lead to the paradoxical situation in which, for example (see Fig. 1), the subarea 6B will be assigned to Stock #1 but the probability that this subarea does not belong to Stock #1 is higher than

50%). For this reason, and for the sake of precaution, the two stocks with the highest probability values (always summing up to more than 50%) were considered for each subarea and graphically combined (i.e. using a color scale) to represent the spatial distribution of stocks by subarea. At a successive level, subareas were combined into GSA and the same procedure was applied.

This precautionary graphical approach was selected to adequately take into account the extensive discussion provided in the Task 4.2. In fact, quoting the Deliverable 3.2, *“different fish and shellfish population units live in areas characterized by different values of driving environmental variables, which in turn can influence the genetic differentiation of the populations, as well as their behavior. Thus, environmental discontinuities might help in distinguishing different stocks. Individuals of different population units may migrate between areas or they may belong to intermediate or possibly under-sampled stocks, thus showing characteristics of more than one population unit.”*

Here, the word “precautionary” should be interpreted as follows. We want to recognize the spatial structures of stocks to better understand how to manage their exploitation and monitoring, pursuing the final goal of sustainability. But given that exploitation means the application of an external mortality (i.e., fishing) to the stocks, it is crucial to identify, if possible, a direct correspondence between fishing grounds and stocks. When it is not possible to univocally attribute a fishing ground (or subarea or GSA, which in general contain more than a fishing ground) to one and one only stock, it is appropriate to consider that the fishing mortality applied in that area can reasonably affect more than a stock.

Merluccius merluccius

According to the Deliverable 4.2:

- Three sources of information, with different levels of spatial coverage of samples with respect to the Mediterranean Sea, were used to investigate the stock structure for this species:
 - Genetic data;
 - Otolith shape;
 - Otolith microchemical composition.
- The Genetic and Otolith shape datasets are those that have the highest numbers of individuals analyzed and represent a high number of sampling locations and subareas, whereas the dataset of Otolith microchemical composition is heavily limited from the point of view of the spatial coverage;
- Different combinations of these input data were inspected, and each combination returned different stock configurations, each of which has been assessed in terms of spatial coverage (where the input dataset with the lower spatial coverage determined the final spatial coverage of the combined model) and amount of variability explained;
- The results obtained from the combination of Genetic data and Otolith shape (without considering the Otolith microchemical composition) offered the highest number of variables considered together at the highest possible spatial coverage. More in details, it was concluded that the configuration with three stocks has a better spatial rationale;
- The higher spatial resolution of the results obtained from the combination of Genetic data and Otolith shape favors a division in three populations:
 - One, and clearly separated is the Eastern Mediterranean.
 - The other two are less well solved, but we may hypothesize that one is the westernmost, possibly related to Atlantic hake populations, and extending along the Alboran and Balearic sea and along the African coast up to the Sicily channel. The other is more characterizing the central-northern part of the Mediterranean, in particular along the Italian coasts, including the Adriatic Sea.
- The spatial coverage of the results obtained from the combination of all the input datasets was not enough to adequately solve these three populations: while the eastern is clearly separated, the central is limited to the Adriatic and Ionian only, while all the rest is brought together into one western Mediterranean cluster. This result also reinforces the idea that the possible two populations of the Western and Central Mediterranean are more intermixed and closer one to the other, than the one in the Eastern Mediterranean.
- Thus, it seems that the main difference is between the East Mediterranean on one side, and all the rest subareas on the other.
- From the result of the analysis performed in the Task 4.2, it was concluded that there is some differentiation in the hake populations of the Mediterranean. There is

a difference between the Eastern Mediterranean hake populations and the populations of the rest of the basin; as well as some indications as to a further differentiation in the Western-Central-Adriatic population. The observed fuzziness might be due to the mobility of the species in its different life phases, and to the absence of strong discontinuities in its habitat. These results come predominantly from the analysis of the two more robust datasets: the combination of Genetic data and Otolith shape, with the highest spatial coverage for the highest number of variables, and the all datasets, i.e., the combination of all three datasets, even though on a reduced sampling grid.

The Fig. 2 represents the spatial configuration of stocks with respect to the FAO GSA partitioning of the Mediterranean Sea. A color was assigned to each of the three stocks identified through the combination of Genetic data and Otolith shape, and each GSA was filled combining these colors and using as weight of each color the values of probability corresponding to different stocks. Only the two stocks with the higher values of probability were considering for each GSA in order to emphasize the stocks' structure.

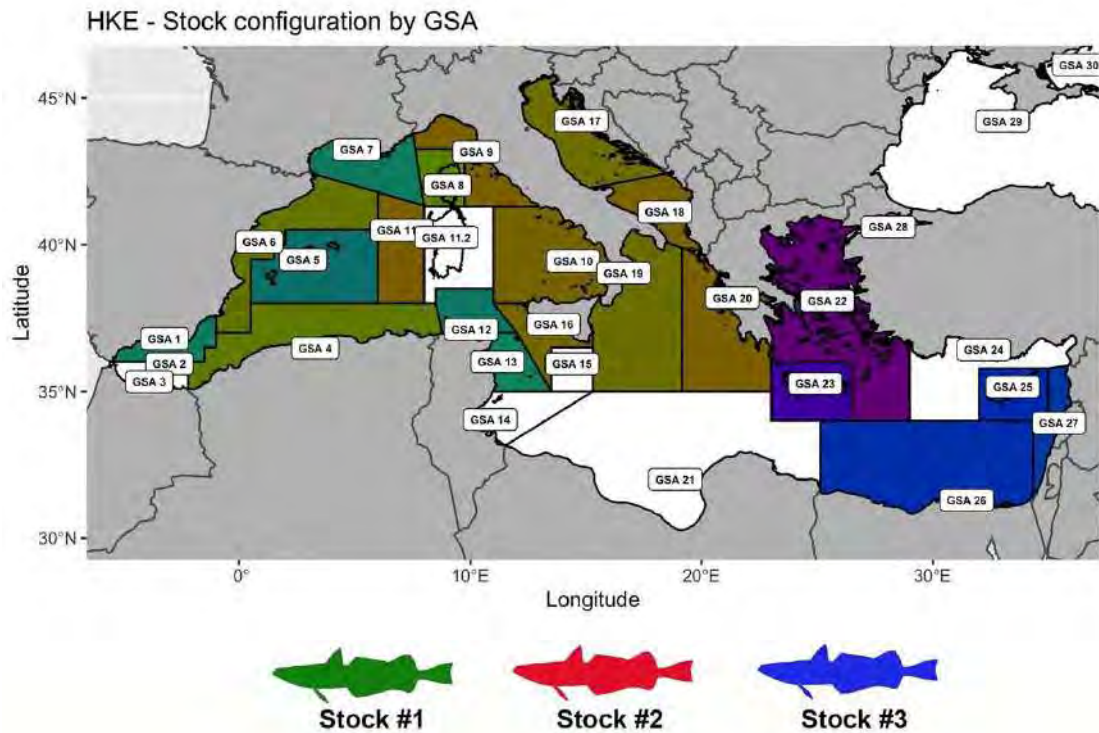
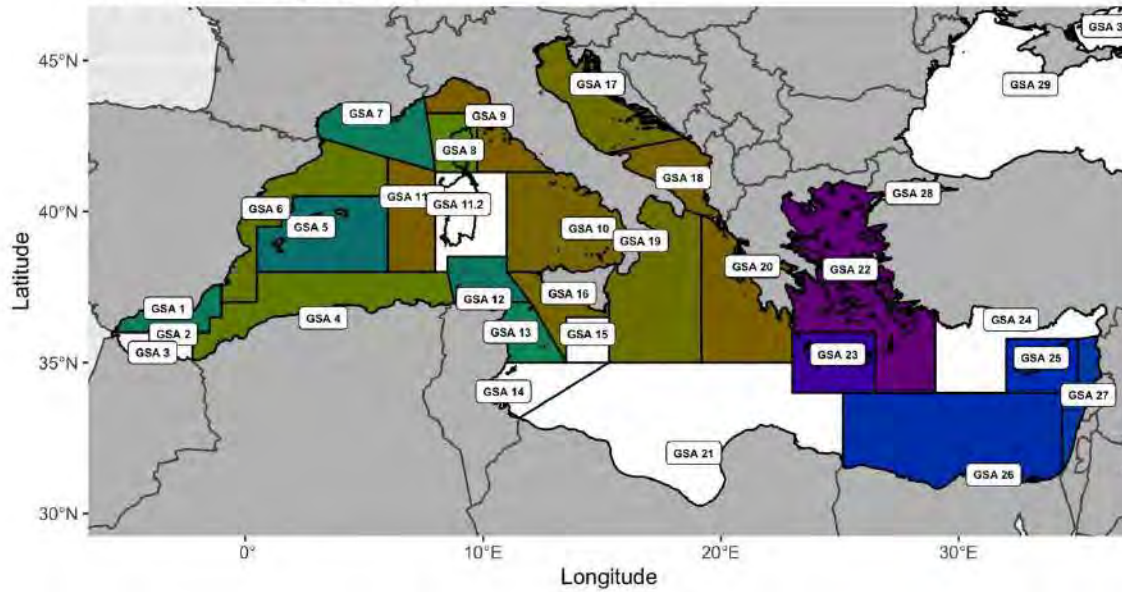


Figure 2 – Representation of the FAO GSA filled with colors generated considering the probability of membership of hake samples to the three stocks identified through the analysis of both Genetic data and Otolith shape data. The color of each GSA was obtained mixing green (stock #1), red (stock #2) and blue (stock #3) using the values of probability as weights.

The Fig. 3 represents the same spatial configuration of stocks with respect to the FAO GSA partitioning of the Mediterranean Sea (top panel), but integrate also, for comparative purposes, the fishing footprint (total fishing effort in days, for the year 2017) for trawlers and longlines, as obtained from the analysis of AIS data (see the Deliverable 3.1 for details).

HKE - Stock configuration by GSA



Stock #1



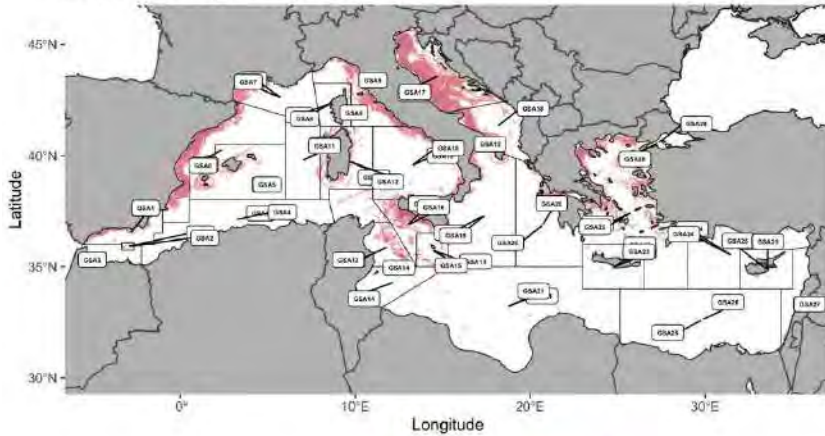
Stock #2



Stock #3



AIS - Trawl fishing effort - year 2017



AIS - Longlines fishing effort - year 2017

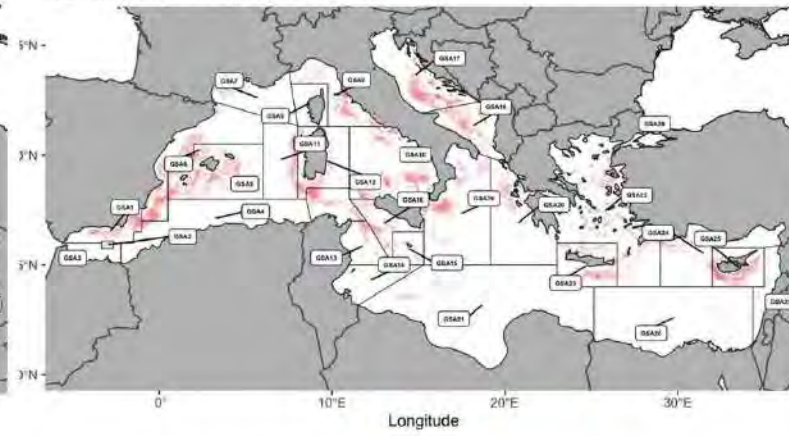


Figure 3 – Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of hake samples to the three stocks identified through the analysis of both Genetic data and Otolith shape data. The color of each GSA was obtained mixing green (stock #1), red (stock #2) and blue (stock #3) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1); Bottom right panel: Fishing footprint (total fishing effort in days, for the year 2017) for longliners from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1)

Nephrops norvegicus

Genetic data were used to investigate the stock structure for this species in the Mediterranean Sea. According to the Deliverable 4.2:

- The results of the fuzzy k-means algorithm and the application of cluster validity indexes indicated that the genetic diversification in *N. norvegicus* is not very strong;
- The spatial distribution in function of GFCM subregions did not give a clear picture; In particular the West and Center subregions are those more distributed among the clusters, while the Adriatic and the East subregions show a clear preference for a specific cluster;
- The interpretation of these results for the Eastern basins has to take into account that only two subareas were sampled in this subregion. We can infer that the fragmentation of the populations is greater in the Western and the Central basin of the Mediterranean, while the Adriatic shows a consistent homogeneity of the population of *N. norvegicus*;
- The most solid result is the separation of the Adriatic populations from the others. Results also indicate the possible coexistence of slightly different sub-populations of *N. norvegicus* in the Adriatic. Nevertheless, the spatial density of sampling locations it is not detailed enough to make further hypothesis on this.
- While the most solid result is the separation of the Adriatic from the Western Mediterranean it has to be noted that this happens only with a number of clusters (stocks) equal to four, not before. This means that the three clusters identified in the Western Mediterranean show a level of differentiation that is the same as the one shown by the Adriatic samples compared to the rest. Thus, while the spatial distribution of the three clusters in the Western Mediterranean seems less easily interpretable, perhaps because of a sparser sampling design in this area is not able to adequately capture this distribution, all three these clusters should be considered if the Adriatic one is considered.
- Globally, the results indicated that the divergences in the *N. norvegicus* genetic data are not very strong. Nevertheless, the partitioning in four stocks was strongly related to spatial and environmental features of the Mediterranean, supporting the case for different stocks.

The Fig. 4 represents the spatial configuration of stocks with respect to the FAO GSA partitioning of the Mediterranean Sea. A color was assigned to each of the four stocks identified through the analysis of Genetic, and each GSA was filled combining these colors and using as weight of each color the values of probability corresponding to different stocks. Only the two stocks with the higher values of probability were considering for each GSA in order to emphasize the stocks' structure.

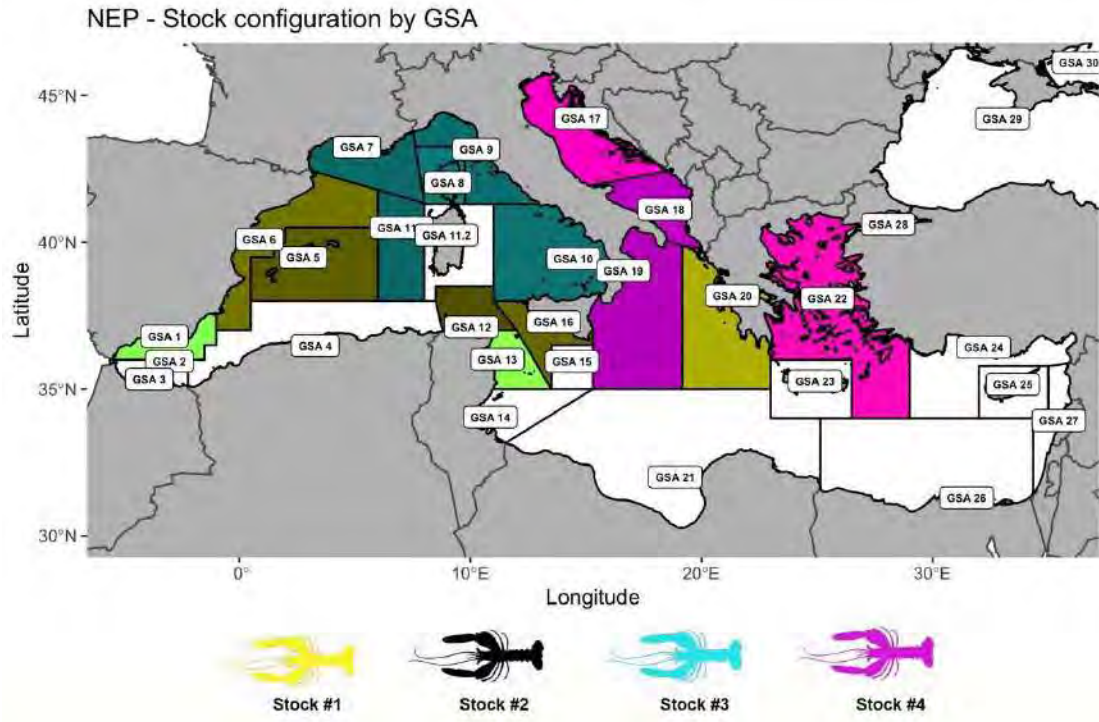
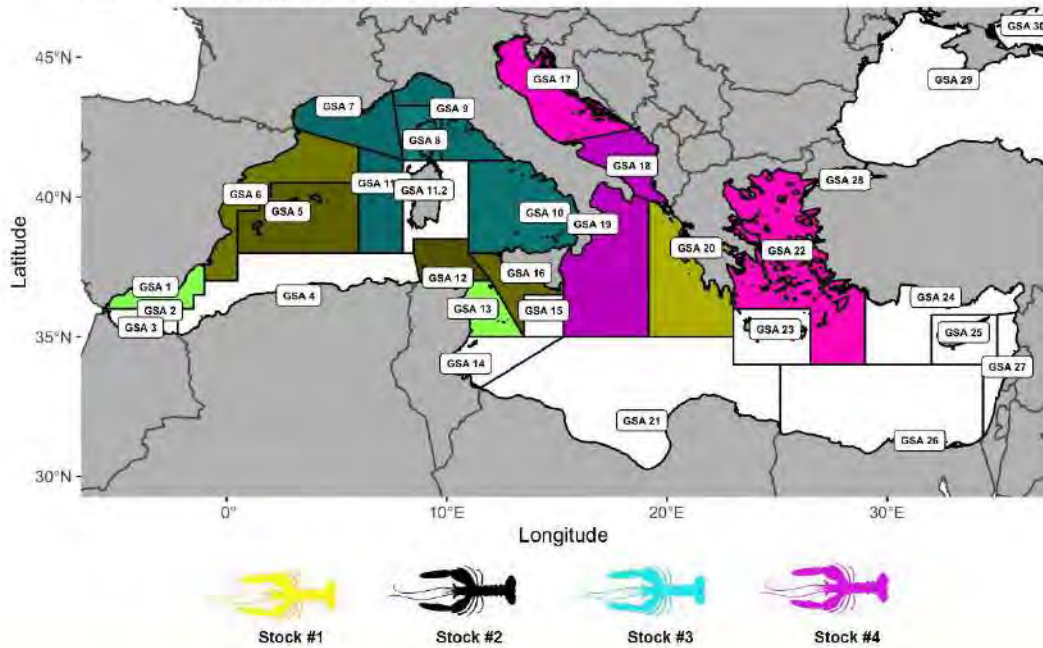


Figure 4 – Representation of the FAO GSA filled with colors generated considering the probability of membership of Norway lobster samples to the four stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing yellow (stock #1), black (stock #2), cyan (stock #3), and magenta (stock #4) using the values of probability as weights.

The Fig. 5, in which the top panel replicates Fig. 4, allows to compare the spatial configuration of stocks with the distribution of fishing effort exerted by trawlers.

NEP - Stock configuration by GSA



AIS - Trawl fishing effort - year 2017

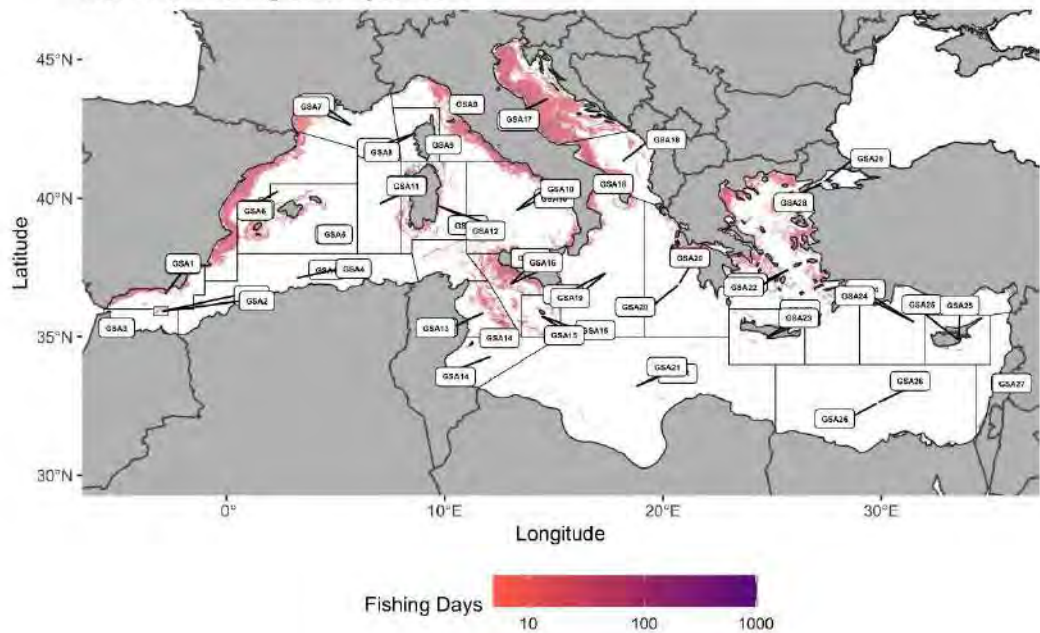


Figure 5 - Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of Norway lobster samples to the four stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing yellow (stock #1), black (stock #2), cyan (stock #3), and magenta (stock #4) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).

Parapenaeus longirostris

Genetic data were used to investigate the stock structure for this species in the Mediterranean Sea. According to the Deliverable 4.2:

- The results clearly show the existence of two neatly separated sub-populations living one in the Western and most of the Central Mediterranean, and the other in the Eastern basin and the Adriatic Sea. Spatial and environmental analysis may further reinforce this outcome. The general pattern of current-mediated dispersal in the depth range of this species and the complex bathymetric and morphological features of the area of the Strait of Sicily-Strait of Messina may be responsible for the observed separation of *P. longirostris* population into two sub-populations;
- The analysis of spatial components shows that the results obtained on the genetic PCAs of *P. longirostris* are rather consistent and robust;
- The results of the internal and external validation procedure were in strong agreement. Thus, for this species to a clear genetic diversification in sub-populations corresponds a clear spatial and environmental diversification of their habitats, i.e.: different stocks do exist in the Mediterranean for *P. longirostris*;
- It was clear already from the internal cluster validity indexes applied that there is a clear separation in the genetic data in two sub-population. The fuzzy membership grades obtained showed clear preferences for one of the clusters in most cases;
- Hence, two stock of *P. longirostris* do co-exist in the Mediterranean, one found mostly in the Western and Central Mediterranean, the other in the Adriatic and in the Eastern Mediterranean. The discontinuity between the two seem located in the 19b subarea of the Ionian. Possible causes of such diversification might be the current patterns influencing dispersal in the depth range of this species, and the complex bathymetry and morphology characterizing especially the area where the discontinuity between the two sub-populations does occur, i.e., between the Strait of Sicily and the Ionian Sea.

The Fig. 6 represents the spatial configuration of stocks with respect to the FAO GSA partitioning of the Mediterranean Sea. A color was assigned to each of the two stocks identified through the analysis of Genetic, and each GSA was filled combining these colors and using as weight of each color the values of probability corresponding to different stocks.

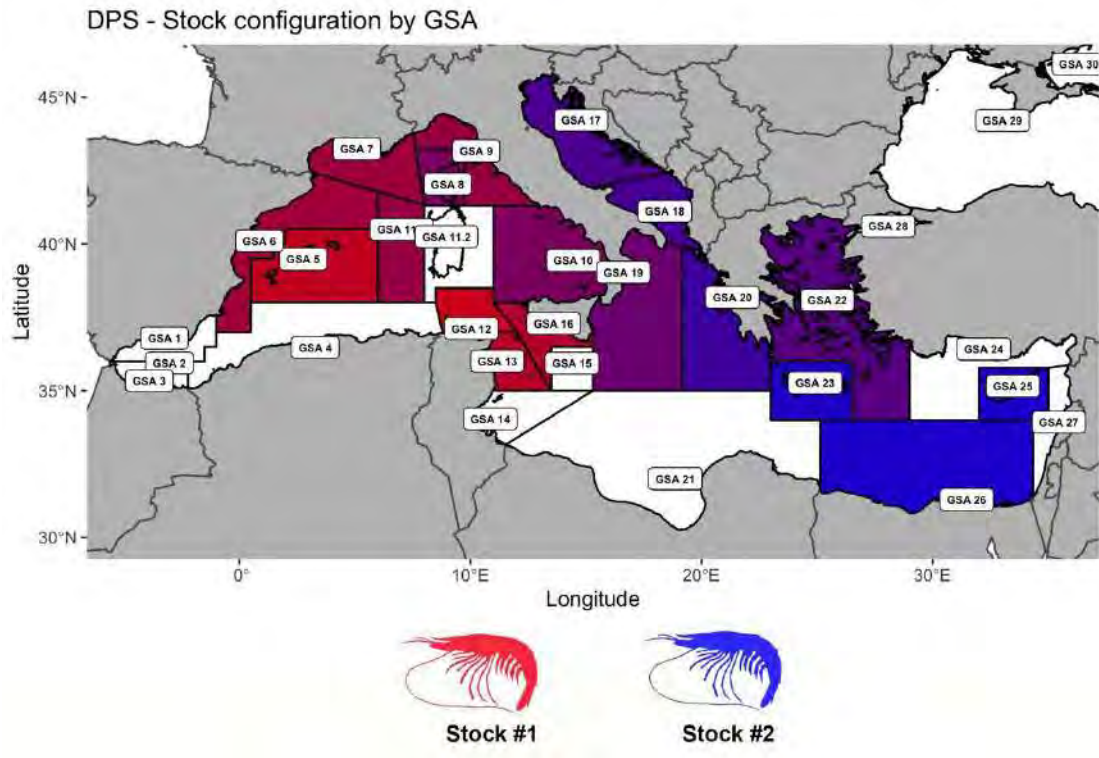
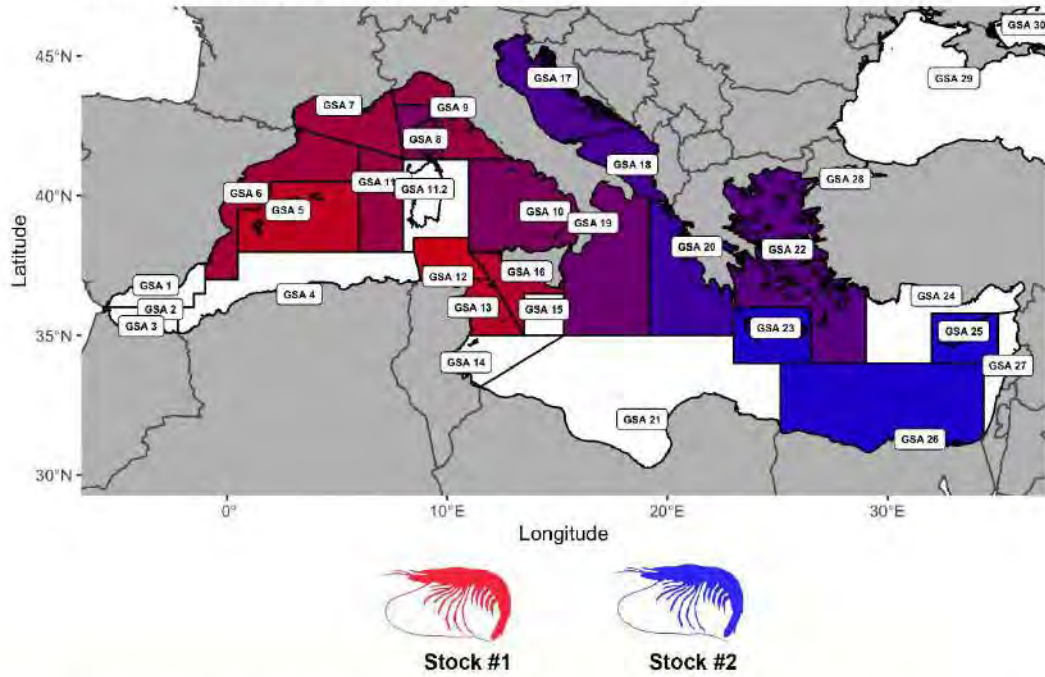


Figure 6 – Representation of the FAO GSA filled with colors generated considering the probability of membership of Deep water rose shrimp samples to the two stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing yellow (stock #1), black (stock #2), cyan (stock #3), and magenta (stock #4) using the values of probability as weights.

The Fig. 7, in which the top panel replicates Fig. 6, allows to compare the spatial configuration of stocks with the distribution of fishing effort exerted by trawlers.

DPS - Stock configuration by GSA



AIS - Trawl fishing effort - year 2017

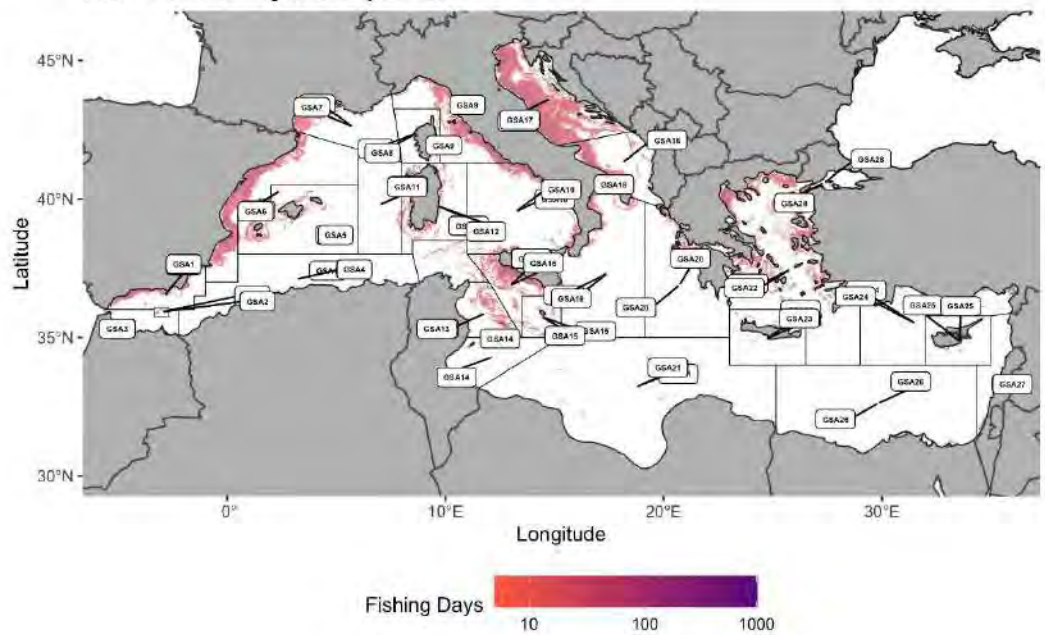


Figure 7 - Top panel: Representation of the FAO GSA filled with colors generated considering the probability of membership of Deep water rose shrimp samples to the two stocks identified through the analysis of Genetic data. The color of each GSA was obtained mixing red (stock #1), and blue (stock #2) using the values of probability as weights; Bottom left panel: Fishing footprint (total fishing effort in days, for the year 2017) for trawlers from AIS data, as obtained using VMSbase platform and the related procedures (see Deliverable 3.1).

Specific Contract No. 03EASME/EMFF/2017/1.3.2.3/01/ SI2.793201 – SC03-
Project MED_UNITS

*“Study on Advancing fisheries assessment and management
advice in the Mediterranean by aligning biological and
management units of priority species”*

WP4 – SYNTHESIS AND PROPOSALS

***Task 4.3 Future improvements for developing adaptive spatial
fisheries management***

D 4.4 Report on the scenario modelled to explore the potential of developing adaptive spatial fisheries management

Responsible: Fabio Fiorentino

Compiled by:

Fabio Fiorentino (CNR – IRBIM)

Tommaso Russo (CoNISMa – Uniroma2)

Isabella Bitetto (COISPA)

Date

July 2021

Version 1

Index

Executive summary	6
Introduction	7
Objectives.....	8
Workplan.....	8
Materials and methods	9
BEMTOOL.....	9
The model	9
Parameterization of BEMTOOL in the hindcasting mode for CS1 and CS3	10
Forecast parameterization of BEMTOOL: assumptions for projections	14
SMART.....	20
Application of the SMART model	20
Results.....	23
BEMTOOL.....	23
CS1 - ARS GSAs 1-12 and GSAs ARA 1-12	23
CS3 – DPS GSAs 17-18-19-20	27
SMART.....	31
CS1 - ARS GSAs 1-12 and GSAs ARA 1-12	31
CS2 - ARS GSAs 15,16,18,19 and 20.....	32
A tentative roadmap with actions required in order to develop adaptive spatial fisheries management	33
Discussion and conclusion.....	35
Deviation from the Workplan.....	37
Remedial actions	37
References	37

Tables index

Table 1 Case Studies selected for task 4.3 of the MEDUNITS project. Stock configurations, justifications, management scenarios and adopted models by case studies are reported. CS1: Reference years Western Mediterranean MAP 2015-2017; Transition year 2020; closures* are the ones of EWG 20-13. CS2: Reference years Adriatic MAP 2015-2018; closures** are the ones of GFCM Recommendation CS3: Reference years for Central Mediterranean 2016-2018; closures*** are those reported in the EU MANTIS project.	8
Table 2 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Results of the assessments from WP3 relevant for BEMTOOL parameterization. The computation of the reduction by stock to reach F0.1 is also reported.	11
Table 3 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12 Fleet segments included in the BEMTOOL simulations and forecast scenarios by GSA for vessel length (VL) segments.	12
Table 4 Case study CS3 - DPS GSAs 17-20: Results of the assessments from WP3 relevant for BEMTOOL parameterization. The computation of the reduction by stock to reach F0.1 is also reported.	13
Table 5 Case study CS3 - DPS GSAs 17-20 Fleet segments included in the BEMTOOL simulations and forecast scenarios by GSA for vessel length (VL) segments.	14
Table 6 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Parameters of the stock recruitment relationships for ARS and ARA.	14
Table 7 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Effort reductions applied by year and fleet segment for scenarios S1, S2, S3 and S4 respect to S0. In the first column are the fishing days in 2020, assumed in line with EU Regulation 2236/2019.	17
Table 8 Case study CS3 - DPS GSAs 17-18-19-20: Parameters of the stock recruitment relationships for CS3 for DPS.	18
Table 9 Case study CS3 - DPS GSAs 17-18-19-20: Effort reductions applied by year and fleet segment for scenarios S1 and S2 respect to S0. In the first column are the fishing days in 2021, assume in line with Recommendation GFCM/43/2019/5.	19
Table 10 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: F, SSB and overall landing (discard null) of ARS and ARA corresponding to the S0 (SQ) scenario and percentage of variation of the S1, S2, S3 and S4 scenarios respect to SQ in 2026.	24
Table 11 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: Total revenues by fleet segment corresponding to the S0 (SQ) scenario and percentage of variation of the Scenarios S1, S2, S3 and S4 respect to SQ in 2026.	26
Table 12 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: CR/BER by fleet segment corresponding to the S0 (SQ), S1, S2, S3 and S4 scenarios in 2026.	27
Table 13 Case study CS3 - DPS GSAs 17-18-19-20: F, SSB and overall catch of DPS corresponding to the S0 (SQ) scenario and percentage of variation of the S1 and S2 scenarios respect to SQ in 2026.	28
Table 14 Case study CS3 - DPS GSAs 17-18-19-20: Total revenues by fleet segment corresponding to the S0 (SQ) scenario and percentage of variation of the S1 and S2 scenarios respect to SQ in 2026.	30
Table 15 Case study CS3 - DPS GSAs 17-18-19-20: CR/BER by fleet segment corresponding to the S0 (SQ) scenario, S1 and S2 scenarios in 2026.	30
Table 16 Stock status and main recent patterns of stock indicators estimated by a4a.	35
Table 17 Main results of MSE according to the BEMTOOL and SMART simulation models and different scenarios as % variation from the status quo (S0; 2020) to 2026. F= fishing mortality, SSB = spawning stock biomass in tons; Y= yield in tons; i= income in millions of euro * referred to the aggregated income of both ARA and ARS. The best results in terms of gain in SSB and increase or maintaining the current yield is marked in bold.	35

Figures index

Figure 1 Case Study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Comparison between stock assessment results with 95% confidence interval (pink) and BEMTOOL estimates (blue dots) on F, SSB and Landings.	12
Figure 2 Case study CS3 - DPS GSAs 17-20: Comparison between stock assessment results with 95% confidence interval (pink) and BEMTOOL estimates (blue dots) on F, SSB and Catch for DPS.	13
Figure 3 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: stock recruitment relationships estimated by Eqsim for CS1 and CS2. ARS on the left and ARA on the right.	14
Figure 4 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Closure areas in EMU2 according to the decree prot. N.9045689 of 06/08/2020 from Italy. Each area is associated to a name as assigned in the cited decree.	16
Figure 5 Case study CS3 - DPS GSAs 17-18-19-20: stock recruitment relationships estimated with Eqsim for DPS. .	18
Figure 6 Schematic representation of the Elman network used in this application. The input layer comprises: neurons related to population size by age, catches by age, and SSB at a given time t. These neurons directly propagate the information to the basic hidden neurons (H). At each step of the training procedure, the updated pattern of the basic hidden neurons is memorized by the context neurons (C) and, at the successive step, propagated to the basic hidden neurons together with the new information in the input neurons. The output layer contains as many neurons as the number of cohorts in the input layer. EMPN is the simplest and most widely used ANN architecture to pursue classification issues when sequential or time-varying patterns are inspected. The basic structure of a multilayer perceptron network (MPN) consists of at least three layers of neurons. Information flows in a unidirectional way from the first (input) layer to the last (output) layer through a second (hidden) layer and is processed in parallel by the neurons of each layer. The input layer contains as many neurons as there are independent variables or descriptors used to predict the dependent variables, which in turn constitute the output layer. Neurons of a given layer are linked to the neurons of the next layer by activation functions: hidden layer neurons compute a weighted sum of the input variables through a first activation function and then send a result to the output neurons through a second activation function. A sigmoid function, which is the most common choice because it is nonlinear and characterized by a very easy to compute derivative was chosen in both cases.	20
Figure 7 Representation of the outputs of the training phase for ARS in the case study #1. The top left panel allows to compare predicted and observed values of SSB. The top right panel presents the same values as time series. The bottom left panel allows to compare predicted and observed values of the population size (number of individuals in log scale) The bottom right panel presents the same values as time series. Values of the adjusted R^2 are provided.	21
Figure 8 Representation of the outputs of the training phase for ARA in the case study #1. The top left panel allows to compare predicted and observed values of SSB. The top right panel presents the same values as time series. The bottom left panel allows to compare predicted and observed values of the population size (number of individuals in log scale) The bottom right panel presents the same values as time series. Values of the adjusted R^2 are provided.	22
Figure 9 Representation of the outputs of the training phase for ARS in the case study #2. The top left panel allows to compare predicted and observed values of SSB. The top right panel presents the same values as time series. The bottom left panel allows to compare predicted and observed values of the population size (number of individuals in log scale) The bottom right panel presents the same values as time series. Values of the adjusted R^2 are provided.	23
Figure 10 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Overall F of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020. Red horizontal solid line corresponds to the $FMSY=F0.1$, and red horizontal dashed lines correspond to Fupper and Flower.	24
Figure 11 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: SSB of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020.	25
Figure 12 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: Overall landing of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020 (discard null).	25
Figure 13 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: Total revenues of the overall fleet for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020.	26
Figure 14 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: CR/BER of the overall fleet for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020.	26
Figure 15 Case study CS3 - DPS GSAs 17-18-19-20: Overall F of DPS for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020. Red horizontal solid line corresponds to the $FMSY=F0.1$, and red horizontal dashed lines correspond to Fupper and Flower.	28
Figure 16 Case study CS3 - DPS GSAs 17-18-19-20: SSB of DPS for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020.	28

Figure 17 Case study CS3 - DPS GSAs 17-18-19-20: Overall catch of DPS for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020. 29

Figure 18 Case study CS3 - DPS GSAs 17-18-19-20: discard of the overall fleet for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020. 29

Figure 19 Case study CS3 - DPS GSAs 17-18-19-20: Total revenues of the overall fleet for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020. 29

Figure 20 Case study CS3 - DPS GSAs 17-18-19-20: CR/BER of the overall fleet for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020. 30

Figure 21 Case study CS1 - GSAs ARS 1-12: SSB of ARS simulated by SMART according to Scenarios S0, S1, S2, S3 and S4. 31

Figure 22 Case study CS1 - GSAs ARA 1-12: SSB of ARA simulated by SMART according to Scenarios S0, S1, S2, S3 and S4. 32

Figure 23 Case study CS2 - ARS GSAs 15, 16, 18, 19 and 20: SSB of ARS simulated by SMART according to Scenarios S0, S1, S2, S3 and S4. 33

Executive summary

This deliverable aims at examining potential development of adaptive spatial fisheries management through simulation approaches. According to the MEDUNITS project, three (3) case studies were selected on the basis of most probable stock configuration resulting from WP3. In agreement with DGMARE, stocks configuration of three high value deep water shrimps, *Aristeus antennatus* (ARA), *Aristaeomorpha foliacea* (ARS), and *Parapenaeus longirostris* (DPS). In particular, ARA and ARS were assessed in the western Mediterranean (GSAs from 1 to 12) (Case study CS1), ARS in the Central Mediterranean (ARS; GSA 15, 16, 18, 19 and 20) (Case study CS2), and DPS (GSA 17,18,19 and 20) (Case Study CS3). Different management scenarios agreed with DGMARE were evaluated by using bio-economic simulation models (BEMTOOL and SMART). These management scenarios considered reduction of fishing effort, improvement of the gear selectivity and spatial closures in areas critical for biological cycles of the targeted species (Essential Fish Habitats).

For CS1 and CS2 five scenarios have been implemented: i) S0: status quo (baseline), ii) S1: 10% reduction in 2021 + closure areas, iii) S2: 10% reduction in 2021 + 20% reduction in 2022 + closure areas, iv) S3: 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023+ closure areas, and v) S4: 10% reduction in 2021+gear selectivity change in 2023+ closure areas..

For CS3 three scenarios have been investigated: i) S0: status quo (i.e. no variations respect to 2021 (transition phase), ii) S1: linear reduction to the ref. point until 2026 +closures, and iii) S2: 10% reduction of fishing day in 2022+Selectivity change in 2023 +closures.

BEMTOOL is a multi-species and multi-gear bio-economic simulation model for mixed fisheries. It follows a multi-fleet approach simulating the effects of management scenarios on stocks and fisheries on a fine time scale (month). Two case studies have been assessed by BEMTOOL:

- CS1 - ARS GSAs 1-12 and GSAs ARA 1-12;
- CS3 – DPS GSAs 17-18-19-20.

BEMTOOL results showed the benefits of improving the exploitation pattern, through spatial closures and/or selectivity improvements: firstly, the discard fraction in the catch is remarkably reduced and secondly, it allows to limit the effort reduction, providing similar economic outcomes and higher biological performances. The implementation of the two case studies with BEMTOOL, using the stock assessment results of the new stock configurations, showed how the available knowledge on spatial dynamics of population components living in a study area can be integrated in a bio-economic model to accommodate the development of an adaptive spatial fishery management. For both ARS and ARA in CS1, S4, combining an effort reduction of 10% with a change in the exploitation pattern, is the best performing scenario, allowing to reduce the F of around 35-40%; this percentage is quite close to the percentage required by ARS but only half of the reduction required by ARA. Although S2 and S3 scenarios are quite equivalent in terms of F, the 20% of effort reduction applied in 2022 in S2 produces increase in SSB and total landing slightly higher than S3, that distributes the 20% reduction in two years. The projections show that S4 increases the SSB of ARS and ARA of 63% and 144 % respectively, with a decrease in landing of 5% for ARS and an increase of 61% for ARA. From an economic point of view S2 and S3 are very similar and allow to reach revenues well above S0 and S1. Although in the short term S4 projections show a sharp decrease in revenues and CR/BER, due to the change in selectivity, this scenario allows to reach, in the long term, values quite comparable to the ones of S2 and S3, reducing the effort of 10% instead of 30% overall.

As the DPS regards, the decrease of the overall F is higher for S1, while the combination of effort reduction and change in exploitation pattern of S2 scenario returns a decrease of 16% that does not allow to reach the F_{msy} range, but allows to increase the SSB of 40% respect to S0 in 2026. In terms of SSB, S1 allows to obtain the best performance, followed by S2; on the other hand, scenario S2 returns the higher overall catch than S1 in the long term, followed by S0. Although both S1 and S2 decrease remarkably the discard respect to S0 in the long term, S2 allows to reduce the discard fraction already from 2022 and to reach the smallest value. From an economic point of view S2 is the scenario best performing in terms of total revenues, allowing to reach values in line with S0, after an initial reduction in the short term; the increase on the overall fleet revenues is of 3%, but varies among the different fleet segments between 2% and 23%.

On the other hand, S1, after the reduction in revenues until 2026, due to implementation of the management measures, allow to reach values slightly lower than S0.

SMART is a spatially- explicit multi-species individual-based model allowing to evaluate the effects on stock dynamics and fishery performance under different time and spatial based management scenarios. The model needs of spatially-resolved data on catch rates by age and fishing effort. Two case studies have been assessed by a modified version of SMART:

- CS1 - ARS GSAs 1-12 and GSAs ARA 1-12;
- CS2 – ARS GSAs 13-16,18-20. The trends of SSB for the ARS in the GSAs 1-12 simulated by SMART indicate that a decreasing of the stock is expected in the present situation and that only some of the explored scenarios are likely to stop or reduce this decline. The most promising approach is represented by the Scenario S4, which integrates spatial closures, effort reduction and changes in the selectivity. The management measures in the Scenarios S1 and S3 seem adequate to slow down the decreasing trend, while those in the Scenario S2 are likely to stabilize the SSB values after 2 years. In the case of ARA in the GSAs 1-12, the model suggested that management measures in the Scenario S1 are not enough to stop the expected decline observed in the S0, whereas those in the Scenario S2 should allow to stop the decline and stabilize the SSB. However, the management measures in the Scenario S3 seem more effective, at least in the middle term, and those in the Scenario S4 are expected to determine a partial recovery of this stock. The predicted trends of SSB for the ARS in the GSAs 15-16 and 18-20 indicate that an increasing of the stock is expected in the present situation and that only the set of management measures considered in S4 could modify this dynamic of the stock and improve its condition. S1-3, however, are associated with a declining phase in the last part of the forecast period. Despite BEMTOOL and SMART assessed the same fisheries under the same scenarios, they were quite different in terms of results. These differences are mainly due to differences in modelled processes and assumptions in MSE. However, both models evaluated as the best management strategy, both in terms of gain in SSB and improve or light decrease of the current yield, the scenario S4. This scenario is characterized by reducing of the 10% the current fishing effort coupled with an improvement of trawl net selectivity and protection of EFHs. Finally, an “ideal” roadmap for developing adaptive spatial fishery management aimed to reduce uncertainty in assessment and management procedures was proposed. The necessity of accurate knowledge of populations units and their connectivity by improving the space and time detailed data collection was highlighted and the main obstacles to the effective adoption of adaptive spatial fishery management was discussed.

Introduction

This deliverable aims at examining potential development of adaptive spatial fisheries management through simulation approaches when knowledge on spatially-defined stock units are available. According to the MEDUNITS project, three (3) case studies was selected on the basis of most probable stock configuration resulting from the results obtained in the WP3. In agreement with DGMARE, three stocks configuration of three high value deep water shrimps (*Aristeus antennatus*-violet shrimp: ARA, *Aristaeomorpha foliacea*-Giant red shrimp: ARS, and *Parapenaeus longirostris*-deep water rose shrimp: DPS). In particular, ARA and ARS were assessed in the western Mediterranean (GSAs from 1 to 12) (Case study CS1), ARS in the Central Mediterranean (ARS; GSA 15, 16, 18, 19 and 20) (Case study CS2), and DPS (GSA 17,18,19 and 20) (Case Study CS3). Different management scenarios, agreed with DGMARE were evaluated by using bio-economic simulation models (BEMTOOL and SMART) developed for Mediterranean fisheries. These management scenarios considered reduction of fishing effort, spatial closures in areas critical for biological cycles of the targeted species (Essential Fish Habitats), and improvement in the gear selectivity.

BEMTOOL simulation platform (e.g. Rossetto et al., 2014, Spedicato et al., 2016; Russo, Bitetto et al., 2017), as recently applied in the DGMARE project SAFENET (Guidetti et al., 2020), was used to simulate management scenarios based on interactions of the different fleets targeting different stocks and areas of stock occupation, in different seasons (months) and under given stock-recruitment relationships.

Initially, the new version of the SMART platform (Russo et al., 2019), employed within the recent DGMARE

project MANTIS (Fiorentino et al., 2020), was proposed to be used to assess the effects of different management scenarios, including regulation of fishing effort, spatial closures and improve of selectivity. Differently from BEMTOOL, SMART is a multi-species individual-based model that needs of spatially-resolved data on catch rates by age and fishing effort. Due to the difficult to obtain these data for the whole countries sharing the investigated stocks, a simplified version of the SMART model using a neural network for simulation (Russo et al., 2014) was implemented.

Objectives

The task 4.3 “Future improvements for developing adaptive spatial fisheries management”, aims at incorporating spatial component of biological processes and fishing pressure into the assessment and management procedures. In particular, different management scenarios (Table 1) of deep water crustacean fisheries were assessed through bio-economic scenario modelling, in line with the Management Strategy Evaluation. The best management scenarios were identified in terms of stock and fishery performance indicators. Finally, a tentative roadmap with actions required in order to develop adaptive spatial fisheries management was proposed.

Workplan

Three case studies were selected on the basis of available knowledge on spatial dynamics of population components of ARS, ARA, and DPS. The management scenarios for which potential effects were evaluated included spatial based measures such as regulation of fishing effort and catch in EFHs and VMEs included effects on both stocks (e.g. in terms of SSB and F) and fisheries (e.g. in terms of income, costs and gains). Both case studies and management scenarios were agreed with DGMARE. The selected case studies were reported in table 1.

Table 1 Case Studies selected for task 4.3 of the MEDUNITS project. Stock configurations, justifications, management scenarios and adopted models by case studies are reported. CS1: Reference years Western Mediterranean MAP 2015-2017; Transition year 2020; closures are the ones of EWG 20-13. CS2: Reference years Adriatic MAP 2015-2018; closures** are the ones of GFCM Recommendation CS3: Reference years for Central Mediterranean 2016-2018; closures *** are those reported in the EU MANTIS project.*

Case Study	Stock Configurations	Justifications	Scenarios	Model
CS 1	ARS GSAs 1-12 and ARA 1-12	Both stocks are part of the Western MAP and target of specific fisheries, both stocks are fished unsustainably, ARA with a higher F_{curr}/F_{msy} ratio	<ol style="list-style-type: none"> 1. 10% reduction in 2021 + closures*; 2. 10% reduction in 2021+20% reduction in 2022 + closures* 3. 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023 +closures* 4. 10% reduction in 2021+Selectivity change (e.g. Gorelli et al. 2017) in 2023 +closures* 	BEMTOOL SMART

CS 2	ARS 13,15,16,18,19,20	ARS in GSAs12-16 is a target of the Recommendation GFCM/43/2019/6 on management measures for sustainable trawl fisheries targeting giant red shrimp and blue and red shrimp in the Strait of Sicily (geographical subareas 12, 13, 14, 15 and 16), and ARS is one of the most important species targeted by the fisheries in the Strait of Sicily. SAC should provide to GFCM specific advice to implement a MAP by 2022. In addition, the Recommendation GFCM/42/2018/4. Recent assessment for GSAs 18-19 resulted in an Fcurr/Fmsy of 1.1, no recent assessments are available for the Strait of Sicily on a multiannual management plan for sustainable trawl fisheries targeting giant red shrimp and blue and red shrimp in the Ionian Sea (geographical subareas 19, 20 and 21).	<ol style="list-style-type: none"> 1. 10% reduction in 2021 + closures*; 2. 10% reduction in 2021+20% reduction in 2022 + closures* 3. 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023 +closures*** 4. 10% reduction in 2021+Selectivity change in 2023 +closures* 	SMART
CS 3	DPS GSAs 17-18-19-20a.	DPS in GSA17-18 is a target of the Recommendation GFCM/43/2019/5 related to the MAP for demersal stocks in the Adriatic. It has become an important target of the fishery in the whole area. The last assessment (2020) indicates that the stock is exploited well beyond Fmsy.	<ol style="list-style-type: none"> 1. status quo, i.e. no variations respect to 2021 (transition phase) 2. linear reduction to the ref. point until 2026 +closures** 3. 10% reduction of fishing day (following the formula in the GFCM recommendation) in 2022+Selectivity change in 2023 +closures** 	BEMTOOL

Materials and methods

BEMTOOL

The model

BEMTOOL is a multi-species and multi-gear bio-economic simulation model for mixed fisheries developed for Mediterranean fisheries (Accadia et al., 2013). It consists of six operational modules characterized by different components: biological (age/length structured dynamic model, Lembo et al., 2009), Impact, Economic, Behavioural, Policy and Multi Criteria Decision Analysis (MCDA) (Rossetto et al, 2014; Spedicato et al., 2016; Russo et.al, 2017). BEMTOOL follows a multi-fleet approach simulating the effects of

management scenarios on stocks and fisheries on a fine time scale (month). The model accounts for length/age-specific selection effects, discards, economic and social performances, effects of compliance with landing obligation and reference points. Compared to existing bio-economic tools, BEMTOOL presents a number of innovations, including the simulation of discard and escape survivability, the estimation of additional costs and, potentially, additional income due to the landing obligation (Spedicato et al., 2018). Six selectivity functions are implemented in BEMTOOL, plus a vector at length/age. The model can consider a large number of fleet groups. The implementation of a decision module (Multi-Criteria Decision Analysis and Multi-attribute utility theory) allows stakeholders to weigh model-based indicators and rank different management strategies.

The model can simulate management scenarios based on changes in fishing pattern, fishing effort, fishing mortality and TAC. A wide set of biological, pressure and economic indicators is the default output. The uncertainty implemented in the model following Monte Carlo paradigm allows a risk evaluation in terms of biological sustainability of the different management strategies accounting for the economic performances. BEMTOOL allows the inclusion of process and parameters errors (on recruitment, individual growth and natural mortality, maturity ogive, and fleet/gear selectivity), crucial to gauge management strategies from a Management Strategy Evaluation (MSE) perspective (Spedicato et.al, 2017). Uncertainty can be applied according to three different probability distributions: normal, lognormal and uniform. BEMTOOLv.3 platform allows also the implementation of a scenario based on a TAC set according to an escapement strategy approach (GFCM, 2018). Further information on the model applications can be found in STECF (2018). Recently BEMTOOL model has been used for simulating scenarios and predict the consequences of the implementation of the management measures foreseen in the MAP of the western Mediterranean in EMU2 (STECF, 2019a; 2019b;2020).

Parameterization of BEMTOOL in the hindcasting mode for CS1 and CS3

The model hindcasting for the two case studies covers the years 2006-2020. The relevant information related to the assessments with biological and fishing mortality components of the model was collected from the Deliverable 3.3, while the effort and the socio-economic information by fleet was collected from FDI data (<https://stecf.jrc.ec.europa.eu/dd/fdi>), AER (<https://stecf.jrc.ec.europa.eu/dd/fleet>) and the MedUnits Data Call.

Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12

BEMTOOL bioeconomic simulation model was implemented for GSAs 1-5-6-9-10-11 (CS1), following the experiences gained in STECF EWG 19-01, EWG 19-14 and EWG 20-13 (STECF, 2019a;2019b; 2020). BEMTOOL model was first parameterized in the hindcasting mode for the *Aristaeomorpha foliacea* (ARS) in GSAs 5-9-10-11 and *Aristeus antennatus* (ARA) in GSAs 1-5-6-9-10-11, covered by the Multiannual Management Plan (MAP, Regulation (EU) 2019/1022) in the areas considered. Landing data of ARS and ARA for GSA 7 were not included in the WP3 assessments, nor in the simulations, because considered negligible.

The transversal data and socio-economic information at GSA level, for the GSAs 9-10-11, was available from the MED_UNITS Data Call.

For GSAs 1, 5 and 6 the landings by stock, the total landing, the revenues by stock (thus also the price data) and the total revenues were available from FDI Data Call at GSA level. The socio-economic data for GSAs 1, 5 and 6, were derived on the basis of the National data in AER (Country: Spain and supra_reg: AREA 37). Specifically, the variable costs were derived as the national variable costs for the fleet segment (e.g. DTS_VL_1218) multiplied by the ratio between the fishing days of the specific fleet segment (e.g. GSA1_DTS_VL1218) and the fishing days of national fleet segment. Analogously, the fixed, the capital and the labour costs were estimated on the basis of the number of vessels and relative ratios.

Landing data from GSAs 8 were not available (neither in FDI nor in DGMARE Data Call, except for few tons of ARA in 2010 and 2011). Data for GSA 12 were not available in FAO statistics; some information on total

landing and on the overall fleet is available in FAO (2020) but the data are not in the appropriate level of aggregation (GSA and fleet segment). For this reason, the data of GSA 8 and 12 have not been included neither in the assessment nor in the simulations.

The fleet selectivity was modeled on the basis of DCF data by area (ARS, a proxy of CL_{50} varying between 28 and 33 mm) and the relevant literature (Gorelli et al., 2017; ARA, $CL_{50} = 22.6$ mm). Assessed fishing mortality, spawning stock biomass and the observed catches were compared with the simulated ones. Discard was not included in the assessment, because negligible, and thus it was not modeled.

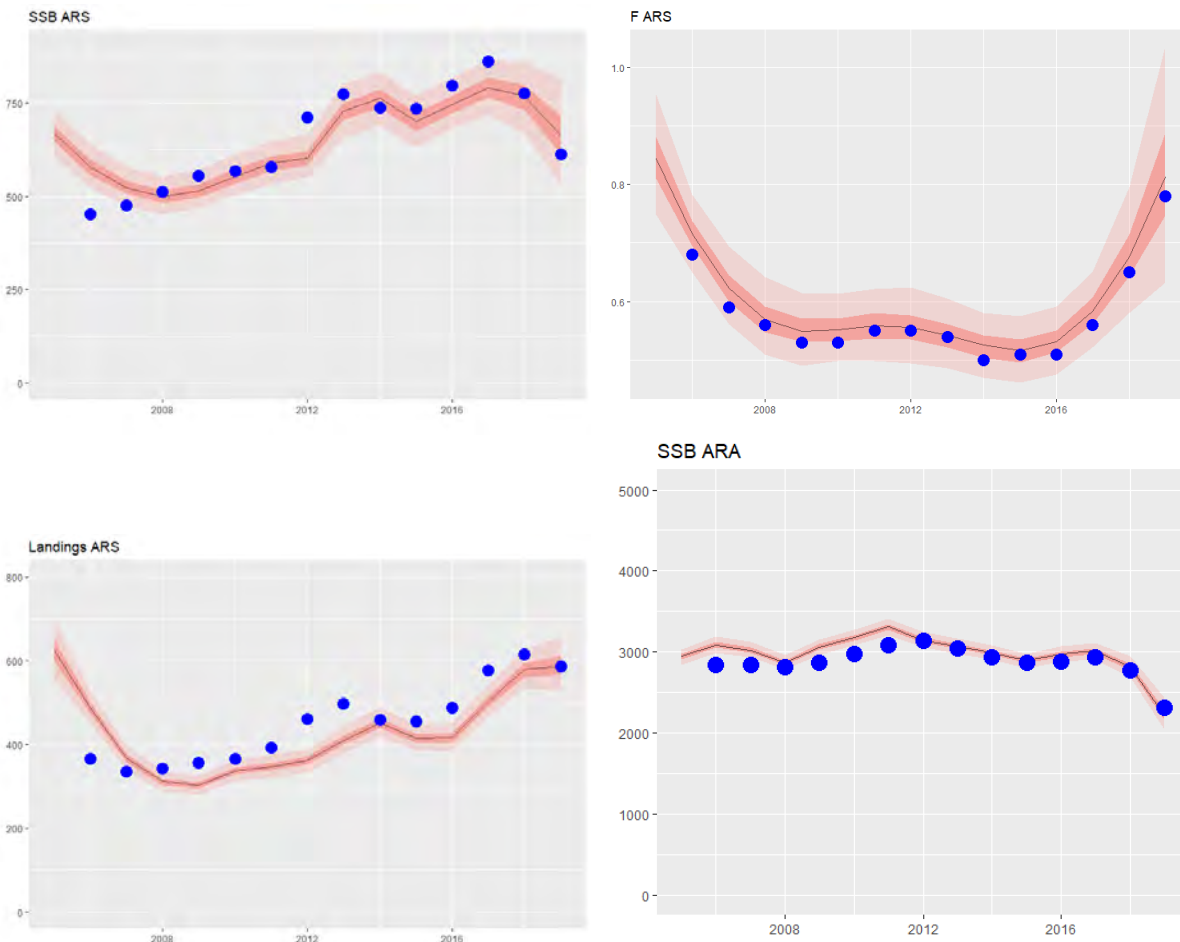
For 2020 the effort was assumed in line with EU Regulation 2236/2019. For the other variables an invariant situation compared to 2019 was assumed.

The relevant results of the assessment for the model parameterization, i.e. the current fishing mortality (F_{curr}) and the reference point ($F_{0.1}$) are reported in the **Errore. L'origine riferimento non è stata trovata..**

Table 2 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Results of the assessments from WP3 relevant for BEMTOOL parameterization. The computation of the reduction by stock to reach $F_{0.1}$ is also reported.

Stock	F_{curr}	$F_{0.1}$	$F_{curr}/F_{0.1}$	$F_{0.1lower}$	$F_{0.1upper}$	% reduction
ARS	0.81	0.46	1.8	0.31	0.63	43%
ARA	0.95	0.163	5.83	0.11	0.23	83%

The results of the stock assessment for the considered stocks have been replicated in BEMTOOL. The comparison of F , SSB and Landings shows a good level of agreement between BEMTOOL and the stock assessment results (**Errore. L'origine riferimento non è stata trovata.**).



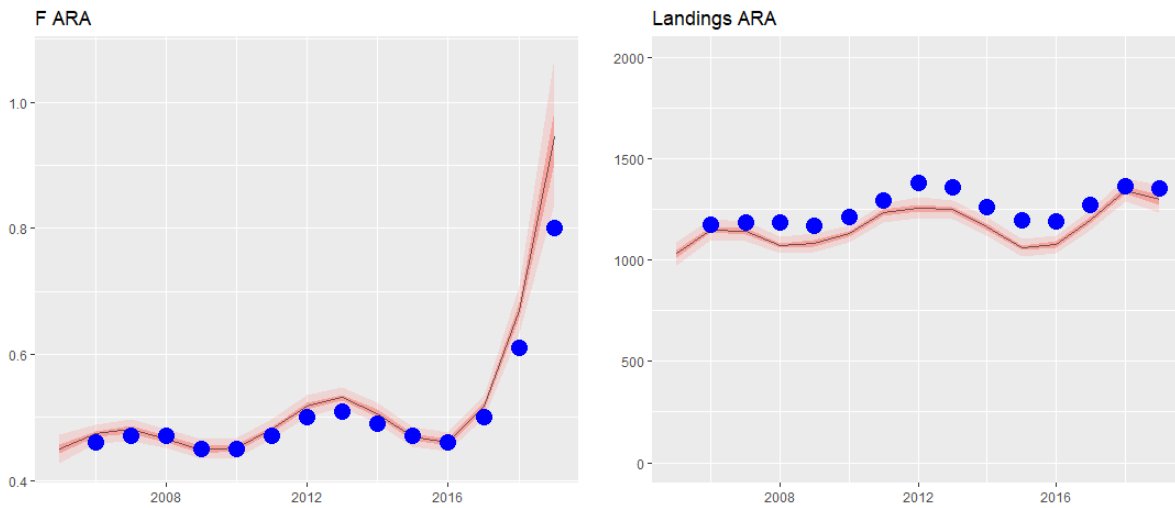


Figure 1 Case Study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Comparison between stock assessment results with 95% confidence interval (pink) and BEMTOOL estimates (blue dots) on F, SSB and Landings.

In the simulation 19 fleet segments, shown in table 3, have been considered. These include trawlers operated by fleet segments that influence both stocks included in the case study. Three fleets targeting ARS and/or ARA are allocated to each of the GSAs 1, 5, 6, 9 and 11 and four in GSA 10. Following the Deliverable 3.3 GSA 5, 9, 10 and 11 fleet segments were assumed to exploit ARS, while all fleet segments interact with ARA.

Table 3 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12 Fleet segments included in the BEMTOOL simulations and forecast scenarios by GSA for vessel length (VL) segments.

GSA 1	GSA 5	GSA 6
GSA1_DTS_VL1218	GSA5_DTS_VL1218	GSA6_DTS_VL1218
GSA1_DTS_VL1824	GSA5_DTS_VL1824	GSA6_DTS_VL1824
GSA1_DTS_VL2440	GSA5_DTS_VL2440	GSA6_DTS_VL2440
GSA 9	GSA 10	GSA 11
GSA9_DTS_VL1218	GSA10_DTS_VL0612	GSA11_DTS_VL1218
GSA9_DTS_VL1824	GSA10_DTS_VL1218	GSA11_DTS_VL1824
GSA9_DTS_VL2440	GSA10_DTS_VL1824	GSA11_DTS_VL2440
	GSA10_DTS_VL2440	

Case study CS3 – DPS GSAs 17-18-19-20

BEMTOOL was implemented for GSAs 17-18-19-20, following the experiences gained in FAIRSEA project (<https://www.italy-croatia.eu/web/fairsea/site>). DCF data (FDI and MED&BS Data Call, landings, discards, fishing effort, biological and economic parameters) as well as the results from the assessment of DPS (*Parapaeneus longirostris*) GSAs 17-20 carried out in WP3 and described in Deliverable 3.3 were used to parameterize the BEMTOOL model. The socio-economic information at GSA level was available from the MED_UNITS Data Call. BEMTOOL was parameterized in the hindcasting mode in GSAs 17-20. As reference for some general management rules the Recommendation GFCM/43/2019/5 for a multiannual plan (MAP) in the GSAs 17-18 was considered. Thus, for 2020 and 2021 the effort was assumed in line with Recommendation GFCM/43/2019/5.

The fleet trawl electivity was modeled on the basis of DCF data and the relevant literature (Brčić et al., 2018, CL50 = 18 mm).

Discard was included in the simulations and modeled according to a reverse ogive model, parameterized

according to DCF data. Assessed fishing mortality, spawning stock biomass and the observed catches were compared with the simulated ones.

The relevant results of the assessment for the model parameterization, i.e. the current fishing mortality (F_{curr}) and the reference point ($F_{0.1}$) are reported in the **Errore. L'origine riferimento non è stata trovata**.4.

Table 4 Case study CS3 - DPS GSAs 17-20: Results of the assessments from WP3 relevant for BEMTOOL parameterization. The computation of the reduction by stock to reach $F_{0.1}$ is also reported

Stock	F_{curr}	$F_{0.1}$	$F_{curr}/F_{0.1}$	$F_{0.1lower}$	$F_{0.1upper}$	% reduction
DPS	1.6	0.44	3.63	0.3	0.605	72.5%

The results of the stock assessment for DPS have been replicated in BEMTOOL. The comparison of F , SSB and Catch showed a good level of agreement between BEMTOOL and the stock assessment results (

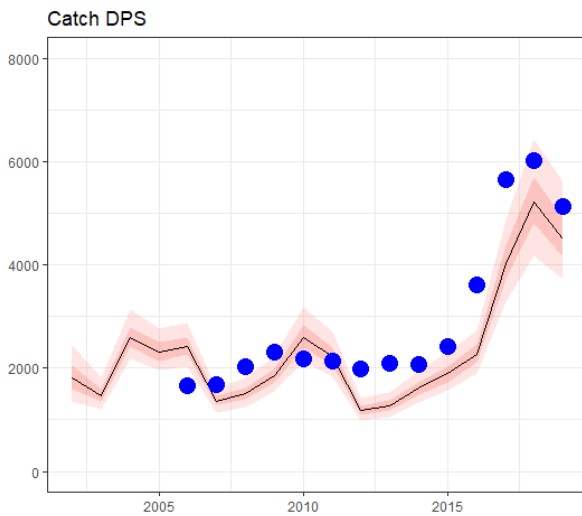
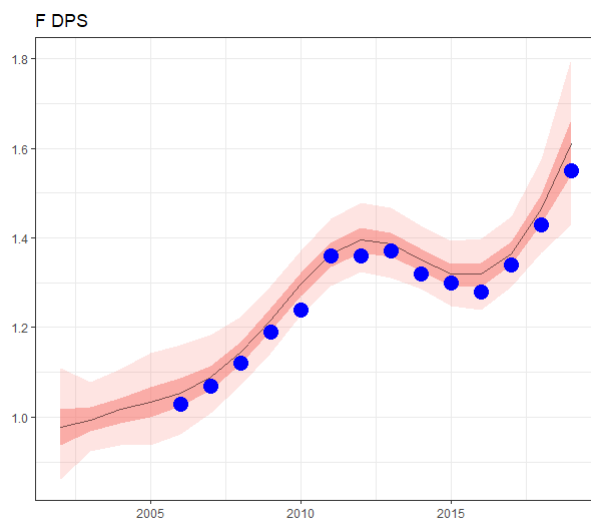
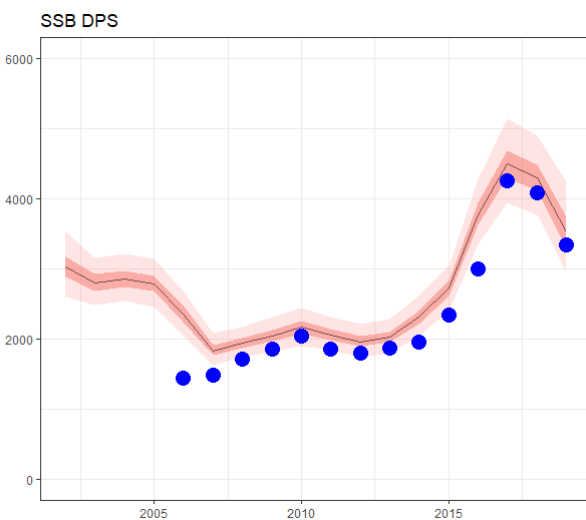


Figure 2 Case study CS3 - DPS GSAs 17-20: Comparison between stock assessment results with 95% confidence interval (pink) and BEMTOOL estimates (blue dots) on F , SSB and Catch for DPS.

).



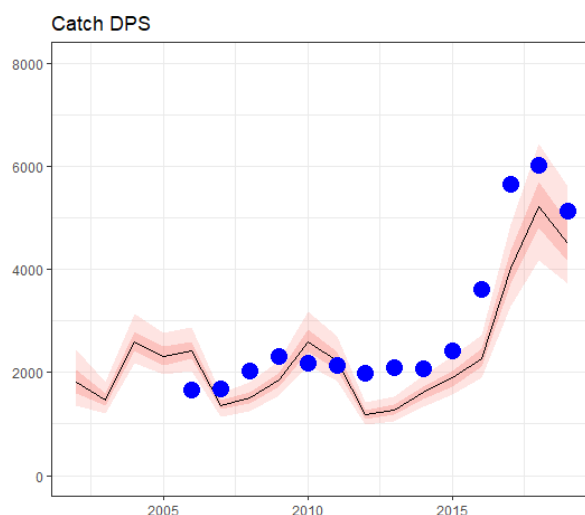


Figure 2 Case study CS3 - DPS GSAs 17-20: Comparison between stock assessment results with 95% confidence interval (pink) and BEMTOOL estimates (blue dots) on F, SSB and Catch for DPS.

In the simulation 15 fleet segments, shown in **Errore. L'origine riferimento non è stata trovata.**, have been considered. These include trawlers operated by fleet segments that influence the DPS in the considered areas. Five fleets targeting DPS are allocated to GSA 17 (with flag Italy and Croatia), six to GSA18 (Italy, Albania and Montenegro), two to GSA 19 (Italy) and two in GSA 20 (Greece). Following the Deliverable 3.3 all fleet segments interact with DPS.

Table 5 Case study CS3 - DPS GSAs 17-20 Fleet segments included in the BEMTOOL simulations and forecast scenarios by GSA for vessel length (VL) segments.

GSA 17	GSA 18	GSA 19	GSA 20
ITA_17_DTS_1218	ITA_18_DTS_0612	ITA_19_DTS_1218	GRC_20_DTS_VL1824
ITA_17_DTS_1840	ITA_18_DTS_1218	ITA_19_DTS_1824	GRC_20_DTS_VL2440
HRV_17_DTS_0612	ITA_18_DTS_1840		
HRV_17_DTS_1218	ALB_18_DTS_1224		
HRV_17_DTS_1840	MNE_18_DTS_0612		
	MNE_18_DTS_1224		

Forecast parameterization of BEMTOOL: assumptions for projections

Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12

The forecasts covered the period from 2021 to 2030 to assess the biological and economic results in the medium term.

The stock-recruitment relationships for the stocks projections have been estimated using Eqsim (Minto et al., 2014) (**Errore. L'origine riferimento non è stata trovata.**). **Errore. L'origine riferimento non è stata trovata.**6 reports the parameters of the stock-recruitment relationships for the ARS and ARA in GSAs 1-12.

The same fleets as in the hindcasting have been considered in the forecast scenarios.

Table 6 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Parameters of the stock recruitment relationships for ARS and ARA.

Stock	Break Point (b)	a
ARS GSAs 1-12	500	527
ARA GSAs 1-12	2000	133

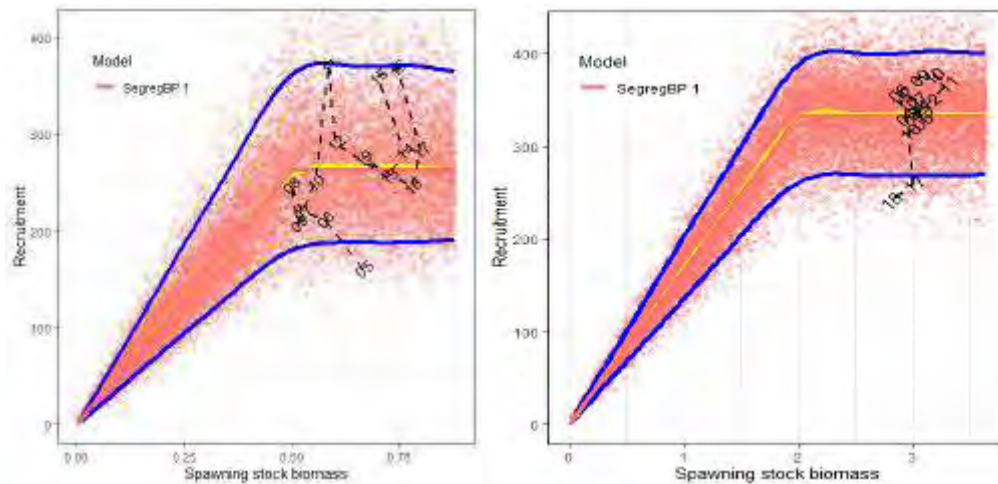


Figure 3 Case study CS1 - ARS GSA 1-12 and GSA ARA 1-12: stock recruitment relationships estimated by Eqsim for CS1 and CS2. ARS on the left and ARA on the right.

The variable costs (fuel and other) have been assumed to vary proportionally to the annual fishing days, while the fixed and the maintenance costs depend on the annual GT on the basis of the historical data. The labour costs have been assumed in line with the crew share system on the difference revenues minus variable costs, and the depreciation costs depending on the annual GT. For GSAs 1-5-6 the opportunity costs have been assumed as null and, in case of inconsistency of the crew share coefficients, 0.5 was assumed.

The employment was assumed not changing for all scenarios. Moreover, the total landing and the total revenues are assumed proportionally changing with the landing of the target stocks, by means of a correction factor based on historical data.

The performance of the scenarios was evaluated on the basis of spawning stock biomass, catch, F, revenues and current revenues to break-even revenues (CR/BER). The latter is an economic indicator that shows how close the current revenue of a fleet is to the revenue required for the economic break even. Ratios > 1 indicate that enough income is generated to cover operational costs (variable and non-variable costs) and therefore break-even. If the ratio is less than 1, insufficient income is generated to cover operational costs and therefore the fleet is in a loss.

For all scenarios the basis was given by the number of fishing days by fleet as the average in the period 2015-2017, following the western Mediterranean MAP (Regulation (EU) 2019/1022).

For CS1 five scenarios have been implemented:

- S0: status quo (baseline);
- S1: 10% reduction in 2021 + closures areas;
- S2: 10% reduction in 2021 + 20% reduction in 2022 + closures areas;
- S3: 10% reduction in 2021+10% reduction in 2022+10% reduction in 2023+ closures areas;
- S4: 10% reduction in 2021+Selectivity change (e.g. Gorelli et al. 2017) in 2023+ closures areas.

For all the scenarios except the scenario S0 the basis for the closure areas for GSA9, 10 and 11 was given by the decree prot. N.9045689 of 06/08/2020, from the Italian Fishery Directorate of the Ministry of the Agriculture Food and Forestry Policy, and the fraction of the fleet activity (whole fleet and by fleet segment to accommodate the closure allocation to the closure areas.

Following this decree 10 areas have been closed all year round (see **Errore. L'origine riferimento non è**

stata trovata.), three of these are located in GSA9, five in GSA10 and two in GSA11, covering an overall surface of approximately 2093 km² (overall area between 50 and 800 m depth covering approximately 89,640 km²). In order to quantify the contribution of these closure areas the same approach followed in the STECF EWG 20-13 was applied (STECF, 2020). The information of effort allocation (Global Fishing Watch, GFW, <https://globalfishingwatch.org/>) was crossed with the spatial information on nurseries in the area (MEDISEH project), and the overlap with closure areas was identified. The fleet selectivity was then translated into a change (increase) of the length at first capture, proportionally shaping the exploitation pattern of ARS by GSA and fleet (2% in GSA9 and 1% in GSA 10).

According to STECF 2020, the Spanish spatial closures (GSAs 1-5-6) impact only on European hake and not on ARS and ARA. Thus, no change was applied in Spanish fleet selectivity due to spatial closures.

For scenario S4 an increase in fleet selectivity was simulated. This was set on the basis of Gorelli et al. (2017) applying an increase of the size at first capture of 24% (from square 40 mm mesh size, S40, to square 50 mm mesh size, S50 in Gorelli et al., 2017) on both stocks (generally inhabiting the same grounds and targeted by the same fleet and gear).

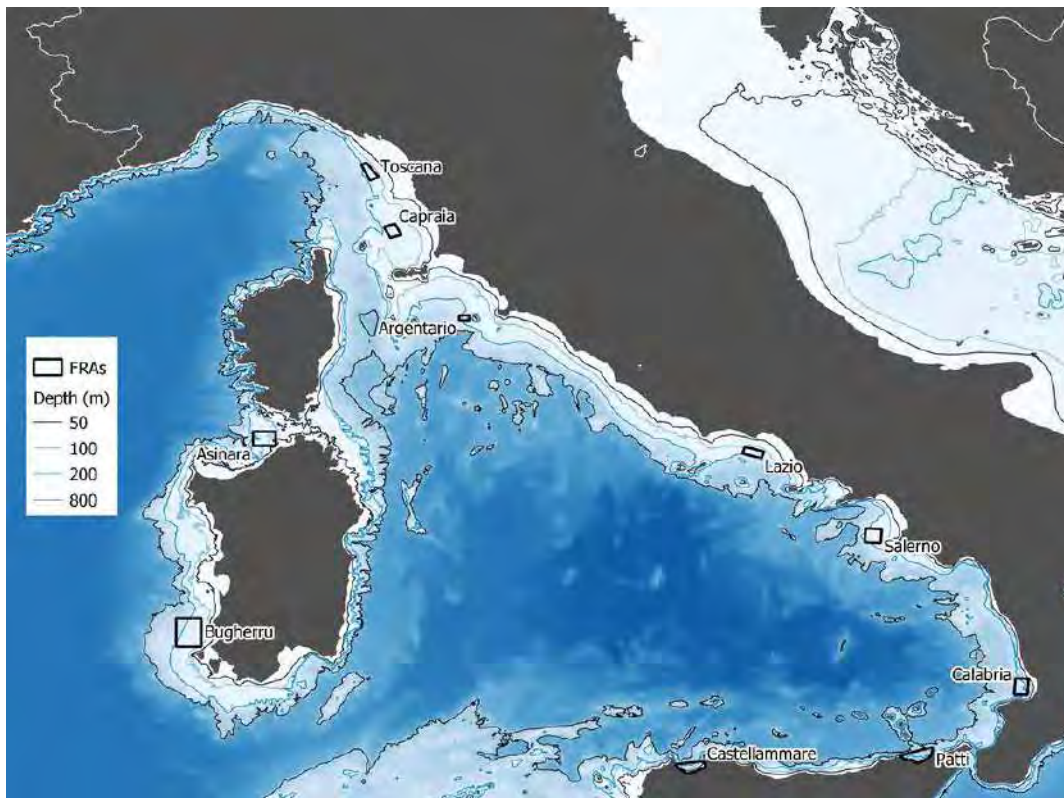


Figure 4 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Closure areas in EMU2 according to the decree prot. N.9045689 of 06/08/2020 from Italy. Each area is associated to a name as assigned in the cited decree.

The following equation, internally applied by the model to recalculate the fishing mortality, was used to reshape the fleet selectivity, acting on the Sel parameter:

$$F_f(a) = (Z_{inp} - \text{mean}(M)) * Sel_f(a) * f_{act,f} * p_f ;$$

where $f_{act,f}$ in the forecast is the ratio between the product of the number of fishing days, the number of vessels and the average GT (or Kw) of the fleet segment f for each month of forecast to the product of the

number of fishing days, the number of vessels and the average GT (or Kw) of the fleet segment f in the last year of the simulation. This quantity is considered as reference for the application of change in fishing effort. $sel_{f(a)}$ is the fleet selectivity at a given length/age; p_f is the monthly ratio between the fleet segment catch to the total catch in the simulation (in the forecast it is fixed as an average of the last (n) years). The effort reductions applied in scenarios S1, S2, S3 and S4 are reported in **Errore. L'origine riferimento non è stata trovata.**

Table 7 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Effort reductions applied by year and fleet segment for scenarios S1, S2, S3 and S4 respect to S0. In the first column are the fishing days in 2020, assumed in line with EU Regulation 2236/2019.

Fleet segment	S0	S1			S2		
	2020	2021	2022	2023	2021	2022	2023
ALL	132195	-10%	-10%	-10%	-10%	-30%	-30%
GSA1_DTS_VL1218	2763	-10%	-10%	-10%	-10%	-30%	-30%
GSA1_DTS_VL1824	4432	-10%	-10%	-10%	-10%	-30%	-30%
GSA1_DTS_VL2440	499	-10%	-10%	-10%	-10%	-30%	-30%
GSA10_DTS_VL0612	1460	-10%	-10%	-10%	-10%	-30%	-30%
GSA10_DTS_VL1218	14935	-10%	-10%	-10%	-10%	-30%	-30%
GSA10_DTS_VL1824	6362	-10%	-10%	-10%	-10%	-30%	-30%
GSA10_DTS_VL2440	1817	-10%	-10%	-10%	-10%	-30%	-30%
GSA11_DTS_VL1218	7076	-10%	-10%	-10%	-10%	-30%	-30%
GSA11_DTS_VL1824	3154	-10%	-10%	-10%	-10%	-30%	-30%
GSA11_DTS_VL2440	2690	-10%	-10%	-10%	-10%	-30%	-30%
GSA5_DTS_VL1218	790	-10%	-10%	-10%	-10%	-30%	-30%
GSA5_DTS_VL1824	4398	-10%	-10%	-10%	-10%	-30%	-30%
GSA5_DTS_VL2440	1034	-10%	-10%	-10%	-10%	-30%	-30%
GSA6_DTS_VL1218	13890	-10%	-10%	-10%	-10%	-30%	-30%
GSA6_DTS_VL1824	26374	-10%	-10%	-10%	-10%	-30%	-30%
GSA9_DTS_VL0612	2523	-10%	-10%	-10%	-10%	-30%	-30%
GSA9_DTS_VL1218	17511	-10%	-10%	-10%	-10%	-30%	-30%
GSA9_DTS_VL1824	18546	-10%	-10%	-10%	-10%	-30%	-30%
GSA9_DTS_VL2440	1939	-10%	-10%	-10%	-10%	-30%	-30%
Fleet segment	S0	S3			S4		
	2020	2021	2022	2023	2021	2022	2023
ALL	132195	-10%	-20%	-30%	-10%	-10%	-10%
GSA1_DTS_VL1218	2763	-10%	-20%	-30%	-10%	-10%	-10%
GSA1_DTS_VL1824	4432	-10%	-20%	-30%	-10%	-10%	-10%

Fleet segment	S0	S1			S2		
	2020	2021	2022	2023	2021	2022	2023
GSA1_DTS_VL2440	499	-10%	-20%	-30%	-10%	-10%	-10%
GSA10_DTS_VL0612	1460	-10%	-20%	-30%	-10%	-10%	-10%
GSA10_DTS_VL1218	14935	-10%	-20%	-30%	-10%	-10%	-10%
GSA10_DTS_VL1824	6362	-10%	-20%	-30%	-10%	-10%	-10%
GSA10_DTS_VL2440	1817	-10%	-20%	-30%	-10%	-10%	-10%
GSA11_DTS_VL1218	7076	-10%	-20%	-30%	-10%	-10%	-10%
GSA11_DTS_VL1824	3154	-10%	-20%	-30%	-10%	-10%	-10%
GSA11_DTS_VL2440	2690	-10%	-20%	-30%	-10%	-10%	-10%
GSA5_DTS_VL1218	790	-10%	-20%	-30%	-10%	-10%	-10%
GSA5_DTS_VL1824	4398	-10%	-20%	-30%	-10%	-10%	-10%
GSA5_DTS_VL2440	1034	-10%	-20%	-30%	-10%	-10%	-10%
GSA6_DTS_VL1218	13890	-10%	-20%	-30%	-10%	-10%	-10%
GSA6_DTS_VL1824	26374	-10%	-20%	-30%	-10%	-10%	-10%
GSA9_DTS_VL0612	2523	-10%	-20%	-30%	-10%	-10%	-10%
GSA9_DTS_VL1218	17511	-10%	-20%	-30%	-10%	-10%	-10%
GSA9_DTS_VL1824	18546	-10%	-20%	-30%	-10%	-10%	-10%
GSA9_DTS_VL2440	1939	-10%	-20%	-30%	-10%	-10%	-10%

Case study CS3 – DPS GSAs 17-18-19-20

The forecasts are covering the period from 2021 to 2030 to check the biological and economic results in the medium term.

The stock-recruitment relationships have been estimated by using Eqsim (**Errore. L'origine riferimento non è stata trovata.**). **Errore. L'origine riferimento non è stata trovata.** reports the parameters of the stock-recruitment relationships for DPS in GSAs 17-20.

The same fleets as in the hindcasting have been considered in the forecast scenarios.

Table 8 Case study CS3 - DPS GSAs 17-18-19-20: Parameters of the stock recruitment relationships for CS3 for DPS.

Stock	Break Point (b)	a
DPS GSAs 17-20	3000	2433

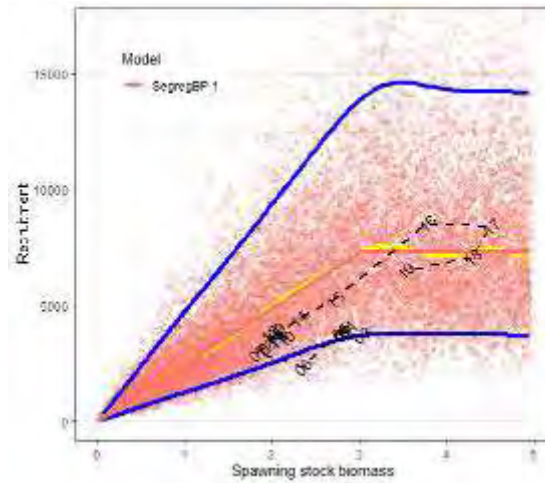


Figure 5 Case study CS3 - DPS GSAs 17-18-19-20: stock recruitment relationships estimated with Eqsim for DPS.

The variable costs (fuel and other) have been assumed to vary proportionally to the annual fishing days, while the fixed and the maintenance costs depend on the annual GT on the basis of the historical data. The labour costs have been assumed in line with the crew share system on the difference revenues minus variable costs, and the depreciation costs depending on the annual GT. For GSA 20 the opportunity and depreciation costs have been assumed as null (because not available) and, in case of inconsistent crew share coefficients, 0.5 was assumed.

The employment was assumed not changing for all scenarios. Moreover, the total landing and the total revenues are assumed proportionally changing with the landing of the target stock, by means of a correction factor based on historical data.

The performance of the scenarios was evaluated on the basis of spawning stock biomass, catch, F, revenues and current revenues to break-even revenues (CR/BER).

For all scenarios the basis was given by the number of fishing days by fleet in the period 2015-2018.

For CS3 three scenarios have been implemented:

S0: status quo (i.e. no variations respect to 2021 (transition phase));

S1: linear reduction to the ref. point until 2026 +closures;

S2: 10% reduction of fishing day (following the formula in the GFCM recommendation) in 2022+Selectivity change in 2023 +closures.

For all the scenarios except the scenario S0 the basis for the closure areas (within 6 nautical miles, until December) was taken in line with the Recommendation GFCM/43/2019/5.

In order to quantify the contribution of these closure areas the same approach followed in the STECF EWG 20-13 was applied in Adriatic Sea within FAIRSEA project. The information of effort allocation (Global Fishing Watch, GFW, <https://globalfishingwatch.org/>) was crossed with the spatial information on nurseries in the area (MEDISEH project), and the overlap with closure areas was identified. The results of this procedure was then translated into a change (increase) of the length at first capture, proportionally shaping the exploitation pattern of the target species applied to all the fleets considered (7% until December).

For scenario S1 a linear reduction of 70% of the fishing days applied to all fleets except Montenegro (fishing less than 1 000 days during the reference period). For scenario S2 the 10% reduction in 2022 was differentiated by fleet according to the allocation formula in the Recommendation GFCM/43/2019/5:

$$CFP\ effort\ reduction_f = Overall\ reduction * \frac{(CFPeffect_f)^2}{\sum_f (CFPeffect_f)^2}$$

For scenario S2 the change in selectivity was cumulatively applied to the scenario S0 increasing the length at first capture of 2 mm, to simulate a 50 mm mesh size (IMPLEMED project).

The effort reductions applied in scenarios S1 and S2 are reported in **Errore. L'origine riferimento non è stata trovata.**

Table 9 Case study CS3 - DPS GSAs 17-18-19-20: Effort reductions applied by year and fleet segment for scenarios S1 and S2 respect to S0. In the first column are the fishing days in 2021, assume din line with Recommendation GFCM/43/2019/5.

Fleet segment	FD	S1					S2				
	2021	2022	2023	2024	2025	2026	2022	2023	2024	2025	2026
ALB_18_DTS_1224	22748	-14%	-27%	-41%	-55%	-70%	-2%	-2%	-2%	-2%	-2%
GRC_20_DTS_VL1824	2422	-14%	-27%	-41%	-55%	-70%	-1%	-1%	-1%	-1%	-1%
GRC_20_DTS_VL2440	2978	-14%	-27%	-41%	-55%	-70%	-1%	-1%	-1%	-1%	-1%
HRV_17_DTS_0612	11797	-14%	-27%	-41%	-55%	-70%	-3%	-3%	-3%	-3%	-3%
HRV_17_DTS_1218	14581	-14%	-27%	-41%	-55%	-70%	-3%	-3%	-3%	-3%	-3%
HRV_17_DTS_1840	6458	-14%	-27%	-41%	-55%	-70%	-3%	-3%	-3%	-3%	-3%
ITA_17_DTS_1218	28202	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
ITA_17_DTS_1840	30060	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
ITA_18_DTS_0612	3142	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
ITA_18_DTS_1218	31274	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
ITA_18_DTS_1840	11963	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
ITA_19_DTS_1218	28644	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
ITA_19_DTS_1824	3664	-14%	-27%	-41%	-55%	-70%	-13%	-13%	-13%	-13%	-13%
MNE_18_DTS_0612	178	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
MNE_18_DTS_1224	983	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
ALL	199093	-14%	-27%	-41%	-54%	-70%	-10%	-10%	-10%	-10%	-10%

The same equation described for CS1 was internally applied by the model to recalculate the fishing mortality.

SMART

Application of the SMART model

The artificial neural network described in Russo et al., 2014 was adapted to be applied on the stock objects produced within the Task 3.3 (Figure 6). These stock objects contain updated and validated information, for the time period 2006-2019, about:

- Catches by species/age;
- Estimated size of population at sea by species/age;
- Mean annual weight at age by species;

The Elman multilayer perceptron network (EMPN) described in Russo et al., 2014 was modified as follows

to process these input data:

- The model explores the data along a time series (i.e. the stock object) and perform the training by tuning the weights in the hidden layer in order to predict population by age and SSB at time t using the information about catches by age, population by age and SSB at the previous time ($t-1$);
- For the forecast period, catches by age at time t are computed using the population by age at the same time and a vector of F by age. This fishing mortality by age class was embedded in the stock objects for the observed time series but was reshaped from the user according to the management scenarios to be evaluated.

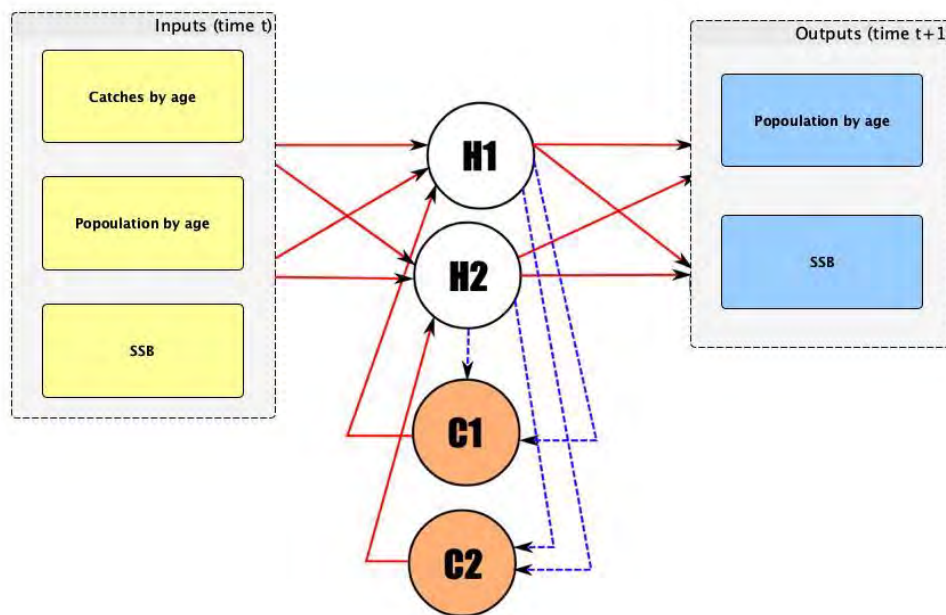


Figure 6 Schematic representation of the Elman network used in this application. The input layer comprises: neurons related to population size by age, catches by age, and SSB at a given time t . These neurons directly propagate the information to the basic hidden neurons (H). At each step of the training procedure, the updated pattern of the basic hidden neurons is memorized by the context neurons (C) and, at the successive step, propagated to the basic hidden neurons together with the new information in the input neurons. The output layer contains as many neurons as the number of cohorts in the input layer. EMPN is the simplest and most widely used ANN architecture to pursue classification issues when sequential or time-varying patterns are inspected. The basic structure of a multilayer perceptron network (MPN) consists of at least three layers of neurons. Information flows in a unidirectional way from the first (input) layer to the last (output) layer through a second (hidden) layer and is processed in parallel by the neurons of each layer. The input layer contains as many neurons as there are independent variables or descriptors used to predict the dependent variables, which in turn constitute the output layer. Neurons of a given layer are linked to the neurons of the next layer by activation functions: hidden layer neurons compute a weighted sum of the input variables through a first activation function and then send a result to the output neurons through a second activation function. A sigmoid function, which is the most common choice because it is nonlinear and characterized by a very easy to compute derivative was chosen in both cases.

For this application of SMART to the case studies of Western Mediterranean (ARA and ARS in CS1; GSAs 1-12) and of Central Mediterranean (ARS in CS2; GSAs 15,16,18, 19 and 20), the fishing footprint exerted by trawlers equipped with AIS data, as reconstructed according to the procedures described in the Deliverable 3.1, was considered. This fishing footprint represented the baseline to quantify the amount of fishing effort in the closures (i.e. Fishery-restricted areas) considered in all the different scenarios. The new values of fishing mortalities by species/age, as determined by these closures, were retrieved by Russo et al., 2019. In addition, the overall fishing effort was reduced according to Table 1.

The results of the training phase were evaluated comparing observed and predicted (i.e. fitted) values for SSB and population size (number of individuals) by age. The outputs of the training phase for the three case studies are represented in Figures 7-9. The neural network predicts N per age at time t based on information relating to the previous year. The graphs of observed vs. predicted N at age show that for the

0+ cohort the uncertainty is higher than for the others age groups.

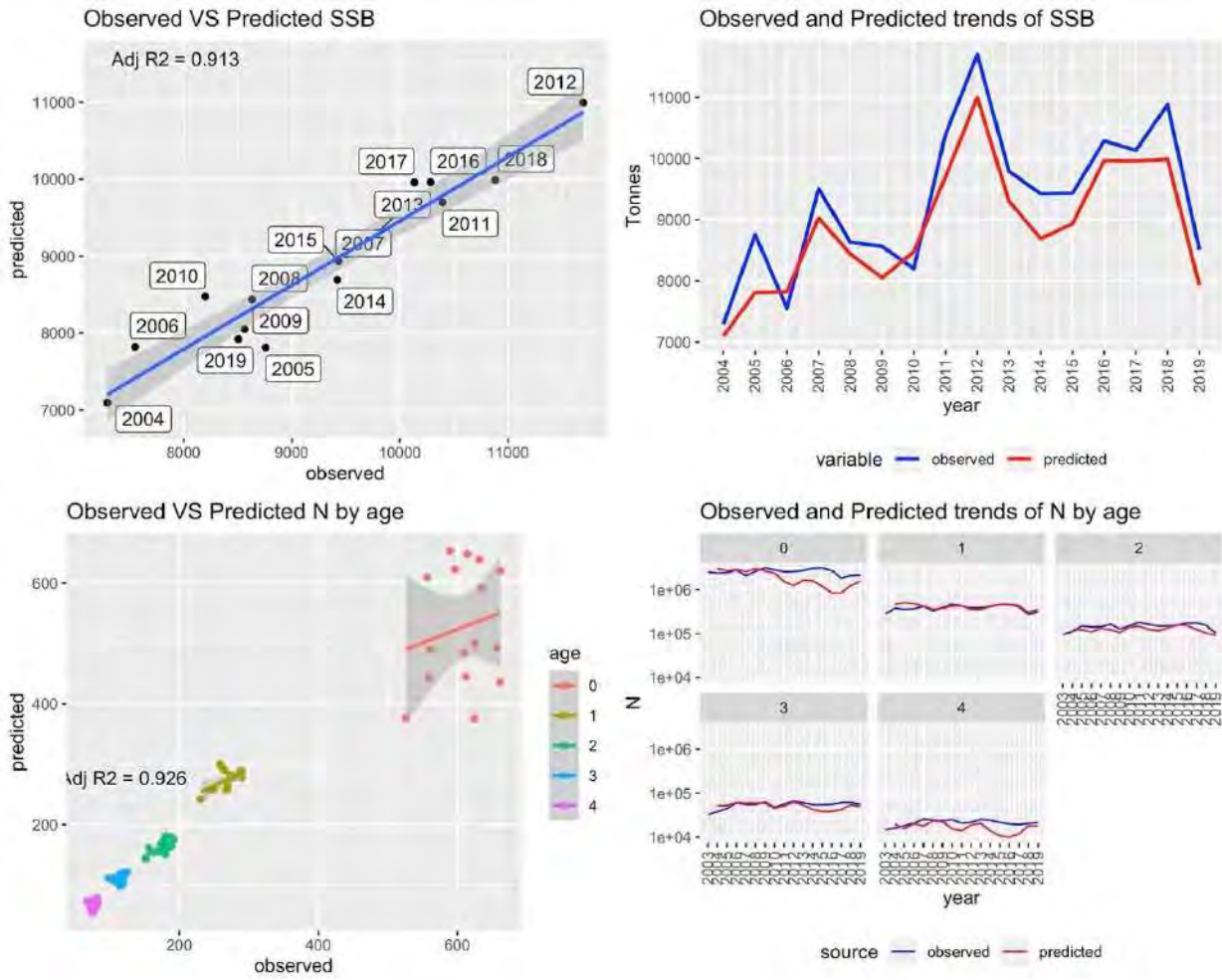


Figure 7 Representation of the outputs of the training phase for ARS in the case study #1. The top left panel allows to compare predicted and observed values of SSB. The top right panel presents the same values as time series. The bottom left panel allows to compare predicted and observed values of the population size (number of individuals in log scale) The bottom right panel presents the same values as time series. Values of the adjusted R^2 are provided.

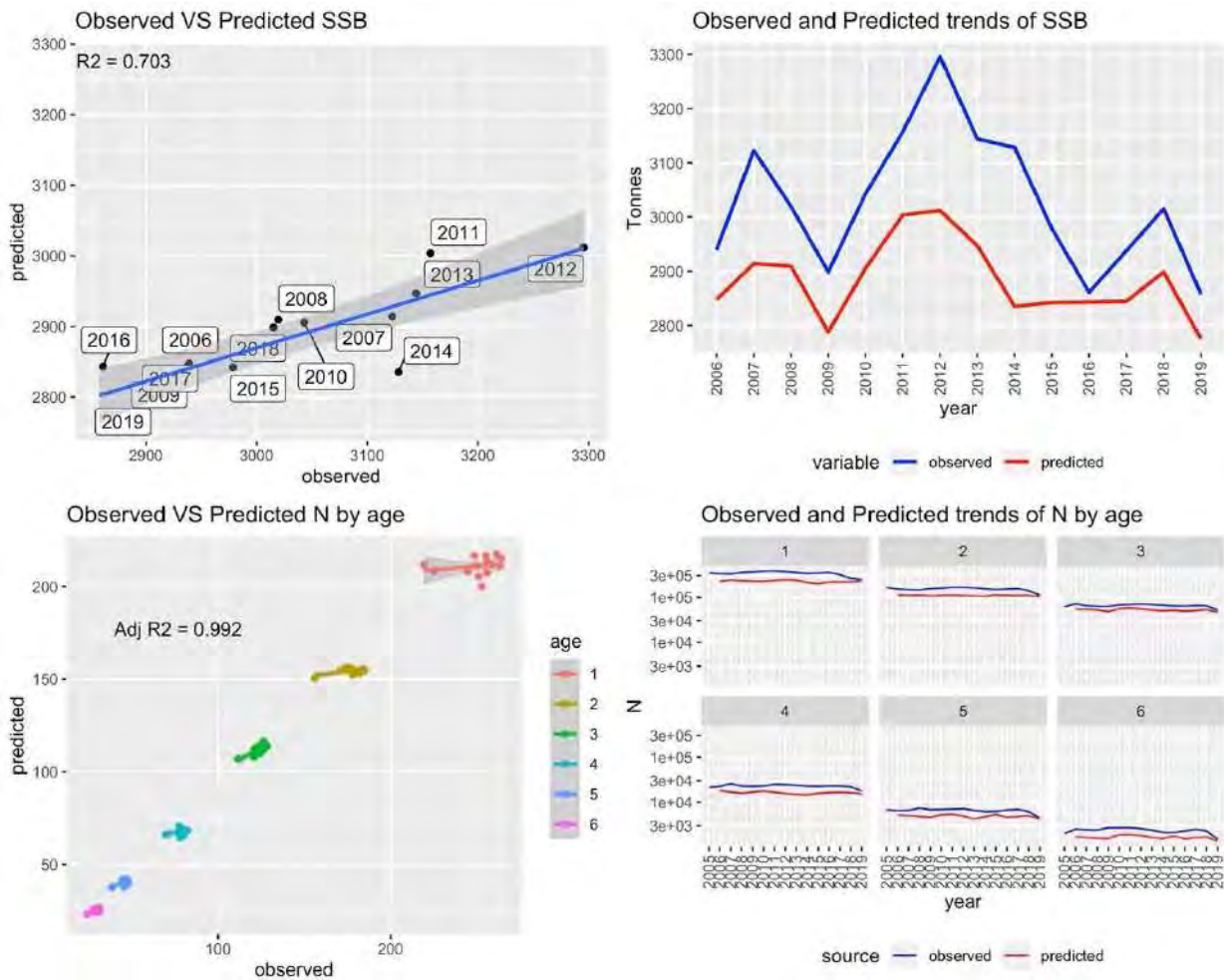


Figure 8 Representation of the outputs of the training phase for ARA in the case study #1. The top left panel allows to compare predicted and observed values of SSB. The top right panel presents the same values as time series. The bottom left panel allows to compare predicted and observed values of the population size (number of individuals in log scale) The bottom right panel presents the same values as time series. Values of the adjusted R² are provided.

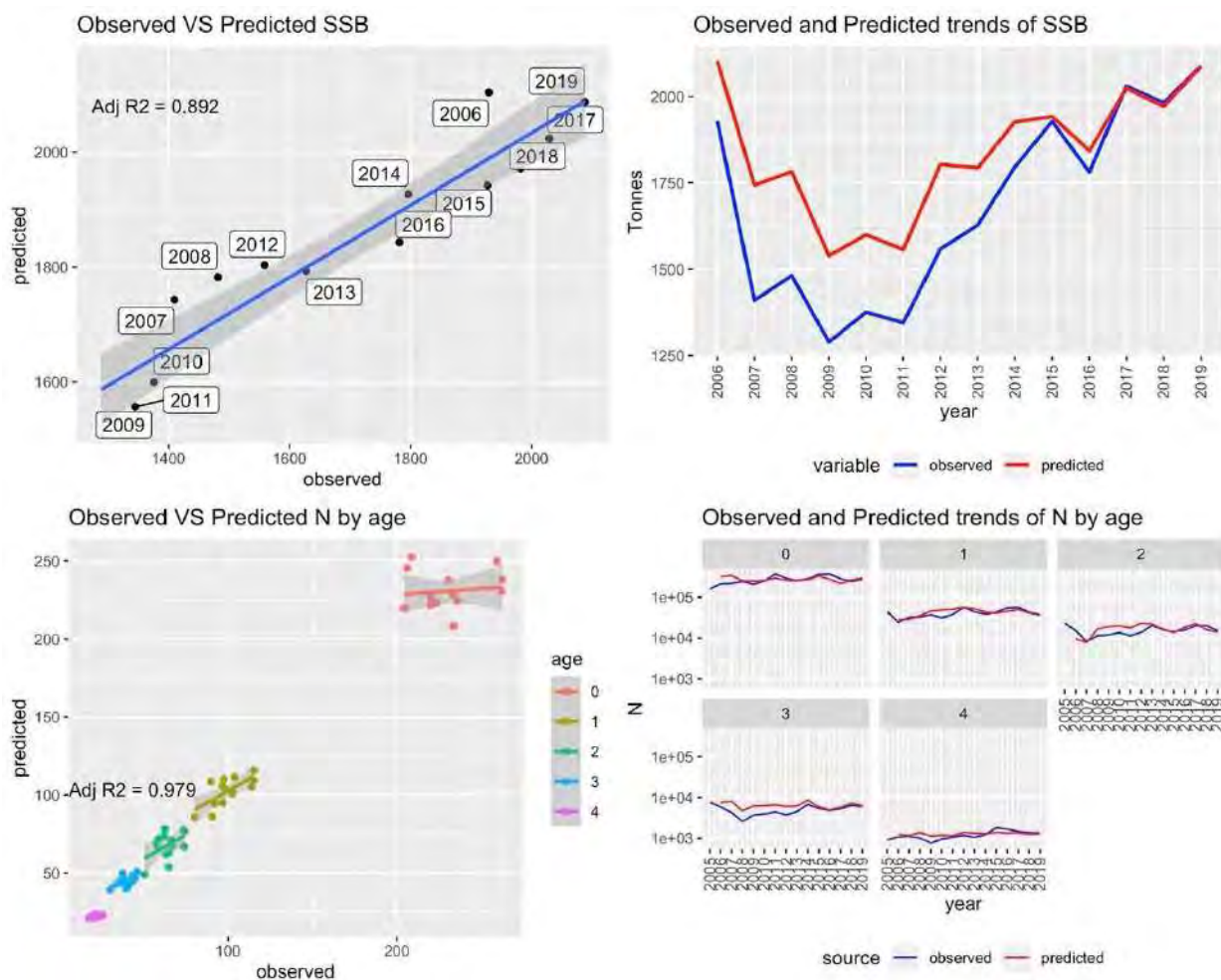


Figure 9 Representation of the outputs of the training phase for ARS in the case study #2. The top left panel allows to compare predicted and observed values of SSB. The top right panel presents the same values as time series. The bottom left panel allows to compare predicted and observed values of the population size (number of individuals in log scale) The bottom right panel presents the same values as time series. Values of the adjusted R^2 are provided.

As a general comment, the model returned a consistent and good fit for all the three case studies, as evidenced by the values of the adjusted R^2 .

Results

BEMTOOL

CS1 - ARS GSAs 1-12 and GSAs ARA 1-12

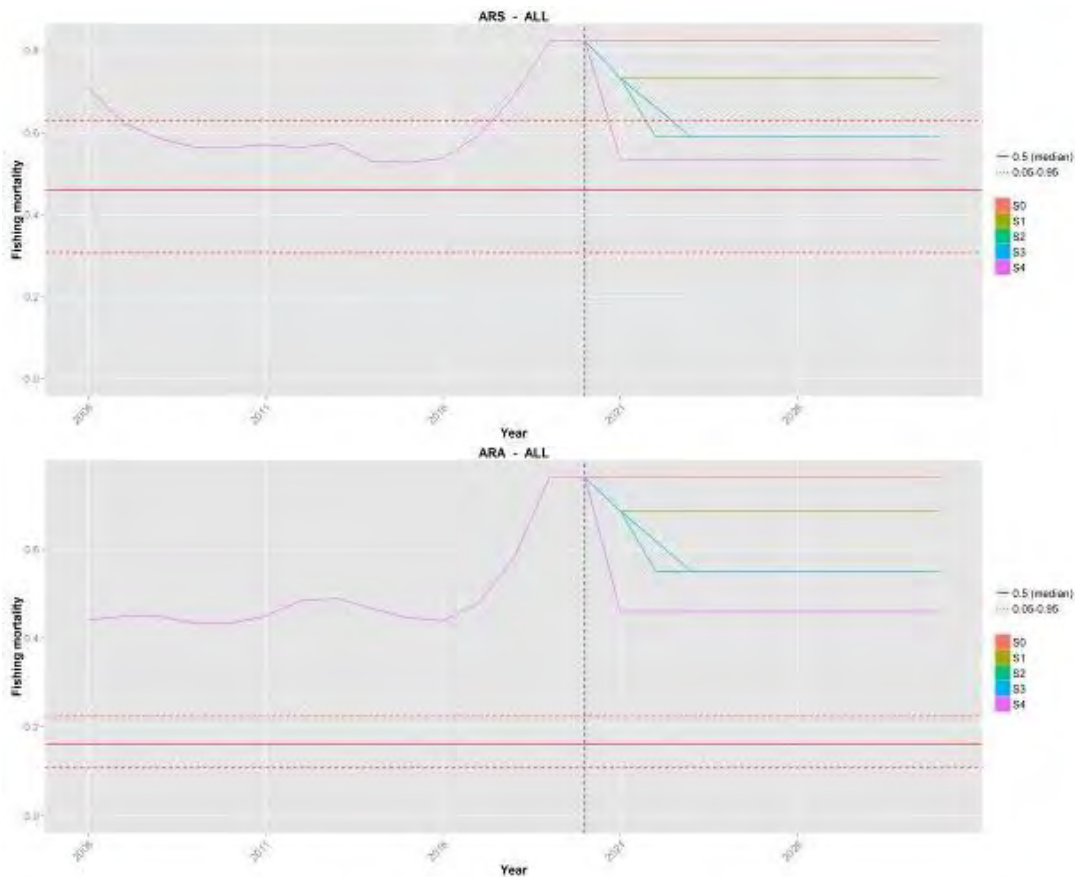


Figure 10 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Overall F of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020. Red horizontal solid line corresponds to the $F_{MSY}=F_{0.1}$, and red horizontal dashed lines correspond to F_{upper} and F_{lower} . and **Errore. L'origine riferimento non è stata trovata.** report the F and SSB for ARS and ARA for each scenario. As ARS concerns, all alternative scenarios allow to reach the upper limit of the F_{msy} , except S1, while for ARA no scenario decreases the F until reaching the reference range for the stock. For both stocks, S4, combining an effort reduction of 10% with a change in the exploitation pattern, is the best performing scenario, allowing to reduce the F of around 35-40%; this percentage is quite close to the percentage required by ARS but only half of the reduction required by ARA (**Errore. L'origine riferimento non è stata trovata.**). Although S2 and S3 scenarios are quite equivalent in terms of F, the 20% of effort reduction applied in 2022 in S2 produces increase in SSB and total landing slightly higher than S3, that distributes the 20% reduction in two years. The projections show that S4 increases the SSB of ARS and ARA of 63% and 144 % respectively, with a decrease in landing of 5% for ARS and an increase of 61% for ARA.

From an economic point of view S2 and S3 are very similar and allow to reach revenues well above S0 and S1 (**Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**). Although in the short term S4 projections show a sharp decrease in revenues and CR/BER, due to the change in selectivity, this scenario allows to reach, in the long term, values quite comparable to the ones of S2 and S3, reducing the effort of 10% instead of 30% overall. In the medium term the revenues and CR/BER are slightly higher, highlighting the benefit of the improvement in the exploitation pattern (**Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**).

In general, despite in the simulations the assumption of no change in the number of employees was made, a reduction in the employment in the short term could be expected, due to a decrease in the effort.

Table 10 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: F, SSB and overall landing (discard null) of ARS and ARA corresponding to the S0 (SQ) scenario and percentage of variation of the S1, S2, S3 and S4 scenarios respect to SQ in 2026.

Stock	Indicator	S0 (SQ)	S1	S2	S3	S4
-------	-----------	---------	----	----	----	----

ARS	F	0.82	-11%	-28%	-28%	-35%
	SSB (tons)	477	15%	36%	37%	63%
	Catch (tons)	469	1%	-3%	-3%	-5%
ARA	F	0.76	-10%	-28%	-28%	-40%
	SSB (tons)	1182	29%	85%	81%	144%
	Catch (tons)	787	18%	45%	40%	61%

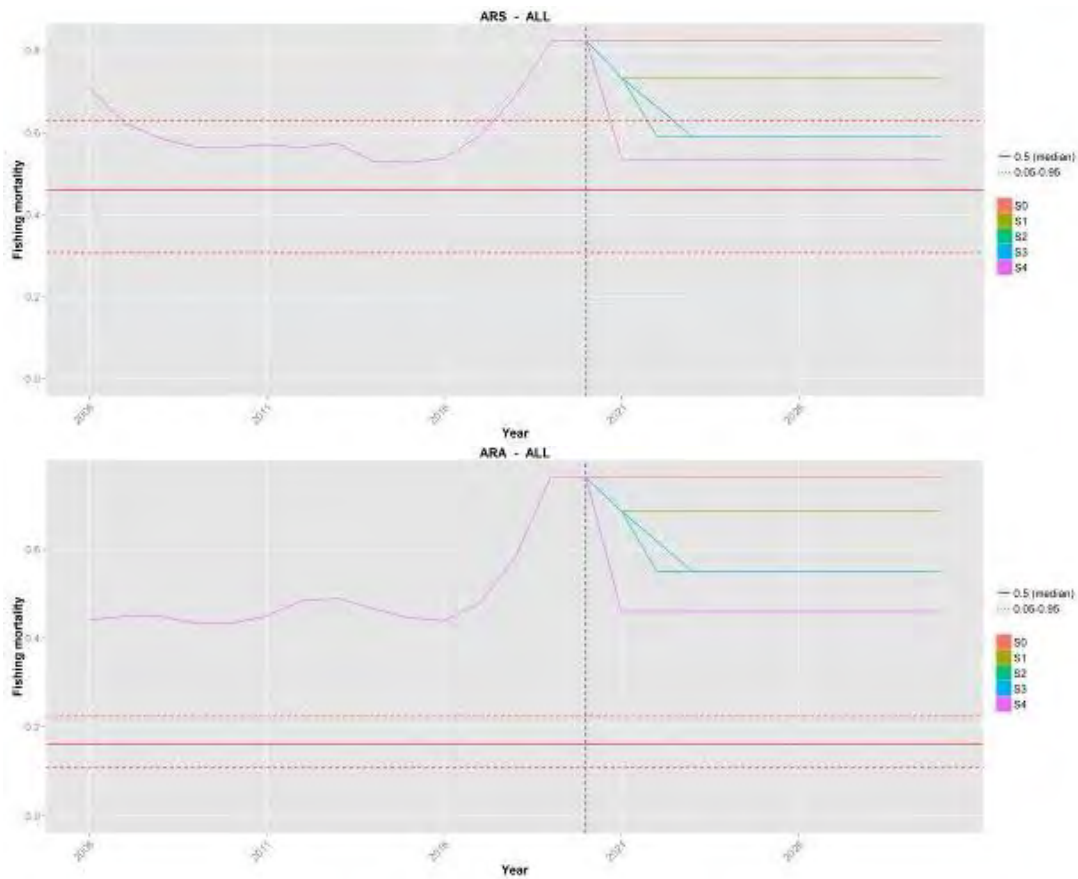
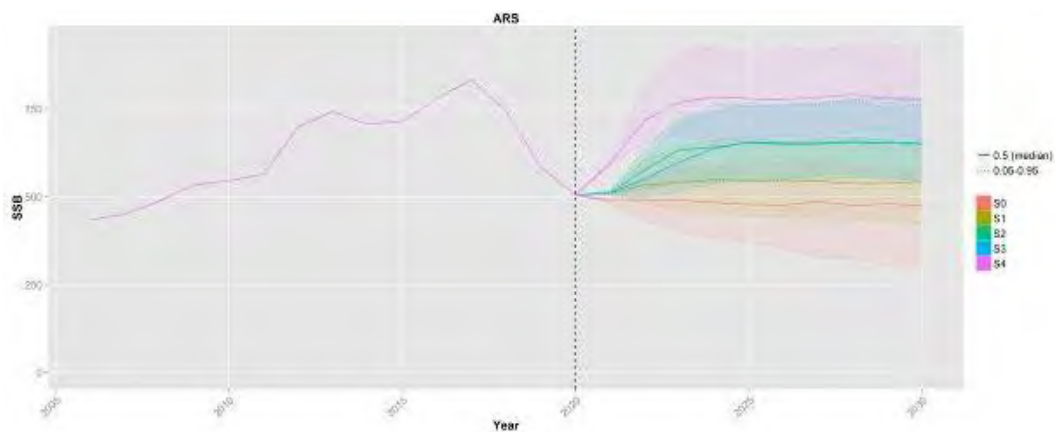


Figure 10 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: Overall F of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020. Red horizontal solid line corresponds to the $FMSY=F0.1$, and red horizontal dashed lines correspond to $Fupper$ and $Flower$.



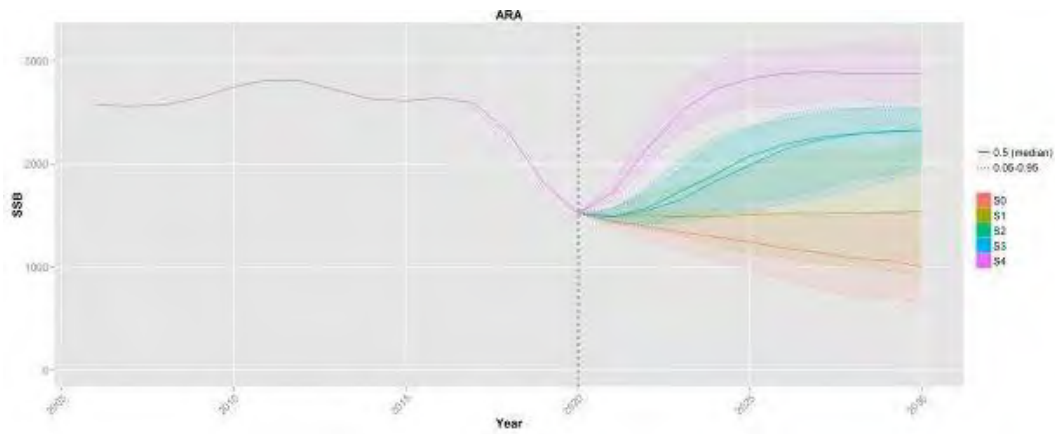


Figure 11 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: SSB of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020.

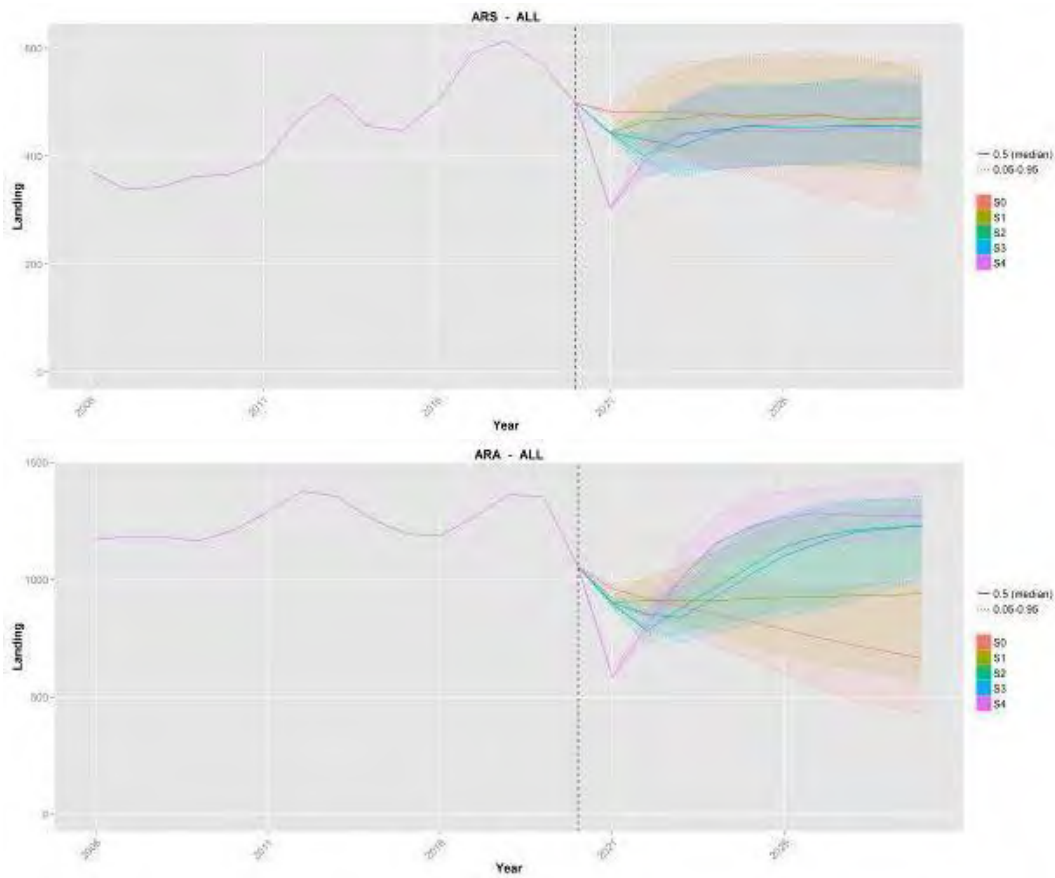


Figure 12 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: Overall landing of ARS and ARA for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020 (discard null).

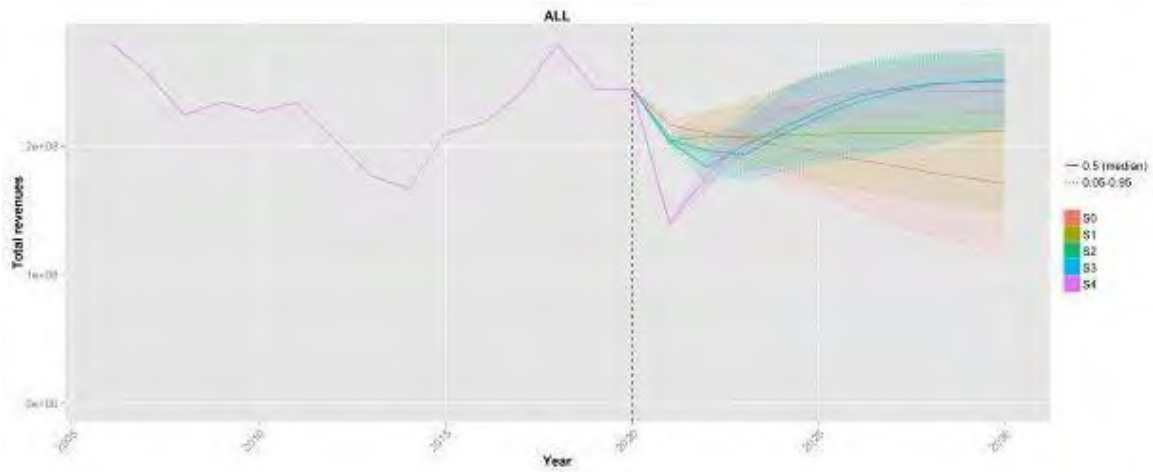


Figure 13 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: Total revenues of the overall fleet for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020.

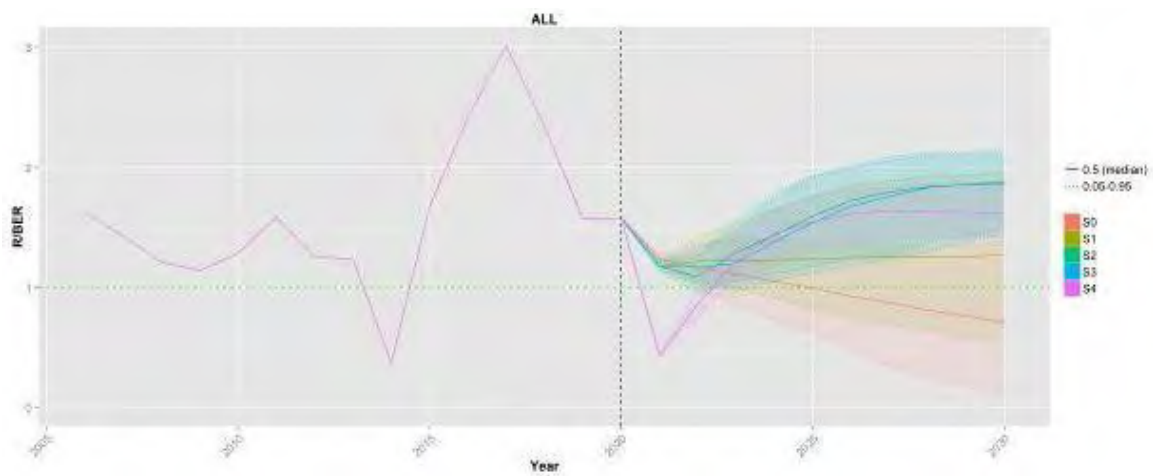


Figure 14 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: CR/BER of the overall fleet for Scenarios S0, S1, S2, S3 and S4. The black vertical dashed lines corresponds to 2020.

Table 11 Case study CS1 - ARS GSAs 1-12 and ARA GSAs 1-12: Total revenues by fleet segment corresponding to the S0 (SQ) scenario and percentage of variation of the Scenarios S1, S2, S3 and S4 respect to SQ in 2026.

Fleet segment	Baseline (S0)	S1	S2	S3	S4
GSA1_DTS_VL1218	6161639	14%	34%	31%	39%
GSA1_DTS_VL1824	11689743	14%	34%	31%	39%
GSA1_DTS_VL2440	2181444	14%	34%	31%	39%
GSA10_DTS_VL0612	393089	5%	7%	6%	12%
GSA10_DTS_VL1218	11277429	4%	6%	5%	4%
GSA10_DTS_VL1824	11164868	4%	6%	5%	3%
GSA10_DTS_VL2440	3885931	3%	5%	4%	0%
GSA11_DTS_VL1218	8100476	9%	19%	17%	20%
GSA11_DTS_VL1824	8220794	9%	19%	18%	21%
GSA11_DTS_VL2440	8273611	10%	22%	21%	26%
GSA5_DTS_VL1218	1005674	14%	33%	30%	38%
GSA5_DTS_VL1824	9670577	14%	34%	31%	38%
GSA5_DTS_VL2440	3874845	14%	34%	31%	39%
GSA6_DTS_VL1218	13978039	14%	34%	31%	39%

Fleet segment	Baseline (S0)	S1	S2	S3	S4
GSA6_DTS_VL1824	46065190	14%	34%	31%	39%
GSA9_DTS_VL0612	1096641	10%	25%	23%	19%
GSA9_DTS_VL1218	15820425	10%	26%	23%	22%
GSA9_DTS_VL1824	24318726	10%	24%	22%	19%
GSA9_DTS_VL2440	2084042	10%	25%	23%	21%
ALL	189263183	11%	26%	24%	27%

Table 12 Case study CS1 - ARS GSAs 1-12 and GSAs ARA 1-12: CR/BER by fleet segment corresponding to the S0 (SQ), S1, S2, S3 and S4 scenarios in 2026.

Fleet segment	Baseline (S0)	S1	S2	S3	S4
GSA1_DTS_VL1218	8.38	10.65	14.08	13.68	14.05
GSA1_DTS_VL1824	2.08	3.01	4.43	4.28	4.30
GSA1_DTS_VL2440	1.52	2.40	3.74	3.61	3.55
GSA10_DTS_VL0612	0.92	0.97	1.00	1.00	1.06
GSA10_DTS_VL1218	1.69	1.90	2.15	2.13	1.91
GSA10_DTS_VL1824	-0.07	-0.02	0.05	0.05	-0.02
GSA10_DTS_VL2440	0.98	1.02	1.06	1.05	0.99
GSA11_DTS_VL1218	2.03	2.46	3.02	2.95	2.87
GSA11_DTS_VL1824	1.26	1.56	1.94	1.90	1.85
GSA11_DTS_VL2440	0.31	0.51	0.79	0.76	0.72
GSA5_DTS_VL1218	5.30	6.60	8.53	8.30	8.58
GSA5_DTS_VL1824	4.15	5.27	6.95	6.75	6.96
GSA5_DTS_VL2440	3.68	4.70	6.24	6.06	6.23
GSA6_DTS_VL1218	0.75	1.19	1.87	1.79	1.85
GSA6_DTS_VL1824	0.60	1.01	1.65	1.58	1.61
GSA9_DTS_VL0612	1.29	1.43	1.65	1.62	1.56
GSA9_DTS_VL1218	0.64	0.84	1.16	1.13	1.00
GSA9_DTS_VL1824	0.74	1.01	1.46	1.42	1.19
GSA9_DTS_VL2440	0.56	0.76	1.11	1.07	0.92
ALL	0.93	1.25	1.73	1.68	1.61

CS3 – DPS GSAs 17-18-19-20

Errore. L'origine riferimento non è stata trovata. and **Errore. L'origine riferimento non è stata trovata.** report the simulated F and SSB for DPS for each scenario. The decrease of the overall F is higher for S1, while the combination of effort reduction and change in exploitation pattern of S2 scenario returns a decrease of 16% that does not allow to reach the F_{msy} range (**Errore. L'origine riferimento non è stata trovata.**, **Errore. L'origine riferimento non è stata trovata.**), but allows to increase the SSB of 40% respect to S0 in 2026. In terms of SSB, S1 allows to obtain the best performance, followed by S2; on the other hand, scenario S2 returns the higher overall catch than S1 in the long term, followed by S0 (**Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**). Although both S1 and S2 decrease remarkably the discard respect to S0 in the long term, S2 allows to reduce the discard fraction already from 2022 and to reach the smallest value, due to the increase of size at first capture and the improvement of exploitation pattern (**Errore. L'origine riferimento non è stata trovata.**).

From an economic point of view S2 is the scenario best performing in terms of total revenues, allowing to

reach values in line with S0, after an initial reduction in the short term; the increase on the overall fleet revenues is of 3%, but varies among the different fleet segments between 2% and 23% (**Errore. L'origine riferimento non è stata trovata.**). On the other hand, S1, after the reduction in revenues until 2026, due to implementation of the management measures, allow to reach values slightly lower than status quo (**Errore. L'origine riferimento non è stata trovata.**). Although for the vessels still remaining in the sector after the 70% reduction, S1 would allow to obtain the highest values in terms of profitability, this is due to the important reduction of fishing days and, thus, of variable costs (**Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**).

In general, despite in the simulations the assumption of no change in the number of employees was made, a reduction in the employment in the short term could be expected, due to a decrease in the effort.

Table 13 Case study CS3 - DPS GSAs 17-18-19-20: F, SSB and overall catch of DPS corresponding to the S0 (SQ) scenario and percentage of variation of the S1 and S2 scenarios respect to SQ in 2026.

Indicator	S0 (SQ)	S1	S2
F	1.4	-69%	-16%
SSB (tons)	3462	93%	40%
Catch (tons)	4950	-23%	9%

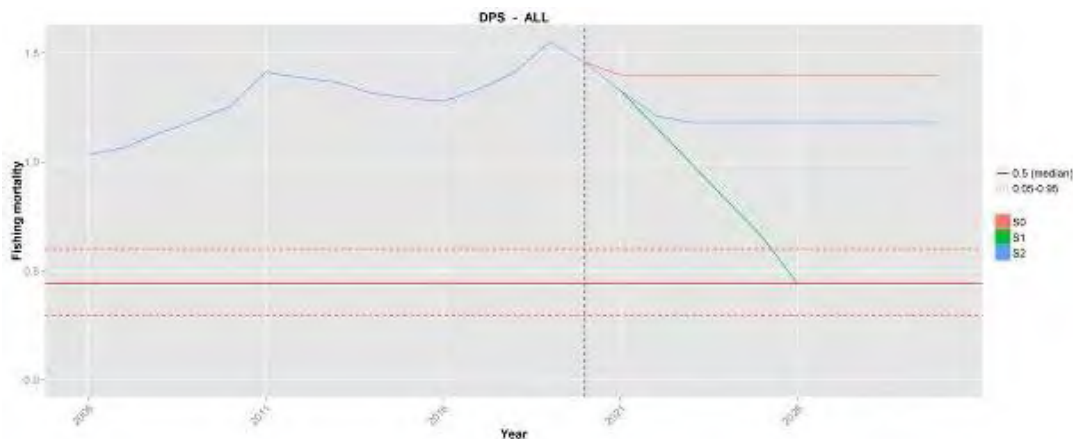


Figure 15 Case study CS3 - DPS GSAs 17-18-19-20: Overall F of DPS for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020. Red horizontal solid line corresponds to the FMSY=F0.1, and red horizontal dashed lines correspond to Upper and Lower.

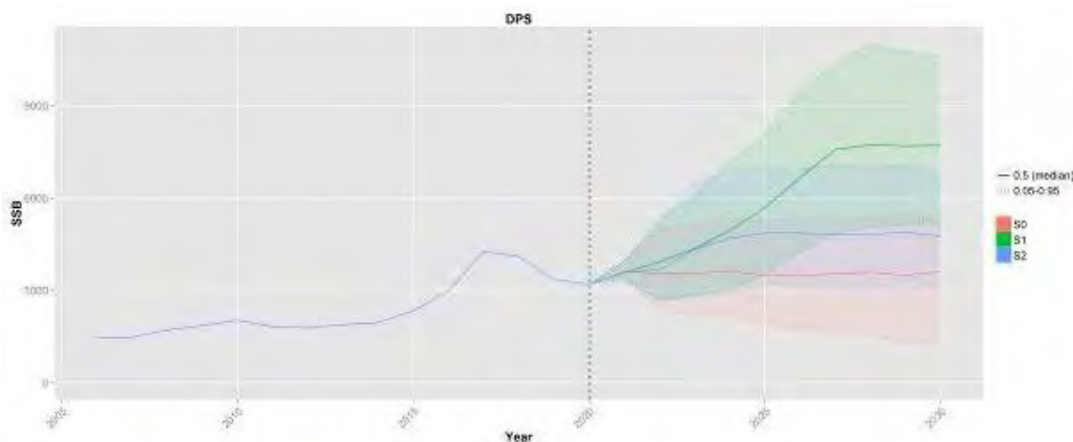


Figure 16 Case study CS3 - DPS GSAs 17-18-19-20: SSB of DPS for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020.

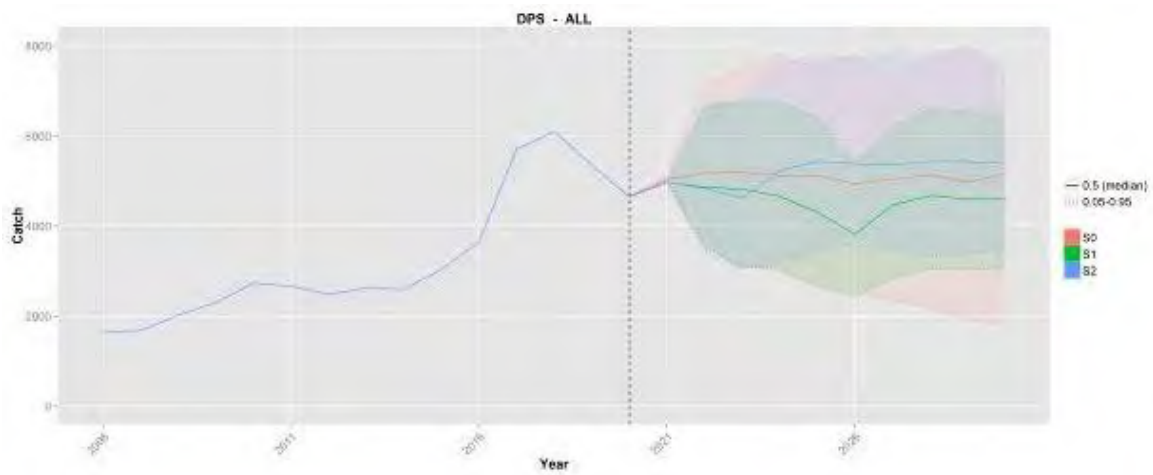


Figure 17 Case study CS3 - DPS GSAs 17-18-19-20: Overall catch of DPS for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020.

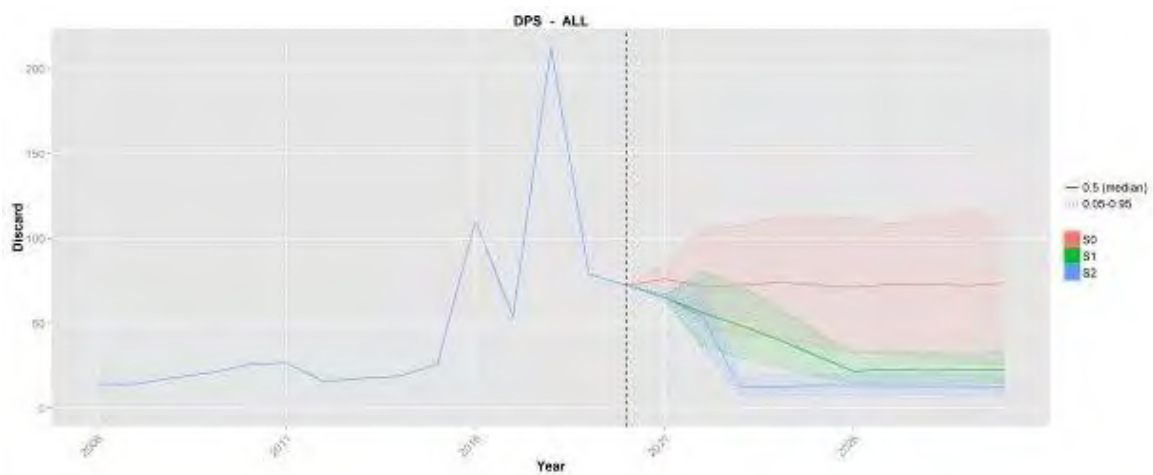


Figure 18 Case study CS3 - DPS GSAs 17-18-19-20: discard of the overall fleet for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020.

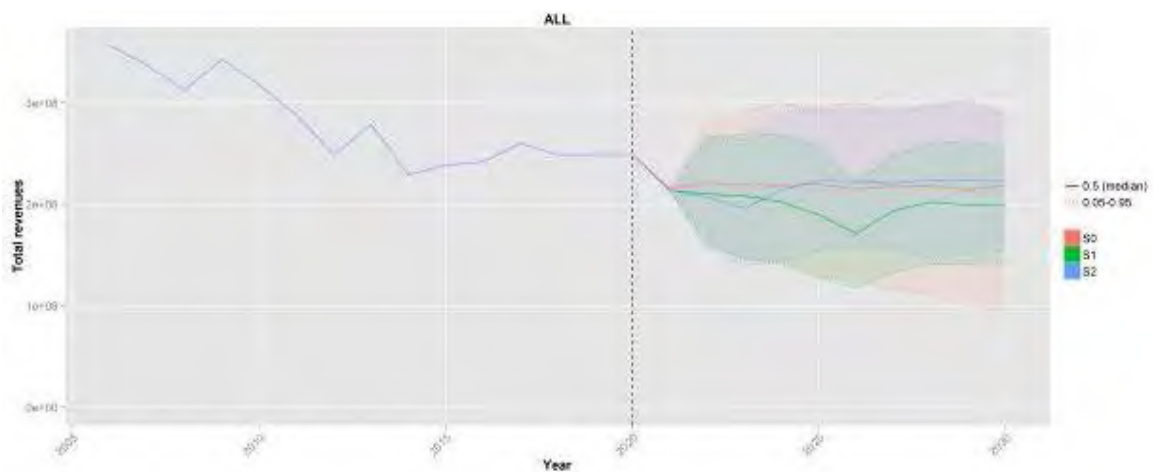


Figure 19 Case study CS3 - DPS GSAs 17-18-19-20: Total revenues of the overall fleet for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020.

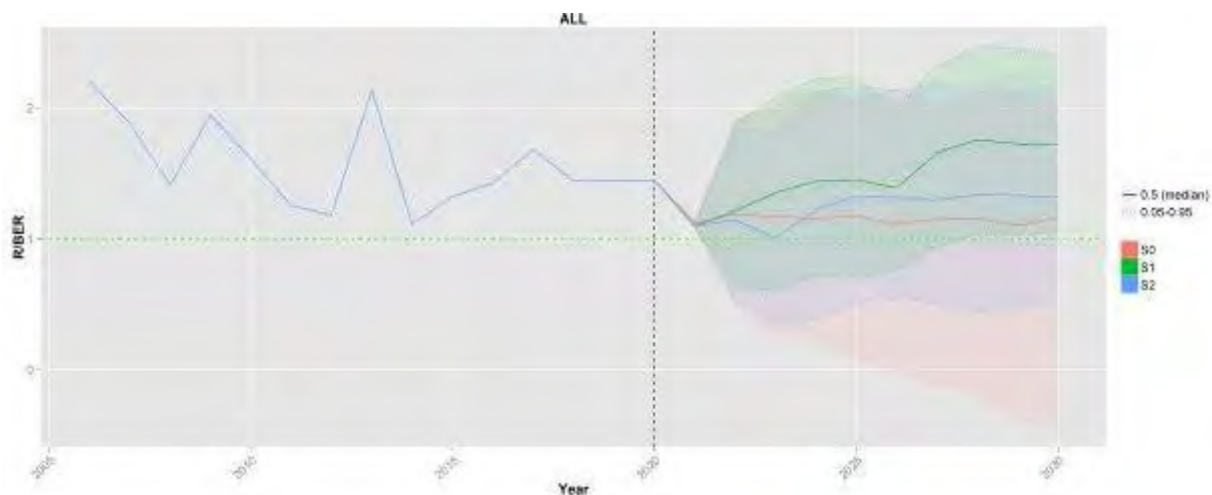


Figure 20 Case study CS3 - DPS GSAs 17-18-19-20: CR/BER of the overall fleet for Scenarios S0, S1 and S2. The black vertical dashed lines corresponds to 2020.

Table 14 Case study CS3 - DPS GSAs 17-18-19-20: Total revenues by fleet segment corresponding to the S0 (SQ) scenario and percentage of variation of the S1 and S2 scenarios respect to SQ in 2026.

Fleet segment	Baseline (S0)	S1	S2
ALB_18_DTS_1224	10029416	-17%	19%
GRC_20_DTS_VL1824	3417532	-16%	23%
GRC_20_DTS_VL2440	4834854	-16%	23%
HRV_17_DTS_0612	2368917	-17%	17%
HRV_17_DTS_1218	6008672	-18%	17%
HRV_17_DTS_1840	5794649	-18%	16%
ITA_17_DTS_1218	31333998	-20%	4%
ITA_17_DTS_1840	61535222	-24%	-2%
ITA_18_DTS_0612	2526614	-19%	9%
ITA_18_DTS_1218	39493681	-23%	-1%
ITA_18_DTS_1840	17958783	-20%	2%
ITA_19_DTS_1218	23317768	-19%	5%
ITA_19_DTS_1824	5111691	-19%	4%
MNE_18_DTS_0612	100672	104%	21%
MNE_18_DTS_1224	949606	104%	21%
ALL	214782077	-20%	3%

Table 15 Case study CS3 - DPS GSAs 17-18-19-20: CR/BER by fleet segment corresponding to the S0 (SQ) scenario, S1 and S2 scenarios in 2026.

Fleet segment	Baseline (S0)	S1	S2
ALB_18_DTS_1224	0.8	2.2	1.6
GRC_20_DTS_VL1824	-2.0	4.2	1.3
GRC_20_DTS_VL2440	4.8	17.1	13.7
HRV_17_DTS_0612	0.0	0.0	0.0
HRV_17_DTS_1218	0.1	0.3	0.3
HRV_17_DTS_1840	0.0	0.2	0.1
ITA_17_DTS_1218	1.8	2.0	2.1

Fleet segment	Baseline (S0)	S1	S2
ITA_17_DTS_1840	0.9	1.1	1.0
ITA_18_DTS_0612	4.6	4.3	5.3
ITA_18_DTS_1218	3.3	3.4	3.5
ITA_18_DTS_1840	0.8	1.1	0.9
ITA_19_DTS_1218	1.4	1.9	1.8
ITA_19_DTS_1824	0.8	1.0	1.0
MNE_18_DTS_0612	1.5	5.6	2.3
MNE_18_DTS_1224	1.5	6.5	2.5
ALL	1.1	1.4	1.3

SMART

CS1 - ARS GSAs 1-12 and GSAs ARA 1-12

The predicted trends of SSB for the ARS in the GSAs 1-12, according to the *Status quo* S0 and to the different management scenarios (Figure 21), indicate that a decreasing of the stock is expected in the present situation and that only some of the explored scenarios are likely to stop or reduce this decline. The most promising approach is represented by the Scenario S4, which integrates spatial closures, effort reduction and changes in the selectivity. The management measures in the Scenarios S1 and S3 seem adequate to slow down the decreasing trend, while those in the Scenario S2 are likely to stabilize the SSB values after 2 years.

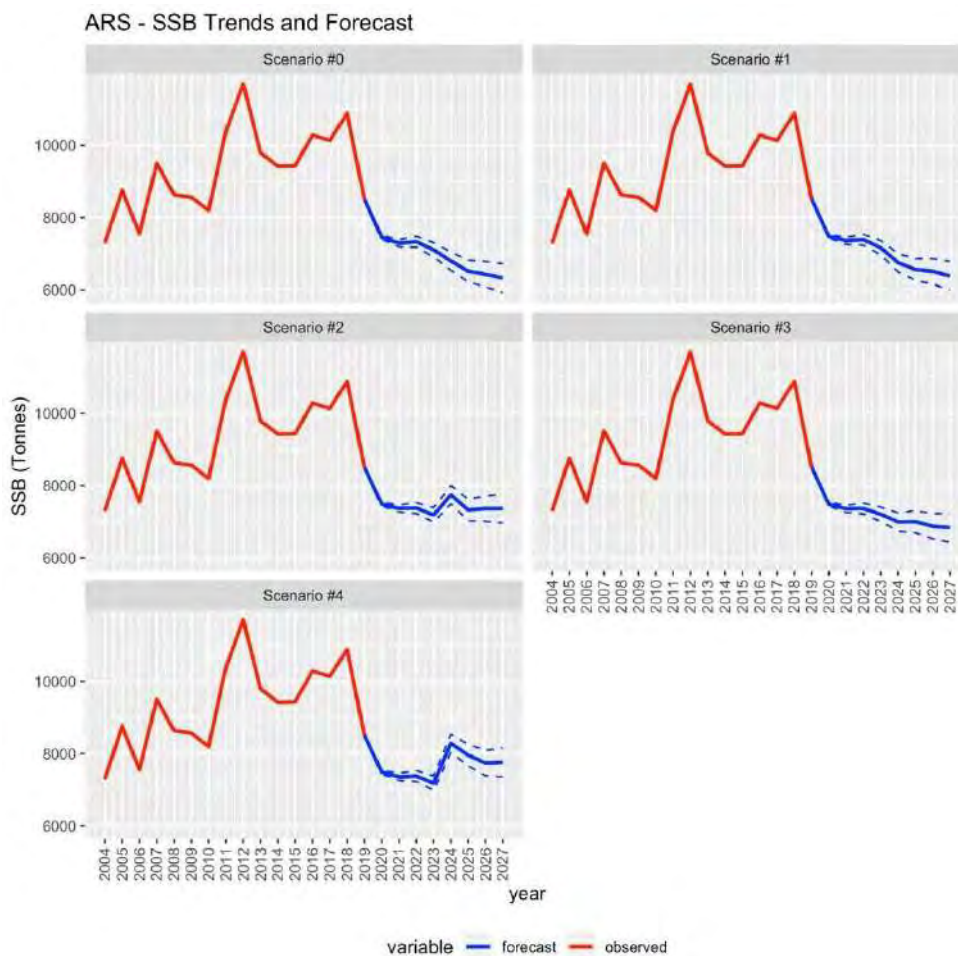


Figure 21 Case study CS1 - GSAs ARS 1-12: SSB of ARS simulated by SMART according to Scenarios S0, S1, S2, S3 and S4.

In the case of ARA in the GSAs 1-12 (Figure 22), the model suggests that management measures in the Scenario S1 are not enough to stop the expected decline observed in the *Status quo*, whereas those in the Scenario S2 should allow to stop the decline and stabilize the SSB. However, the management measures in the Scenario S3 seem more effective, at least in the middle term, and those in the Scenario S4 are expected to determine a partial recovery of this stock.

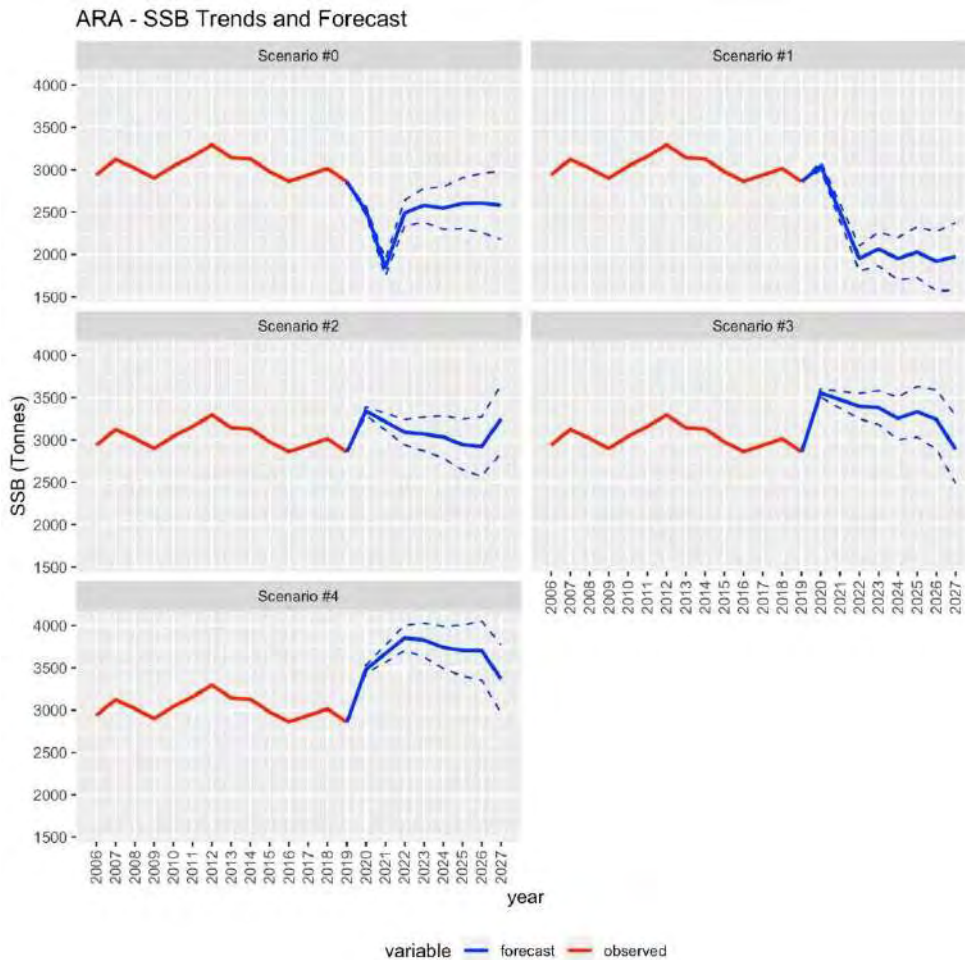


Figure 22 Case study CS1 - GSAs ARA 1-12: SSB of ARA simulated by SMART according to Scenarios S0, S1, S2, S3 and S4.

CS2 - ARS GSAs 15,16,18,19 and 20

The predicted trends of SSB for the ARS in the GSAs 15-16 and 18-20, according to the *Status quo* and to the different management scenarios (Figure 23), indicate that an increasing of the stock is expected in the present situation and that only the set of management measures considered in the Scenario #4 could modify this dynamic of the stock and improve its condition. Scenarios #1-3, however, are associated with a declining phase in the last part of the forecast period. This decline is expected to return the system to its initial (present) condition, whereas the management measures in the Scenarios #4 should significantly support an initial recovery that, although followed by a decline phase, is expected to guarantee a relevant increase of the SSB.

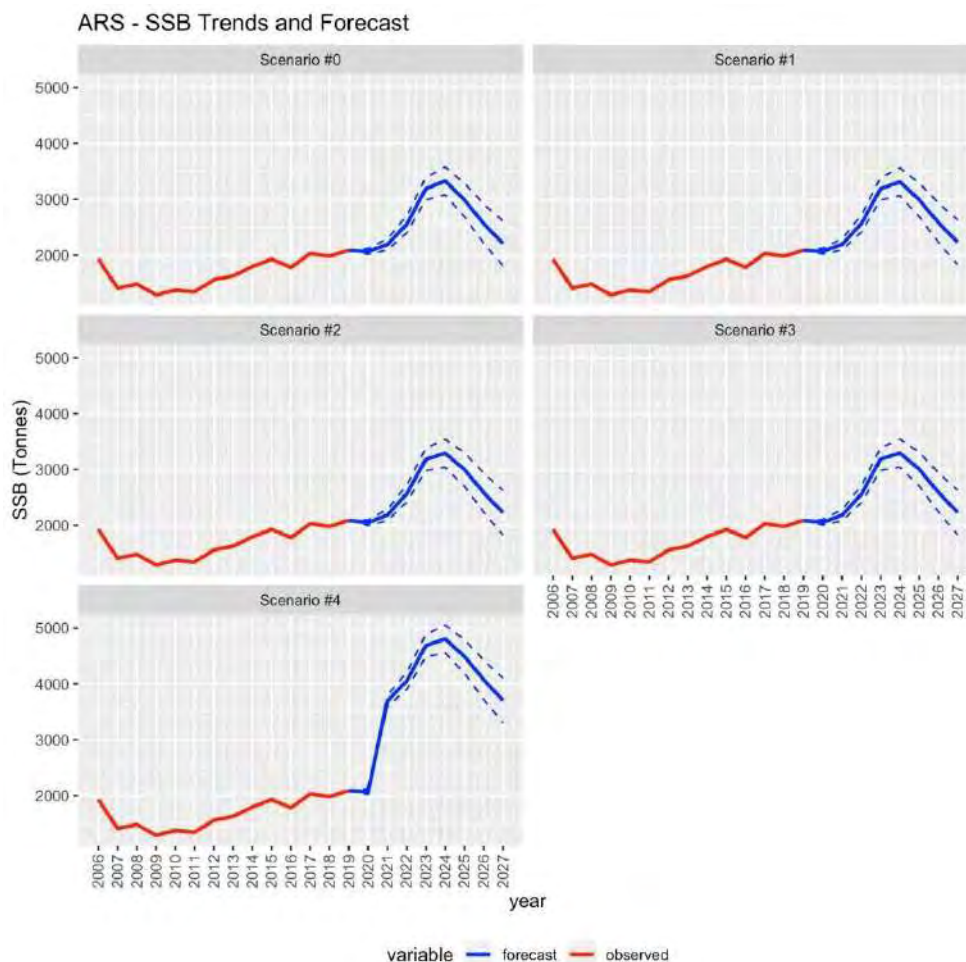


Figure 23 Case study CS2 - ARS GSAs 15, 16, 18, 19 and 20: SSB of ARS simulated by SMART according to Scenarios S0, S1, S2, S3 and S4.

A tentative roadmap with actions required in order to develop adaptive spatial fisheries management

Adaptive fisheries management has been widely recommended as a way to deal with high uncertainty in stock status and management decisions (Holling 1978, Walters 1986). This approach emphasizes the identification of critical uncertainties concerning fishery resource dynamics and the design of diagnostic management experiments and large scale monitoring systems to reduce these uncertainties (Walters 2007). Uncertainty on stock spatial structure and impact of fishing effort on spatial distribution of exploited populations is considered one of the main factor that can made fisheries management ineffective to maintain population persistence throughout time (McGilliard et al., 2015; Goethel et al., 2016; Carson et al., 2017; Kerr et al., 2017; Khoukh & Maynou, 2018). Recently, considerable progress in improving knowledge of spatial structure of populations has been achieved by simulation modelling. It is a useful approach to evaluate the consequences of spatial structure and connectivity in explaining better the variation and reduce uncertainty of productivity and dynamics of local and regional populations (Kerr et al., 2014).

A recent review by Kerr et al. (2017) has shown that spatial structure and connectivity (both in terms of larval dispersal and fish migrations) within and between populations affected strongly the productivity (spawning-stock biomass, SSB), stability (variation in SSB), resilience (time to rebuild SSB after environmental disturbance), and sustainability (maximum sustainable fishing mortality and yield) of exploited populations. Overall, studies reported that stock sub units with high productivity contributed

disproportionately to the resilience of systems to fishery. On the other hand, increased synchrony of subunits responses to environmental drivers decreased the stability of the overall system.

Reviewing main aspects linked to address biological structure of fishery resource for stock assessment and management purposes, Kerr et al., (2017) evidenced some limits to be considered. In particular, the following aspects should be considered: i) the expected benefits by adding spatial complexity in stock assessment and management, ii) the variation in species distribution due to climate change and iii) the difficult to identification of correct population structure in migratory species.

Excluding the last aspect, that can be considered negligible for the species targeted by the MEDUNITS project, limitations in adding spatial information to stock assessment are due to the spatiotemporal scale at which data are collected. Collection of spatial detailed fishery dependent data which are essential for applying spatially-explicit models, such as the extended version of SMART, could be difficult mainly when investigated stocks are shared by countries with different level of technological skills, sampling capabilities and fishery management framework. This is the situation of deep waters fisheries in which fleet belonging to European, African and Middle-East countries exploit the same stocks (Fiorentino & Vitale, 2021). Thus, the choice about what kind of operating model has to be used should consider the biological characteristics of the species, the scale at which the species can practically be assessed and managed, and the willing of stakeholders to be involved in highly detailed in time and space data collection (Walters, 2007; McBride, 2014).

Moreover, it should be mentioned that variations in fish distributions due to climate change can cause further difficulties for designing short and long-term stock assessments, management measures and related sampling programs (Pinsky et al., 2013; Kleisner et al., 2016). Within a framework of an adaptive fishery management, modifications of the spatial distribution and quality of EFHs due to climate change should be monitored to reduce uncertainty in stock assessment and effectiveness of implemented conservation measures (Shackell et al., 2014; Kritzer et al., 2016). Variation in stock distribution may also result in a change in dynamics of the fishing fleets, make difficult stock assessment and determination of management target.

Developing stock assessment and simulations models including information on population structure and connectivity of different life phases, can provide a basis for a more realistic MSE under more flexible management scenarios (Kerr and Goethel, 2014; Goethel et al., 2016).

Taking into account Kerr and Goethel (2014) suggestions, the ideal key steps in the development of MSE for evaluating the implications of spatial stock structure in assessment and management could include:

- i. Development of operating models that represent the most probable configuration of population structure of the fishery resource, using the best available data and finding the trade-off between completeness and resolution (in terms of space and population structure/dynamic);
- ii. Simulation of alternative management strategies
 - a. Generation of spatial detailed data on stock demography from operating models and application of stock assessment methods
 - b. Application of alternate management strategies that integrate information on population structure and spatial impact of fishing fleet
 - c. Projection for a fixed time period of the operating model given the advice from management strategies on stock status indicators and related catch
- iii. Evaluation of performance of alternative management strategies against performance criteria (including biological, economic, and social objectives) to determine the best alternative taking into account biological, economic, and social trade-offs

Within an adaptive fishery framework, fishery management should take a proactive role in adopting the best available stock unit identification information and apply this knowledge to determine potential mismatches between biological and management structure (Kerr et al., 2017). This because the degree of spatial isolation or overlap between populations and harvest stocks are important to identify the appropriate strategy, to assure the renewability of exploited population and reduce uncertainty in assessment and management.

Discussion and conclusion

The simulation of stocks dynamics and fleet economic performance in three case studies by BEMTOOL and SMART, using the stock assessment results of the new stock configurations, showed how the available knowledge on spatial dynamics of population components living in a study area and fishing effort can be integrated in a bio-economic model to develop an adaptive spatial fishery management. The integration of the new stock configurations in the models allowed to take into account the available knowledge on the distribution of fishing effort and of the best biological information on stock configuration.

The updated stock status and main recent patterns of stock indicators, assessed by the a4a and reported in Deliverable 3.3 “Report on the stock assessments with the new stock configurations for the 6 target species of the study area” are synthetized in table 16. These assessments, which not include at all spatial aspects, clearly showed a current state of overfishing assessed for three of the four evaluated stocks.

Table 16 Stock status and main recent patterns of stock indicators estimated by a4a.

Stock	Stock status	Total biomass (t)	Recruitment (millions)	SSB (t)	Catch (t)	F bar
DPS in 17-20	Overfishing	18465 in 2019; decreasing since 2017	6526 in 2019; Decreasing since 2016	3516 in 2019; Decreasing since 2017	5672 in 2019; Decreasing since 2018	1.60 in 2019; Increasing since 2015
ARS in 5, 8-12	Overfishing	990 in 2019; decreasing since 2016	275 in 2019; Decreasing since 2016	666 in 2019; Decreasing since 2017	592 in 2019; Decreasing since 2015	0.81 in 2019; Increasing since 2015
ARA in 1, 5, 6, 9-11	Overfishing	3293 in 2019; decreasing since 2016	241 in 2019; Decreasing since 2016	2214 in 2019; Decreasing since 2017	1298 in 2019; Decreasing since 2017	0.95 in 2019; Increasing since 2016
ARS in 15, 16, 18-20	Sustainable exploited	9335 in 2019; decreasing since 2015	2201 in 2019; Increasing since 2017	3177 in 2019; Decreasing since 2017	1593 in 2019; Decreasing since 2017	0.41 in 2019; Decreasing since 2017

By using the same data set used for a4a assessments, integrated with available information on existing Fishery Restricted Areas and selectivity studies, the analyses carried out and reported in this deliverable highlighted the feasibility to explore a wide set of management scenarios, including spatial based management measures (e.g. regulation of fishing effort in space/time), and their potential effects on both stocks and fleets. A synthetic view of the main result obtained by the scenarios carried out with BEMTOOL and SMART models are reported in table 17.

*Table 17 Main results of explored scenarios according to the BEMTOOL and SMART simulation models and different scenarios as % variation from the status quo (S0; 2020) to 2026. F= fishing mortality, SSB = spawning stock biomass in tons; Y= yield in tons; i= income in millions of euro * referred to the aggregated income of both ARA and ARS. The best results in terms of gain in SSB and increase or maintaining the current yield is marked in bold.*

Stock	Model	scenario	F	SSB	Y	I
ARS 1-12	BEMTOOL	S0	0.82	477	469	189,3*
		S1	-11%	15%	1%	11%
		S2	-28%	36%	-3%	26%
		S3	-28%	37%	-3%	24%
		S4	-35%	63%	-5%	27%
ARS 1-12	SMART	S0	nc	6434	6278	nc
		S1	nc	+1%	-11%	nc
		S2	nc	+14%	-9%	nc
		S3	nc	+6%	-23%	nc

		S4	nc	+20%	-16%	nc
ARA 1-12	BEMTOOL	S0	0.76	1182	787	189,3*
		S1	-10%	29%	18%	11%
		S2	-28%	85%	45%	26%
		S3	-28%	81%	40%	24%
		S4	-40%	144%	65%	27%
ARA 1-12	SMART	S0	nc	2606	1737	nc
		S1	nc	-26%	-7%	nc
		S2	nc	+12%	-9%	nc
		S3	nc	+24%	-12%	nc
		S4	nc	+42%	-16%	nc
DPS 17-20	BEMTOOL	S0	1.4	3462	4950	214, 8
		S1	-69%	93%	-23%	-20%
		S2	-16%	40%	9%	3%
ARS 15-16, 18-20	SMART	S0	nc	2562	2433	nc
		S1	nc	+1%	-5%	nc
		S2	nc	+1%	-6%	nc
		S3	nc	+1%	-8%	nc
		S4	nc	+58%	-3%	nc

When comparing the results of BEMTOOL and SMART simulations, it is worth noting that, although the two models assessed the same fisheries under the same scenarios, they are quite different in terms of results. Overall, SMART neural network lightly underestimated the observed yield in the case of ARA 1-12 and ARS 15, 16, 18-20, while overestimated that of ARS 1-12. This could explain the more pessimistic prevision of response to the adoption to the management scenarios in terms of yield.

These differences are mainly due to differences in modelled processes and assumptions in scenario modelling. However, while they are not directly comparable, it is useful to highlighted the both models indicated as the best management strategy, both in terms of gain in SSB and improve or light decrease of the current yield, the scenario characterized by reducing of the 10% the current fishing effort coupled with an improvement of trawl net selectivity and protection of EFHs.

These results are in line with literature drawing the importance to adopt management strategies targeted not only to decrease the current fishing effort but also to improve the selectivity and, more in general, the exploitation patter of demersal stocks to increase the sustainability of trawl fisheries in the Mediterranean (Colloca et al., 2013; Russo et al., 2014; Gorelli et al., 2017; Brčić et al, 2018; Khoukh & Maynou, 2018; Russo et la., 2019).

Concerning the steps to develop adaptive spatial management, it is clear that a shift from a reactive to a proactive approach from managers should be pursued (Kerr et al., 2017). This proactive management should start from adjusting stock boundaries to the scientific information, and include a more in deep understanding of recruitment processes and connectivity between EFHs. An adaptive management aimed at protecting the spatial structure of the stock should consider as main tool the identification and protection of nurseries together spawning areas, in order to mitigate both growth and recruitment overfishing. Since most of marine species show complex spatial structure, not all aspects of population structure are necessary to be integrated into assessment and management. Although it is difficult to delineate a generalized and prescriptive roadmap for adaptive fishery management, the critical element for management should be to remain flexible to protect the spatial structure of the exploited stock when new information becomes available (Goethel et al., 2015; Kerr et al., 2017). Both BEMTOOL and SMART are able to integrate an increasing amount of information as data on the space-time dynamics of the stocks and fleets that exploit them increases, improving the model ability to perform MSE within an adaptive fishery framework. However, Walters (2007), when dealing with main obstacles to the success of adaptive fishery management mentioned three main institutional problems which are still current: i) lack of management resources for the expanded monitoring needed to carry out large-scale evaluations; ii) unwillingness by decision makers to admit and embrace uncertainty in making policy choices; and iii) lack of leadership in the form of individuals willing to do all the hard work needed to plan and implement new

and complex management programs.

Deviation from the Workplan

The present deliverable has been delayed due to the late availability of WP3 outcomes, due to COVID-19 difficulties that affected the MEDUNITS activities.

The implicit-spatially SMART model was implemented in a simplified approach due to the difficulties to obtain: i) seasonal georeferenced data of CPUE by age/length class; ii) spatially resolved information on fishing effort for non EU countries exploiting shared stocks; iii) spatial distribution of EFH and sensitive habitat not covered by the Mediseh program; iv) information on fish price by fleets and cost values by vessel for European and non-EU countries.

Remedial actions

Most of the activities have been carried out remotely soon after the availability of WP3 outputs by target species.

References

- Accadia P., Bitetto L, Facchini M.T., Gambino M., Kavadas S., Lembo G., Maynou F., Melià P., Maravelias C., Rossetto M., Sartor P., Sbrana M., Spedicato M.T.(2013). BEMTOOL Deliverable DIO: BEMTOOL FINAL REPORT: 46 pp.
- Brčić J, Herrmann B, Sala A (2018). Predictive models for codend size selectivity for four commercially important species in the Mediterranean bottom trawl fishery in spring and summer: Effects of codend type and catch size. *PLoS ONE* 13(10): e0206044. <https://doi.org/10.1371/journal.pone.0206044>
- Carson, S., Shackell, N., Mills Flemming, J. (2017). Local overfishing may be avoided by examining parameters of a spatio-temporal model. *PLoS one*, 12(9), e0184427.
- D'Andrea L, Parisi A, Fiorentino F, Garofalo G, Gristina M, Russo T, Cataudella S. (2020). smartR: a R package for spatial modelling of fisheries and simulation of effort management. *Methods in Ecology and Evolution* 00:1-10. [10.1111/2041-210X.13394](https://doi.org/10.1111/2041-210X.13394)
- FAO, GFCM (2020). The state of mediterranean and black sea fisheries. FAO, Rome (2020).
- Fiorentino F., Calleja D., Colloca F., Perez M., Prato G., Russo T., Sabatella R., Scarcella G., Solidoro C., Vrgoč N. (2020) MANTIS: Marine protected Areas Network Towards Sustainable fisheries in the Central Mediterranean. Final report. November 2019. EUROPEAN COMMISSION. Directorate-General for Maritime Affairs and Fisheries. Directorate D — Fisheries Policy Mediterranean and Black Sea. Unit D.1 — Fisheries Management Mediterranean and Black Sea. ISBN 978-92-76-17170-6. ISBN 978-92-76-17170. doi: 10.2771/33931 . 333 p.
- Fiorentino F., Vitale S. (2021). How Can We Reduce the Overexploitation of the Mediterranean Resources? *Front. Mar. Sci.* 8:674633. doi: 10.3389/fmars.2021.674633
- Goethel, D. R., Legault, C. M., and Cadrin, S. X. (2015). Demonstration of a spatially-explicit, tag-integrated stock assessment model with application to three interconnected stocks of yellowtail flounder off of New England. *ICES Journal of Marine Science*, 72: 164–177.
- Goethel, D. R., Kerr, L. A., and Cadrin, S. X., (2016). Incorporating spatial population structure into the assessment-management interface of marine resources. In *Management Science in Fisheries: An Introduction to Simulation-based Methods*, pp. 319–347. Ed. by C. T. T. Edwards, and D. J. Dankel. Routledge, New York. 460 pp.

- GFCM. (2018). Report of the Workshop on the assessment of management measures (WKMSE). Zagreb, Croatia 9-11 April. 165 pp.
- Gorelli, G., Bahamón, N., & Sardà, F. (2017). Improving codend selectivity in the fishery of the deep-sea red shrimp *Aristeus antennatus* in the northwestern Mediterranean Sea.
- Guidetti P., Di Franco A., Calò A. (2020) SafeNet: Sustainable Fisheries in EU Mediterranean waters through network of MPAs. Final report. November 2019. EUROPEAN COMMISSION. Directorate-General for Maritime Affairs and Fisheries. Directorate D — Fisheries Policy Mediterranean and Black Sea. Unit D.1 — Fisheries Management Mediterranean and Black Sea. ISBN 978-92-76-17166-9. doi: 10.2771/6478. 59 p.
- Holling, C. S. (1978). Adaptive Environmental Assessment and Management. Chichester, UK: John Wiley and Sons.
- Kerr, L. A., and Goethel, D. R. (2014). Simulation modeling as a tool for synthesis of stock identification information. In *Stock Identification Methods: Applications in Fishery Science*, 2nd edn, pp. 501–534. Ed. by S. Cadrin, L. Kerr, and S. Mariani. Elsevier Academic Press, Burlington. 566 pp.
- Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L., Worsøe Dickey-Collas, M., Goethel, D. R., Hatfield, E. M.C., Kritzer, J. P., and Nash, R.D.M. (2017). Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. – *ICES Journal of Marine Science*, 74: 1708–1722.
- Khoukh, M., Maynou, F. (2018). Spatial management of the European hake *Merluccius merluccius* fishery in the Catalan Mediterranean: Simulation of management alternatives with the InVEST model. *Sci. Mar*, 82, 175-188.
- Kleisner, K. M., Fogarty, M. J., McGee, S., Barnett, A., Fratantoni, P., Greene, J., Hare, J. A. et al. (2016). The effects of sub-regional climate velocity on the distribution and spatial extent of marine species assemblages. *PLoS One*, 11: e0149220.
- Kritzer, J. P., DeLucia, M., Greene, E., Shumway, C., Topolski, M. F., Thomas-Blate, J., Chiarella, L. A. et al. (2016). The importance of benthic habitats for coastal fisheries. *BioScience*, 66: 274–284.
- Lembo, G., Abella, A., Fiorentino, F., Martino, S., and Spedicato, M.-T. (2009). ALADYM: an age and length-based single species simulator for exploring alternative management strategies. *Aquatic Living Resources*, 22: 233–241.
- McGilliard, C. R., Punt, A. E., Methot Jr, R. D., & Hilborn, R. (2015). Accounting for marine reserves using spatial stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(2), 262-280.
- Minto, C., Flemming, J.M., Britten, G.L., and Worm, B. (2014). Productivity dynamics of atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences* 71 (2): 203-216.
- Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. (2013). Marine taxa track local climate velocities. *Science* 341: 1239–1242
- Rossetto M., Bitetto L, Spedicato M. T., Lembo G., Gambino M., Accadia P., Melià P. (2014). Multi-criteria decision-making for fisheries management: A case study of Mediterranean demersal fisheries. *Marine Policy*, 53, 83-93.
- Russo T, Parisi A, Garofalo G, Gristina M, Cataudella S, Fiorentino F. (2014) SMART: A Spatially Explicit Bio-Economic Model for Assessing and Managing Demersal Fisheries, with an Application to Italian Trawlers in the Strait of Sicily. *PLoS ONE* 9(1): e86222. doi:10.1371/journal.pone.0086222
- Russo, T., Bitetto, I., Carbonara, P., Cariucci, R., D'Andrea, L., Facchini, M.T., Lembo, G., Maiorano, P., Sion, L., Spedicato, M.T., Tursi, A. and Cataudella, S. (2017). A Holistic Approach to Fishery Management: Evidence and Insights from a Central Mediterranean Case Study (Western Ionian Sea). *Front.Mar.Sci.*, 4:193,

- Russo T, Morello EB, Parisi A, Scarcella G, Angelini S, Labanchi L, Martinelli M, D'Andrea L, Santojanni A, Arneri E, Cataudella S (2018). A model combining landings and VMS data to estimate landings by fishing ground and harbor. *Fisheries Research* 199, 218-230.
- Russo T, D'Andrea L, Franceschini S, Accadia P, Cucco A, Garofalo G, Gristina M, Parisi A, Quattrocchi G, Sabatella RF, Sinerchia M (2019). Simulating the effects of alternative management measures of trawl fisheries in the Central Mediterranean Sea: application of a multi-species bio-economic modelling approach. *Frontiers in Marine Science*, 6, p.542.
- Shackell, N. L., Ricard, D., and Stortini, C. 2014. Thermal habitat index of many northwest Atlantic temperate species stays neutral under warming projected for 2030 but changes radically by 2060. *PLoS One*, 9: e90662.
- Spedicato et al. (2016). European Commission, Study on the evaluation of specific management scenarios for the preparation of multiannual management plans in the Mediterranean and the Black Sea, EASME/EMFF/2014/1.3.2.7/SI2.703 193, June 2016.
- Spedicato M.T. (coord), I. Bitetto, R. Cariucci, S. Cataudella, M.T. Facchini, F. Fiorentino, G. Lembo, P. Maiorano, A. Mariani, C. Piccinetti, T. Russo, A. Santojanni, M. Scardi. (2017). - *Basi scientifiche e strumenti a supporto dei Piani di Gestione delle risorse della pesca nell'ambito della Politica Comune della Pesca e delle politiche ambientali ed economiche" - (Rete3). CoNISMA, Roma. 129 pp.*
- Spedicato, M. T. , Bitetto, I. , Lembo, G. , Sartor, P. , Accadia, P. (2018). Research for PECH Committee - Discard ban, landing obligation and MSY in the Western Mediterranean Sea - the Italian case.
- STECF (2018) Scientific, Technical and Economic Committee for Fisheries (STECF) – Fishing effort regime for demersal fisheries in the western Mediterranean Sea – Part II (STECF-18-13). Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-79396-7, doi:10.2760/509604, JRC114702.
- STECF. (2019a). Scientific, Technical and Economic Committee for Fisheries (STECF) – Methods for developing fishing effort regimes for demersal fisheries in Western Mediterranean-Part III (STECF19-01). Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-08330-6, doi:10.2760/249536, JRC116968.
- STECF. (2019b). Scientific, Technical and Economic Committee for Fisheries (STECF) – Evaluation of fishing effort regime in the Western Mediterranean – part IV (STECF-19-14). Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-14097-9, doi:10.2760/295779, JRC119061.
- STECF (2020) – Scientific, Technical and Economic Committee for Fisheries (STECF) Evaluation of fishing effort regime in the Western Mediterranean – part V (STEC-20-13). EUR 28359 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-27701- 9, doi:10.2760/143313, JRC122924.
- Walters, C. J. (1986). *Adaptive Management of Renewable Resources*. Macmillan, New York.
- Walters, C. J. (2007). Is Adaptive Management Helping to Solve Fisheries Problems? *Ambio* 36: 304-307.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at:

https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from:

<https://publications.europa.eu/en/publications>

Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.



doi: 10.2926/909535
ISBN 978-92-95225-33-6