# Improving the cost-efficiency of fisheries research surveys and fish stocks assessment using next-generation genetic sequencing methods - FishGenome

## Final Report

Written by Fran Saborido-Rey (IIM-CSIC), Rosa Fernández-Otero (CETMAR), Laura Casas (IIM-CSIC), Rebeca Rodríguez-Mendoza (IIM-CSIC), María Pérez (CETMAR), Francesc Piferrer (ICM-CSIC), Dafni Anastasiadi (ICM-CSIC), Beatriz Guijarro (IEO-CSIC), Sergio Ramírez-Amaro (IEO-CSIC), Bàrbara Terrasa (UiB), Antonia Picornell (UiB), Cori Ramon (UiB), Erik Eschbach (Thünen Institute), Yassine Kasmi (Thünen Institute), Reinhold Hanel (Thünen Institute), Christoph Stransky (Thünen Institute)

March 2023

This report should be cited as:

*Saborido-Rey, F., Fernández-Otero, R., Casas, L., Rodríguez-Mendoza, R., Pérez, M., Piferrer, F., Anastasiadi, D., Guijarro, B., Ramírez-Amaro, S., Terrasa, B., Picornell, A., Ramon, C., Eschbach, E., Kasmi, Y., Hanel, R., Stransky, C. Improving the cost-efficiency of fisheries research surveys and fish stocks assessment using next-generation genetic sequencing methods – FishGenome. Final Report. Publications Office of the European Union, 2023, doi: 10.2926/803359.*

Spanish National Research Council (CSIC)

Institute of Marine Research (IIM-CSIC), Vigo, Spain

Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain

Spanish Institute of Oceanography (IEO-CSIC), Palma, Spain

Centro Tecnológico del Mar (CETMAR), Vigo Spain

Thünen Institute

Institute of Sea fisheries, Bremerhaven, Germany

Institute of Fisheries Ecology, Bremerhaven, Germany

University of the Balearic Islands, Palma, Spain

*"The information and views set out in this study are those of the authors and do not necessarily reflect the official opinion of CINEA or of the Commission. Neither CINEA nor the Commission can guarantee the accuracy of the data included in this study. Neither CINEA nor the Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein."*

**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*

# Improving the cost-efficiency of fisheries research surveys and fish stocks assessment using next-generation genetic sequencing methods - FishGenome

**Final Report**

EASME/EMFF/2017/1.3.2.10/ SI2.790889

European Climate, Infrastructure and Environment Executive Agency

***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, 2023

| PDF | ISBN 978-92-95231-54-2 | doi 10.2926/803359 | Catalogue number HZ-04-23-476-EN-N |
|---|---|---|---|

© European Union, 2023
Reproduction is authorised provided that the source is acknowledged.

## ABSTRACT

In response to the EC call for tenders EASME/EMFF/2018/015, the work done in the FishGenome service contract enabled the evaluation of the potential of three High Throughput Sequencing genetic tools: Close Kin Mark Recapture, epigenetic Age Determination, and environmental DNA. Contractors rigorously explored their potential for use in the assessment of fish stocks and in the production of the scientific advice needed for a sustainable management of EU fisheries. The service approach combined an exhaustive update and critical review of the state of the art with field tests of an exploratory nature for three species of major commercial interest (cod, hake and ballan wrasse) in the North Sea, in the northern waters of Iberian Peninsula, and in the Mediterranean. The project delivered, among other outputs, detailed application protocols and guidelines and outlined possible implementation scenarios. For the latter, FishGenome proposes a detailed roadmap for implementation with specific objectives, strategic pillars, and intervention initiatives with a total of 45 specific actions of a diverse and supplementary nature. Those aimed at involving all interested parties in a coordinated approach for a progressive adoption of these new technologies, envisioning their implementation at full scale in the long term.

# TABLE OF CONTENTS

TABLE OF CONTENTS .......... II
LIST OF TABLES .......... III
LIST OF FIGURES .......... IV
LIST OF ABBREVIATIONS .......... VII
EXECUTIVE SUMMARY .......... 1
RÉSUMÉ EXÉCUTIF .......... 7
DELIVERABLES PRODUCED .......... 14
INTRODUCTION .......... 15
Project Structure .......... 18
FishGenome Consortium .......... 19
1. STATE OF THE ART AND CRITICAL ASSESSMENT OF THE METHODS .......... 21
1.1. Main achievements .......... 21
1.2. Work carried out .......... 22
*1.2.1. Fishery independent data collection procedures .......... 22*
*1.2.2. State-of-the-art and critical review of the genomic methods .......... 26*
*1.2.3. Critical assessment of the methods .......... 44*
*1.2.4. Identification of barriers and risks .......... 49*
1.3. Lessons learnt .......... 52
2. PILOT STUDIES .......... 53
2.1. Main achievements .......... 53
2.2. Work carried out .......... 54
*2.2.1. Design of the pilot studies .......... 54*
*2.2.2. Fieldwork .......... 58*
*2.2.3. CKMR .......... 59*
*2.2.4. RAD-Seq for connectivity and stock boundaries .......... 65*
*2.2.5. RAD-Seq for stock substructure .......... 67*
*2.2.6. RAD-Seq for sex assignment .......... 70*
*2.2.7. Epigenetic Age determination .......... 73*
*2.2.8. Environmental DNA .......... 74*
2.3. Lessons learnt .......... 81
*2.3.1. CKMR .......... 81*
*2.3.2. RAD-Seq for connectivity and stock boundaries .......... 82*
*2.3.3. RAD-Seq for stock substructure .......... 83*
*2.3.4. RAD-Seq for sex assignment .......... 84*
*2.3.5. Epigenetic age determination .......... 85*
*2.3.6. Environmental DNA .......... 86*
3. ANALYSIS AND STRATEGIC PLANNING .......... 88
3.1. Main achievements .......... 88
3.2. Work carried out .......... 89
*3.2.1. Cost-benefit analysis .......... 89*
*3.2.2. SWOT .......... 99*
*3.2.3. Roadmap for the implementation of genomic methods .......... 108*
*3.2.4. Long term prospects .......... 118*
3.3. Lessons learnt .......... 127
4. REFERENCES .......... 130
5. GLOSSARY .......... 137

## LIST OF TABLES

Table 1. Epigenetic clocks for age prediction in vertebrates ............................................... 33

Table 2. Information provided by the traditional research surveys, compared with the potential information provided by HTS methods (CKMR, eDNA and epigenetic age determination). 47

Table 3. Cases studies addressed as pilot studies within FishGenome, by ecoregion, species and genomic tool (the three main tools are shown in bold) ........................................ 55

Table 4. Summary of samples collected for the pilot studies. Objective indicates the use of the samples: CKMR, epigenetic analysis (DNAm), environmental DNA (eDNA), connectivity studies (CONN), stock substructure analysis (SUB) and sex assignment (SEX). N indicates the number of fishes sampled for cod, hake and wrasse and number of stations sampled for water and sediment. Length is given in cm and mean and standard deviation are indicated in brackets. ........................................................................................ 58

Table 5. Summary of pairwise comparisons between plain–spotted phenotypes of Labrus bergylta across two sites in the Galician shelf. Ma indicates Malpica, Vi indicates Vigo. Pheno indicates phenotype –plain or spotted. Outflank and Bayescan indicate the number of outliers found in the comparison. Differentiation between both phenotypes is indicated by $F_{ST}$ and Dist denotes distance (km) between the sites. .......................................... 70

Table 6. Number and characteristics of the private alleles found in the North Sea hake population (number of individuals, variance, standard error, mean frequency of the most frequent allele at each locus in this population) ....................................................... 72

Table 7. Species (where NS indicates North Sea, BI corresponds to Balearic Islands and GS denotes Galician shelf), N samples (number of samples analysed), N kinship (number of kinship pairs found), N required 50 (number of samples required to achieve 50 POPs), N required 100 (number of samples required to achieve 100 HSPs). For cod NS, hake NS and GS, no POPs were found so we used a rough approximation to calculate the N required. We estimated that the frequency of POPs is approx. half the frequency of HSPs (*indicates that HSPs not provide useful information for CKMR) .................................. 82

Table 8. Comparison of the performance of the cod epigenetic clock with other piscine clocks. Abbreviations: yr, yeas; wk, weeks; r, Pearson correlation value................................. 86

Table 9. Atlantic Ocean calculation of the cost-efficiency of independent demersal fishery trawl surveys for stock assessment purposes. Summary table. S.A. Number refers to the number of stocks assessed in the specified subregion. Total cost represents the total cost (in euros) of deploying the fishing survey (taking as reference year 2013). S.A. Cost reports the average cost of a single stock assessment in the corresponding FAO region. Survey Cost / Day represents the survey cost per day. Finally, S.A. Cost / Day reports the average stock assessment cost per fishery-independent survey and per day. ........... 93

Table 10. Mediterranean Sea calculation on the cost-efficiency of independent demersal fishery trawl surveys for stock assessment purposes. Summary table. S.A. Number refers to the number of stocks assessed in the specified subregion. Total cost represents the total cost (in euros) of deploying the fishing survey (taking as reference year 2013). S.A. Cost reports the average cost of a single stock assessment in the corresponding marine region. Survey Cost / Day represents the survey cost per day. Finally, S.A. Cost / Day reports the average stock assessment cost per fishery-independent survey day ........................ 93

Table 11. Cost components of the MEDITS GSA5 and IBTS Q3 expressed in euro.................. 94

Table 12.- List of proposed actions related to Objective 1, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets]......................................................................................................... 111

Table 13.- List of proposed actions related to Objective 2, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets]......................................................................................................... 113

Table 14.- List of proposed actions related to Objective 3, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets]......................................................................................................... 114

Table 15.- List of proposed actions related to Objective 4, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets]......................................................................................................... 115

Table 16.- List of proposed actions related to Objective 5, with indication of corresponding strategic pillar and timeframe [years in brackets]. .................................................. 116

## LIST OF FIGURES

Figure 1. Figure illustrating the case studies in FishGenome. The project covered several stocks: in the North Sea, cod and hake (Northern stock) as well as environmental samples, consisting of water and sediment samples and represented by a drop of water in the figure. In the Galician shelf, hake (Southern stock) and ballan wrasse. In the Balearic Mediterranean, hake (Mediterranean stock, GFCM Geographical Sub Area 5) and environmental samples. .............16

Figure 2. PERT chart showing the connections among Work packages and tasks..........................19

Figure 3. Surveys at a glance: Surveys in Europe use different methods to obtain data for assessing the condition of exploited fish stocks and other marine resources. Numbers and types of surveys in this figure. .......................................................................................23

Figure 4. Map showing the 14 mandatory bottom trawl surveys that were reviewed. ....................24

Figure 5. Overview of survey effort (in days), for mandatory bottom trawl surveys, in the different regions. For each survey, the effort by all participating countries was included. Data from 2016-2018. No data were available for the Black Sea Beam trawl survey (BS-BTS). ...............25

Figure 6. Number of participating countries in each of the mandatory bottom trawl and beam trawl surveys. ..................................................................................................................25

Figure 7. Description of the facilitation team appointed by the FishGenome Consortium for the 1st Virtual workshop carried out within the project framework on the 28th May 2020. .................27

Figure 8. Illustration of the CKMR principle; adults (big fish) and juveniles (small fish) are sampled (dark blue) from the total population (dark and light blue). Each juvenile "tags" two fish: each of its parents (solid and dashed lines) in the adult population; but only sampled fish provide us kinship information —POPs (solid lines). The absolute abundance of adults (10) can be estimated from the number of sampled adults and juveniles (5 and 6 respectively) and the number of POPs found (6) (figure redrawn from Bravington et al., 2016a). ..........................28

Figure 9. RAD-Seq derivatives published in literature (figure adapted from Campbell et al., 2018). .30

Figure 10. Epigenetic clocks are based on finding loci the DNA methylation of which increases (A) or decreases (B) with age. A combination of several of them allows clock building (C). Intrinsic or extrinsic factors may accelerate (red line) or slow down (green line) epigenetic age. ..........33

Figure 11. Bioinformatics pipeline for epigenetic age determination. A) Workflow for bioinformatics analysis of bisulfite sequencing data. B) Equation for penalized regression analysis where λ2=0; λ1= λ ~ Ridge Regression and λ1=0; λ2= ~ LASSO. The equation on top shows the cost function which penalized regressions try to minimize. The values in front of the CpGs (β1, β2... βN) show the amount of change in age as a response to change of methylation. C) Workflow for applying a machine learning algorithm. .......................................................36

Figure 12. Overview and general workflow for environmental DNA studies. .................................39

Figure 13. Flowchart describing bioinformatics steps for the analyses of eDNA metabarcoding. .......42

Figure 14. Schematic representation of the FishGenome biobanking process. ...............................57

Figure 15. Schematic representation of the processing of the samples for the biobank. .................57

Figure 16. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, square –male, triangle – immature) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). Two kinship pairs were found among the 235 specimens of cod analysed, one FSP and one HSP. ......................................................................................61

Figure 17. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, triangle –male) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), red – parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). Two kinship pairs (HSPs) were found among the 94 specimens of hake analysed. .....62

Figure 18. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, triangle –male) while their

length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), red – parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). Three kinship pairs were found among the 281 specimens of hake analysed, one POP and two HSPs. ....63

Figure 19. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, triangle –male) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), red – parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). One kinship pair, a HSP, was found among the 142 specimens of hake analysed. ......64

Figure 20. Population structure of hake populations from the North Sea, the Balearic Islands and the Galician shelf detected by principal component analysis along axes PC1 and PC2 that are used to infer the number of clusters of genetically related individuals, based on 12957 SNP markers. Each dot represents an individual and three different colours are used to differentiate the populations: red, green and blue represent the Galician shelf, Balearic Islands and North Sea populations, respectively. Ellipses illustrate the distribution of individuals within groups. ..66

Figure 21. Plot of posterior probabilities of assigning individual membership to their original populations obtained by Bayesian clustering. Each vertical line represents an individual and the colours refer to the different clusters. ....66

Figure 22. Population structure of North Sea cod stock detected by principal component analysis along axes PC1 and PC2 that are used to infer the number of clusters of genetically related individuals, based on 25,571 SNP markers. Each dot represents an individual. No distinctive groups are detected indicating a unique genetic stock. ....68

Figure 23. Bayescan plot of 25,571 SNPs according to FST and log10 (q-value) in genome scan analysis of 235 individuals of cod from the North Sea. Each dot represents a SNP locus and no outliers were identified. ....68

Figure 24. Bayescan plots of 11,608 (left), 10,084 (center) and 10,360 (right) SNPs according to FST and log10 (q-value) in genome scan analysis of 94, 142 and 281 individuals of hake from the North Sea, the Galician shelf and the Balearic Islands, respectively. Each dot represents a SNP loci and no outliers were identified in any of the three stocks. ....68

Figure 25. Population structure of ballan wrasse detected by principal component analysis along axes PC1 and PC2 that are used to infer the number of clusters of genetically related individuals, based on 35,096 SNP markers. Each dot represents an individual. Two distinctive groups are detected indicating two genetic stocks. Samples were collected in two locations, Vigo (indicated Vi) and Malpica (indicated Ma) that are separated by ~120 km. The two differential components correspond to two differential color phenotypes (rigth panel, plain (above) and spotted (below)). ....69

Figure 26. Individual ancestry coefficients of 60 individuals of ballan wrasse from the Galician shelf for K = 2 are shown in the figure. Each bar represents an individual. Vi indicates Vigo, Ma indiates Malpica, P denotes plain color morphotype and S spotted color morphotype. ....70

Figure 27. Heat map of RADSex markers in 33 males and females of the North Sea hake population showing a higher coverage. Positions with significant association with sex are indicated with a red box on the lower right corner and upper left corner for male and female respectively. The intensity of the blue color corresponds to the number of markers. ....72

Figure 28. The Atlantic cod epigenetic clock for age estimation. Number of CpGs retained: 26. Penalized regression: LASSO. Accuracy (Pearson correlation): r = 0.979, P = 2.2. e-16. Precision: leave-one-out-cross-validation (LOOCV). Maximum absolute error (MAE) = 0,691 years ( = 252 days; 8,2 months). ....74

Figure 29. Overview of the strategy followed to develop a highly specific qPCR test for Atlantic cod (*Gadus morhua*). ....75

Figure 30. The OTU assignment for water samples presenting the 35 most common species. The control is based on ultra-pure water. ....77

Figure 31. Venn diagram showing the number of common and distinguishable species in trawl catches versus eDNA from sediment and water samples. 8 stations are presented. Sharks were principally not detected. Numbers for Station 1 are to be explained as: 9 species were only detected via eDNA analysis of water sample, 5 detected only by trawling and 4 species were

detected by metabarcoding from water and also trawling, 1 species each was detected in trawling and in both eDNA metabarcoding techniques. 78

Figure 32. Panel a) Standard curve of threshold cycle number (Ct values) plotted against the log concentration (copy number). Dark green dots represent 9 replicates for each dilution. In the black square (Figure -a-), the results from different qPCR assays are shown. Panel b) Detail of a) including standard (dark green dots) - the last dilution represented coincides with LoD value - non-target samples (red dots), eDNA water samples in triplicates (blue dots). Dotted lines indicate LoB, LoD and LoQ values. In this assay the standard equation is Ct=-3,57*(concentration)+39,73. 79

Figure 33. eDNA sampling results from water (a) and sediment (b) in the North Sea (mission number WH428). In the WH428 cruise (in 2019), we sampled water and sediment from 10 stations in the North Sea, with three different positions (S: start, M: middle and E: end). The sample name is readable as: Cruise number (WH428), followed by the box titled by a letter (e.g., "A1a", where "A" is the box, "1" is the position during trawling, and "a" for the Niskin Bottle replicates. 80

Figure 34. Comparison between output of Catch results and output of GPR model for eDNA copies. (R = 96%, Error MAE = 0,002). The plot shows the correlation between predicted data by the model (Predicted response) and the real data collected from trawl (True response). 81

Figure 35. Relationship between vessel cost per day and no. of days at sea, and the total cost per survey. 91

Figure 36. Efficiency. Cost per stock assessed by country/survey. 2013. Codes in the Y-axis indicates in anonymous manner the 41 surveys conducted by EU countries. 92

Figure 37. Aggregated cost per stage for the application of HTS methodologies to hake and cod. 95

Figure 38. Disaggregation of the costs of samples processing at the laboratory. 95

Figure 39. Aggregated cost per category for the application of HTS methodologies to hake and cod 96

Figure 40. Aggregated cost per method for the application of HTS methodologies to hake and cod 96

Figure 41. Cost per assessed Stock vs. vessel cost per day (by survey) 98

Figure 42. Average effort per stock reported vs. average vessel cost per day (by survey) 98

Figure 43. The FishGenome Roadmap at a glance 117

## LIST OF ABBREVIATIONS

| Term | Description |
|---|---|
| Bis-RAD-Seq | bisulfite-converted restriction site associated sequencing |
| BITS | Baltic International Trawl Survey |
| CETMAR | Centro Tecnológico del Mar (Technological Center of the Sea) |
| CINEA | European Climate, Infrastructure and Environment Executive Agency |
| CKMR | Close Kin Mark Recapture |
| COI | Cytochrome Oxidase I |
| CPUE | Catch Per Unit Effort |
| CSIC | Consejo Superior de Investigaciones Científicas (Spanish National Research Council) |
| CytB | Cytochrome B |
| DCF | Data Collection Framework |
| DGMARE | Directorate-General for Maritime Affairs and Fisheries |
| DNAm | Epigenetic Age Determination by DNA methylation |
| EC | European Commission |
| eDNA | Environmental DNA |
| EEP | External Experts' Panel |
| EU-MAP | European Union Multi Annual Programme |
| GSA 5 | Geographical subarea Balearic Islands |
| HTS | High-Throughput Sequencing |
| IBTS | International Bottom Trawl Survey |
| IEO | Instituto Español de Oceanografía (Spanish Institute of Oceanography) |
| MEDITS | International Bottom Trawl Survey in the Mediterranean |
| NS-IBTS | North Sea International Bottom Trawl Survey |
| qPCR | quantitative PCR |
| RAD-Seq | Restriction site Associated DNA Sequencing |
| SNP | Single Nucleotide Polymorphism |
| SoA | State of the Art |
| STECF | Scientific, Technical and Economic Committee for Fisheries |
| Thuenen | Thünen Institute |
| UiB | Universitat de les illes Balears (University of the Balearic Islands) |
| WP | Work Package |

## EXECUTIVE SUMMARY

The Tender contract "FishGenome: Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods" was developed to assess the ability and readiness of several new emerging genomic technologies to enhance fisheries assessments. These assessments are needed to monitor the status of fish stocks and ensure fishing practices that exploit them at sustainable levels. They rely on the collection of data from multiple sources, that include fishery- dependent data -catches, landings, biological information - as well fishery-independent data obtained from research surveys. These surveys provide valuable and systematised information regarding exploited fish populations, marine biodiversity and their environment, and are essential piece in stock assessment and scientific advice.

However, research surveys are confronted with a high economic cost coupled with complex logistics and a long time is required for treating and analysing the collected data. In addition, regardless of the sampling method, conventional methods to determine biological parameters have some limitations, such as the inability to determine the sex structure of the younger juveniles or to provide information on stocks connectivity, both of which are crucial for stock management. Survey design and technology are constantly progressing to cope with these limitations and to reduce uncertainty in stock assessment and scientific advice on the status of the harvested stocks and ecosystems. Thus, recent progress in the field of genomics is expected to improve efficiency and help to mitigate some of the shortcomings of traditional methodologies, such as high costs and complex logistics.

The fast growth in genomic techniques over the last decades provides today the potential to resolve some of these challenges and complement traditional methods to assist fisheries management in the mid and long term. The term "genetics" refers to the study of a group of genes or other regions of the genome, while "genomic" is used for studies involving the whole genome aided by high-throughput genetic sequencing methods. The genome of any living-being encodes most of its potential characteristics and in combination with the environment determines the appearance, behaviour, and physiology of the organisms. In the context of fisheries, genomic tools can offer useful inputs to improve cost-efficiency with respect to traditional procedures in the mid-term and increase the accuracy and spatial resolution of data used in fish stock assessment. Nonetheless, they have seldom been applied into fisheries management and, so far, their ability to solve fisheries-specific questions has not been sufficiently assessed.

The overall purpose of FishGenome was to evaluate the suitability of three genomic techniques - **Close Kin Mark Recapture (CKMR), epigenetic Age Determination by DNA methylation (DNAm)** and **environmental DNA (eDNA)** to estimate various essential parameters for fisheries stock assessments, including absolute abundance, survival, age and biomass. Additionally, we assessed the potential and ripeness of a fourth genomic technique - **restriction site Associated DNA Sequencing (RAD-Seq)** - to estimate a series of parameters that are also important to evaluate stocks but have been mostly neglected in fisheries assessments, i.e. fine-scale stock substructure, connectivity and molecular sexing.

To this end, we performed a comprehensive and systematic review of literature and other sources of relevant knowledge (grey literature, reports and working documents) on these genomic tools and on the main bioinformatics processing required to analyse and understand the information delivered by them. Additionally, two other reviews were carried out: a review of the mandatory trawl-based research surveys in EU waters that were in place in 2020 and a review on cost-efficiency of the use of genomic methods for stock assessments were carried out. Finally, based on the reviews above we conducted a critical assessment of the potential of genomic methods to enhance fisheries stock

assessment and we identified the barriers and risks, and analysed the impact and mitigation of such implementation.

The resulting State of the Art reviews served as the foundation to design a series of Pilot studies to test the genomic tools of interest in a relevant context.

The Pilot studies focused on three commercially important marine fishes with different levels of exploitation; two demersal species, Atlantic cod (*Gadus morhua*) and European hake (*Merluccius merluccius*) as well as on a coastal species targeted by small-scale fisheries, the ballan wrasse (*Labrus bergylta*). The rationale behind this strategy was to cover a wide spectrum of life-histories. Cod and hake are the most important demersal species caught by EU fleet in terms of landings, economic value and food consumption and both are assessed using bottom trawl surveys and fishery-dependent data. The ballan wrasse is an important target species for small-scale and recreational fisheries in the Atlantic European waters. Despite the suspected poor status of ballan wrasse stocks, no assessment is in place due to lack of information on stock structure and population dynamics. The Pilot studies covered three different regions: the North Sea, the North-West Iberian Peninsula and the Balearic Islands in the Mediterranean.

The first step of the Pilot studies consisted of the development of a series of tailored protocols for each of the genomic techniques, describing the experimental design, laboratory work and equipment needed to collect, analyse and preserve the samples. These consisted of cod, northern hake stock and environmental samples (specifically, water and sediment) from the North Sea; hake and environmental samples from the Balearic Islands; and hake and ballan wrasse from the North-West Iberian Peninsula. The latter were part of a tissue biobank owned by the FishGenome consortium whereas the samples from the first two regions were collected during two fisheries research surveys, namely the North Sea IBTS-Q3 and the Mediterranean MEDITS-GS5. This strategy aimed at gaining insight not only on the suitability of each technique, but also on the feasibility of accommodating their requirements to surveys' characteristics. This information is key to assess the potential of the genomic techniques tested to contribute to fisheries assessments, as such potential relies not only on the capacity of the methodologies to produce accurate data, but also on their technical and biological requirements.

The first technique tested was Close Kin Mark Recapture (**CKMR)**, a method that has recently gained considerable attention, due to the increasing affordability of more powerful sequencing technologies that can boost its application. This methodology requires the inference of kinship relationships (parent-offspring, half-siblings) among individuals using their genetic information. It is based on a simple concept: the larger the population the less likely to find relatives and *vice versa*. The method is **useful to characterize the demography of wild populations.** It can estimate essential parameters of the exploited populations for the sustainable management of fisheries, **including abundance, population trend, survival rates and fecundity**. Many of these parameters are often extremely challenging to estimate, especially in abundant, highly mobile and dispersed organisms, such as fishes.

The proof of concept for CKMR was provided by a 2016 study that estimated the absolute abundance of southern bluefin tuna, revealing a less-depleted and more productive stock than the conventional assessment with fishery data. Despite this successful application, and several more that followed, mostly targeting small populations of freshwater fish and elasmobranchs[1], it remains questionable whether the method can be scaled-up to accommodate species that are more abundant. The methodology demands an extensive sampling effort and the high costs associated to the analysis of

[1] Atlantic salmon, Artic grayling, white shark, brown trout, thornback ray and blue skate

large populations might render the approach unfeasible for the majority of commercially exploited fish. Our Pilot study focused on three populations of European hake from the North Sea, the North-West Iberian Peninsula and the Balearic Islands and a population of cod from the North Sea. We used a powerful genomic technique called RAD-seq that is able to sequence regions across the whole genome, to reveal thousands of genetic markers called single nucleotide polymorphisms (**SNPs**), which consist in the substitution of one nucleotide in the DNA sequence. The markers provided a high prediction accuracy of kinship among the individuals of each population. All the samples collected during the research surveys were analysed, but the number of kinship pairs detected was very low. According to our results, sampling sizes of populations from all three locations and for both species should be increased by, at least, 50- to 100-fold, to allow an accurate inference of absolute abundance and other demographic parameters. Thus, a clear bottleneck for the implementation of CKMR into current fisheries assessments lies on the collection of the required number of specimens. **The applicability of CKMR to exploited species with large population sizes would require an intensification of current sampling on research surveys, the collection of samples across several years and/or the involvement of vessels from the commercial fleet.** Multi-year samplings add uncertainty to the estimations, but CKMR could still be a valuable tool as a periodic independent estimate of population parameters to cross-validate population estimates provided by regular stock assessments.

The data we produced to assess CKMR was re-analysed to test the capacity and readiness of genomic analysis (**RAD-Seq**) to delineate **stock boundaries, determine fine-scale population structure/substructure and estimate connectivity**. Although an adequate management of fish stocks inherently relies on a precise estimation of their spatial distribution, their structure and the migration movements among stocks, these key parameters are not quantitatively used in current fisheries assessments.

Our **results prove the robustness, accuracy and technical power of the methodology used to explore stock boundaries, population structure and quantify connectivity in hake**. Additionally, its ability to uncover stock sub-structuring was evident in the ballan wrasse population, where a strong genomic differentiation was detected. The distinct genomic signatures were strongly correlated with two different colour phenotypes (plain and spotted), indicating a strong reproductive isolation between them that demands a separate management of both phenotypes.

We demonstrated that well-established genomic methods are readily available to identify stocks and explore structure, substructure and connectivity, all critical parameters for stock management. Importantly, their application requires a small number of individuals, which can be easily collected during current fisheries research surveys. Considering that climate change is already shifting the distribution range and the migratory patterns of many fish stocks, there is a pressing need to adopt a methodology for the estimation of stock structure and connectivity, in order to ensure the persistence, productivity and resilience of exploited stocks.

The data derived from CKMR from all three regions, was also re-analysed to perform a sex-marker search across the genome of hake. Although we could only isolate a marker that was specific for the North Sea population, there is ample evidence across the scientific literature of the **utility of RAD-seq to characterize sex markers** in species with simple chromosomal sex determination systems. Genetic sexing offers important advantages compared to classical histological assignment of sex, as it allows sexing of early stages of development (eggs and young juveniles) and only requires a small piece of tissue that can be collected non-invasively. The development of genetic tools to identify sex in exploited species is relevant as they provide a fast, cheap and easy way

to determine sex but, more importantly, because they can be used to incorporate sex-specific information of juveniles into fisheries stock assessments. This information is currently not considered but is critical for understanding the structure and resilience of fish stocks.

The second HTS technique assessed in FishGenome was the epigenetic Age Determination by DNA methylation (**DNAm)**. DNA methylation refers to the chemical process of adding a tag known as a methyl group to one of the bases of the DNA. It is a mechanism that cells use to control gene expression as it serves to switch genes on and off by altering the DNA's 3D structure. Genomic DNA methylation patterns change with the age of the organisms and thus, the methylation status has the potential to be a good indicator of the biological age. Epigenetic clocks are mathematical models that combine the methylation levels at specific sites in the DNA to estimate age.

In fisheries management, accurate estimates of age are essential to infer life-history traits and for effective stock assessment. Knowledge about the age composition of fish populations provides information about stock structure, age at maturity, life span, mortality and their growth rate. Inaccurate determination of age also likely increases the uncertainty in assessment and scientific advice. Age estimation in fishes has traditionally relied on the analysis of growth marks in hard structures such as otoliths, but this requires well-trained personnel, is time-consuming, lethal and has low accuracy in some species, highlighting the need for new methodologies.

The FishGenome Pilot study was designed to develop an epigenetic clock in cod and to test its potential for determining age with the accuracy and precision needed in stocks assessment. To this end, a set of individuals collected in the North Sea, which covered all age classes (dated with otoliths), from both sexes were analysed using a genomic technique (bis-RAD-seq) to scan across the entire cod genome. The process rendered an epigenetic clock consisting of 26 genome locations or loci with methylation profiles highly correlated with age. It **has the capacity to estimate biological age with a precision of ~8 months**. The results obtained **provide evidence of the potential of this tool** for age estimation, although the clock should be further validated on samples from different geographic origins and always calibrated with chronological age. The new technique offers a non-lethal tool for a rapid, non-subjective estimation of age and is suited to automation.

The third technique evaluated in FishGenome was the environmental DNA (**eDNA**) based analysis, an increasingly popular genomic tool for monitoring ecosystems. eDNA refers to the traces of DNA shed by organisms (via skin, faeces, gametes, etc.) into their environment. It provides a non-invasive alternative for surveying aquatic communities and has seen a massive increase in application during the last decade. The analysis of eDNA has a demonstrated capacity to provide information on biodiversity and distribution of species. Nonetheless, its application is not free of challenges, mainly derived from the low-quantity of DNA present in the samples that makes it prone to contamination and biases. The capacity of the eDNA to reflect abundance and/or biomass has been suggested by a few studies, but a large body of research also reports a poor relationship between biomass and eDNA concentration. These parameters are highly relevant for stock management, while the estimation of biodiversity is of fundamental importance in ecosystem-based approaches. Current fisheries research surveys regularly collect data to estimate both.

The design of the FishGenome Pilot study involved the collection of water and sediment, right before the trawling, to compare the eDNA based-estimates with those obtained from standardized trawl catches. We tested the potential of two different genomic analyses, **metabarcoding**, which is useful to analyse all the species present in the eDNA at the same time, and the quantitative PCR (**qPCR**), which is species-specific. Our results indicate that the **metabarcoding analysis of eDNA from water samples is highly accurate, as it revealed the presence or absence of a given fish species**

**with 90% reliability, compared to trawl catches**. Moreover, the method unveiled a large number of species undetected by trawling, in line with other studies, which possibly correspond to species with the ability to escape nets and small fish of young developmental stages that pass through the mesh. Nonetheless, some relevant limitations were also detected, as the method was unable to detect some species, like sharks, due to amplification incompatibility, while the detection of others was hindered by the incompleteness of public genomic databases. Additionally, the successful implementation of the method required extensive fine-tuning of the protocols used. The results obtained from sediments where less consistent with trawling catches, possibly because they preserve the DNA traces for longer periods, acting as repositories of ancient, rather than contemporary, biodiversity.

The suitability of the qPCR to detect the presence of Atlantic cod from water eDNA was also tested, revealing a highly concordant detection of the species between eDNA and trawl catches. Importantly, cod could be detected from eDNA at very low abundances, even at stations where it did not appear in trawl catches. **The capacity of the qPCR to quantify biomass was also remarkable, as a significant positive correlation was found between the Atlantic cod biomass and the quantity of eDNA.**

Our results support the ability of eDNA to reliably reflect biodiversity and abundance of aquatic macroorganisms, but also highlights some hurdles that need to be overcome, before the tool can be used to assist fisheries management.

he last stage of the FishGenome project directly contributing to its ultimate goal consisted of the development of a **Roadmap** for the implementation of genomic-based approaches in fishery stock assessment. This required a **cost-benefit** analysis of these approaches and the factors that may drive, directly or indirectly, their implementation (Strengths, Weaknesses, Opportunities and Threats (**SWOT**) analysis). The approach we followed for the cost-benefit analysis was to evaluate whether the use of genomic methodologies can yield, at least, equal or equivalent information outputs at a lower cost, than methodologies currently used for scientific advice. It consisted of a contextualised cost analysis and comparison of the research surveys at sea and of the implementation of the HTS techniques. This enabled the identification of some potential pathways and criteria for efficiency gains. However, several assumptions were necessary, due to lack of past references and scarce data for the analysis.

The information obtained on cost-efficiency was combined with findings from the State-of-the-art reviews and with the insights from the pilot studies. All the above was used to feed the SWOT analysis for the future use of these techniques in data collection and fisheries assessments. Identified factors that may directly or indirectly affect the implementation of the techniques in fishery assessments included internal and external ones. Internal factors were classified as **S**trengths and **W**eaknesses and referred to features such as robustness and accuracy, reliability, versatility, coverage, cost efficiency and added value for each of the techniques. External factors reflected on **O**pportunities and **T**hreats and included implementation capacity, social, political, legal, financial and technological trends. Therefore, advantages and limitations of the genomic techniques, as well as the needs and conditions required for their implementation in stocks assessment were classified and synthesized to feed into the final step.

A final compilation of all the information served as an input for the Roadmap design and prior to it, was used for the definition of plausible implementation scenarios (considering technical, logistic, financial, scientific and environmental aspects, among others). The Roadmap provides detailed information on whether and how the analysed genomic-based approaches could become part of the regular research surveys. It describes the steps, the pathways, and the timeline for a progressive implementation in fisheries assessments and management, proposing a long-term ambition for full-scale implementation.

This **complex but comprehensive roadmap** is designed to achieve five specific objectives in the short- mid- and long-terms, each requiring the involvement of different stakeholders to varying degrees. These objectives and pillars enabled the definition and organisation of a total of 45 actions around the following five strategic challenges:

1. **Towards a progressive uptake of the genomic information** aiming to ensure the progressive uptake of genomic information for stock assessment and scientific advice. The various genomic techniques have different levels of maturity and readiness, while the research surveys differ in capacity to implement the demanding routines, and some species or stocks are more suitable to embrace genomic approaches than others. Thus, a thorough stepwise approach is defined to achieve a successful implementation.
2. **Continuous scientific improvement of the methodologies** is required. Genomic technology is still progressing very fast and refinements and optimizations should be taken from a scientific perspective. This includes adjusting and standardizing protocols, lab intercalibration and incorporating further scientific findings and technology developments. At the same time, the actions under this strategic challenge will explore the use of other genomic methods towards the same goal, as well as the use of the same methods to estimate other biological parameters of interest in scientific advice and fisheries management
3. **Fostering a coordinated roadmap** to engage all the relevant disciplines and stakeholders required for a successful integration of the HTS methods into fisheries stock assessment and management. This relies on existing initiatives for coordination and cooperation in fisheries management within the legal framework in EU. A key driver for the success of the roadmap will be its capacity to harness and underpin the potential of the already existing initiatives, minimising the new structures and avoiding unnecessary overlaps and duplications.
4. **Developing capacities for a successful implementation** is essential in two major areas. First, infrastructures across Europe must be reinforced and developed, always keeping in mind the overarching priority of maximizing the use of existing ones. Second, it is crucial to improve skills and knowledge, to build engagement through training networks and to develop close communication and common understanding between scientist from different disciplines by using a common language.
5. **Ensuring value for money** aims at assessing and demonstrating the benefits of the methods more systematically, through gathering of relevant data for evaluating cost and investment efficiency across objectives and actions and doing so beyond stock assessment. Applying genomics methods in fisheries assessments is quite innovative and there is scarce information available on the costs and investment needs derived from this specific application.

The proposed roadmap is inevitably complex due the number of actions required, and the high number and diversity of actors that need to be involved. Yet, it defines the needs for the implementation in the most plausible scenario along with a precise definition of the criteria to select case studies. The **roadmap is designed** to provide **innovative tools** to the scientific community to improve the **scientific advice** necessary for a **sustainable exploitation of fisheries** resources, but also to increase our knowledge on ecosystem functioning and biodiversity that contributes to a **better ecosystem management**.

Finally, it is noteworthy that the study outcomes at different stages were subject to evaluation and feedback from external experts. In this way, the consortium capacities were enriched on technologies, research surveys, how these feed into the fisheries stock assessment, scientific advice and management, in compliance with established regulations and policies.

## RESUME EXECUTIF

Le contrat d'appel d'offres "FishGenome : Améliorer le rapport coût-efficacité des campagnes de surveillance halieutique et des évaluations des stocks de poissons à l'aide de méthodes génétiques de séquençage à haut débit (next-generation sequencing)" a été développé pour évaluer la capacité et le niveau de maturité de plusieurs nouvelles technologies génomiques émergentes pour améliorer les évaluations des stocks de poissons. Ces évaluations sont nécessaires pour surveiller l'état des stocks de poissons et garantir des pratiques de pêche qui les exploitent à des niveaux durables. Elles s'appuient sur la collecte de données provenant de sources multiples, qui comprennent des données liées à la pêche - captures, débarquements, informations biologiques - ainsi que des données indépendantes de la pêche obtenues à partir de campagnes de surveillance halieutique. Ces campagnes fournissent des informations précieuses et systématisées sur les populations de poissons exploitées, la biodiversité marine et leur environnement.

Cependant, les campagnes de surveillance halieutique ont également des limites importantes telles qu'un coût économique élevé couplé à une logistique complexe (ce qui produit des données insuffisantes dans l'espace et dans le temps), et du temps est nécessaire pour traiter et analyser les données collectées. Ces limitations produisent des données éparses dans l'espace et dans le temps et réduisent la précision de l'estimation de paramètres clés dans l'étude des populations. En outre, les méthodes conventionnelles de détermination de paramètres biologiques présentent certaines limites, telles que l'incapacité de déterminer la structure de sexe des juvéniles ou de fournir des informations sur la connectivité, deux éléments cruciaux pour la gestion des stocks. La conception et la technologie des campagnes en mer progressent constamment pour faire face à ces limitations et réduire l'incertitude des évaluations de stocks et des avis scientifiques sur l'état des stocks et des écosystèmes exploités. Ainsi, les progrès technologiques récents dans le domaine de la génomique devraient améliorer l'efficacité et aider à atténuer certaines des lacunes des méthodologies traditionnelles telles que des coûts élevés et une logistique complexe.

La croissance rapide des techniques de génomique au cours des dernières décennies offre aujourd'hui le potentiel de résoudre certains de ces défis, ainsi que de compléter les méthodes traditionnelles pour aider la gestion des pêches à long terme. Le terme «génétique» fait référence à l'étude d'un groupe de gènes ou d'autres régions du génome, tandis que «génomique» est utilisé pour des études impliquant l'ensemble du génome aidées par des méthodes de séquençage génétique à haut débit. Le génome de tout être vivant encode la plupart de ses caractéristiques et détermine ainsi l'apparence, le comportement et la physiologie des organismes. Dans le contexte des pêches, les outils génomiques peuvent contribuer à améliorer le rapport coût-efficacité par rapport aux procédures traditionnelles et augmenter la précision et la résolution spatiale des données utilisées dans l'évaluation des stocks de poissons. Néanmoins, ils ont rarement été appliqués à la gestion des pêches et, jusqu'à présent, leur capacité à résoudre des questions spécifiques à la pêche n'a pas été suffisamment évaluée.

L'objectif global de FishGenome était d'évaluer la pertinence de trois techniques génomiques – **l'approche marquage-recapture basée sur l'identification génétique des paires des individus apparentés (**en anglais : **Close Kin Mark Recapture (CKMR)), la détermination épigénétique de l'âge par méthylation de l'ADN (ADNm) (**en anglais : **epigenetic Age Determination by DNA methylation (DNAm))** et **l'ADN environnemental (ADNe)** pour estimer divers paramètres essentiels pour les évaluations des stocks de pêche, y compris l'abondance absolue, la survie, l'âge et la biomasse. De plus, nous avons évalué le potentiel et la maturité d'une quatrième technique génomique - le séquençage d'ADN associé au site de restriction (en anglais : **restriction site Associated DNA Sequencing (RAD-Seq))** - pour estimer une série de paramètres qui sont également importants pour évaluer les stocks mais qui ont été pour la plupart négligés dans les évaluations des pêcheries, c'est-à-dire, la sous-structure du stock à échelle fine, la connectivité et sexage moléculaire.

Pour cela, nous avons effectué une revue exhaustive et systématique de la littérature et d'autres sources de connaissances pertinentes (littérature grise, rapports et documents de travail) sur ces outils génomiques et sur les principaux traitements bioinformatiques nécessaires à l'analyse et à la compréhension des informations délivrées par ceux-ci. En outre, une revue littéraire sur les campagnes de surveillance halieutiques menées sur des chalutiers dans les eaux européennes, ainsi qu'une revue sur le rapport coût-efficacité de l'utilisation des méthodes génomiques pour l'évaluation des stocks ont été effectués.

Ces revues produites sur l'état de l'art ont servi de base à la conception d'une série d'études pilotes pour tester les outils génomiques d'intérêt dans un contexte pertinent.

Les études pilotes se sont concentrées sur trois poissons marins d'importance commerciale avec différents niveaux d'exploitation ; deux espèces démersales, la morue franche (*Gadus morhua*) et le merlu commun (*Merluccius merluccius*) ainsi que sur une espèce côtière ciblée par la pêche artisanale, la grande vieille (*Labrus bergylta*). L'objectif de cette stratégie était de couvrir un large éventail d'histoires de vie. La morue franche et le merlu commun sont les espèces démersales les plus importantes capturées par la flotte de l'UE en termes de débarquements, de valeur économique et de consommation alimentaire, et les deux espèces sont évaluées à l'aide de campagnes de surveillance halieutiques menées sur des chaluts de fond. La grande vieille est l'une des espèces cibles les plus importantes pour la pêche artisanale et récréative. Malgré le mauvais état présumé des stocks de grandes vieilles dans les eaux européennes, aucune évaluation n'est en place. Les études pilotes ont couvert trois régions différentes : la mer du Nord, le nord-ouest de la péninsule ibérique et les îles Baléares en Méditerranée.

La première étape des études pilotes a consisté à développer une série de protocoles sur mesure pour chacune des techniques génomiques, décrivant le design expérimental, le travail de laboratoire et le matériel nécessaire pour collecter, analyser et conserver les échantillons. Les échantillons proviennent de morues franches, de merlus du stock nord européen, et d'échantillons environnementaux (en particulier d'eau et de sédiments) de la mer du Nord ; de merlus et d'échantillons environnementaux des îles Baléares ; et de merlus et grandes vieilles du nord-ouest de la péninsule ibérique. Ces derniers faisaient partie d'une biobanque de tissus appartenant au consortium FishGenome, tandis que les échantillons des deux premières régions ont été collectés lors de deux campagnes de recherche halieutique, à savoir l'IBTS-Q3 en mer du Nord et le MEDITS-GS5 en Méditerranée. Cette stratégie visait à produire des informations précises non seulement sur la pertinence de chaque technique, mais aussi sur la possibilité d'adapter leurs exigences aux caractéristiques des campagnes de recherche. Ces informations sont essentielles pour évaluer le potentiel des techniques génomiques testées pour améliorer les évaluations des pêcheries, car elles reposent non seulement sur la capacité des méthodologies à produire des données précises, mais également sur leurs exigences techniques et biologiques.

La première technique testée a été celle de l'approche marquage-recapture basée sur l'identification génétique des paires des individus apparentés (**CKMR**), une méthode qui a récemment gagné une attention considérable en raison de l'accessibilité croissante de technologies de séquençage plus puissantes pouvant stimuler son application. Cette méthodologie nécessite l'inférence des relations de parenté (parent-descendant, demi-frères et sœurs) entre les individus en utilisant leur information génétique. Cette technique est basée sur un concept simple ; plus la population est importante, moins il y a de chances de trouver des liens de parenté et *vice versa*. La méthode est **utile pour caractériser la démographie des populations sauvages**. Elle peut estimer des paramètres essentiels pour la gestion durable des populations exploitées, **y compris l'abondance, les tendances des populations, les taux de survie et la fécondité**. Bon nombre de ces paramètres sont souvent extrêmement difficiles à estimer, en particulier chez les organismes abondants, très mobiles et dispersés tels que les poissons.

La preuve de concept pour la méthode CKMR a été fournie par une étude de 2016 qui a estimé l'abondance absolue du thon rouge du sud, révélant un stock moins exploité et plus productif qu'en utilisant une évaluation conventionnelle avec des données de pêche. Malgré cette application réussie, et plusieurs autres qui ont suivi, ciblant principalement de petites populations de poissons d'eau douce et d'élasmobranches, on peut se demander si la méthode peut être étendue pour s'adapter aux espèces plus abondantes. La méthodologie exige un effort d'échantillonnage important et les coûts élevés associés à l'analyse de grandes populations pourraient rendre l'approche irréalisable pour la majorité des poissons exploités commercialement. Notre étude pilote a porté sur trois populations de merlu commun de la mer du Nord, du nord-ouest de la péninsule ibérique et des îles Baléares, ainsi qu'une population de morue de la mer du Nord. Nous avons utilisé une technique génomique puissante appelée RAD-Seq qui est capable de séquencer des régions sur l'ensemble du génome, pour révéler des milliers de marqueurs génétiques appelés polymorphismes nucléotidiques simples (**SNPs**), qui consistent en la substitution d'une seule base dans la séquence d'ADN. Les marqueurs ont fourni une grande précision de prédiction de la parenté entre les individus de chaque population. Tous les échantillons collectés lors des campagnes en mer ont été analysés mais le nombre de paires de parenté détectées était très faible. Selon nos résultats, les tailles d'échantillonnage des populations des trois sites et pour les deux espèces devraient être multipliées par 50 à 100 pour permettre une inférence précise de l'abondance absolue et d'autres paramètres démographiques. Ainsi, un goulot d'étranglement évident pour la mise en œuvre de la méthode CKMR dans les évaluations actuelles des pêches réside dans la collecte du nombre de spécimens requis. **L'applicabilité de la méthode CKMR aux espèces exploitées avec de grandes tailles de population nécessiterait la collecte d'échantillons sur plusieurs années et une intensification des campagnes de surveillance halieutique actuelles ou, alternativement, l'implication de bateaux/navires de la flotte commerciale.** Les échantillonnages pluriannuels ajoutent de l'incertitude aux estimations, mais la méthode CKMR pourrait tout de même être un outil précieux en tant qu'estimation indépendante périodique des paramètres de populations pour valider les estimations de populations fournies par les évaluations régulières des stocks.

Les données que nous avons produites pour évaluer la méthode CKMR ont été de nouveau analysées pour tester la capacité et le niveau de maturité technologique de l'analyse génomique (**RAD-Seq**) à délimiter **les limites des stocks, déterminer la structure/sous-structure de la population à petite échelle et estimer la connectivité**. Bien qu'une gestion adéquate des stocks de poissons repose intrinsèquement sur une estimation précise de leur répartition spatiale, de leur structure et des mouvements migratoires entre les stocks, ces paramètres clés ne sont pas utilisés quantitativement dans les évaluations actuelles des pêches.

Nos **résultats prouvent la robustesse, la précision et la puissance technique de la méthodologie utilisée pour explorer les limites des stocks, la structure de la population et quantifier la connectivité chez le merlu**. De plus, sa capacité à déterminer la sous-structuration du stock était évidente chez la population de grandes vieilles, où une forte différenciation génomique a été détectée. Les signatures génomiques distinctes étaient fortement corrélées avec deux phénotypes différents, indiquant un fort isolement reproductif entre eux qui exige une gestion séparée des deux sous-stocks.

Nous avons démontré que des méthodes génomiques bien établies sont facilement disponibles pour faciliter la gestion des stocks. Il est important de noter que leur application nécessite un petit nombre d'individus, qui peuvent être facilement collectés lors des campagnes de recherche halieutique en cours. Considérant que le changement climatique modifie déjà l'aire de répartition et les schémas migratoires de nombreux stocks de poissons, il est urgent d'adopter une méthodologie pour l'estimation de ces paramètres afin d'assurer la persistance, la productivité et la résilience des stocks exploités.

Les données dérivées de la méthode CKMR ont également été ré-analysées pour effectuer une recherche de marqueur de sexe dans le génome du merlu. Bien que nous n'ayons pu isoler qu'un marqueur spécifique à la population de la mer du Nord, il existe de nombreuses indications dans la littérature scientifique de l'**utilité de la méthode RAD-Seq pour caractériser les marqueurs de sexe**. Le sexage génétique offre des avantages importants par rapport à l'attribution histologique classique du sexe, car il permet le sexage des premiers stades de développement (œufs et jeunes juvéniles) et ne nécessite qu'un petit morceau de tissu pouvant être collecté de manière non invasive. Le développement d'outils génétiques pour identifier le sexe des espèces exploitées est pertinent car ils fournissent un moyen rapide, bon marché et facile de déterminer le sexe mais, plus important encore, parce qu'ils peuvent être utilisés pour incorporer des informations spécifiques au sexe des juvéniles dans les évaluations des stocks de pêche. Ces informations ne sont actuellement pas prises en compte mais sont essentielles pour comprendre la structure et la résilience des stocks de poissons.

La deuxième technique de séquençage à haut débit évaluée dans FishGenome était la détermination épigénétique de l'âge par méthylation de l'ADN (**ADNm**). La méthylation de l'ADN fait référence au processus chimique consistant à ajouter une étiquette connue sous le nom de groupe méthyle à l'une des bases de l'ADN. C'est un mécanisme que les cellules utilisent pour contrôler l'expression des gènes car il sert à activer et désactiver les gènes en modifiant la structure 3D de l'ADN. Les schémas de méthylation de l'ADN génomique changent avec l'âge des organismes et, par conséquent, l'état de méthylation a le potentiel d'être un bon indicateur de l'âge biologique. Les horloges épigénétiques sont des modèles mathématiques qui combinent les niveaux de méthylation à des sites spécifiques de l'ADN pour estimer l'âge.

Dans la gestion des pêches, des estimations précises de l'âge sont essentielles pour déduire les traits d’histoire de vie et pour une évaluation efficace des stocks. La connaissance de la composition en âge des populations de poissons fournit des informations sur la structure des stocks, l'âge à maturité, la durée de vie, la mortalité et leur taux de croissance. Une détermination inexacte de l'âge peut aussi augmenter l'incertitude des évaluations et des avis scientifiques. L'estimation de l'âge des poissons reposait traditionnellement sur l'analyse des marques de croissance dans les structures dures telles que les otolithes, mais cela nécessite un personnel bien formé, prend du temps, est létal pour les poissons et a une faible précision chez certaines espèces, ce qui souligne la nécessité de nouvelles méthodologies.

L'étude pilote FishGenome a été conçue pour développer une horloge épigénétique chez la morue et pour tester son potentiel pour déterminer l'âge biologique avec l'exactitude et la précision nécessaires à l'évaluation des stocks. Pour cela, un ensemble d'individus collectés en mer du Nord, couvrant toutes les classes d'âge (datées par des otolithes) et les deux sexes, ont été analysés à l'aide d'une technique génomique (bis-RAD-Seq) pour scanner l'ensemble du génome de la morue. Le processus a rendu une horloge épigénétique composée de 26 emplacements du génome avec des profils de méthylation fortement corrélés avec l'âge. **Il a la capacité d'estimer l'âge biologique avec une précision d'environ 8 mois**. Les résultats obtenus **mettent en évidence le potentiel de cet outil** comme alternative à l'estimation traditionnelle de l'âge basée sur les otolithes, même si l'horloge doit encore être validée sur des échantillons d'origines géographiques différentes. La nouvelle technique offre un outil non létal pour une estimation rapide et non subjective de l'âge et est adaptée à l'automatisation.

La troisième technique évaluée dans FishGenome était l'analyse basée sur l'ADN environnemental (ADNe), un outil génomique de plus en plus populaire pour la surveillance des écosystèmes. L'ADNe fait référence aux traces d'ADN rejetées par les organismes (via la peau, les fèces, les gamètes, etc.) dans leur environnement. Il fournit une alternative non invasive pour l'étude des communautés aquatiques et a connu une augmentation massive de son utilisation au cours de la dernière décennie. L'analyse de l’ADNe a démontré sa capacité à fournir des informations sur la biodiversité et la répartition des espèces. Néanmoins, son application n'est pas exempte de défis,

principalement dus à la faible quantité d'ADN présente dans les échantillons qui la rend sujette à la contamination et aux biais. La capacité de l'ADNe à refléter l'abondance et/ou la biomasse a été suggérée par quelques études, mais un grand nombre de recherches rapportent également une mauvaise relation entre la biomasse et la concentration d'ADNe. Ces paramètres sont très pertinents pour la gestion des stocks tandis que l'estimation de la biodiversité est d'une importance fondamentale dans les approches écosystémiques. Les campagnes de surveillance halieutique actuelles collectent régulièrement des données pour estimer les deux.

La conception de l'étude pilote FishGenome impliquait la collecte d'eau et de sédiments, juste avant le chalutage, pour comparer les estimations basées sur l'ADNe avec celles obtenues à partir de captures au chalut standardisées. Nous avons testé le potentiel de deux analyses génomiques différentes, **le métabarcoding**, qui est utile pour analyser toutes les espèces présentes dans l'ADNe en même temps, et la PCR quantitative (**qPCR**), qui est spécifique à l'espèce. Nos résultats indiquent que **le métabarcoding de l'ADNe à partir d'échantillons d'eau est très précise, car elle a révélé la présence ou l'absence d'une espèce de poisson donnée avec une fiabilité de 90%, par rapport aux prises au chalut**. De plus, la méthode a dévoilé un grand nombre d'espèces non détectées par le chalutage, conformément à d'autres études, qui correspondent peut-être à des espèces ayant la capacité d'échapper aux filets et aux petits poissons de jeunes stades de développement qui passent au travers des mailles du filet.

Néanmoins, certaines limitations pertinentes ont également été détectées, car la méthode n'a pas pu détecter certaines espèces comme les requins, en raison d'une incompatibilité d'amplification, tandis que la détection d'autres a été entravée par l'incomplétude des bases de données génomiques publiques. De plus, la mise en œuvre réussie de la méthode a nécessité d'importants ajustements. Les résultats obtenus à partir des sédiments étaient moins cohérents avec les captures au chalut, possiblement parce qu'ils préservent les traces d'ADN pendant de plus longues périodes, agissant comme des dépositaires de la biodiversité ancienne plutôt que contemporaine.

La pertinence de la qPCR pour détecter la présence de la morue de l'Atlantique à partir de l'ADNe de l'eau a également été testée, révélant une détection hautement concordante de l'espèce entre l'ADNe et les captures au chalut. Il est important de noter que la morue a pu être détectée à partir de l'ADNe à de très faibles abondances, même aux stations où elle n'apparaissait pas dans les captures au chalut. **La capacité de la qPCR à quantifier la biomasse était également remarquable, car une corrélation positive significative a été trouvée entre la biomasse de morue et la quantité d'ADNe.**

Nos résultats confirment la capacité de l'ADNe à refléter de manière fiable la biodiversité et l'abondance des macro-organismes aquatiques, mais mettent également en évidence certains obstacles qui doivent être surmontés avant que cet outil puisse être utilisé pour aider la gestion des pêches.

La dernière étape du projet FishGenome contribuant directement à son objectif ultime consistait en l'élaboration d'une **feuille de route** pour la mise en œuvre d'approches basées sur la génomique dans l'évaluation des stocks halieutiques. Cela a nécessité l'analyse d'informations sur le rapport **coût-bénéfice** de ces approches et sur les facteurs qui peuvent conduire, directement ou indirectement, à leur mise en œuvre (analyse des forces, des faiblesses, des opportunités et des menaces (en anglais : Strengths, Weaknesses, Opportunities and Threats (**SWOT**) analysis). L'approche que nous avons suivie pour l'analyse coûts-bénéfice consistait à évaluer si l'utilisation de méthodologies génomiques peut produire des résultats d'information au moins égaux ou équivalents, et à un moindre coût, aux méthodologies actuelles utilisées pour les avis scientifiques. Elle a consisté en une analyse et une comparaison contextualisée des coûts des campagnes de recherche en mer et de la mise en œuvre des techniques séquençage à haut débit. Cela a permis d'identifier certaines voies et critères potentiels

de gains d'efficacité. Cependant, plusieurs hypothèses étaient nécessaires en raison des contraintes liées au manque de références et à la rareté des données pour l'analyse.

Les informations obtenues sur le rapport coût-bénéfice ont été combinées avec les résultats des revues de l'état de l'art et avec les informations des études pilotes. Tout cela a été utilisé pour alimenter l'analyse SWOT pour l'utilisation future de ces techniques dans l'EU-MAP pour la collecte de données et les évaluations de la pêche. Les facteurs identifiés susceptibles d'affecter directement ou indirectement la mise en œuvre des techniques dans les évaluations des pêcheries comprenaient des facteurs internes et externes. Les facteurs internes ont été classés en tant que forces et faiblesses, et ont fait référence à des caractéristiques telles que la robustesse et la précision, la fiabilité, la polyvalence, la couverture, la rentabilité et la valeur ajoutée pour chacune des techniques. Les facteurs externes ont porté sur les opportunités et les menaces et ont ajouté des informations sur la capacité de mise en œuvre, les tendances sociopolitiques, juridiques, financières et technologiques, etc. Par conséquent, les avantages et les limites des techniques génomiques, ainsi que les besoins et les conditions nécessaires à leur mise en œuvre dans l'évaluation des stocks ont été classés et synthétisés pour alimenter la toute dernière étape.

Une compilation finale de toutes les informations a servi de contribution pour la conception de la feuille de route et avant celle-ci, a été utilisée pour la définition de scénarios de mise en œuvre plausibles (en tenant compte des aspects techniques, logistiques, financiers, scientifiques et environnementaux, entre autres). La feuille de route fournit des informations précises et spécifiques sur la question de savoir si et comment les approches génomiques analysées pourraient faire partie des campagnes de recherche régulières. Elle décrit les étapes, les voies et le calendrier d'une mise en œuvre progressive des évaluations et de la gestion des pêches, en élaborant une ambition à long terme pour une mise en œuvre à grande échelle.

Cette **feuille de route complexe mais complète** est conçue pour atteindre cinq objectifs spécifiques à court, moyen et long terme, à travers cinq domaines ou piliers stratégiques, chacun nécessitant dans une certaine mesure l'implication de différentes parties prenantes. Ces objectifs et piliers ont permis de définir et d'organiser un total de 45 actions autour des initiatives suivantes :

1. **Vers une adoption progressive des informations** génomiques visant à assurer l'assimilation progressive des informations génomiques pour l'évaluation des stocks et les avis scientifiques. Les diverses techniques génomiques ont différents niveaux de maturité technologique et de disponibilité tandis que les campagnes de recherche diffèrent dans leur capacité à mettre en œuvre les routines exigeantes, et certaines espèces ou stocks sont plus adaptés pour adopter des approches génomiques. Ainsi, une approche prudente et détaillée est définie par étapes pour parvenir à une mise en œuvre réussie.
2. **Une amélioration scientifique continue des méthodologies** est nécessaire. La technologie génomique progresse très rapidement et des améliorations et optimisations doivent être prises d'un point de vue scientifique. Cela comprend l'ajustement et la standardisation des protocoles, une calibration entre laboratoires et l'intégration d'autres découvertes scientifiques et de développements technologiques. Parallèlement, cette initiative explorera l'utilisation d'autres méthodes génomiques dans le même but, ainsi que l'utilisation des mêmes méthodes pour estimer d'autres paramètres biologiques d'intérêt pour les avis scientifiques et la gestion des pêches.
3. **Favoriser une feuille de route coordonnée** pour engager toutes les disciplines et parties prenantes pertinentes nécessaires à une intégration réussie des méthodes séquençage à haut débit dans l'évaluation et la gestion des stocks halieutiques. Cela s'appuie sur des initiatives de coordination et de coopération en matière de gestion de la pêche existantes dans l'ensemble de l'UE. L'un des principaux moteurs du succès de la feuille de route sera sa capacité à exploiter et

à étayer le potentiel des initiatives déjà existantes, en minimisant les nouvelles structures et en évitant les chevauchements et duplications inutiles.

4. **Le développement des capacités pour une mise en œuvre réussie** est essentiel dans deux domaines majeurs. Premièrement, les infrastructures à travers l'Europe doivent être renforcées et développées, en gardant toujours à l'esprit la priorité primordiale de maximiser l'utilisation des infrastructures existantes. Deuxièmement, il est très important d'améliorer les compétences, les connaissances et de renforcer l'engagement par le biais de réseaux de formation et de développer une communication étroite et une compréhension commune entre les scientifiques de différentes disciplines en utilisant un langage commun.
5. **Garantir l'optimisation des ressources** en vue d'évaluer et de démontrer plus systématiquement les avantages des méthodes, en rassemblant des données pertinentes pour évaluer l'efficacité des coûts et des investissements à travers les objectifs et les actions et ce, au-delà de l'évaluation des stocks. L'application des méthodes génomiques dans les évaluations des pêches est assez innovante et il existe peu d'informations disponibles sur les coûts et les besoins d'investissement découlant de cette application spécifique.

La feuille de route proposée est nécessairement complexe en raison du nombre d'actions requises, du nombre élevé et de la diversité des acteurs qui doivent être impliqués. Pourtant, elle définit les besoins de mise en œuvre dans le scénario le plus plausible ainsi qu'une définition précise des critères de sélection des études de cas. La **feuille de route** est conçue pour fournir des **outils innovants** à la communauté scientifique pour améliorer les **avis scientifiques** nécessaires à une **exploitation durable des ressources halieutiques**, mais aussi pour accroître nos connaissances sur le fonctionnement des écosystèmes et la biodiversité qui contribuent à une **meilleure gestion des écosystèmes**.

Pour finaliser ce résumé, il est remarquable que les résultats de l'étude à différentes étapes aient été soumis au jugement et aux commentaires d'experts externes. Par conséquent, les capacités du consortium ont été renforcées par une solide connaissance des technologies elles-mêmes, mais aussi de la manière dont les campagnes de recherche sont conçues et réalisées, et comment celles-ci alimentent l'évaluation des stocks de pêche, les avis scientifiques et la gestion, conformément aux réglementations et politiques établies.

## DELIVERABLES PRODUCED

Below is the list of the deliverables produced that are referred along this report. Click on the deliverable name to open or download the deliverable document.

| Code | Deliverable name |
|---|---|
| D1.1 | Fishery-independent data collection procedures |
| *D1.2* | *State of the art and critical review of the genomic methods* |
| D1.2a | State of the art review of genomics of Close Kin Mark-Recapture |
| D1.2b | State of the art review of environmental DNA genomics |
| D1.2c | State of the art review of age prediction in fishes using epigenetic clocks |
| *D1.3* | *State of the art and critical review of the bioinformatics tools* |
| D1.3a | State of the art review of bioinformatics analysis of CKMR |
| D1.3b | State of the art review of bioinformatics analysis of Environmental DNA |
| D1.3c | State of the art review of bioinformatics analysis for age prediction in fishes using epigenetic clocks |
| *D1.4* | *Integrated critical assessment of the methods* |
| D1.4a | The state of art in cost-benefit of HTS methods for stock assessment: an overview |
| D1.4b | Critical assessment of the potential of genomic methods to enhance fisheries stock assessment |
| D1.5 | Identification of barriers and risks, impact and mitigation of the implementation of genomic methods into fisheries stock assessment |
| D2.1 | Experimental design and protocols for conducting pilot studies to assess the implementation of genomic methods into fisheries research surveys |
| D2.2 | Implementation of genomic methods into fisheries stock assessment: a comparative analyses of pilot studies |
| D2.3 | Technical guidelines to integrate genomic-based approaches into fisheries data collection |
| D3.1 | Cost-efficiency of the application of High Throughput Sequencing (HTS) methods on fisheries research surveys and stock assessment |
| D3.2 | Implementation of genomic methods into fisheries stock assessment: A SWOT Analysis |
| D3.3 | Roadmap for the implementation of genomic-based approaches in fish stock assessment |
| D3.4 | The road ahead in fisheries science genomics: long-term prospects |

## INTRODUCTION

The status of marine fish stocks needs to be assessed to ensure fishing practices that exploit the stocks at sustainable levels. This assessment is based on multiple data types that include fishery-dependent data –catches, landings, biological information – as well fishery-independent data obtained from research surveys. This is a well-established traditional methodology serving to different policy and management purposes at the EU level and also internationally. Research surveys provide valuable and systematised information regarding exploited fish populations, marine biodiversity and their environment.

However, research surveys are confronted with a high economic cost coupled with complex logistics and a long time is required for treating and analysing the collected data (Stamatopoulos, 2002). These limitations produce sparse data in space and time and reduce accuracy in the estimation of key population parameters. In addition, regardless of the sampling method, conventional methods to determine biological parameters have some limitations, such as the inability to determine the sex structure of the younger juveniles or to provide information on connectivity, both of which are crucial for stock management. Survey design and technology are constantly progressing to cope with these limitations and to reduce uncertainty in stock assessment and scientific advice on the status of the harvested stocks and ecosystems. Thus, recent progress in the field of genomics is expected to improve efficiency and help to mitigate some of the shortcomings of traditional methodologies such as high costs and complex logistics.

DNA high-throughput sequencing (HTS) methodologies provide now the possibility to address some of these challenges more efficiently and complement traditional methods to assist fisheries management in the long term. Some recently developed methods could offer useful inputs to improve cost-efficiency with respect to traditional procedures, increasing the accuracy of fish stocks´ assessments and, eventually, reducing the time needed to carry those out. **The overall purpose of FishGenome was to evaluate the suitability of several novel HTS genomic techniques to estimate essential parameters for fisheries stock assessments. These parameters include absolute abundance, growth and biomass, stock boundaries and connectivity, fine-scale population structure and molecular sexing**. They have the potential to improve gradually stock assessment and fisheries management, but at present are not being considered (mostly) in evaluation. This reasons behind include the complexity of obtaining data, the effort needed to evolve the current system co-ordinately and without disruptions, and the still relatively low Technology Readiness Level[2] (TRL) of some of the techniques for their use in fisheries assessments.

**Box 1: DNA sequencing**

Refers to the technique for determining the exact sequence of nucleotides, or bases, in a DNA or RNA molecule. Sequencing information has traditionally been elucidated using a low throughput technique, until **high-throughput sequencing** (HTS) technologies were developed. These are capable of sequencing multiple DNA molecules in parallel, enabling hundreds of millions of DNA molecules to be sequenced at a time and lowering the cost of DNA sequencing.

FishGenome focused on three commercially important marine fishes with different levels of exploitation: two demersal species, Atlantic cod - *Gadus morhua* – and European hake - *Merluccius merluccius* –, as well as a coastal species targeted by small-scale fisheries, ballan wrasse - *Labrus bergylta*. The rationale behind this strategy was to cover a wide spectrum of reproductive modes that included gonochorism-hermaproditism, capital-income breeders and high-low productivity. Cod and hake are the most important demersal species caught by EU fleet in terms of landings, economic value, and food

[2] Technology Readiness Level (TRL) is a scale for measuring or indicating the maturity of a given technology. TRL 8 – system complete and qualified and TRL 9 – actual system proven in operational environment

consumption and both are assessed using bottom trawl surveys. Ballan wrasse is a target species for several activities, including small-scale fishery, recreational fishery, and alive extraction for use as a cleaner fish in the aquaculture industry. This species is one of the most important coastal species targeted in European waters, where it inhabits coastal areas difficult or impossible to access by bottom trawl surveys. Despite the suspected poor status of ballan wrasse stocks across European waters, no assessment is in place.

FishGenome covered three different regions – North Sea, North-West Iberian Peninsula, and Balearic Islands in the Mediterranean, targeting the following stocks (see Figure 1):

1) North Sea cod (*Gadus morhua*)
2) North Sea hake (*Merluccius merluccius*, Northern stock)
3) Galician shelf hake (*M. merluccius*, Southern stock)
4) Balearic Island hake (*M. merluccius*, GFCM Geographical Sub Area GSA-5)
5) Galician shelf ballan wrasse (*Labrus bergylta*)

Samples from the North-West Iberian Peninsula were obtained from a tissue biobank owned by the Consortium. Specimens from the North Sea and the Balearic Islands were collected during two bottom trawl research surveys, specifically the North Sea IBTS-Q3 survey and the Mediterranean MEDITS-GS5 Survey (Balearic Islands), that cover regularly these two ecosystems to evaluate the status of cod and hake, respectively. Additionally, environmental samples, consisting of water and sediment, were obtained during the surveys, to enable the analysis of environmental DNA (Figure 1).

**Figure 1. Figure illustrating the case studies in FishGenome. The project covered several stocks: in the North Sea, cod and hake (Northern stock) as well as environmental samples, consisting of water and sediment samples and represented by a drop of water in the figure. In the Galician shelf, hake (Southern stock) and ballan wrasse. In the Balearic Mediterranean, hake (Mediterranean stock, GFCM Geographical Sub Area 5) and environmental samples.**

Several novel genomic techniques were analysed within FishGenome, including Close Kin Mark Recapture (**CKMR**), epigenetic Age Determination by DNA methylation (**DNAm**), environmental DNA (**eDNA**) and Restriction Site Associated DNA Sequencing (**RAD-Seq**) for determining stock structure and connectivity, sub-structure and sex.

**CKMR** relies on identifying related individuals with different degrees of kinship using genetic markers and then analyse the number and pattern of pairs in a mark-recapture framework, which considers that an animal is "captured" if is present in a sample and that is "recaptured" if the sample also contains a close relative. The method can be used to estimate absolute abundance and other key demographic parameters – adult survival rate, size-specific fecundity, and selectivity (Bravington et al., 2015, 2016a). This method has only been proven on a large-scale in a handful of marine fishes characterized by very small population sizes and its potential applicability to other fish species has yet to be demonstrated. This tool was tested on cod and hake (three stocks, Northern, Southern and Balearic Islands).

**DNAm** offers a new alternative to traditional ageing techniques, such as otoliths and is based on the changes in DNA methylation (DNAm) that occur across the genome in response to aging (Heyn et al., 2012). In vertebrates, a gradual loss of methylation at some genes has been shown, but there is still little DNAm age data for wild fish. Its potential use as a biomarker for age for species of interest in fisheries remains to be assessed. We tested the accuracy, robustness, and power of this tool in cod.

**eDNA** uses the DNA that has been shed from organisms into their surrounding environment to harness information encoded in marine waters (Lodge et al. 2012; Taberlet et al. 2012; Bohmann et al. 2014; Creer and Seymour 2017). It offers an extremely powerful non-invasive method for estimating diversity and inventorying fish, but its capability of determining abundance and biomass is controversial. In FishGenome, we tested the potential of this tool to quantify abundance of cod in the North Sea and hake in the Balearic Islands, by analysing water and sediment collected in these two ecosystems.

**RAD-Seq** offers a potential tool to estimate accurately several parameters that could significantly improve stock assessments, but are not considered, due to the difficulty for their estimation. Such parameters include stock structure and connectivity, stock sub-structure and sex assignment in young individuals. All of them affect population persistence, productivity and response to exploitation. An accurate definition of stock structure is fundamental to define management units and stock boundaries, while the degree of connectivity among adjacent stocks affects recruitment. Information about finer scale population structure, or stock substructure, in harvested marine fish is essential since, if undetected, may lead to overfishing of local sub-stocks and a subsequent decline in biomass. All these parameters require regular and systematic monitoring, especially in the current context of global change, which is known to affect the distribution of marine fish species. We tested RAD-Seq to determine stock structure and connectivity in hake; sub-structure in all three species –hake, cod and ballan wrasse– and sex determination in hake. Determining the sex of individuals in any exploited species is essential for fisheries management. In species characterized by different growth rates between sexes, commercial catches often target exclusively one of them, having a direct impact on egg production and fertilization rates. Moreover, incorporating sex-specific information of juveniles in fisheries stock assessments is critical in understanding the structure and resiliency of fish stocks. The possibilities of this tool were tested in all three species (cod, hake and ballan wrasse).

FishGenome combined a comprehensive and systematic review of literature and other sources of relevant knowledge (grey literature, reports, and working documents) on these tools, with pilot studies that tested them in practice. The outcomes from these activities enabled **an integrated analysis of all the factors hampering or facilitating the implementation of genomic tools in fisheries assessments**. Such analysis allowed the partnership to devise and produce a roadmap and guidelines for the progressive implementation of the molecular tools in regular assessments, to

envision the future of stock evaluation accordingly, and to harness the advantages from the expected evolution of HTS techniques.

### PROJECT STRUCTURE

FishGenome responded to seven specific demands made by CINEA and DGMARE in the Tender Specifications for this study:

1. Providing a critical, state of the art, review of the genomic analyses on marine resources applying any of the three HTS methods.
2. Providing a critical, state of the art, review of the main bioinformatics tools required to analyse and understand the information delivered by the three DNA HTS methods.
3. Carry out at least two pilot studies to test all three genetic techniques and covering at least two fish stocks in at least the North Sea and the Mediterranean Sea.
4. Perform a SWOT analysis of the three DNA HTS methods and their related bioinformatics tools.
5. Perform a cost-benefit (in terms of efficiency and effectiveness) analysis comparing the DNA HTS methods to traditional research trawl surveys for demersal and benthic resources.
6. Identify which trawl-based research surveys and targeted stocks are the most likely to benefit from the genomic-based approaches.
7. Identify and comment as adequate other genomic-based approaches not listed in Task 1, which could be of high potential in monitoring and assessments of fish stocks.

The study work plan was aligned with a logical operational framework. It was aimed to deliver the best available knowledge to formulate advice on whether and how next-generation DNA-HTS methods can support improved stock assessment of marine resources. Following a multidisciplinary approach, it pursued the answers to the queries above. Therefore, the project work plan was structured mirroring the project conceptual framework and its overall objectives. This conceptual framework was based on: i) producing state of the art reviews of the knowledge concerned for setting the scene and producing the **state of the art and critical assessment of the methods** (WP1, addressing demands 1 and 2); ii) testing four HTS methods to estimate six parameters in several geographical areas through **pilot studies** (WP2, addressing demand 3), which results produced the data needed for posterior analyses of the feasibility of implementing HTS methods in the short and long terms; and iii) to learn about the cost-efficiency and efficacy of trawl-based research surveys for demersal and benthic stocks and stock assessments we conducted its **analysis and formulated a strategic planning** (WP3, addressing demands 4 to 7). Considering also coordination and management (WP4), the project was thus divided into four work packages, and 22 tasks, whose linkages are shown in the PERT diagram (Figure 2).

This report is divided in three main sections for each technical work package (WP).

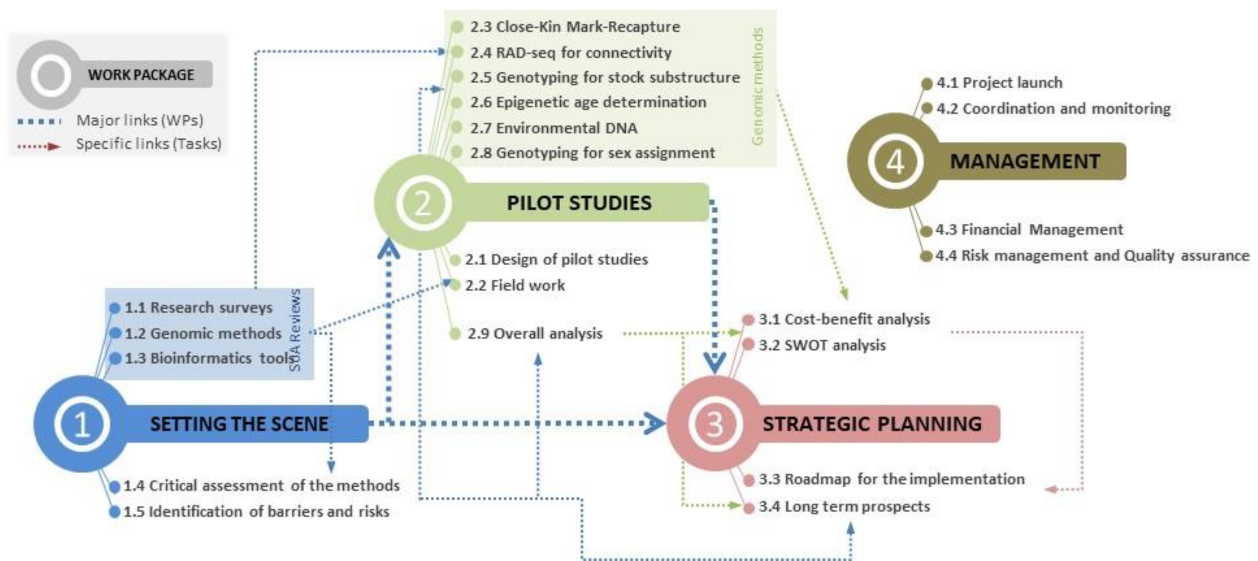



**Figure 2. PERT chart showing the connections among Work packages and tasks**

## FishGenome Consortium

FishGenome was carried out by a consortium made up of five organisations, namely the Spanish Council for Scientific Research (IIM-CSIC, ICM-CSIC and IEO-CSIC), the CETMAR Foundation, THÜNEN Institute and the University of Balearic Islands (UiB). IIM-CSIC acted as coordinator.

CETMAR is a public Foundation established in 2001 in Galicia. It specialises in research, innovation, and knowledge management activities in the marine realm, with a special focus on strengthening the links among the different stakeholder groups and reinforcing the research and innovation capacity of the blue economy sectors. CETMAR's contribution to this service contract has mostly concentrated in WPs 3 and 4. In WP3 CETMAR addressed the cost-benefit analysis with the support of the Fisheries Economy Group of the University of Santiago de Compostela; they also formulated the SWOT with the insights from the consortium and external experts and have also played a relevant role in the design of the FishGenome Roadmap. In WP4 CETMAR shared all the coordination tasks with IIM-CSIC.

The Spanish National Research Council (CSIC) is the largest public institution in Spain dedicated to scientific and technical research, as well as one of the most prominent in the European Research Area. CSIC has more than 120 Research Centres and Joint Research Units with universities or other research institutions.

The IIM-CSIC (Institute of Marine Research), located in Vigo, performs multidisciplinary marine research enabling a comprehensive and global understanding of marine ecosystems and the scientific and technological aspects of the fishery and aquaculture sectors. IIM-CSIC acted as the coordinator of this service contract. IIM-CSIC contributed to WP1 in the production of the State-of-the-Art Reviews of Fishery-independent data collection, Genomic methods and Bioinformatics tools. In WP2, IIM-CSIC contributed significantly to setting up the pilot studies and was in charge of implementing Close-kin Mark-recapture and RAD-Seq derived analysis (population, connectivity, substructure and sex analyses). In WP3, IIM-CSIC worked closely with CETMAR to elaborate the FishGenome Roadmap and was responsible for elaborating the long-term prospects of the genomic methods implemented in FishGenome. In WP4, IIM-CSIC worked hand in hand with CETMAR to perform the coordination tasks.

The ICM-CSIC (Institute of Marine Sciences), located in Barcelona, is the largest marine research institute in the Mediterranean Sea. The centre conducts frontier research and foster both knowledge and technology transfer on topics related to ocean and climate interactions, conservation and sustainable use of marine life and ecosystems, and impact mitigation of natural and anthropogenic hazards. ICM fosters research groups dedicated to marine physics, marine geology, marine biology, marine chemistry and technological development. In WP1, ICM-CSIC contributed to the State-of-the-Art Reviews of genomic methods and bioinformatics tools. In WP2, ICM-CSIC was in charge of carrying out the analyses for epigenetic age determination.

The IEO-CSIC (Spanish Institute of Oceanography) is a public institution dedicated to research in marine science, especially in relation to scientific knowledge of the oceans, sustainability of fishing resources and the marine environment. The IEO-CSIC's activities include representing Spain in international forums related to oceanography and fisheries as well as developing, coordinating, carrying out and managing research programs on marine resources. The Oceanographic Centre of the Balearic Islands is focused on research of marine ecosystems, their functioning and biodiversity, the ecology of the species on those ecosystems and their interactions with the environment and fisheries exploitation, from a multidisciplinary and integrated approach. IEO-CSIC contributed to WP1 in the production of the State-of-the-Art Review of Fishery-independent data collection. In WP2, IEO-CSIC was responsible for carrying out the field work related to the collection of samples from Mediterranean Sea that were used in the pilot studies.

The University of the Balearic Islands (UiB) was founded in 1978 and is located in Palma on the island of Majorca, Spain and it is composed of seven research institutes. The University is funded by the autonomous Government of the Balearic Islands. The UIB is an institution dedicated to the public service of higher education, research, knowledge transfer and innovation. At the UIB, there are almost 150 research groups, covering very diverse knowledge areas, from Social Sciences to Environmental Science and Technology. In WP2, UiB was involved in the field work related to the collection of samples in the Mediterranean Sea for the pilot studies and for analysing environmental DNA in collaboration with the Thünen Institut.

The Thünen Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries, is a German research institute under the auspices of the German Ministry of Food and Agriculture (BMEL). It consists of 15 institutes that carry out research within the fields of science, technology and socio-economy. The Thünen Institute of Sea fisheries focuses its activity on marine living resources, marine ecosystems, operational observation systems, marine spatial management and economic analyses related to fisheries. The Institute of Fisheries Ecology explores and monitors the marine environment to identify early changes and to assess their impacts on living resources. The research focuses on the sustainable use of marine resources, the preservation of genetic diversity in seas and inland waters and the analysis, evaluation and optimisation of aquaculture systems. In WP2, Thünen Institut was responsible for carrying out the sample collection in the North Sea for the pilot studies and for the analysis of environmental DNA in collaboration with the University of the Balearic Islands.

# 1. STATE OF THE ART AND CRITICAL ASSESSMENT OF THE METHODS

The first stage of the FishGenome project aimed at setting the scene for later conducting the pilot studies. It consisted of a comprehensive review of the trawl-based research surveys in EU waters, followed by a critical, state of the art (SoA) review of genomic analyses targeting marine resources and the main bioinformatics tools required to analyse and understand the information delivered by those techniques. We addressed each of the three main HTS methods tested: CKMR, DNAm and eDNA.

## 1.1. MAIN ACHIEVEMENTS

The main achievements of the first stage of the project are summarized below:

- ***A review of fishery-independent data collection procedures.*** A comprehensive review of the bottom trawl research surveys in EU waters was carried out. This review provided critical information to understand the diversity of approaches in the EU surveys within the context of stock or ecosystem assessment conducted in each region. It allowed us to understand methods and protocols, as well as targeted species and survey objectives used in each trawl-based fishing survey, to characterise the type of data collected, and determine those that could be, potentially, complemented/replaced by genomic approaches. The goal of this review was the identification of those research surveys and targeted stocks, which are most likely going to benefit from genomic-based approaches.
- ***State-of-the-art and critical review of the genomic methods and bioinformatics tools for their analysis***. A series of SoA reviews of the three main genomic methods considered in FishGenome (CKMR, DNAm and eDNA) and the bioinformatics tools required for their analysis were performed. The reviews analysed scientific literature, reports, patents, and grey literature, along with the Consortium´s own know-how and cutting-edge protocols, to describe thoroughly the characteristics of each genomic method, including the bioinformatics pathways needed to analyse genomic data, along with their advantages and limitations.

  The SoAs identified the best candidates among the biological parameters estimated in research surveys, to be addressed by each of the genomic method and evaluated its potential application in fisheries stock assessments.
- ***A State-of-the-art review on cost-efficiency of the use of genomics methods for stock assessments.*** This SoA consisted in/ a review of the existing literature on cost-efficiency of the application of the three genomic methods (i.e. CKMR, eDNA and DNAm) for providing biological data for stock assessments. This report includes the identification of the different steps and processes needed for the assessment of commercial fish stocks and a preliminary analysis of the costs (effort) of the methods in terms of such processes. These include sampling intensity, annual/seasonal replication needed, and ability to obtain samples as well as any other factors that are relevant about the current survey and stock assessment systems and plans. This task integrated the most relevant information from all three SoA reviews of the genomic methods.
- ***A critical assessment of the genomic methods and their potential implementation in research surveys.*** The objective of this report was to integrate and analyse the information from the above-mentioned state of art reviews regarding the potential of the genomic methods to produce equivalent or improved estimates of stock parameters, compared to those currently estimated in research surveys. This report compares the type and quality of the data obtained in research surveys and through genomic methods. The implementation of genomic methods in the research surveys are discussed, including the possible consequences of redesigning surveys to accommodate sampling for genomic methods, along with the barriers and risks (e.g., technical, economic, etc.) that have been identified in the FishGenome project.

- ***Identification of barriers and risks for the implementation of genomic methods in stock assessment.*** A virtual workshop, held on May 28th 2020, gathered a significant representation of experts from the genetics and fisheries assessment communities to discuss on the state-of-the-art and to identify barriers and advantages for the implementation of genomic methods in stock assessment. A total of 54 participants from 13 countries addressed each technique individually to identify application experiences, knowledge gaps, advantages, and drawbacks, as well as potential barriers for a widespread implementation. Convened experts included the FishGenome consortium and its External Experts Panel; the European Commission, the ICES Working Group on the Applications of Genetics on Fisheries and Aquaculture (WGAGFA); the fisheries assessment community engaged in EFARO network; and the EC DCF Regional Coordination Groups.

## 1.2. Work carried out

Below we provide a brief description of the methodology used to prepare the SoAs and critical assessments, followed by a summary of each document.

### 1.2.1. Fishery independent data collection procedures

#### Methodology

We first prepared a documented list of all bottom-trawl research surveys carried out across the EU, considering the mandatory surveys listed in Commission Decision 2021/1168, establishing the list of mandatory research surveys and thresholds for the purposes of the multiannual Union programme for the collection and management of data in the fisheries and aquaculture sectors (EU-MAP). A total of 14 bottom trawl surveys were selected for further review. However, international bottom trawl surveys such as BITS, NS-IBTS IBTS and MEDITS are integrated by several national surveys operating in coordination and under harmonized protocols, which is essential because they cover extensive areas. Thus, these national research surveys were considered as well.

Several sources of information were consulted, such as current legislation regarding the Data Collection Framework (DCF), the report of the EGW 19-05 on the Evaluation of mandatory surveys under the DCF (STECF, 2019), the Member States' Annual Work Plans (2018-2019), various reports of the International Council for the Exploration of the Seas (ICES) Working Groups and the Series of ICES Survey Protocols (SISP).

The final review considered, thus, the 14-mandatory bottom trawl research surveys in EU waters and allowed an in-depth understanding of the methods and protocols used in each trawl survey, as well as the targeted species and survey objectives. In addition to that, we were able to characterise the type of data collected, and determine those that can be, potentially, complemented/replaced by genomic approaches.

Moreover, the survey information from the review was analysed to identify key parameters. This analysis aimed at characterizing the different types of surveys based on their temporal and spatial coverage, the target species and assessed stocks. Key biological information of the main target species was identified and included in the review. A database with key parameters, such as area covered, year, period, effort (days), number of hauls and target species by country, was created to classify the surveys.

#### Results

A summary of the Deliverable 1.1, Fishery independent data collection procedures is presented below.

A research survey at sea is defined by the EU-MAP as the activities involving the monitoring of fish stocks and/or marine biological resources and the ecosystem, carried

out on a vessel dedicated to such scientific research and designated for this task by a Member State. The Member States regularly conduct research surveys of marine fish resources to provide fundamental data for assessing the condition of exploited fish stocks and for monitoring general conditions of the marine ecosystem. A number of these surveys are included in the DCF (Figure 3). Since 2014, these surveys have been financially supported by the European Maritime and Fisheries Fund (EMFF)[3]. Most of the surveys use protocols and methods that have been unified across Europe following standards and agreements. Yet, it is necessary to identify the existing differences in methodology, approaches and outputs among surveys. Thus, this review aimed at determining the main features of each survey, such as design and spatio-temporal coverage, target species and data collection and usage.

The scope of this review is restricted to the 14 bottom-trawl research surveys within the EU DCF (Commission Implementing Decision 2019/909[4]). These surveys are carried out in several regions (Figure 4). Specifically, the Baltic Sea, the North Sea, the North Atlantic (ICES Areas 5-14 and NAFO) and the Mediterranean and Black Sea. Thus, different organizations are responsible for coordinating and standardizing these surveys.

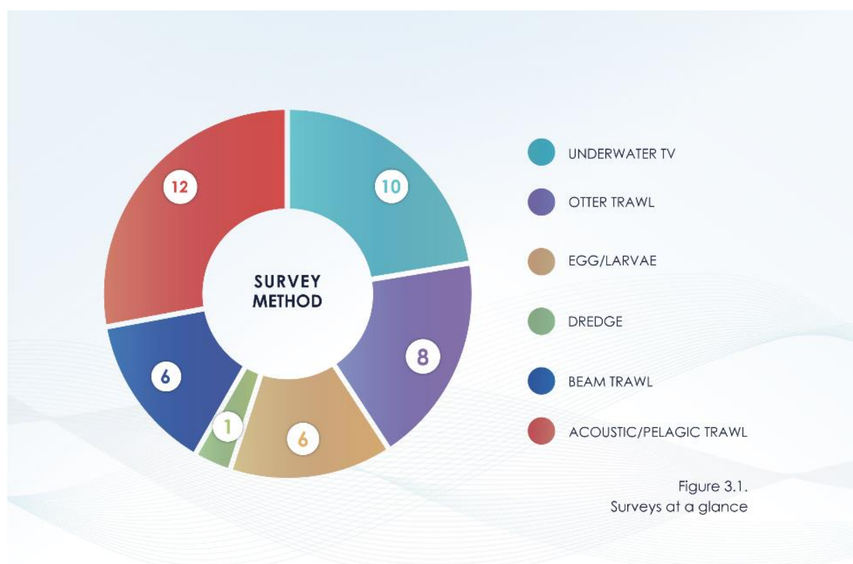



**Figure 3. Surveys at a glance: Surveys in Europe use different methods to obtain data for assessing the condition of exploited fish stocks and other marine resources. Numbers and types of surveys in this figure.**

This review clearly highlights that mandatory bottom trawl surveys are diverse. Primarily, this is a consequence of the different regions covered by each survey, which implies that different ecosystems with different species of marine organisms are surveyed. In addition to these differences, some surveys assess only a few species (for example, the Sole Net Survey in the North Sea, which targets 0-4 group sole, plaice and turbot) while others assess dozens of species (e.g. MEDITS or IBTS). For target species, biological parameters such as length, weight, sex, maturity, and age are determined in all surveys. For the rest of species, taxonomic identification is carried out and, in many cases, the length of the specimens is recorded. Additional information is collected on other biological components of the ecosystem (e.g. marine mammals, birds, benthic invertebrates and plankton) as input to an ecosystem approach to fisheries. Oceanographic data such as temperature and salinity are recorded as well, together with marine litter data.

[3] Since 2021 DCF is supported by the European Maritime, Fisheries and Aquaculture Fund (EMFAF)

[4] This review was conducted when the former COM Decision 2019/909 was still in force. The new COM Dec. 2021/1168 has increased the number of mandatory surveys to 51.

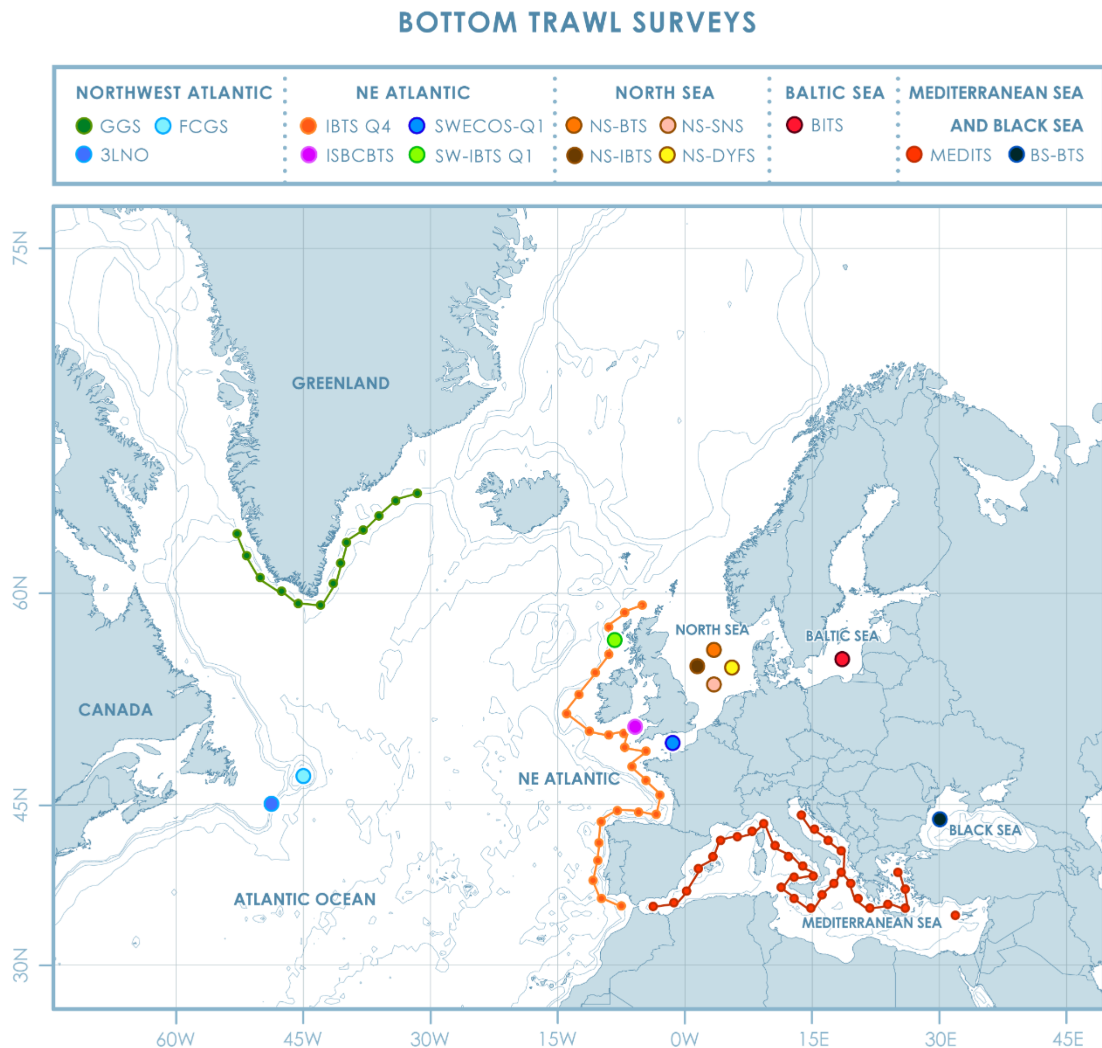



**Figure 4. Map showing the 14 mandatory bottom trawl surveys that were reviewed.**

According to methodological aspects, bottom trawl surveys can be divided based on fishing gear, survey duration and distance, and international participation.

Fishing gear. There are two types of fishing gears used in bottom trawl surveys: otter trawl (e.g. IBTS) and beam trawl surveys (e.g. NS-BTS)

Duration and distance. Surveys can be divided into in-shore surveys and off-shore surveys. In-shore surveys such as the DYFS and SNS in the North Sea cover small regions close to the coast. On the other hand, off-shore surveys are carried out further away from the coast and may cover extensive areas (e.g. IBTS and MEDITS). Among off-shore surveys, those that are carried out in NAFO areas can be considered long-distance (i.e. GGS, FCGS and 3LNO). Survey duration is very variable (Figure 5). For example, the Sole Net in-shore survey lasts for 8 or 9 days, while others like the IBTS or MEDITS last more than 250 days (when the effort from all participating countries is considered).

International participation. About half of the surveys are carried out by one or two countries but there are surveys, such as MEDITS or the NS-IBTS that involve many countries (Figure 6).

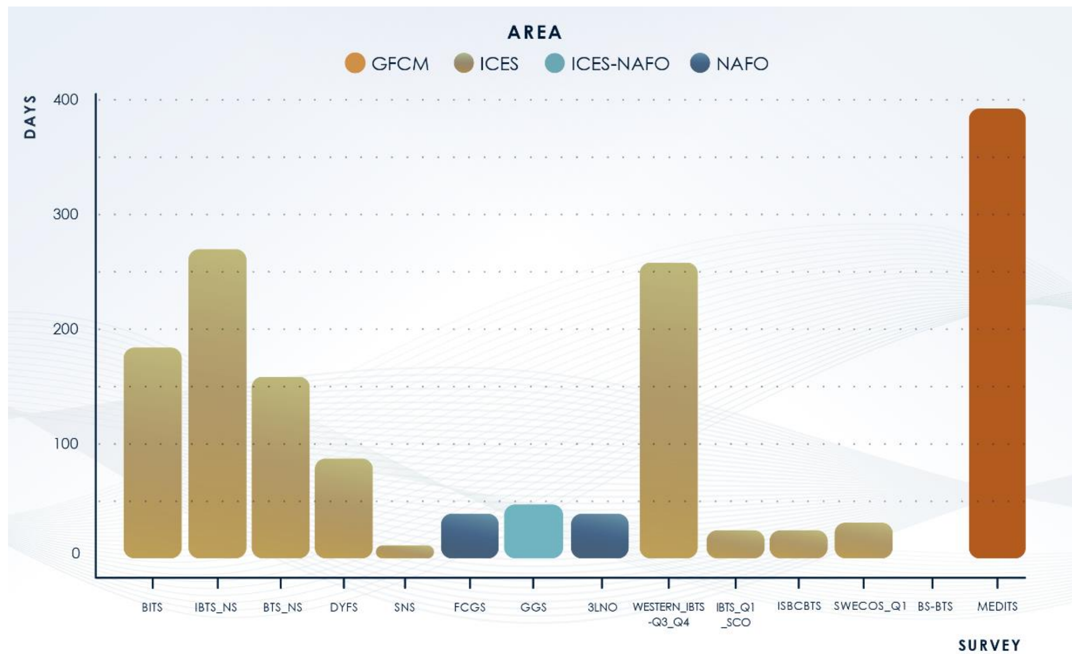



**Figure 5. Overview of survey effort (in days), for mandatory bottom trawl surveys, in the different regions. For each survey, the effort by all participating countries was included. Data from 2016-2018. No data were available for the Black Sea Beam trawl survey (BS-BTS).**

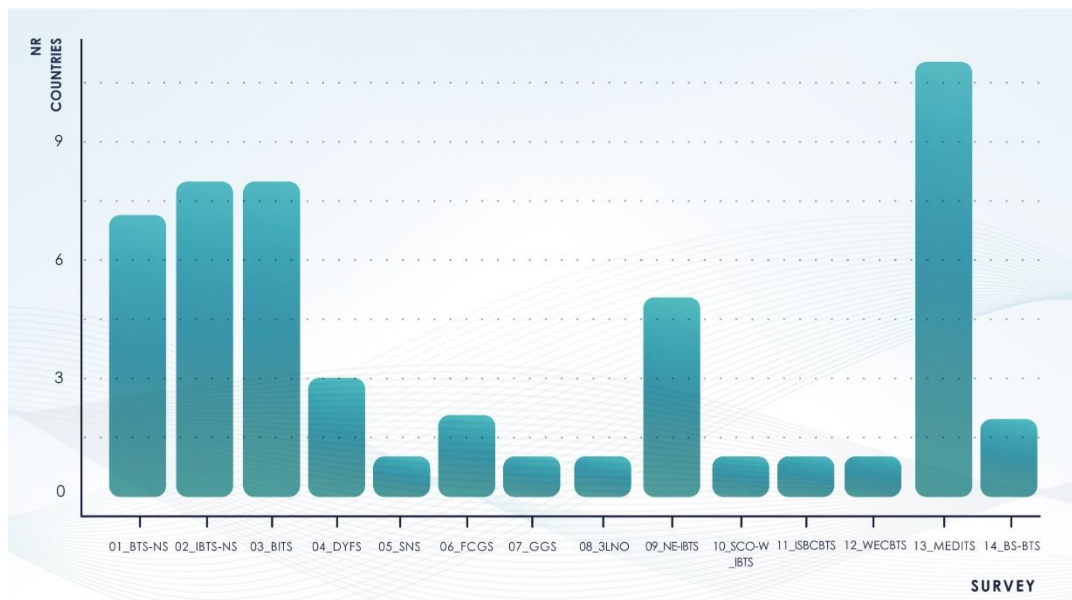



**Figure 6. Number of participating countries in each of the mandatory bottom trawl and beam trawl surveys.**

### 1.2.2. State-of-the-art and critical review of the genomic methods

#### Methodology

The State-of-the-art reviews were focused on the three main genomic methods considered in FishGenome, that is:

- Close-kin Mark-Recapture (CKMR)
- Epigenetic Age Determination by DNA methylation (DNAm)
- Environmental DNA (eDNA)

For the three genomic methods, a comprehensive, state of the art review (SoA) was carried out following a procedure consisting of several steps.

First, an exhaustive collection of bibliographic documents (scientific literature, reports, patents and grey literature) was used to produce preliminary SoAs that was shared and discussed with the project External Expert Panel (EEP, see Box 2 ) to improve the contents of the SoA. The experts critically reviewed the documents and presented their opinions and recommendations, followed by further discussions to decide the final content and the optimal structure and format presentation for the SoA reviews. Experts and project members' inputs were used to produce a consensus SoA report of CKMR, DNAm and eDNA genomics and bioinformatics methods and their potential application to fisheries assessments.

The last step involved the organization of a workshop, hosted by the Consortium and the EEP to gather feedback from other external experts, aimed at enhancing the consensus SoA reports. The workshop was originally planned as an in-person meeting, but due to the COVID-19 crisis, had to be held in a virtual format. Finally, the 1st Virtual Workshop was held on May 28, 2020, and titled *Fisheries research surveys and stock assessments using HTS genetic sequencing methods. State of the art, foreseen advantages and barriers for practical implementation*. A total of 54 participants from 13 countries from the genetics and fisheries assessment communities attended the workshop.

**Box 2: The External Expert Panel**

An External Experts Panel was created for bringing unpublished information and their know-how, assess outputs and outcomes produced by the Consortium and lead thorough discussions on the different subjects.

The panel consisted of five renowned experts on genomics and bioinformatics:

- Prof. Gary Carvahlo (Bangor University, UK)
- Prof. Lazslo Orban (Unversity of Pannnonia, Hungary)
- Dr. Allan Tucker (Brunel University London, UK)
- Dr. Julian Catchen (University of Illinois, USA)
- Dr. Sissel Jentoft (University of Oslo, Norway)

A set of documents was provided in advance so that participants could have access to the project and the SoA reports, to allow them preparing their contributions in advance:

- Workshop agenda;
- Briefing of the FishGenome project;
- Executive summary of SoA reports;

A facilitation team (Figure 7) was appointed to ensure a smooth running of the workshop (see chart below).

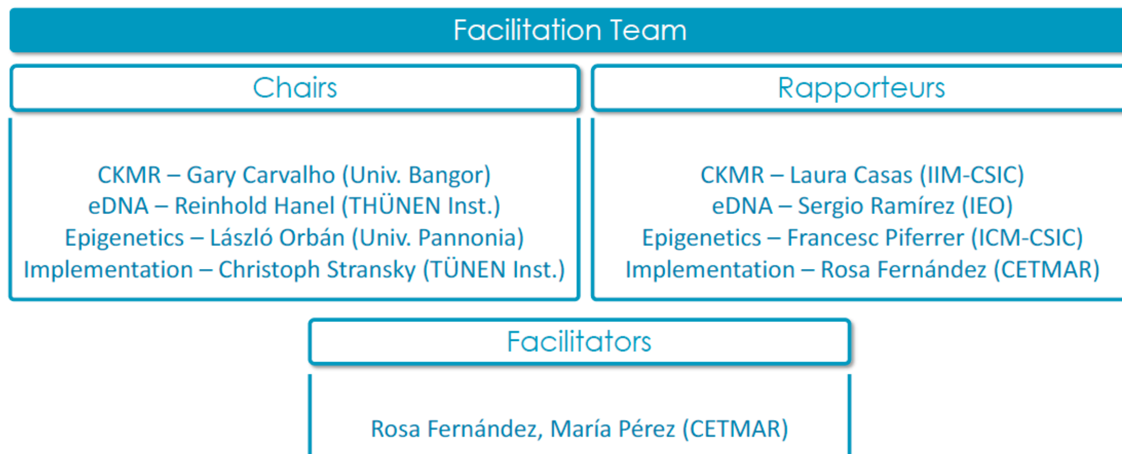



**Figure 7. Description of the facilitation team appointed by the FishGenome Consortium for the 1st Virtual workshop carried out within the project framework on the 28th May 2020.**

The workshop was organised in two parts:

Presentations of the Data Collection Framework, project briefing and state-of-the-art reports for each of the main techniques – CKMR, DNAm, eDNA, followed by a round of questions and debate on their contents;

Open debate aimed at highlighting pros and cons of each of the HTS techniques, identifying application experiences and knowledge gaps, as well as barriers and advantages of their implementation within the context of stock assessment and scientific advice for fisheries management.

All sessions were recorded for reporting purposes. A report of the Workshop was produced and shared with all participants for revision and additional contributions. Finally, new inputs from the workshop were integrated to produce the consolidated SoA reports.

### Results

Two SoAs were produced: Deliverable 1.2 State of the art and critical review of the genomics methods and Deliverable 1.3, State of the art and critical review of the bioinformatics tools. A summary of these deliverables by HTS technique is presented below:

### *Close-kin Mark-Recapture (CKMR)*

The advent of high-throughput genomic approaches has opened new possibilities for the two-decade old Close-Kin Mark-Recapture (CKMR) methodology (Nielsen et al., 2001; Skaug, 2001). In this new genomic framework, CKMR can potentially offer a direct method to estimate abundance and other demographic parameters of wild fish stocks. Using modern genetics is now possible to identify close relatives amongst large sample sizes of fish (Bravington et al., 2015). This information can be then used to make demographic inferences about the adult stock from the number and pattern of pairs found (Figure 8). Although CKMR has been already implemented in several applications, it is questionable whether the extensive sampling efforts and high costs associated to the analysis of large populations of mobile and dispersed species make this approach feasible for most commercially exploited fish (Casey et al., 2016).

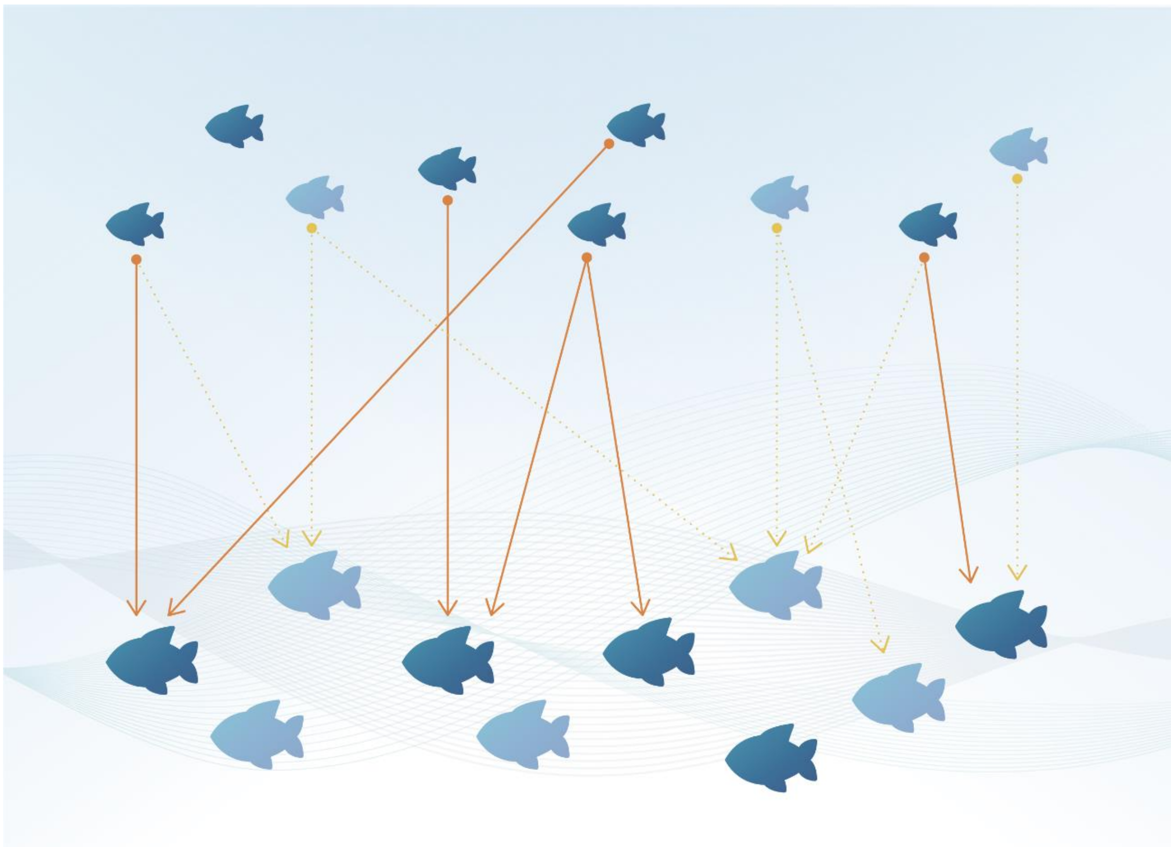


**Figure 8. Illustration of the CKMR principle; adults (big fish) and juveniles (small fish) are sampled (dark blue) from the total population (dark and light blue). Each juvenile "tags" two fish: each of its parents (solid and dashed lines) in the adult population; but only sampled fish provide us kinship information —POPs (solid lines). The absolute abundance of adults (10) can be estimated from the number of sampled adults and juveniles (5 and 6 respectively) and the number of POPs found (6) (figure redrawn from Bravington et al., 2016a).**

### *Considerations for the application of CKMR genomics methodology*

To apply CKMR successfully to a given population, a series of parameters need to be taken into consideration. A precise sampling strategy that accounts for patterns of social structure (random or non-random association of individuals) and habitat-use is essential. **Sample size** is critical since an accurate parameter estimation relies on obtaining reasonable rates of recapture (Mills et al., 2000). Thus, CKMR requires a number of samples that increases as the population size enlarges, highlighting one of the major limitations of CKMR that might prevent its application to large populations. Moreover, a solid understanding of the underlying population biology is advisable to make biologically reasonable assumptions according to the species under study. This constitutes another potential limitation for the implementation of CKMR in many exploited fish, since there is a shortage of biological knowledge about a vast number of species.

A reliable **genetic identification** of the kinship relationships among individuals is also crucial to ensure precise and accurate estimation of demographic parameters. Several marker types can be used for kin-pair finding, mainly microsatellites and single nucleotide polymorphisms (SNPs) (Ovenden et al., 2015). Although both have pros and cons for CKMR studies, SNP markers seems the sensible choice for any project initiated today, due to their higher accuracy and power, together with their lower rate of genotyping errors compared to microsatellites.

Similarly, different **sequencing technologies** can be applied to infer relatedness among individuals. Although the majority of CKMR studies published so far have used classical genotyping methods (e.g. microsatellites) (Ruzzante et al., 2019), their limitations compared to high-throughput sequencing (HTS) techniques make advisable to use the latter in future CKMR projects.

A combination of both, SNPs and HTS techniques, allow resolving more distant kinship relationships – parent-offspring-pairs (POPs) and half-sibling pairs (HSPs) – compared to microsatellites and classical methods (that might only allow the reliable inference of POPs), reducing the required sample sizes in CKMR studies.

The generation of the data sets for genetic identification of kinship pairs requires strict **quality control** steps for reliable identification of genetic relatedness among individuals, including the detection of null alleles or the removal of duplicate samples among others, which can lead to spurious parentage- and sibship- exclusions or inclusions, resulting in biased CKMR estimates.

Moreover, it is essential to diagnose the **quality of the CKMR assessment model** to detect possible sources of errors before validation and estimation of the key biological parameters of interest. Besides abundance and effective population sizes, CKMR studies via POPs and HSPs can be used to accurately calculate adult survival rates, provided that information of length/age-compositions is available. If only POPs are used, it is necessary to incorporate additional information on female daily fecundity to estimate statistically this parameter. Moreover, size-specific fecundity can be inferred by comparing the length distributions of identified mothers to that of adult females. Adult selectivity can complicate the estimation of all the previous parameters and should be considered to avoid skewed calculations.

Only four studies involving CKMR in aquatic animals have been published to date. All of them have focused on species with very small population sizes, except the first study, focused on Southern Bluefin tuna, a medium size population compared to most commercially exploited fish. Thousands of single-nucleotide polymorphisms (SNPs) are necessary to estimate heritability with the same accuracy, as when using pedigree relatedness (Gay et al., 2013; Bérénos et al., 2014). Among the three main categories of pedigree reconstruction methods - exclusion methods, relatedness-based methods and likelihood-based methods – the latter are preferred for CKMR analysis, due to their higher precision. However, it remains essential to assess confidence in kinship analysis, to avoid assignment errors that can bias population genetic statistics and ultimately lead to incorrect conclusions. Autosomal markers can aid the identification of false-positive kinship assignments (Koops et al., 2015).

Several HTS methods for SNP marker discovery and genotyping coexist these days. Among them, the most promising for CKMR studies - i.e., for the study of wild populations with no reference genomes available – are those that use **restriction enzymes**. They produce an unbiased set of markers distributed all throughout the genome that share a few common design characteristics. This group of HTS methods uses restriction enzyme digestion of target genomes to reduce their complexity and a sequencer to read DNA fragments. Polymorphisms (SNPs4) in the resulting sequenced fragments are used as genetic markers to infer relatedness. Single digested Restriction site-associated DNA sequencing (sdRAD-Seq) was one of the first methods developed and it has been widely applied in population genomic studies (Miller et al., 2007; Baird et al., 2008; Etter et al., 2012). Several derivatives have been published but the majority consist of only minor and subtle modifications of this parent protocol (Figure 9). With the notable exception of double-digested RAD-Seq (ddRAD-Seq), most variants have only been marginally used and tested, preventing their application in CKMR studies. Technical differences among the methods lead to important considerations for the types of bias and error inherent in the resulting data (Andrews et al., 2016) and these are much better understood in sequences generated by sdRAD-Seq and ddRAD-Seq techniques. Any of these two techniques should be the method of choice for CKMR studies in the near future, when analysing sample sizes of a few thousand individuals

or less. Larger sample sizes would benefit, in terms of costs and time, of first using one of the mentioned RAD-Seq methods for SNP discovery with a subset of samples, and then using a SNP chip for genotyping of the remaining samples.

A new generation of methods combining RAD-Seq and hybridization-capture technologies has emerged as a gateway genomic approach to enable the analysis of low-quality DNA samples. Although this could facilitate to a great extent obtaining samples for CKMR studies (from fish markets or museum specimens), our lack of understanding and control of sources of error advises against their use.

Poor DNA quality, together with low sequencing coverage, PCR duplicates and genotyping errors and allele dropout, and null alleles are among the main sources of problems for population genomics analyses. All these artefacts can produce genotyping errors, skewing allele frequency estimates and cause false positive alleles. To ensure reliable genomic data that can produce reliable estimations in CKMR studies, the first golden rule is to obtain high-quality DNA samples. It is also essential to achieve a 20X minimum depth coverage – the number of times a nucleotide is read during the sequencing process - to minimize sequencing errors (Rivera-Colón and Catchen, 2022). A high coverage increases confidence in the obtained sequences, as it aids in differentiating sequencing errors from real SNPs. Moreover, it is important to follow a well-established bioinformatics pipeline that identifies and removes potentially problematic sequences.

Although numerous bioinformatics software packages and workflows have been developed, Stacks v27 is undoubtedly the most complete and more widely used for marker discovery. Pairwise relationships prediction and kinship pedigree reconstruction should be performed using a maximum likelihood framework and possibly several different programs, such as ANGSD and polyRAD) to assess the confidence of kinship assignments.

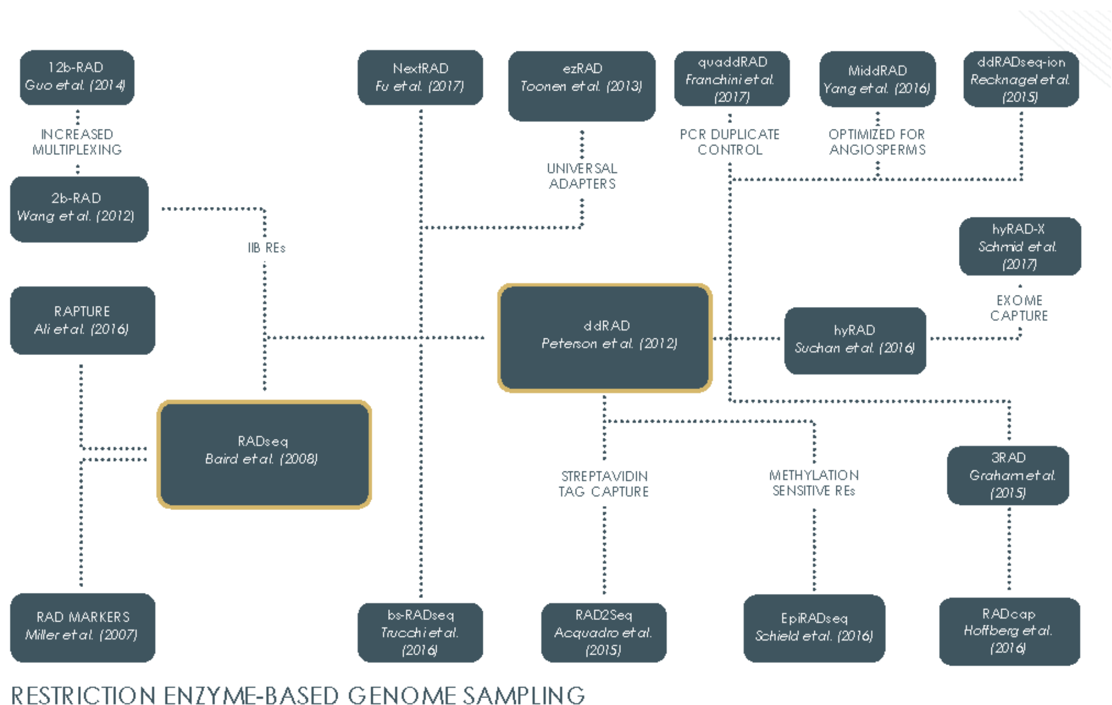



**Figure 9. RAD-Seq derivatives published in literature (figure adapted from Campbell et al., 2018).**

### 1st Virtual Workshop

The discussions held during the 1st Virtual Workshop (see methodology in page 25, above) highlighted the following questions:

1) Biomass estimation, i.e. whether CKMR can be used to estimate the maximum biomass that can be sustainably exploited, and used as biological reference point (BRP). Although CKMR is not actually providing the abundance, but the effective spawning stock biomass (i.e., those individuals that reproduce successfully, instead of the whole adult population), it could be used as a BRP, which is not the case so far. When estimating fishable biomass, this is reflecting the capacity of the population to reproduce within a year. As not all the fish reproduce, effectively, every year, recruitment is not coming from all the population, and CKMR can be a measure of the effective spawning stock. This parameter is currently not used in stock assessment and it can be a good proxy for the precautionary approach biomass.

2) A better estimation of effective spawning stock should contribute to improve stock-recruitment relationship (SRR). SRR is fundamental to estimate BRP, Maximum sustainable yield and stock projections, and hence key to provide sound scientific advice. The effective spawning stock biomass is actually the portion of the population that is reproducing and actually contributing to recruitment each year, which often is not the whole adult population.

3) The convenience and difficulties of implementing HTS tools in the existing methodologies, models and assessment practices arose at different points of the debate. It was highlighted that genetic data constitute an additional source of information to feed into those models that could improve, but not replace, current assessment models. Moreover, data delivered by traditional research surveys are still necessary.

4) Regarding practical aspects, the difficulties in establishing kinship were discussed. One of the workshop attendants explained the hurdles that were encountered by her research team in a CKMR study, which consisted in difficulties in detecting differences between parents, full siblings, half siblings and unrelated individuals, even when using a relatively large number of SNPs (600 SNPs chip derived from another study that included a large sampling). However, this seems to be due to the use of an insufficient number of SNP markers. Another attendant described a study carried out by his research group where the successful reconstruction of the real pedigree of 44 turbot families required the use of 18,000 SNPs. It is strongly recommended to use RAD-Seq for selecting SNPs in a subset of individuals and then replace this technique with SNP chips to lower the cost and complexity of the analysis. The cost of RAD-Seq in the previous study was about 30 euro/sample for 1,000 SNPs, and 60 euro/sample for 25,000 SNPs. By contrast, of the use of a single SNP chip with markers of different species multiplexed can further diminish the cost to about 10-15 euro/sample for 10,000 SNPs/species. Although RAD-Seq has been useful, especially for non-model species with no genomic resources available, i.e., no genomic information in the public databases, having an assembled genome for the species of interest is always an advantage. Nowadays, it is possible to assemble reliable genomes at a reasonable cost (6,000€ to 9,000€, although this figure would be higher for large size genomes). Thus, an alternative to RAD-Seq could be to obtain whole genomes for a subset of individuals for about 6000 euro/individual (although this, again, would clearly vary according to the species). This information would be then utilized to select SNP panels to develop high-density SNP chips at a cost of 10-15 euro. This could be an efficient approach if using the same set of SNPs for many years for fish stocks assessment, since whole genome sequencing offers far more information than RAD-Seq.

   The selection of SNPs should be performed very carefully, since SNPs specific to the individuals surveyed can become obsolete in the years to come and the exploratory

approach would need to be repeated over the years. For example, SNPs alleles – the alternative form (A,G,T,C) underlying the polymorphism – that are under strong selection can reach fixation over time, causing the loss of the polymorphism.

5) Another issue discussed was the way connectivity and substructure affect CKMR studies. Whilst there is no problem in panmictic or very strongly structured populations, for those that are weakly separated there should be a way to account immigrants. This is a key issue that also affects traditional surveys and stock assessment and it is precisely the reason why genotyping for substructure and the connectivity tools using the same technique (RAD-Seq) were included in the FishGenome project, so the same samples can be used to combine these elements and provide an approach to this parameter.

The discussion during the 1st Virtual Workshop regarding CKMR bioinformatics showed that *Stacks* has been the software of choice to analyse the RAD-Seq reads among the workshop attendees. On the other hand, SNP selection and kinship analysis have been addressed using either self-written code in R and hand calculations or a specific software developed by Mark Bravington that is not publicly available.

### *Epigenetic Age determination (DNAm)*

In fisheries management, accurate estimates of age are essential to infer life-history traits and for effective stock assessment. Knowledge about the age composition of fish populations provides information about stock structure, age at maturity, life span, mortality and their growth rate (Pardo et al., 2013). Further, inaccurate determination of age is also likely to lead to errors in the estimation of stocks assessment models input data, including catch and stock weights-at-age, maturity-at-age and any age-structured catch per unit effort (CPUE) indices, generating uncertainty in assessment and scientific advice.

Age estimation in fishes has traditionally relied on the analysis of growth marks in hard structures such as scales and otoliths. These methods require well-trained personnel, are time-consuming, and have low accuracy in some species, including some exploited fish of high commercial value (e.g. hake). As fish populations continue to decline globally, due to exploitation, it is imperative to further validate these methods and develop new ageing techniques.

DNA methylation is a chemical modification of the DNA without change in the nucleotide sequence whereby the 5′ carbon atom of cytosine is replaced by a methyl group, becoming 5′-methylcytosine (5mC). DNA methylation is influenced by a variety of external factors such as diet, stress and environmental cues, as well as internal factors such as sex, tissue and age (Jung & Pfeifer., 2015). In general, with increasing age there is a progressive genomic hypomethylation (Heyn et al., 2012). However, aside this *epigenetic drift*[5] there are DNA methylation changes that are of a *clock-like* nature (Paoli-Iseppi et al., 2019). Chronological age predictors built based on DNA methylation are termed *epigenetic clocks* (Zhang et al., 2019) and are based on a carefully selected group of loci across the genome, the methylation of which is linked with chronological age (Figure 10). These clocks build on the fact that aging is associated with changes in DNA methylation in specific cytosine-guanine (CpG) loci.

[5] Epigenetics includes modifications to histone proteins, noncoding RNAs, and DNA methylation. In this context, epigenetic drift is the alteration of epigenetic patterns during aging.

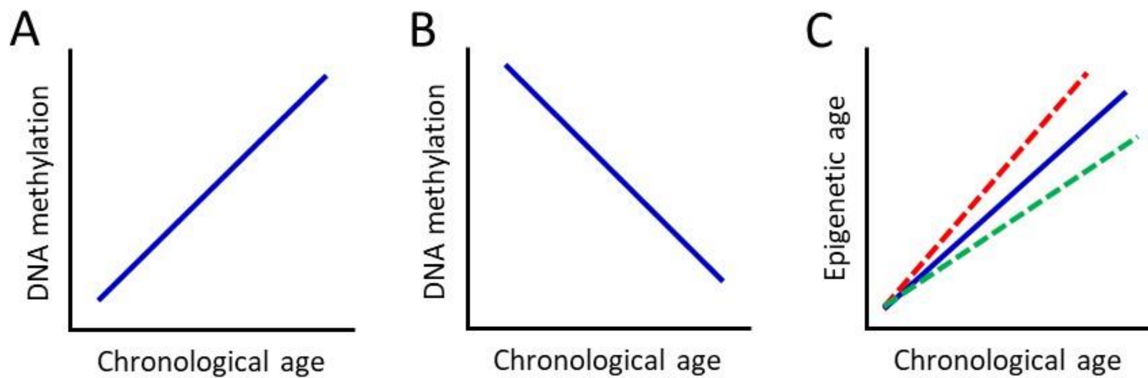



**Figure 10. Epigenetic clocks are based on finding loci the DNA methylation of which increases (A) or decreases (B) with age. A combination of several of them allows clock building (C). Intrinsic or extrinsic factors may accelerate (red line) or slow down (green line) epigenetic age.**

### *Considerations for the application of the DNA methylation genomics methodology*

The first epigenetic clock was developed in humans in 2013 and ever since, a few epigenetic clocks were published, mostly in mammals, but their levels of accuracy are highly variable, as reflected in Table 1.

The potential for developing epigenetic clocks in fish species is still underexplored. The first attempt to develop a piscine epigenetic clock was carried out in the European sea bass (*Dicentrarchus labrax*) by a member of the FishGenome Consortium (Anastasiadi and Piferrer, 2020). Although its performance compares favourably with that of the previous clocks developed in other vertebrates, it is unknown whether the technique might provide a highly accurate tool for measuring biological age in other fish species.

**Table 1. Epigenetic clocks for age prediction in vertebrates**

| Species | Tissue | Method | Initial CpGs | Final CpGs | Accuracy | Precision | Reference |
|---|---|---|---|---|---|---|---|
| ***Homo sapiens (M)*** | Multi-tissue | Microarrays | 21369 | 353 | 0.96 | 3.60 | Horvath (2013) |
| ***Megaptera novaeangliae*** | Skin | Pyroseq. | 37 | 3 | 0.79 | 2.99 | Polanowski et al (2014) |
| ***Mus musculus (M)*** | Blood | Pyroseq. | 71 | 3 | 0.95 | 0.10 | Han et al (2018) |
| ***Pan troglodytes (M)*** | Blood | Pyroseq. | 14 | 4 | 0.73 | 5.43 | Ito et al (2018) |
| ***Myotis bechsteinii (M)*** | Wing | Pyroseq. | 7 | 7 | 0.58 | 2.08 | Wright et al (2018) |
| ***Canis familiaris (M)*** | Blood | RRBS | 252,240 | 41 | 1.00 | 0.05 | Thompson et al (2017) |
| ***Canis lupus (M)*** | Blood | RRBS | 252,240 | 67 | 0.97 | 0.04 | Thompson et al (2017) |
| ***Ardenna tenuirostris (A)*** | Blood | DREAM | 2338 | 7 | 0.59 | 2.81 | Paoli-Iseppi et al (2019) |
| ***Dicentrarchus labrax (F)*** | Muscle | MBS | 48 | 16 | 0.82 | 2.15 | Anastasiadi & Piferrer (2020) |

Abbreviations: RRBS, Reduced Representation Bisulfite Sequencing; DREAM, Digital Restriction Enzyme Analysis of Methylation; MBS, Multiplex Bisulfite Sequencing; M: mammal; A: avian, F: fish.

In principle, it should be possible to build an epigenetic clock for any fish species, since clock-like DNA methylation changes in some loci seem to be a conserved feature in all vertebrate genomes (Jung and Pfeifer, 2015). Ideally, epigenetic clocks should work well in both short and long-lived species. Since DNA methylation can be influenced by both genetics and the environment, clock construction should take this into account. For clock development, all age classes should be targeted with a sufficient number of individuals per class. Since the construction of modern epigenetic clocks involves the use of machine learning procedures, large training sets produce a more accurate prediction of the chronological age, so large sample sizes should be used whenever possible. Data to be collected during sampling would need to include weight, length and sex of the specimens. The tissue of choice for clock development can be the fin clip as it is easy to obtain and is already used for many genetic studies, although other tissues (liver, muscle, etc.) could be used. Dissected tissues can be cold-stored in ethanol. DNA of high quality should be obtained for downstream clock development.

For actual clock development, a non-targeted approach is preferred. The possibility of finding loci with methylation patterns that are not only strongly correlated with age but also conserved across many species should be considered, because this could enable the development of a multi-species clock. We propose a three-step approach, consisting of: 1) the use of a whole-genome or genome-wide methods to assess DNA methylation levels of hundreds or thousands of candidate sites, 2) identify candidate loci highly correlated with age, prioritizing those located in genomic regions that are known to have a function in aging. Test the identified loci using targeted approaches in a large number of independent samples, 3) develop a targeted assay based on the selected loci that can be automated with the possible lowest cost per sample. The ultimate goal would be to have a multi-species array for age determination in fish that could provide reliable results at low cost and able to process a large number of samples coming from research surveys.

As a measuring device, an epigenetic clock should have characteristics common to all measuring devices that are indicative of their reliability. These should be well-defined and include:

- Accuracy (months/years): this parameter refers to how close is the measured age to the biological age. Bias (if any) should be known and little as possible. Differences between estimated and true values can be corrected by calibration (e.g., by using a different method of measure such as otoliths).
- Precision (months/years): this parameter refers to repeated measures on the same sample. Error should be estimated. Repeatability (same measure on the same sample) and reproducibility (same measure taken with different method) are inherent parts of precision and should also be considered.
- Intra-assay and inter-assay errors: these parameters can be calculated with repeated measures of a set of samples at once (coefficient of variation) or at different times (standard deviation).
- Resolution: this parameter needs to be determined. It is related to precision and refers to the minimum difference (in age) that the clock can detect.
- Sensitivity: is the minimum value (age) that can be detected. This is likely to be a challenge in the context of fisheries/conservation. For epigenetic clocks developed in vertebrates so far, age determination around one year seems difficult.

Expertise in epigenetics, machine learning methods and fisheries management are essential to bringing the development of epigenetic clock for fisheries applications closer to success. Developing piscine epigenetic clocks for target species could have a major impact, since it will likely provide an accurate method for age assessment in fish and circumvent the limitations of the current methods. Advances in techniques aimed at measuring DNA methylation will make it possible to estimate age in large sample sizes

at a very low cost. Challenges to be resolved include whether there are specific loci with age-related methylation changes that are conserved across species to facilitate the development of multi-species epigenetic clocks. Another important aspect is to determine how changes in the environment may affect the tick rate[6] of piscine epigenetic clocks. Epigenetic age estimation could contribute to stock assessment and fisheries management in a significant manner in the years to come.

### *Critical review of the bioinformatics tools related to DNAm*

Epigenetic clocks to predict the age of animals have been constructed on the basis of DNA methylation levels analysed. The methods used to analyse DNA methylation can be categorized using three broad levels (Anastasiadi, 2016; Barros-Silva, Marques, Henrique, & Jerónimo, 2018). At level 1, methylated loci are identified either by: a) use of restriction enzymes with graded sensitivity to the methylation status of the cytosines (Cs), b) use of antibodies that show affinity with specific methylation status, or c) use of the properties of bisulfite that converts unmethylated Cs to thymines (Ts) but leaves intact the methylated Cs. At level 2, the resolution in the information obtained is categorized as: a) low, b) medium or c) high resolution, depending on whether it provides information on global methylation or at single nucleotide resolution. Level 3 refers to what portion of the genome is targeted, and it can be: a) locus-specific, b) genome-wide or c) whole genome. Next generation sequencing (NGS) approaches that make use of bisulfite at single nucleotide resolution at a genome-wide or whole genome basis are ideal for building epigenetic clocks. Thus, these techniques can target a genome-wide part, like the Reduced Representation Bisulfite Sequencing (RRBS) or the Bisulfite-converted Restriction site Associated DNA sequencing (bs-RAD-Seq) or locus-specific parts such as the BisPCR2 or the Multiplex Bisulfite Sequencing (MBS).

All NGS data produced from methods that use bisulfite conversion share some common characteristics. A summary workflow includes steps of quality controls, filtering/trimming, alignment/mapping, methylation extraction and analysis (Figure 11A).

The processing of NGS data always starts with the appropriate quality controls of the raw sequencing data, followed by filtering of the data that fall below the specified thresholds. The sequences of adapters or indices usually added during the preparation of the libraries are removed. Then, the reads are aligned against the genome which, importantly, must have been previously bisulfite converted using specific kits. These steps consist of a series of chemical reactions that enables the differentiation between methylated and unmethylated cytosines (C). Then, the information of the methylation status has to be extracted at each C position of the genome, a process called methylation extraction or methylation calling. This output is adequate for subsequent statistical or bioinformatics analyses.

The data on methylation extraction are used with the objective to identify CpGs correlated with the age of the individual, i.e., CpGs with methylation levels that show a linear decrease or increase with age. The methylation of each correlated CpG will be given a specific weight (coefficient) to allow their combined use. Such combination boost the power of the individual CpGs to predict age. The dataset consists of biological samples that cover a defined age range for which other variables such as weight, length and sex have also been obtained. The application of genome-wide methods to explore methylation levels, usually renders hundreds of correlated CpGs per sample. This is a large multivariate dataset, where the number of variables (the different CpGs) is much higher than the number of samples (biological samples). A way to circumvent the structure of these types of datasets is to apply penalized regressions (Kassambara, 2018). The penalization occurs via the addition of a constraint in the equation (Bruce & Bruce, 2017) (Figure 11B). The methodology to achieve this is the shrinkage or

[6] The rate of change in DNA methylation at age-dependent specific sites represents the ticking rate of the epigenetic clock

regularization, which results in the shrinkage of some coefficients values to zero. This allows to retain the minimum number of CpGs that provide valuable information for age prediction. The three most commonly used methods of penalized regression are the ridge, lasso and elastic net, and typically all three are tested when constructing an epigenetic age prediction clock for a new species.

Penalized regression models are built as any other machine learning models. Thus, the typical workflow of building would include (Figure 11C): 1) data preparation and pre-processing, 2) data splitting into training, testing and potentially validation sets, 3) variable selection, 4) evaluation of the model, and 5) assembling the predictions. At the end of the process, the final model is built and can be used to predict data and thus, for epigenetic age prediction (Rauschert et al., 2020).

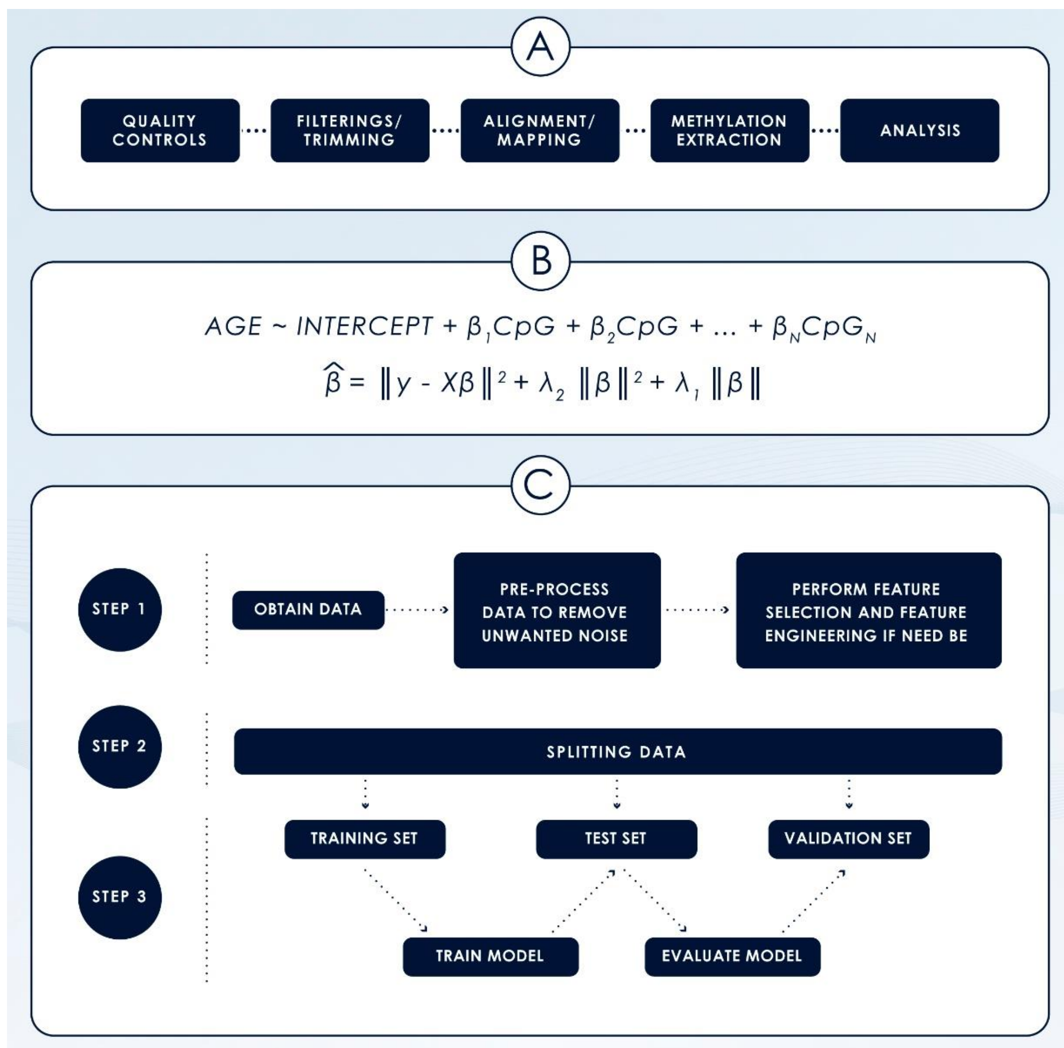



**Figure 11. Bioinformatics pipeline for epigenetic age determination. A) Workflow for bioinformatics analysis of bisulfite sequencing data. B) Equation for penalized regression analysis where λ2=0; λ1= λ ~ Ridge Regression and λ1=0; λ2= ~ LASSO. The equation on top shows the cost function which penalized regressions try to minimize. The values in front of the CpGs (β1, β2… βN) show the amount of change in age as a response to change of methylation. C) Workflow for applying a machine learning algorithm.**

### *1st Virtual Workshop*

The discussions held during the 1st Virtual Workshop highlighted the following questions:

Robustness of the method in terms of potential environmental effects was the first issue under discussion, particularly the potential effect of contamination in methylation levels. However, only the influence of temperature on young fish has been tested and the results showed no effect on age prediction (Anastasiadi and Piferrer, 2020). Although lab and field experiments in model organisms have shown an effect of contaminants on DNA methylation, experimental data on from wild animal populations are largely lacking. Yet, modifications are not expected in such a generalized manner that there would be no CpG loci left for clock construction. Therefore, this is not considered a major issue for building a good epigenetic clock, even though it needs to be verified experimentally.

Some apparent contradiction is perceived from the double role of epigenetics as a detection tool of environmental effects and age determination. It was explained that two different sets of CpG loci are in play: environmentally sensitive ones which can be used for the former purpose, whereas the non-sensitive ones are suitable for epigenetic clock construction.

Another question raised had to do with the suitability of this kind of analysis depending on the age of the individuals used to build the clock. Thus, the European seabass clock was built with an over-representation of young individuals, and was agreed that an epigenetic clock can be used only to the age range of the fish used to build it.

Although accuracy is an essential issue for all the methods addressed in FishGenome, it becomes crucial in the case of epigenetics. Age estimation needs to be highly precise, as uncertainty in age determination is magnified when abundance-at-age is estimated with important consequences in the stock assessment quality. When considering wild populations, epigenetic clocks can be built with a precision of <1 year, thus able to identify year classes. Therefore, there is potential to become an alternative to traditional methods. For most of the relevant species in Europe 1-year precision is sufficient for stock assessment purposes, less than that is irrelevant.

The relevance of having a good quality reference genome arose again related to the epigenetic clock, as two of the three species targeted by the FishGenome project are yet to have a high-quality reference genome. This is an advantage, but not a requirement as RAD-Seq provides enough data to find the best CpGs. Moreover, draft versions of the hake and ballan wrasse genomes are already available.

The minimum number of loci tested is also a matter of discussion, as some studies cited in the SoA review had extremely low numbers (e.g. three for mice). Theoretically, there is no minimal number. Although it is tempting to think the more the better, what really matters is the strength of the relationship between the level of DNA methylation and age for that particular locus. In theory, one CpG very strongly correlated with age could be enough. In the vertebrate epigenetic clocks published so far, there is no relationship between the number of final informative CpG loci making the clock and the accuracy or precision of the clock. However, it can be assumed that a very robust clock can be obtained with carefully selected 10-50 CpG.

It is important to start with as many CpGs per sample as possible. Ideally, whole-genome bisulfite sequencing would be the best approach, as it would yield millions of CpGs to choose from. But bis-RAD-Seq approach is a good trade-off to balance the number of CpGs to start with, with the amount of labour and the cost per sample.

The minimum setup costs for analysing the methylome to find the best CpGs for a single species is estimated around 25.000–30.000 euros, but it depends on the technique used and the hours needed for multiplexing, library construction and data analysis. The cost of testing an epigenetic clock on different stocks will depend on the stock size. For a 100 individuals' stock it would amount from 10.000 to 15.000 euros (without labour costs) at current prices, before the clock is scaled, to reduce the cost per sample. Regarding the specific case of European seabass, the cost per otolith is 15-20

euro/sample, while the target for a successful epigenetic clock would be around 10 euro/sample. This can be considered a realistic price if a sufficient number of samples is analysed, as scaling up the method to produce a custom microarray-based chip requires a minimum volume of a few thousand samples (Harrison and Parle-McDermott, 2011). With sequencing prices going down constantly, eventually the cost per sample would not be an issue.

With regard to effort in terms of time, number of stocks, number of specimens, etc. in relation to the testing of the potential effects of environmental conditions on the clock, experiments in the lab would need to last for the same amount of years for which the clock is used. This would be a highly time-consuming process. In addition, it is debatable whether if they would bring realistic information to match the actual situation in nature because, in such a different environment, other variables could influence the clock's tick rate. If done by samplings in the field, extreme years could be chosen (e.g., very cold or very hot), but a good alternative would be to sample the same species along a thermal cline (e.g. North-South), comparing populations living in different water temperatures.

This would apply just to temperature. Besides, it is necessary to clarify: (i) how many species are aged each year; (ii) how many individuals in total; and (iii) how much is the current cost and time per sample. There are no numbers available on how many fish are aged nowadays in the world in fisheries research, but an estimate from 2009 put the figure to "well over a million" each year.

The discussion during the 1st Virtual Workshop regarding DNAm bioinformatics suggested that bis-RAD-Seq data should be used to extract population data structure and thus, to compare clock performance across populations. In this regard, contacting the Earth Biogenome Project (https://www.earthbiogenome.org/) and suggesting them to add FishGenome targeted species to the list of those scheduled for sequencing could be a good approach. The main problem to deal with in bioinformatics is related to small sample sizes and very high dimensionality, which is not new in this field. Although some epigenetic clocks have been built manually, for most of them penalized regression, e.g., elastic net regression, has been used.

### *Environmental DNA (eDNA)*

In recent years, eDNA coupled with metabarcoding methodologies has emerged as a promising tool for rapid, non-invasive and cost-efficient biodiversity monitoring with enormous potential to inform aquatic conservation and management (Bohmann et al., 2014). Studies based on eDNA analyse complex mixtures of genomic DNA from many organisms isolated from environmental samples, without requiring access to the target animals (Lodge et al., 2012; Taberlet et al., 2012). In general, the eDNA approach involves a series of steps, which include eDNA capture, preservation, extraction, amplification and sequencing to ensure match to target species (Figure 12).

Because of its versatility, eDNA has been applied in many environments, including water, soil, faeces, pollen, or air (Ficetola et al., 2008). Nevertheless, there are unique challenges associated with using this tool in marine environments. Extreme water-volume to biomass ratio, the effects of sea currents and waves on dispersion and dilution of eDNA, and the impact of salinity on the preservation and extraction of eDNA can influence capture and detection (Thomsen et al., 2012).

An appropriate use of the technique in the marine realm requires a better understanding of the mechanisms that influence eDNA presence and concentration through time and space, mainly in four domains (Harrison et al., 2019): a) origin (the sources of eDNA), b) state (the physical form of eDNA, e.g., dissolved, particle adsorbed, intracellular, and organellar), c) transport (the way DNA moves in the environment) and fate (how eDNA degrades). Moreover, a robust interpretation of eDNA patterns requires the consideration of several parameters that characterize its surrounding environment (currents, temperature, etc.) (Turner et al., 2015). The DNA in environmental samples

is typically highly degraded into fragments limiting the scope of eDNA studies, as often only small segments of genetic material remain. Mitochondrial DNA is typically targeted because of greater number of copies, compared to nuclear DNA, and it is known to be highly effective for identifying organisms to the species level (Rees et al., 2014).

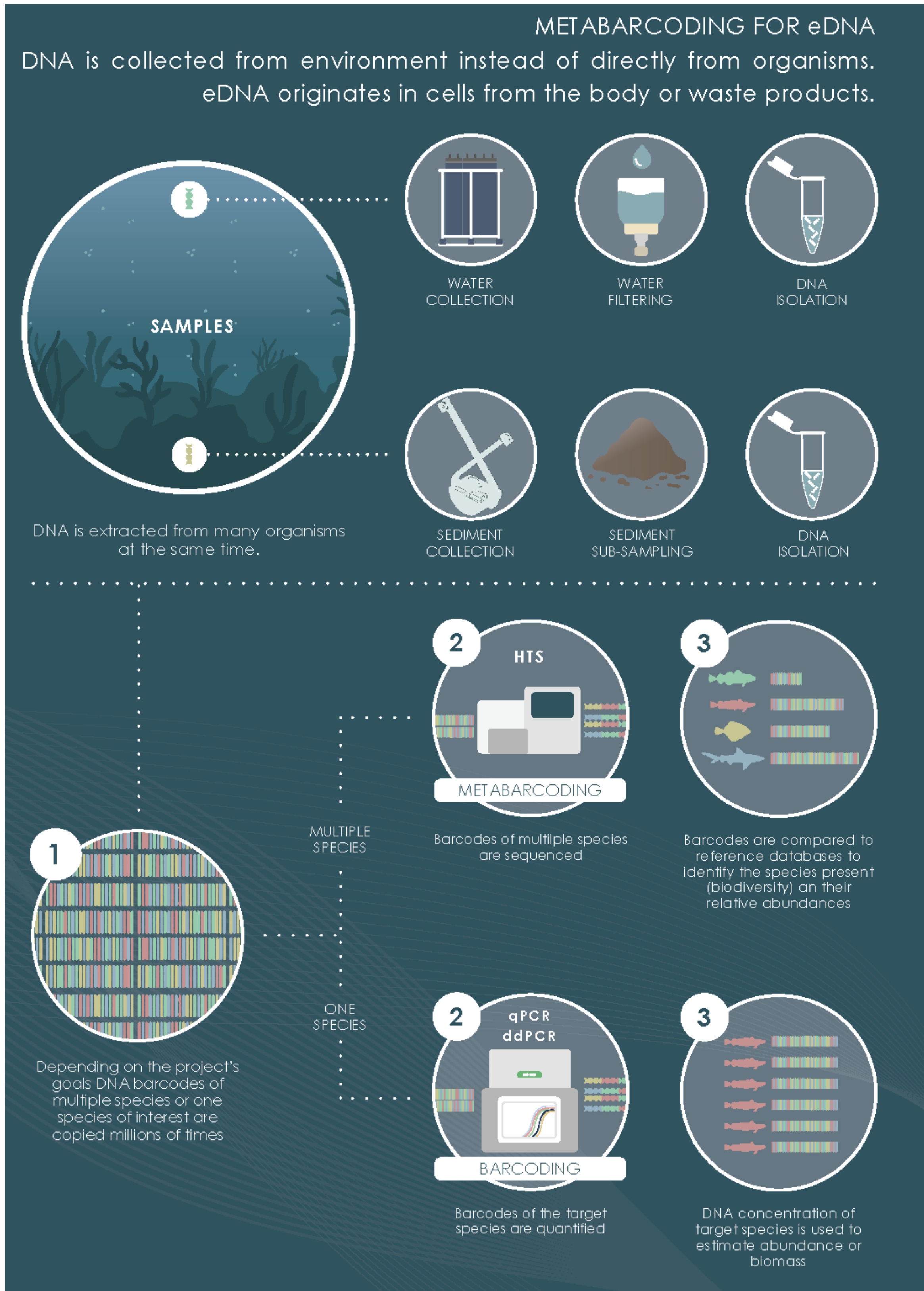



**Figure 12. Overview and general workflow for environmental DNA studies.**

Concentration of eDNA is dependent on biomass, age and feeding activity of the organisms as well as their physiology, life history and spatial behaviour (Barnes et al., 2014; Goldberg et al., 2016). A crucial step in the eDNA workflow is DNA capture. Several studies have focused on sampling design optimization, capture and extraction methods (e.g. Turner et al., 2014; Deiner et al., 2015, Eichmiller et al., 2016). Collection methods typically have sought to identify organisms at low densities and, thus, should be optimized for detection sensitivity. There are multiple capture protocols in the literature that were developed for different types of samples (e.g. Turner et al. 2014; Deiner et al. 2015), making the selection of a protocol a difficult task that needs to be carefully considered, depending on the goals of the study and the type of sample being analysed. eDNA extraction protocols that are being optimized within the frame of FISHGENOME project target three main applications: single species detection, estimation of abundance and biomass of target species, and biodiversity assessments.

*Considerations for the application of eDNA genomics methodology*

The FishGenome project reviewed two methods to analyse the eDNA: **High-Throughput Sequencing** (HTS) for biodiversity assessment, and **quantitative Polymerase Chain Reaction** (qPCR) for the quantification of a target species. For the first method, both universal and species-specific primers may be used, depending on the goal of the study. The power of detection is determined by the affinity to the targeted taxa sequences and the availability of DNA reference collection databases needed for species identification. HTS is mostly used to detect multiple species and for biodiversity assessment. On the other hand, qPCR is widely used for gene expression analysis, due to its large dynamic range – i.e., the high tolerance to input DNA amount (maximum and minimum DNA amount limits) that preserves the precision and robustness of the method, tremendous sensitivity, high sequence specificity, simplicity of the analysis, and sample throughput, i.e., the volume of samples that can be analysed per day (Lodge et al., 2012). This method is usually employed for species detection and implies the use of species-specific primer sets, allowing the quantification of the target species DNA, which has been shown to correlate with species abundance and biomass in the environment (Lodge et al. 2012; Thomsen et al. 2012).

Since eDNA is a sensitive method, there are many potential sources of "errors". Some of these errors are associated to collection, laboratory and bioinformatics procedures and include contamination, inhibition, amplification and sequencing errors, computational artefacts and inaccurate taxonomic assignment (Thomsen et al. 2016; Barnes and Turner 2016). From these errors, the most serious is probably the risk of contamination and hence the possibility of false positive results. The use and sensitivity of HTS has further complicated the contamination issue, as they produce a very high throughput of DNA sequences (Ficetola et al. 2016). Therefore, understanding the potential sources of errors and translating these into methodological protocols and interpretations of the results is crucial for obtaining reliable outcomes.

Along these lines, eDNA offers a potential method to revolutionize marine biomonitoring by significantly augmenting spatial and temporal biological monitoring in aquatic ecosystems, due to the ease of collecting water samples (Thomsen and Willerslev, 2015; Sassoubre et al. 2016). eDNA has also the potential to advance fisheries monitoring and conservation by improving the detection-probabilities for the rare fishes that often comprise a large proportion of the total species richness found in species assemblages. The non-invasive nature of eDNA analysis may provide advantages over traditional catch-based sampling, by making it possible to determine the presence or absence of species without disturbing the fish or their environment. This approach could be particularly beneficial in situations of endangered species, where there is significant risk of injury to the fishes or damage to their critical habitat (Evans and Lamberti, 2018). More investigations are required to understand how well the eDNA method will work for aquatic species, to evaluate the effect of species abundance on detection efficiency and to upscale species detection from local water samples to larger spatial areas, such as drainage basins.

### *Critical review of the bioinformatics tools related to eDNA*

Combined with fieldwork, laboratory procedures, and molecular tools, bioinformatics and computational analysis are important to perform an adequate analysis using eDNA samples. In eDNA metabarcoding method, eDNA samples are sequenced using HTS producing thousands or millions of raw DNA sequence reads (sequencing libraries) that must be processed in a standardized way in order to answer the initial question or hypothesis. This massive amount of data requires multiple computationally intensive steps to produce an appropriate analysis (Figure 13). However, to date, there is no single universal processing workflow that provides a unified and streamlined manner for satisfactorily treating eDNA data from raw sequences to taxonomic identification and diversity analysis. On the contrary, there are many bioinformatics pipelines that have been separately developed and are being used and improved by the eDNA research community. Moreover, mainly due to the emergence of novel technologies, the bioinformatics considerations are constantly evolving and protocols must be constantly adapted.

The main goal of eDNA HTS data processing is to generate reliable data that can provide the building blocks to answer ecological and environmental questions, starting from the raw sequences and most commonly involving the comparison of taxonomic diversity among samples from different environments and/or conditions. Results taken from eDNA metabarcoding data must be interpreted with caution, given that some taxa could be present in the final dataset by erroneous assignments due to contaminations, mistagging, or PCR and sequencing errors (false positives); and some other taxa can remain undetected, due to partial sampling, DNA extraction, PCR amplification, or HTS bias (false negatives). In this sense, an adequate analysis of any eDNA metabarcoding experiment should include the following steps/bioinformatics considerations:

1. **Quality control and pre-processing of the raw data**: HTS technologies could produce "sequencing errors" causing point mutations – changes in a single base pair - and chimeric fragments –artefacts that derive from the fusion of similar DNA sequences that belong to different genomic locations, and incorrect base calling – inference of the order of nucleotides of a DNA sequence. These errors affect the resulting read composition in different ways depending on the sequencing technology and can contribute to overestimation/underestimation of sample diversity. This step is key to perform a proper subsequent analysis for eDNA metabarcoding.
2. **Clustering:** This step produces a simplified but comprehensive list of unique sequences grouped by common attributes that ideally cannot be further subdivided. To do this, sequencing can be classified using reference genetic databases that in some cases allow the classification of sequences by species. Public genomic databases are built with the input of researchers from all over the world, and despite their exponential increase over the last decades, they are still incomplete and contain some errors. Although imperfect, they have proved very useful for the characterization of marine fish biodiversity. Similar sequences based on a similarity threshold are retained as a unique representative sequence named molecular Operational Taxonomic Unit (OTU), which theoretically represents the same species. In cases for which no sufficient annotation and database information exists, it might only be possible to group sequences by nucleotide similarity, using clustering methods and infer the genus, family or order of the analysed specimen. One solution to overcome this problem is the construction of a private database where the sequences, species labelling and geographic origin are carefully verified; however, it is a costly and laborious task. Proper sequence classification is necessary to minimize the inflation of biodiversity estimates caused by intraspecific and interspecific polymorphism (Alberdi et al. 2017) or errors produced during the PCR amplification and sequencing (not removed during the quality control step) (Taberlet et al. 2018).

3. **Taxonomic assignment**: Here, OTUs are assigned within a taxonomic group. Sequence database searches often contain spurious results that need to be filtered to keep only the set of matches that are most plausible. The quality and completion of the public reference databases for taxonomic assignment is crucial for an eDNA metabarcoding study. This step frequently involves text and table processing methods using custom descriptors such as the percentage identity, the alignment length, query coverage, and many other options that are according to the goals of each study. This assignment reduces the search results into a higher confidence subset that can be further analysed.

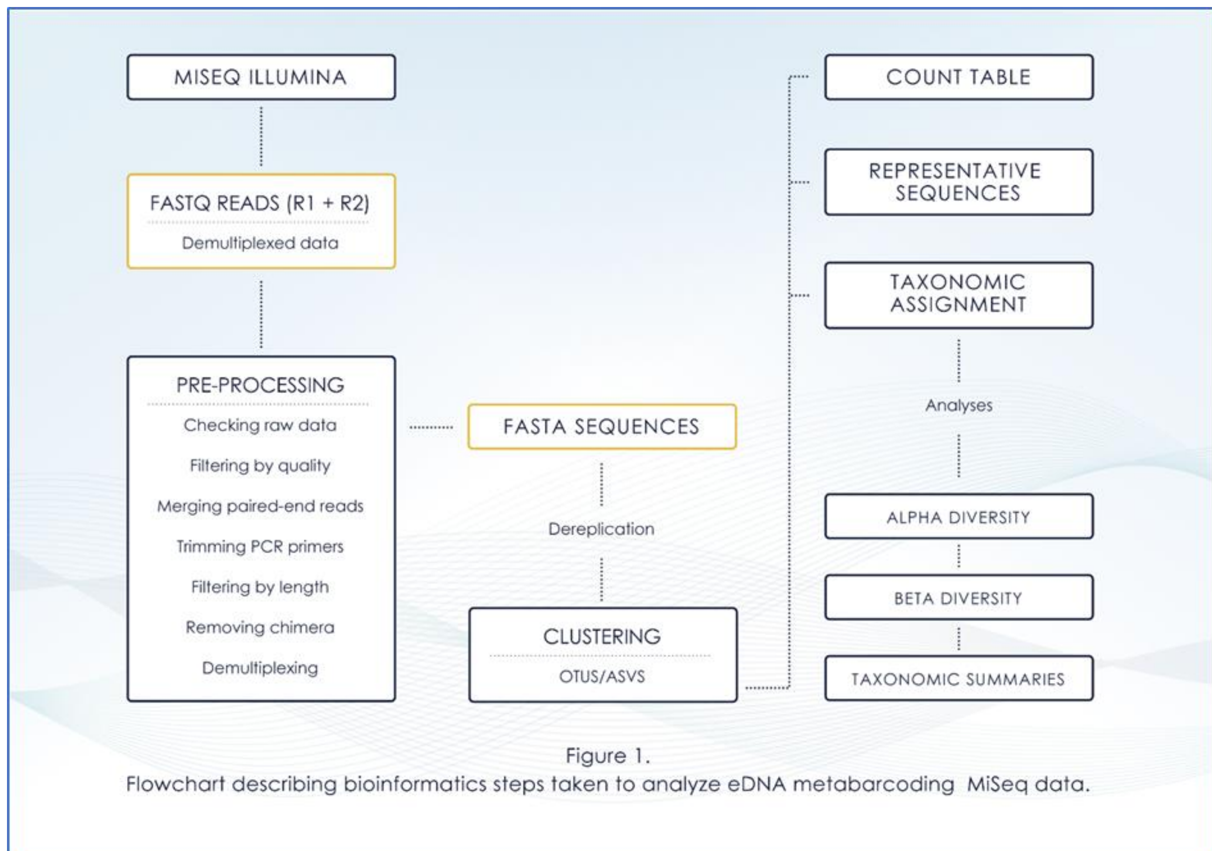



**Figure 13. Flowchart describing bioinformatics steps for the analyses of eDNA metabarcoding.**

Many tools are available to process the aforementioned bioinformatics steps. Most of these tools are integrated into bigger packages, such as MOTHUR or Qiime2, arguably the two most used in eDNA studies and selected for the analysis within the frame of FishGenome.

On the other hand, the quantitative Polymerase Chain Reaction (qPCR) tool does not require a specialized bioinformatics handling. qPCR is a technique capable of detecting and quantifying tiny amounts of DNA present in a sample by contrasting the data obtained to those of a standard curve. Although several studies have found a positive relationship between eDNA concentration and abundance or biomass, there is still substantial variability surrounding this relationship. In particular, the correlation between density and eDNA in flowing water remains unclear due to contrasting results. The variation in the relationship between eDNA and density could be due to the differences in the movement and retention of eDNA in the systems. This correlation is performed by a qPCR standard curve that is a simple statistical procedure, where the quantification cycle values are plotted versus DNA concentration for different dilutions of a DNA sample of known concentration. This produces a linear relationship between quantification cycle and the logarithm of the initial amount of total template DNA. It is

useful to calibrate the qPCR and interpolate the data from our samples of unknown concentrations. It is the most significant parameter that can be obtained from a qPCR experiment and inversely correlates with the amount of amplifiable target that is present in the original sample.

Additionally, there is another technique called digital-droplet PCR (ddPCR) that also allows for the quantification of minute amounts of DNA. This technique does not need a standard curve for quantification, making quantification possible even when a standard sample is not available. Obtaining target eDNA concentration values from a ddPCR experiment is rather straightforward, as the proprietary software from each ddPCR machine manufacturer will directly estimate that parameter. The data that this system generates is gathered from a binary signal that, after applying a Poisson correction to consider droplets with more than one molecule, can be used to directly count the number of targets eDNA molecules in the original sample.

To date, there is no clear consensus about the correlation of the number of molecules estimated by qPCR or ddPCR and the actual abundance or biomass of fish in a given sample. However, this relationship needs to be tuned up for each study if we seek to completely pivot away from traditional capture techniques into molecular methods.

#### *1st Virtual Workshop*

The discussions held during the 1st Virtual Workshop highlighted the following questions:

The first issue with regard to eDNA was if, in terms of stock assessment, it would make more sense to focus on a few target species or be more ambitious and follow a metagenomics approach. Experience on the application of metabarcoding for the assessment of biodiversity in the Bay of Biscay using anchovy surveys (Fraija-Fernandez et al., 2020) allowed the identification of some patterns, despite the difficulty to infer spatial distributions from eDNA at sea. It is hard to determine the current presence of certain species or if they were there before, and how long ago, or if the DNA has been transferred there by some other means (e.g.: effluent of a nearby industry). In general, the most abundant species take over most sequencing reads, so deeper sequencing might be necessary to identify the less abundant ones, similarly to what happens in trawling: the more is captured, the more likely to find less abundant species. In this regard, alternative approaches, such as using blocking primers to avoid amplification of the most abundant species were mentioned, but it is complicated by the shortness of the barcode and has yielded no success for the moment. However, the same sample can be sequenced again to get enough sequences.

The experience from the Bay of Biscay delivered the expected results, about 70-80% of the reads corresponded to anchovy and sardine, in accordance with the surveyed area and the time of the year. Patterns related to shallow and deeper water species were also coherent with expectations, as well as the spatial distribution.

Another issue, especially in estuaries, has to do with the effect of tides. Tidal bore getting in twice a day entails a force that brings the sediment up, so eDNA that was bound in the sediment gets again to the surface. Often, but not necessarily, related to this, is the finding of DNA of species that clearly do not correspond to the surveyed area, such as tuna in very shallow estuaries or cod upstream in rivers. The origin of this genetic material cannot be easily determined, and its impact can usually be corrected and buffered with repetition and a proper design of the sampling.

The poor quality of existing public databases was also pointed out as a problem, since they contain significant errors that can mislead the analyses interpretations. For this reason, building a specific database, containing the species inhabiting the surveyed environment, can be recommended as an alternative. This has been the strategy adopted by AZTI researchers. They built their own database, they curated it and periodically update it.

Comparing sediment and water samples from the same station was also a topic of interest. A summary of available studies regarding this matter (e.g. controlled

environments in canals; studies for macroinvertebrates, microbes and bacteria) was discussed. A higher number of fish species was identified from sediment compared to water, but this extended biodiversity did not represent the species inhabiting the area according to electrofishing survey data. It was concluded that sediment usually represents biodiversity more spread in time, while water could be more representative of the present moment.

Regarding the possibility of eDNA use to estimate biomass abundance, it seems that further research is necessary to calibrate the tool. Once the quantity of DNA is determined, it is necessary to calibrate such amount with the true biomass, which can depend on many factors. However, some relationships have been found in metabarcoding, indicating that qPCR could work even better for individual species, considering to the superiority of the second method in terms of robustness, specificity and precision. If the objective is to quantify one species and a continuous monitoring is performed, a qPCR might be able to produce relevant information (there are previous studies that have shown the capacity of the method to detect peaks of presence of a specie of interest). It is still not known how close we are from the objective of real biomass quantification that could be integrated in the assessment models using metabarcoding.

Beyond the measurement of species presence and distribution, the implementation of eDNA could provide information on the co-occurrence of species and relationships through ecological networks, that could be related to environmental parameters. This would enhance ecological quantifiable information that is of broader interest for biodiversity management and assessment.

An additional aspect that can greatly benefit from eDNA based research is the study of inaccessible or very difficult to reach environments, such as deep sea.

The discussion during the 1st Virtual Workshop regarding eDNA bioinformatics showed that read clustering is not considered necessary when doing a taxonomic assignment. It does depend on the kind of analysis performed, but for the identification at the species level, clustering will lead to missing information and to the need to make assumptions. Thus, in this case, clustering is not recommended if a good reference database is available.

### 1.2.3. Critical assessment of the methods

#### Methodology

This task was divided into two subtasks:

In the first subtask, a review of the state of the art on the cost-benefit/cost-efficiency of the application of genomic methods for providing biological data for stock assessments was carried out. Based on the requirements of the FishGenome project, this review focused on three relevant genomic methods: 1) close-kin mark-recapture (CKMR) 2) epigenetics for age determination based on DNA methylation (DNAm) and; 3) environmental DNA (eDNA). The results of this review were presented in Deliverable 1.4a, *The state of art in cost-benefit of HTS methods for stock assessment: an overview*.

The accepted view in scientific literature of genomics studies applied to fisheries assumes that HTS genomics methods' application will be cost-effective, faster and more efficient than the traditional survey methods used in fish stocks' assessments. However, the underlying research question of whether the genomic techniques are indeed cost-effective compared to traditional survey techniques or not, still remains.

The SoA review of the existing literature on HTS methods was performed through the Snowballing systematic reviewing approach, i.e., the references cited in the papers found through the search where are also systematically reviewed. The search of literature shown in this report has been conducted using Google Scholar and Thomson

Reuters' Web of Science. A search on these academic platforms was performed between 15 April to 20th of June 2019 using the following core concepts and terms: i) Next Generation Sequencing (NGS); ii) Epigenetic age determination method (NGS1); iii) Environmental DNA studies (NGS2); iv) Close-kin mark-recapture studies (NGS3); v) Cost-effectiveness; vi) NGS1, NGS2, NGS3 combined with "Cod", "Hake", "Wrasse"; vii) NGS1, NGS2, NGS3 combined with "North Sea", "North-West Iberian Peninsula", "Balearic Islands", "Mediterranean"; viii) NGS1, NGS2, NGS3 combined with "Trawl", "Trawlers" and "Demersal"; ix) Fisheries research surveys /Traditional Surveys and, finally; x) Fish stocks assessments.

The second subtask was to integrate into a single analysis the critical assessments of the current potential of the genomic methods (i.e., including bioinformatics) to produce equivalent or improved estimates of stock parameters, currently estimated in research surveys, and which allow an enhanced stock assessment. The results of the integrated analysis were presented in Deliverable 1.4b, *Critical assessment of the current potential of the genomic methods for its use in stock assessment*.
The critical assessment focuses on the implementation of HTS genomic methods in research surveys for their use in stock assessment. It critically analyses deliverables 1.1 (Fishery-independent data collection procedures), 1.2 (State of the art and critical review of the genomic methods), 1.3. (State of the art and critical review of the bioinformatics tools), 1.4a (The state of art in cost-benefit of HTS methods for stock assessment) and 1.5 (Identification of barriers and risks, impact and mitigation).
For this integrated analysis, the first step was to compare the type and quality of the data that can be obtained in traditional research surveys and through HTS genomic methods. This was done by: i) determining the degree of equivalence between traditional methods and HTS methods in their capacity to provide data for stock assessments and ii) determining the advantages and disadvantages of both approaches. Then, several aspects regarding the implementation of HTS methods in the traditional research surveys are discussed, including the possible consequences of restructuring surveys to accommodate HTS methods in the sampling programs, as well as and the barriers (e.g., technical, economic, etc.) that have been identified in the FishGenome project. These barriers have been summarized in deliverable 1.5, *Identification of barriers and risks, impact and mitigation*.

### Results

Following is a summary of the results of the cost-benefit/cost-efficiency of the application of genomic methods for providing biological data for stock assessments (Deliverable 1.4a):

- HTS methods have been claimed to be cost-efficient, nevertheless, very few publications have systematically and accurately addressed the issue (Rodríguez-Rodríguez et al., 2022).
- Most cases that claim cost-efficiency are not referred to stock assessments but to other objectives, such as biodiversity observation, traceability of fishes, etc. hence, cost-efficiency in stock assessment cannot be directly inferred from those. As a general trend, it was common to find research about HTS methods focused on species such as reptiles, amphibians, birds, earthworms, mammals, invertebrates, phytoplankton and fish, which were analysed in different habitats as terrestrial, air, freshwater or marine systems (Deiner et al., 2017). However, very few papers or reports addressed specific conditions closer to those typical of stock assessment and specifically, to the conditions selected for the FishGenome project: trawling techniques and demersal representative species such as hake, cod and wrasse.
- Most of the research on HTS methods that claims to be cost-effective is about eDNA metabarcoding in rivers, lakes or ponds.
- Even those very few cases referred to stock assessment are focusing on species that differ from commercially exploited species.

- In terms of information outputs, traditional surveys provide a broader scope of variables needed for stock assessment, while HTS methods provide more accurate data for very specific variables. Therefore, in this context, both groups of methodologies seem to be more complementary than substitutes. The guideline for future substitution could be based on the evolution of the cost-efficiency.
- Despite of the fact that the responsible authorities of research surveys systematically collect cost-related data, there is still a clear shortage of information, not only of published cost-efficiency studies on the use of HTS methods for stock assessments, but also a general lack of published systematic cost analysis, both for traditional and new methodologies. Therefore, specific research on cost-efficiency is encouraged.
- HTS methods can provide additional valuable information outputs for managing, not only the fisheries, but also the marine ecosystems.

The main conclusion reached in this research is that more information should be gathered to evaluate whether HTS methods can reduce costs of the assessment processes, by being quicker, more efficient and/or time-saving methods, with respect to the traditional evaluation techniques. Undoubtedly, the combined use of traditional and genomic tools will offer a broader picture about some of the marine ecosystems' core characteristics, such as biodiversity, stock status, age, sex, maturity and fertility than each one of the approaches separately.

In summary, considering the recent surge of HTS methods, their dependency on the information gathered during traditional research surveys, and considering a short-term scenario, the balance in the use of both complementary approaches should be the critical factor that improves the efficiency of the processes and allows taking advantage of potential cost reductions and scale economies.

Regarding the results of the second subtask, i.e., the critical assessment (Deliverable 1.4b), the most relevant results of this integrated analysis were two facts: i) as mentioned above, there is a dependency of the genomics methods on the collection of samples during research surveys and ii) the type of data that can be collected by using different methodologies varies. For example, the size structure of a fish population can be determined in a traditional survey by measuring the length of fish samples, however, no HTS method is capable of determining size structure because the length of a fish cannot be determined by genomic analysis. The comparison of the type of data obtained with both approaches is shown in the Table 2.

From this comparison, the first direct observation is that HTS methods do not provide all the parameters that traditional methods are able to provide for the stock assessment of targeted species (i.e., abundance, biomass and demographic structure). Traditional surveys, in addition, provide information for monitoring the general conditions of the marine environment (e.g., marine litter and pollutants). Most notably, HTS methods do not provide information on size structure and maturity, and there are difficulties for estimating abundance-at-age. However, CKRM can provide key parameters for stock assessment, such as abundance, eDNA can accurately provide information on biodiversity (and potentially on stock biomass) (Bohmann *et al.* 2014; Goldberg *et al.* 2015) and epigenetics can provide age data. What is more, HTS methods could provide a better solution in some situations where survey approaches fail (see table 2 above) and may produce more reliable data than research surveys in such cases. For instance, in trawl surveys, the catch is not necessarily representative of the true abundance and biomass of fish present in a surveyed area of the ocean (Thomsen et al. 2016). This is because no trawl gear samples all the individuals present in its path, and catch rates of fish of different species and size in a given fishing gear vary considerably. The availability of fish to the trawl gear is affected by several factors, such as: daily variations of the vertical distributions that occur in many species, the behaviour of fish ahead of the trawl gear (some are herded into the path of the net by the action of the

otter doors while others show net-avoidance behaviour), the size and shape of the fish, their swimming endurance, etc. (Fraser et al. 2007 and references therein). When considering all this, it is possible that HTS methods could provide more accurate information than traditional research surveys.

**Table 2. Information provided by the traditional research surveys, compared with the potential information provided by HTS methods (CKMR, eDNA and epigenetic age determination).**

| Type of data | Traditional<br>Y/N; comment | HTS methods<br>Y/N (Method); comment |
|---|---|---|
| **Demographic/ biological data** | | |
| **Abundance** | Yes; Refers to the number of fish in a given fish population. Abundance estimations are based on the numbers of sampled fish for a species.<br>However, in a number of situations research surveys are not able to produce reliable estimations (e.g., widely distributed stocks, benthopelagic stocks, where catchability is an issue) or directly cannot be applied (in coastal/littoral areas, rocky bottoms, etc.). | Yes (CKMR); probably not affordable for all target species (e.g., species with very large population sizes may require a very large number of samples).<br>However, it is still possible to be used in those situations where surveys do not produce reliable data. |
| **Biomass** | Yes; Refers to the total weight of the fish in a given fish population. Biomass estimations are based on the weight of sampled fish for each species. Total biomass of a certain species during the survey is calculated using weight data and the trawled area (e.g., using the swept area method). | Yes (eDNA – using qPCR); Estimating biomass using eDNA for some species might not be possible though (e.g., in low abundance species). |
| **Size structure** | Yes; size structure is determined by measuring fish length of a sample of fish. | No. |
| **Age** | Yes; usually determined by analyzing calcified structures of fish (e.g., otoliths and illicia) to count growth rings. | Yes (Epigenetics); When epigenetic clocks are made available for the species of interest. |
| **Sex** | Yes; sex is usually determined by visual inspection of the animals. | Yes; When/if sex markers are available. |
| **Maturity** | Yes; maturity is determined by visual inspection or histological examination of the gonads. | No. |
| **Stock structure** | Could be possible (e.g., using stock identification methods such as analysis of parasites or using morphometric characters). | Yes |
| **Diet** | Yes; diet is determined by analyzing stomach contents. | No |
| **Other data** | | |
| **Marine litter** | Yes | No |
| **Biodiversity** | Yes; although with some limitations. | Yes (eDNA) |
| **Oceanographic data** | Yes; Oceanographic data include seawater temperature and salinity, for example. | No |

Moreover, abundance estimates based on bottom trawling are difficult to obtain for fish stocks closer to the shore, since rocky coasts or shallow seas are not accessible to trawling. In these cases, eDNA monitoring could greatly contribute to existing monitoring programs of fish stocks (Knudsen et al. 2019). Other studies have demonstrated that eDNA was able to detect species that were missed by trawling. Mostly, these were species that are anadromous, pelagic, small, rare, or those inhabiting rocky and muddy areas. This highlights the limited ability of trawl to capture taxa in particular types of habitats, or fish with different sizes and behaviours, while eDNA could theoretically detect fish in any type of habitat, with different swimming behaviours and sizes when the metabarcoding protocols are well-established. In addition, the eDNA-based approach can detect organisms at different life stages (different sizes) compared with net-based traditional methods that only catch mature individuals with specific size ranges. (Afzali et al. 2021).

Similarly, epigenetic age determination may offer a better solution for obtaining the age structure of monitored fish stocks in many cases. We must consider that not all teleost fish species exhibit otolith growth increments or other phenotypic age characteristics, making it more difficult to monitor the population dynamics for those species (Mayne et al. 2021 and references therein). In addition, the extraction of otoliths for age estimation is a lethal process, making it undesirable for application to threatened or endangered species. In addition, estimating fish age by counting of otolith increments can result in large biases and uncertainties, due to the combination of processing and interpretation errors. Both error types affect the estimates of growth, mortality and other demographic rates required for population dynamics models (Dortel et al. 2013). This is the current situation for the Atlantic cod in the eastern Baltic Sea, for which increasingly uncertain ageing has led to failed analytical stock assessment with substantial consequences for management between 2014 and 2019 (Heimbrand et al. 2020 and references therein). Therefore, developing epigenetic clocks for target species could have a major impact, since it will likely provide an accurate method to assess age in fish and circumvent the limitations of the current methods. In this sense, epigenetic age determination could also open the possibility of using advanced stock assessment models in species where age determination has been shown to be difficult (e.g., hake, or monkfish) (ICES, 2021b). Moreover, epigenetic age determination is non-lethal, which makes it very attractive in the case of threatened species, such as sharks.

Finally, in many situations survey time may be reduced, if some of the data are estimated by HTS methods. For example, there is no need to perform many hauls to obtain a good size structure of a stock; if a survey performs many hauls, it is with the goal of covering a wide area (normally stratified) of the stock distribution in order to obtain a good abundance estimation. The latter effort can be substantially reduced using HTS methods.

Nevertheless, there are still few studies where the efficacy of traditional methods (research surveys) versus genomic methods has been formally compared. At present, most of these studies have focused on eDNA metabarcoding (i.e. biodiversity approach).

The set of traditional methods is the outcome of a long process of adaptation to the goals and needs of stock assessments, while HTS methodologies are scientific developments, which still need to follow further innovation steps for fitting them to the stock assessment specific needs. It is expected that HTS methods, once fully developed and tuned, will be able to provide more accurate data on their fields of application than traditional methods. Nevertheless, the improved accuracy of the HTS methods regarding traditional approaches is yet to be demonstrated in a variety of scenarios. Data obtained from HTS methods cannot be implemented in stock assessment, if their accuracy is lower than that from traditional methods. In such a case, its implementation would incorporate a great uncertainty in the stock assessment. Thus, thorough research on HTS accuracy and precision, in comparison with traditional methods, is required for each of the stocks where the HTS methods can be expected to be implemented. Moreover, detailed studies assessing the costs of producing the data of interest by both methods

are needed. The rapid replacement of genomic methods and constant drop in their cost, demand frequent regular assessments.

In summary, as explained in this section, research surveys show a number of drawbacks in relation to stock assessment that HTS methods may help to overcome. However, research surveys provide a considerable amount of information beyond that used in stock assessment, especially that related with environmental monitoring and ecosystem status. Thus, substituting surveys with HTS techniques would lead to an important loss of information. In consequence, both methodologies seem to be rather complementary than substitutes. The guideline for future implementation of HTS methods in research surveys could be based on the evolution of the cost-efficiency and on further evidence of precision and accuracy gains but should always consider the relevance of the potential loss of information.

### 1.2.4. Identification of barriers and risks

#### Methodology

The objective of this task was to identify any technical, legal, environmental and economic barriers and risks for the practical implementation and roll out of the HTS techniques.

The first were extracted from the SoA reports for the different HTS techniques and also for the research surveys and then subjected to a more systematic analysis, implementing a participatory process, to visualise the impact each of the barriers and risks can have in the future implementation of the techniques.

Throughout the state-of-the-art revision process, both the members of the partnership and the External Experts collaborating with FishGenome have identified the main hurdles and difficulties for the implementation of the HTS genomic tools, as well as some of the advantages, which can derive from their use. This information can be found in the State-of-the-Art Reports and highlighted across the infographics prepared to illustrate the techniques. These barriers have also been discussed in the specific context of the fisheries assessments, in particular as one of the points for discussion during the 1st Virtual Workshop on the 28th of May, 2020.

Therefore, the purpose of this deliverable is not to present de novo information, but to extract and organize information about relevant risks and barriers for implementation, which can also be found scattered through the different deliverables and reports presented so far.

The resulting report is deliverable D.1.5 *Identification of barriers and risks, impact and mitigation*, made as an extract of identified risks and barriers regarding the suitability and feasibility of the use of the genomic methods in fisheries assessments. Apart from technical barriers some others have been considered: policy and regulation related, cost-efficiency and economics, cultural and institutional barriers (aversion to change), etc.

#### Results

In this task, different types of barriers and risks for the implementation of genomic methods for stock assessment were identified:

- Barriers and risks inherent to the genomic methods themselves
- Barriers and risks from the economic and legal point of views and
- Barriers and risks related to research surveys and stock assessment

In this deliverable, the impact of those barriers and possible mitigation strategies were also proposed.

The main barriers and risks identified for the genomics methods are detailed below for each of them:

For CKMR:

- The methodology has been developed using fish species characterized by very small population sizes and undergoing plans of conservation and recovery.
- Appropriate sample size is required to use CKMR successfully. Because of the dependency of the sample size requirements on the expected population size, the method is likely not applicable to species with very large population numbers.
- Requires a solid knowledge of the biology of the species and its population structure.
- The method requires specialized knowledge, mainly on stock assessment modelling and mathematics, that is usually outside of the experience of population geneticists, who have the skills to generate and analyse the genetic data.
- Bioinformatics analyses may not be straightforward. Parentage analysis becomes markedly more challenging in situations where neither parent is known by observation, which is the case for most marine exploited fish populations.

For eDNA:

- This method is still at an emerging scientific stage. Multiple independent research groups have developed eDNA analysis techniques, leading to a variety of protocols for eDNA detection of aquatic organisms across various taxa and environments. Currently, there are diverse approaches for sampling and interpreting eDNA data (Goldberg et al. 2016).
- The sampling strategy can strongly influence the amount of eDNA found in the samples.
- The amount of eDNA that can be recovered under field conditions is influenced by many factors and can vary across different study areas. Aspects such as persistence, dilution and distribution of eDNA may affect the quality of the samples.
- The eDNA method is mainly capable of detecting the presence or absence of a species (Herder et al. 2014; Evans and Lamberti 2018).
- There is a concern regarding eDNA degradation: eDNA is usually highly degraded and the fragment size rarely exceeds 150 bp (Deagle et al. 2006).
- For eDNA studies, it is difficult to develop a quantitative PCR assay (qPCR) for a species like cod in an environment that is inhabited by closely related species. It seems to be particularly difficult, if the target species is relatively rare.

For DNAm:

- Lack of sufficient research on epigenetics for age determination in fish. No literature reviews exist on the subject of epigenetic clocks to estimate age in fishes. For example, the only commercial species for which such type of clocks has been developed so far is the European sea bass (Anastasiadi and Piferrer, 2020).
- Calibration of the epigenetic clock in fish is only possible for species where the "true" age is known or can be estimated accurately.
- Sensitivity of the epigenetic clock may not be good enough depending on the minimum age of the fish used to initially build the clock. Sensitivity is the minimum value (age) that can be detected using an epigenetic clock.
- In epigenetic studies, there might be potential issues with bisulfite conversion step of DNA samples. Although this is nowadays routinely done with commercial kits for this specific purpose, it is possible that conversion rates are lower than needed.
- Bis-RAD-Seq is believed to be one of the best techniques for epigenetic age determination, because it combines single nucleotide resolution (a must for building and epigenetic clock) and a reasonable number of interrogated methylated loci. Unfortunately, it appears that bis-RAD-Seq is not offered as a regular service in companies dedicated to genomic services.

The following economic barriers and risks were identified:

- Funding should be allocated to data collection if HTS methods are incorporated to the DCF. At least in the short-term, if HTS methods were implemented an increase of costs for data collection can be expected.
- Cost-efficiency of HTS methods has not been proven. HTS methods have been claimed to be cost-efficient but very few publications have systematically and accurately addressed the issue.
- All the analyses claiming the cost-efficiency of HTS methods are not referred to stock assessments but to other objectives (e.g., biodiversity monitoring, fish products traceability, etc.) and hence, cost-efficiency in stock assessment cannot be directly inferred from those.
- The genomic methods explored in FishGenome have not been routinely used before either for the species we are targeting nor in the context of stock assessments and thus, there is lack of standardisation.
- While the costs and their possible evolution for the use of the techniques can be estimated within an acceptable margin of error, it will be more difficult to make a quantitative estimation of the benefits, as those may be of a very diverse nature.

Legal, policy related and institutional barriers and risks:

- .
- The Data Collection Framework is already oriented to respond to the needs and demands of its stakeholders, i.e. the data to be collected is selected on the basis of needs clearly substantiated by end-users of scientific data, considering the scientific relevance and usefulness of those data. These needs are discussed and adopted within Regional Coordination Groups (RCG). However, RCGs are not fully aware of the outcomes from FishGenome at the moment but they will need to properly understand the pros and cons of the HTS for a future adoption scenario.
- However, the outcome of FishGenome do not involve only data collection, but also survey design, monitoring programmes, stock modelling, stock assessment and scientific advice, fisheries management, environmental management and scientific developments. This means the involvement of institutions at national, European and international level.
- 2010 Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity. In the EU, all Member States, as well as users of genetic resources in the EU, are bound by these regulations.
- Not all fisheries research institutions have facilities to carry out genomics analysis. There may be a lack of specialised personnel too and sequencing services for a scaled-up implementation process will still largely depend on out-sourced services (which may lead to the need to go through public procurement mechanisms).
- Some specific components of the techniques are subject to intellectual property rights (IPR) protection. For research and experimental purposes, there is no restriction of use, however, it needs to be thoroughly assessed, if this may have consequences in a potential routine and large-scale implementation in the context of the DCF.

Barriers and risks related to research surveys and stock assessment

- Lack of research related to the applications of genomic methods for stock assessment.
- In terms of information outputs, traditional surveys provide a broader scope of variables needed for stock assessment, while genomic methods provide data for only very specific variables requested by stock assessment.
- The improved accuracy of the genomic methods in comparison to traditional approaches, has yet to be demonstrated in a variety of scenarios.
- Personnel and space limitations on board the survey vessel.
- There is a limited amount of funding and time during surveys for data collection, creating reluctancy to modify the type of data collected and procedures to collect

them during research surveys by fishery scientists. This is especially important because it is required keeping consistency in data collected throughout the time, to avoid disruptions in the time-series[7].

### 1.3. Lessons learnt

The overall perspective of the applicability of genomics tools in fisheries assessments is promising. The three tools discussed, namely CKMR, eDNA and epigenetics for age determination, have the **potential to provide an important range of data to be used in stock assessment and improving scientific advice**. The use of genomic data may also increase cost-efficiency of research surveys and stock assessments. FishGenome is a timely study, under a service contract, as the evolution of these techniques is fast and it is important to set the basis and to perform a foresight exercise to assess the potential of these technologies in the mid-term.

The application of HTS tools in fisheries research is still limited scarce. Yet, FishGenome has identified the main drawbacks and advantages of the implementation of these techniques in the context of data collection in research surveys and in stock assessment. Many of the **drawbacks detected can be overcome with more investment in research and development activities** for this purpose. For instance, it could be important to develop more reference genomes; public databases for genomics need further curation; in other cases, there is also need to leverage the available information about the environment, the biology and ecology of the species, etc.

In the particular case of epigenetics for age determination, the epigenetic clock built in European seabass has paved the way for similar developments in other teleost species and have shown its potential use in fisheries science. It may replace at mid-term the need of collect fish otoliths or similar structures and the use of the time-consuming age determination methodology used nowadays. CKMR and eDNA also seems to provide important potential to improve stock assessment and scientific advice, delivering abundance data and other key biological parameters. Yet the level of maturity and readiness of the techniques clearly require effort and research before implementation.

The genomic techniques still rely in the collection of samples onboard research surveys. They may contribute to increase cost efficiency of the surveys, but this is only a hypothesis not demonstrated. There is potential, however, for reducing survey time by using HTS methods. These techniques, nevertheless, will not replace the need of research surveys, but will increase the utility of the surveys providing complement data to improve stock assessment as well environmental management. Genomic techniques, while providing valuable information on stock and ecosystem status, are not able to provide all parameters required by stock assessment.

The main lesson from the analysis conducted in this section is the existence of high expectations but with of very few applications of HTS techniques in fisheries research. However, in spite of the barriers and risks identified it cannot be denied that these techniques can contribute to improve stock assessment if the proper track is followed.

The flexibility of the **EU DCF gives a great opportunity for testing to what extent the implementation of HTS methods will improve stock assessment**. This is so because DCF is oriented to respond to the needs and demands of the end-users of scientific data, considering the scientific relevance and usefulness of those data, by discussions within RCGs. However, the outcome of FishGenome do not involve only data collection, but also survey design, monitoring programmes, stock modelling, stock assessment and scientific advice, fisheries management, environmental management and scientific developments. This means the involvement of institutions at national, European and international level, implying a big challenge.

[7] This is a legal obligation under the DCF Regulation (EU) 2017/1004 of the European Parliament and of the Council, Article 5, point 5.

## 2. PILOT STUDIES

The overall goal of this work package was to assess the potential of several novel techniques, based on HTS data, to complement traditional methods and improve fish stocks assessments. Pilot studies were designed alongside fisheries research surveys, specifically the North Sea IBTS-Q3 survey and the Mediterranean MEDITS-GS5 Survey (Balearic Islands). This design facilitates producing results and achieving conclusions about their potential implementation that are realistic in the context of current surveys. This is relevant because the potential of a novel technique relies not only on its capacity to produce accurate data but also on its implicit requirements (number of specimens needed, time (in days at sea) needed to reach this number, etc). If such requirements exceed the capacity of regular surveys, the potential of the technique is low compared to a method with needs that can be fitted easily within present survey operations. Four HTS techniques were tested: i) CKMR to estimate stock abundance, ii) RAD-Seq for connectivity, substructure and sex marker, iii) DNAm for age determination, and iv) eDNA to estimate stock abundance and biodiversity.

### 2.1. MAIN ACHIEVEMENTS

A series of ***protocols describing the experimental design, lab work and equipment used to analyse data for Close Kin Mark Recapture, connectivity, substructure and sex marker search studies using RAD-Seq***. These protocols are presented in D2.1a "Experimental design and protocols for CKMR, connectivity, substructure and sex marker search studies using RAD-Seq". The purpose of this document was to describe the protocols used by the FishGenome Consortium, including the sampling protocol used to collect biological samples, the protocol for isolating DNA, the laboratory protocol for applying restricted site associated DNA sequencing (RAD-Seq) and a description of the bioinformatic pipelines used to analyse the data.

A series of ***protocols describing the experimental design, lab work and equipment used to analyse data for Bis-RAD-Seq epigenetic studies for age determination***. These protocols are presented in D2.1b "Experimental design and laboratory protocols for Bis-RAD-Seq epigenetic studies for age determination". The purpose of this document was to describe the protocols used by the FishGenome Consortium, including the sampling protocol used to collect biological samples, the protocol for isolating DNA, the laboratory protocol for applying bisulfite restriction-site associated DNA sequencing (bis-RAD-Seq) using Illumina high throughput technology to identify DNA methylation patterns and the protocol to conduct the bioinformatic analysis and perform the statistical analysis required to build the epigenetic clock for age prediction.

A series of ***protocols describing the experimental design, lab work and equipment used to analyse data for eDNA studies for biodiversity and biomass***. These protocols are presented in D2.1c "Experimental design and laboratory protocols for eDNA studies for biodiversity and biomass". The purpose of this document was to describe the protocols used by the FishGenome Consortium, including the sampling protocol used to collect biological samples, the protocol for isolating DNA, the laboratory protocol for applying metabarcoding eDNA using Illumina HTS to assess fish communities and fish biomass in seawater and sediment samples, the protocol for conducting the bioinformatics analysis and the protocol for applying qPCR eDNA using a real-time thermal cycler system to detect cod in seawater and sediments from the North Sea basin.

***Collection of samples for the pilot studies.*** The fieldwork carried out in FishGenome involved the collection of samples in two surveys, the International Bottom Trawl Survey in the North Sea (July 2019) and the International Bottom Trawl Survey in the Mediterranean (MEDITS Spain in June 2019).

***Analysis of the results of the pilot studies.*** The results of the pilot studies are presented in Deliverable 2.2. "Pilot studies comparative analysis". This document describes the results obtained by the application of four HTS genomic techniques in the Pilot studies performed by the FishGenome Consortium. The suitability, potentiality, pros and constraints of these novel genomic techniques to complement and improve current fisheries assessments are discussed.

## 2.2. WORK CARRIED OUT

### 2.2.1. DESIGN OF THE PILOT STUDIES

#### Methodology

A set of pilot studies were designed to test the genomic methods considered in FishGenome, namely, Close Kin Mark Recapture (CKMR), RAD-Seq for connectivity, stock substructure, and sex assignment, epigenetic age determination by DNA methylation (DNAm), environmental DNA (eDNA). The pilot studies aimed at covering the following fish stocks in EU waters: Atlantic cod (*Gadus morhua*) in the North Sea, European hake (*Merluccius merluccius*) in both the Galician shelf and Mediterranean and ballan wrasse (*Labrus bergylta*) in the Galician shelf, as well biodiversity studies based on eDNA in two ecoregions. The combination of techniques and stocks results in 14 different case studies, as summarised in Table 3.

The samples analysed in the Pilot studies had two origins. Samples from the Atlantic – southern hake stock and ballan wrasse - were already available, as part of ongoing research or as part of a tissue biobank owned by the FishGenome participant institutes. Samples from the North Sea – cod, northern hake stock and environmental – and the Mediterranean – hake and environmental were collected during two fisheries research surveys, specifically the North Sea IBTS-Q3 and the Mediterranean MEDITS-GS5 Survey. This strategy was followed to obtain precise information on the feasibility of accommodating the requirements of each technique to current survey characteristics. Specifically, we gathered information on the number of samples that can be collected and processed during current surveys vs. the sample requirements of each technique. Additionally, we evaluated the difficulties of incorporating this additional sampling – tissues, water and sediment - within the regular work performed on board. This information is vital to assess the potential of the tested genomic techniques as it relies not only on the capacity of the methodology to produce accurate data but also on its technical and biological requirements. . Therefore, the first step was to produce tailored sampling protocols for the collection of samples for genomic studies during the surveys.

First, an ad hoc technical meeting was held in May 2019 at the premises of IEO and UiB. This was necessary because the pilot studies involved the collection of a series of tissue and environmental samples by different teams and in different surveys. Thus, it was crucial to ensure the consistent application of good standards across the various source sites. During this meeting, the standard work methods used in the surveys (i.e., the North Sea IBTS and MEDITS protocols) were explained. Then, the FishGenome Experimental design and sampling protocols were discussed considering how to integrate them into the workflow of both surveys. A single agreed protocol was established for the two surveys to standardise procedures that ensure traceability and comparability of results between surveys. It included details on experimental design, sampling intensity, sample collection, preservation and data gathering.

**Table 3. Cases studies addressed as pilot studies within FishGenome, by ecoregion, species and genomic tool (the three main tools are shown in bold)**

| | North Sea | | | Galician shelf | | Balearic Isl. | |
|---|---|---|---|---|---|---|---|
| **Genomic tool** | Cod | Hake | Biodiv | Hake | Wrasse | Hake | Biodiv |
| **Close-Kin Mark-Recapture** | X | X | | | | X | |
| **RAD-Seq for connectivity** | | X | | X | | | |
| **Genotyping for stock substructure** | | | | X | X | | |
| **Genotyping for sex assignment** | | | | X | | | |
| **Epigenetic age determination** | X | | | | | | |
| **Environmental DNA** | X | | X | X | | X | X |

In addition to the sampling protocols for the surveys, it was necessary to create a protocol on the storage and archiving of collected samples in the participating laboratories within the FishGenome project. The samples were mainly used in the pilot studies of this Project, but their long-term integrity/ preservation needs to be ensured, to allow for future research and collaborations. Considering this, a biobank and data protocol was also developed. This protocol was issued before the surveys took place, to ensure the integrity and traceability of the samples, especially regarding collection, preservation and storage.

### Results

The protocol to collect samples during the surveys is specified in Section 1 of the Deliverables 2.1a (Experimental design and protocols for CKMR, connectivity, substructure and sex marker search studies using RAD-Seq), D2.1b (Experimental design and laboratory protocols for Bis-RAD-Seq epigenetic studies for age determination), and D2.1c (Experimental design and laboratory protocols for eDNA studies for biodiversity and biomass). The main aspects regarding the collection of samples for the different genomic methods are described below.

For CKMR, the ideal sampling design for estimating population size should have an even distribution of both reproductively mature and juvenile groups and both sexes evenly represented. For connectivity and fine-scale substructure, two or more different stocks/populations/groups need to be sampled, while the search of a sex marker requires the presence of individuals of both sexes. The goal of the sampling is to collect a minimum total number of 600 specimens in the North Sea, which corresponds to sampling virtually all the fish captured, considering the 2018 catches (490 specimens of North Sea cod and 188 specimens of North Sea hake). Grouping the individuals by 5 cm sizes classes, up to 150 individuals by size range should be collected. In the Balearic Islands, the goal is to collect in total 750 individuals, between 150-200 individuals by size range, grouping the sizes by 5-cm classes (the MEDITS-GS5 2018 survey accounted for 1042 hake specimens in 2018). Genetic information for CKMR, connectivity, fine scale-substructure and isolation of a sex marker is obtained through the extraction of DNA from any tissue sampled from individual fish. However, fin clips are generally used due to their non-invasive nature. Large numbers of individuals are routinely sampled during the course of traditional research trawl surveys and our proposal adds a simple, inexpensive, and time efficient addition to most sampling protocols already in place. The method to obtain the samples is explained in detail, together with a list of the necessary equipment.

For epigenetic age determination (DNAm), the ideal sampling design should have an even distribution of age groups and both sexes evenly represented. In total, five time-points represented by 10 individuals (five males + five females) each, are going to be

analysed. Since at this stage the age is still unknown (otoliths are collected but not processed), the size is going to be used as a proxy for the time-points. The goal of the sampling is to collect the fin clip, liver, branchial arch and muscle of 20 fish by size range and sex. Although some of the smaller undifferentiated specimens will not allow sexing, they should be collected anyway. Genetic information for DNAm age is obtained through the extraction of DNA from any tissue sampled from individual fish. At least two tissues should be collected for comparison, in our case, four tissues will be collected: fin clip, a piece of muscle, branchial arch and liver. Tissues should be dissected perimortem, i.e., at or near the time of death, to minimize degradation. The method to obtain the samples is explained in detail, together with a list of the necessary equipment.

The eDNA method consists of the analysis of environmental samples to capture DNA from aquatic organisms. In FishGenome, two types of samples were collected, and the protocol established the following procedures:

- Water samples: Should be collected at trawl depth before bottom trawls, l to and minimizing the disturbance of the standard procedures carried out on board during the regular fisheries research survey. Two replicas (2 Niskin 5L botlles) will be processed per sampling site.
- Sediment samples: Should be collected after the water to avoid resuspended sediment particles that may interfere with the analysis of the water. The sediment will be collected from the superficial strata in duplicates (2x per station) using a common corer. The corer must be rinsed with sterile water in between samples.

This protocol also details how to filter the water to capture the DNA from aquatic organisms. A list of the necessary equipment is provided.

All the samples collected during the Pilot studies were archived following the biobanking protocol, which establishes how the handling and storage of samples shall be done in FishGenome. An overview of the biobanking process can be found in Figure 14. The first section of the protocol describes sample handling and storage conditions, according to the samples' nature (e.g., tissue type, sediment or water). The protocol of processing of samples, establishes that, once the samples arrive in the lab, they should be processed to obtain subsamples for the analysis of genetic information and for storage in the biobank. Each sample should be split into two sub-samples. The first one should be archived and stored. The second sub-sample should be processed for DNA extraction, according to the protocol for each genomic technique (e.g., for CKMR, see Section 2 of Deliverable 2.1a "Experimental design and protocols for CKMR, connectivity, substructure and sex marker search studies using RAD-Seq"). Then, the obtained DNA stocks should be quantified and two aliquots at 100 ng/µl should be prepared for each of the samples. The DNA from the first aliquot is intended to be used to analyse genetic information (e.g., CKMR, DNAm, eDNA, etc.), while the second DNA aliquot is to be archived, for further analyses if needed. The rest of the DNA stock should be archived. It is important to note that new aliquots can be obtained from the DNA stock, if required by researchers in future collaborations. The schematic representation of this process can be found in Figure 15.

In addition to this, the protocol describes how the information of each sample should be recorded and stored in a database, including collection details, material type and attributes, to enable information retrieval via multi-criteria queries. The fields to be recorded in the database are described in the protocol.

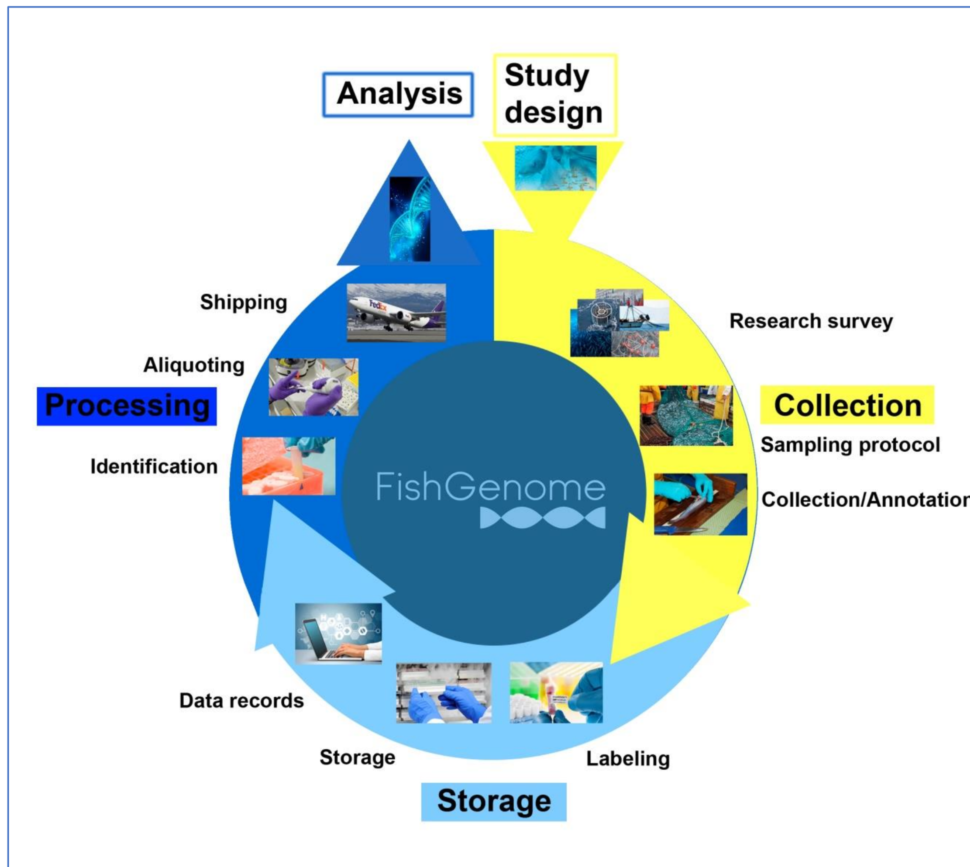



**Figure 14. Schematic representation of the FishGenome biobanking process.**

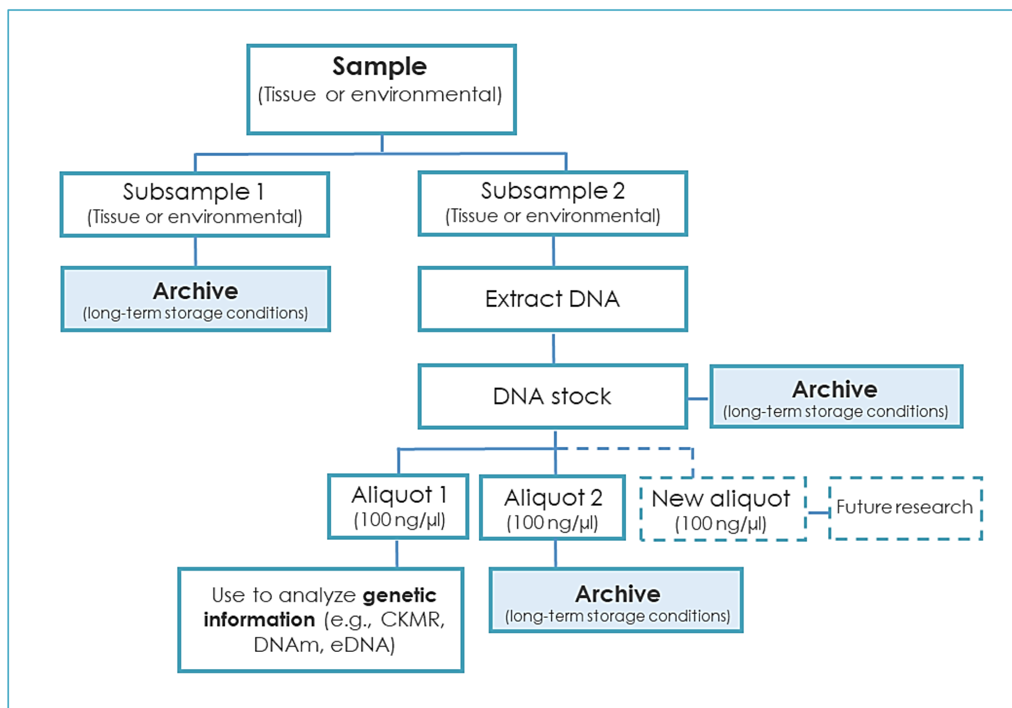



**Figure 15. Schematic representation of the processing of the samples for the biobank.**

### 2.2.2. Fieldwork

#### Methodology

As mentioned in the previous section, the collection of new samples for the pilot studies was done during the regular scientific research surveys carried out as part of the fisheries data collection programs in the North Sea and the Mediterranean Sea (both within the EU DCF).

For the MEDITS survey, the institute in charge of the survey was the IEO while the Thünen-Institut was responsible for the North Sea survey. All the logistics were planned in the previous task (2.2.1 Design of the pilot studies) and put in practice on May 2019. These included the inclusion of the FishGenome sampling tasks in the survey programmes, the distribution of work, material required, etc. IEO agreed on the participation of researchers from Thünen-Institut and UiB in the survey for the FishGenome specific tasks.

The MEDITS and North Sea IBTS surveys were carried out between June-August 2019. During these surveys, i) water and sediment samples were collected to obtain eDNA and ii) hake and cod tissue samples were collected for CKMR and epigenetic analysis. DNA has been isolated from water, sediment and tissue samples.

#### Results

The MEDITS survey took place from 11 to 25 June 2019 in the GSA5 area around Mallorca Island. During this survey, eDNA was collected from water and sediment in a total of 18 stations (replicated). 250 hake individuals (muscle, fin, gill and liver) were collected for epigenetic analysis. 330 hake fin clips were collected for CKMR analysis. A total of 36 sediment and water samples were collected for eDNA during the survey.

**Table 4. Summary of samples collected for the pilot studies. Objective indicates the use of the samples: CKMR, epigenetic analysis (DNAm), environmental DNA (eDNA), connectivity studies (CONN), stock substructure analysis (SUB) and sex assignment (SEX). N indicates the number of fishes sampled for cod, hake and wrasse and number of stations sampled for water and sediment. Length is given in cm and mean and standard deviation are indicated in brackets.**

| Area | Species/Type | Objective | N | Length range |
|---|---|---|---|---|
| North Sea | Cod | CKMR, DNAm | 275 | 8-99 (42.1±18.9) |
| | Hake | CKMR, CONN | 109 | 29-88 (58.2±13.2) |
| | Water/Sediment | eDNA | 33 | - |
| Galician Shelf | Hake | CKMR, CONN, SUB, SEX | 501 | 5.3-78.1 32.9±25.4) |
| | Wrasse | DNAm | 60 | 25.5-42.5 (33.9±3.8) |
| | Wrasse | SUB | 110 | 25.5-42.5 (33.9±3.8) |
| Balearic Islands | Hake | DNAm | 176 | 6.5-53.5 (25.1±7.6) |
| | Hake | CKMR | 348 | 6.5-53.5 (21.7±7.9) |
| | Water/Sediment | eDNA | 18 | - |

The North Sea IBTS survey took place from 8th July-3th August 2019 in the North Sea as planned. A total of 275 cod samples and 109 hake samples for CKMR were collected during the survey. For eDNA, 33 water and sediment samples were collected. For epigenetics, 275 cod samples were obtained. The collected 109 hake samples were also be used for connectivity analyses, as several populations (North Sea, Galician shelf and Balearic Islands) were explored in FishGenome. This analysis was prevented in cod, as a unique population (North Sea) was available. The number of available samples is

shown in Table 4. The number of hake samples was lower than expected, so additional effort was placed to collect more numbers from other surveys in the North Sea. The following year, 39 additional samples were collected during the North Sea IBTS survey that took place from the 16th of July to the 14th of august. The number of individuals was too small to generate a pool for RAD-Seq but no more samples could not be obtained.

A total of 32 sediment and water samples were collected from 16 stations in the North Sea during a sampling mission on FRV Walther Herwig (mission number WH428) from 8 July to 3 August 2019, conducted by the Thünen-Institut in the framework of the FishGenome Project. Briefly, water samples from single Niskin bottles were directly filtered after each CTD cast on 0.45 µm filters. Filters were stored at -20°C until DNA extraction. Accordingly, sediment samples were stored in sterile 100 ml tubes filled with ethanol (99%) (Chemsolute, Renningen, Germany) at -20°C until extraction within 3 months after collection. Regarding labelling, we refer to the sampling regions as "boxes" (i.e., box A) and the location within the box as "station" (i.e., ST1, ST2, etc.) with the date of collection (DD/MM/YY).

The samples obtained from the surveys were stored at the IEO (Spain) and Thünen-Institut (Germany), following the biobank protocol.

### 2.2.3. CKMR

#### Methodology

Here, we used restriction site associated sequencing (RAD-Seq), a genome-wide technique, designed to obtain an optimum number of SNP markers specifically in teleosts. Starting DNA samples were obtained from fin clips of hake collected at three different locations, 1) the North Sea, Balearic Islands and Galician shelf and 2) from cod collected in the North Sea.

RAD libraries were prepared in house and sequenced on a HiSeqX platform (Illumina) in 150-bp paired reads. Obtained sequences were processed in Stacks 2.0 through two different pipelines in cod and hake. A reference genome-based pipeline was used for cod as the species has a high-quality reference genome available in public databases. For hake, only a highly fragmented genome is available with insufficient quality to guide the analysis. Accordingly, a de novo approach was adopted. For both species, data was filtered to keep only loci that was present in at least 80% of samples.

Selected SNP markers obtained after filtering were used to estimate kinship among samples at each location for each species. Two different statistical models were used; the method of moments (MoM) (Purcell et al., 2007) and maximum likelihood estimation (MLE) (Milligan, 2003; Choi et al., 2009).

Relatedness was measured by estimating kinship coefficients for all pairs of individuals using four kinship categories (POP: parent–offspring-pairs, FSP: full sibling-pairs, HSP: half sibling pairs, UN: unrelated). FSP and POPs share, on average, half their genes while HSP share one quarter. Kinship pairs were validated only if computed by both models and for siblings, only cross-cohort comparisons were considered “true” as only cross-cohort sibling comparisons are suitable for estimating abundance (Maunder et al., 2021). This is because the probability of finding intra-cohort siblings is affected by the variance in offspring number among individuals of the same age and sex. Sibling pairs from different cohorts share a parent who has to be alive at the time of birth of each sibling. The larger the difference in age between the cohorts relating to the siblings, the longer the parent has to survive. As difference in age increases, fewer HSPs are expected. The rate of decline is related to adult survival but is unaffected by year-to-year fluctuations in the individual´s reproductive output or the overall survival probability of juvenile cohorts. In contrast, if there is variation between litters in number of surviving offspring— due, say, to variations in the initial litter size caused by different reproductive outputs, or in early-life-history survival rates per litter— then there will be

a systematic excess of intra-cohort sibs relative to cross-cohort sibs. Ignoring this phenomenon, i.e., interpreting the likely excess of intra-cohort sibs in the same way as the cross-cohort sibs, could lead to negative bias in the abundance estimates. This is especially true in teleosts fish, where litter sizes are large and consequently, the effects can be very large. Moreover, HSPs are not as informative as POPs/FSP as they provide information about a single parent vs. both parents.

### Results

The results in this section have been summarized from different sections of Deliverable 2.2., *Pilot studies comparative analysis*.

CKMR ANALYSIS OF NORTH SEA COD POPULATION/S

The initial goal of the FishGenome project was to collect a total number of 250 cod specimens in the North Sea. The Consortium sampled all the fish captured during the IBTS survey, which amounted to 275 samples. Of these, 240 were included in the RAD libraries and 235 (98%) passed the quality criteria. Our analysis produced 17 946 648 loci of which 25 571 passed sample/population constraints and were polymorphic. These variable markers were used to infer kinship coefficients by comparing the genotype of each individual against all other specimens. A total of 27 495 comparisons were performed, revealing three kinship pairs, two Parent-offspring pairs (POPs) and one Half-sibling pair (HSP). The next step consisted on determining, based on their biological data (size, weight, age if known), if these related specimens belong to the same or different cohorts. Since the age of cod can be determined very reliably using their otoliths, we used this parameter for the analysis. The first kinship pair detected corresponds to a Full-sibling-pair (FSP) intra-cohort and should not be considered for CKMR. Maturity at age, in cod is 3.5 years (Stock Assessment ICES Report 2021), so individuals younger than four years are unlikely to have had progeny. Thus, two kinship pairs were found among the 235 specimens analysed, one FSP and one HSP inter-cohort (Figure 16).

CKMR ANALYSIS OF NORTH SEA HAKE

The initial goal of the FishGenome project was to collect a total number of 250 hake specimens in the North Sea. Although the Consortium sampled all the fish captured during the IBTS survey, only 109 samples could be collected. Of these, 96 were included in the RAD libraries and 94 (98%) passed the quality criteria.

Our analysis produced 8 782 197 loci of which 11 608 passed sample/population constraints and were polymorphic. A total of 4371 comparisons between sampled individuals revealed the presence of six kinship pairs, two POPs/FSPs and HSPs.

The next step consisted on determining, based on their biological data (size, weight, age if known) if these related specimens belong to the same or different cohorts. Since the age of hake cannot be determined reliably using their otoliths, these were not analysed at any of the locations included in the Pilot studies. We based our analysis on the size and length of the individuals, considering the known data about the biology of the species. European hake males grow faster than females up to 3 years old, when their grow rate decreases, whereas females grow faster from that age onwards. Females reach a larger size and grow older than males resulting in a grow difference of up to 15cm for total lengths ≥40cm between females and males of the same cohort (same year) and this difference increases with age (Figure 17). Large differences in weight between males and females are also observed.

Considering this information, several kinship pairs detected in our analysis clearly correspond to the same cohort and should not be considered for CKMR. Only two of the initial HSP pairs belong to different cohorts.

**Figure 16. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, square –male, triangle –immature) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). Two kinship pairs were found among the 235 specimens of cod analysed, one FSP and one HSP.**

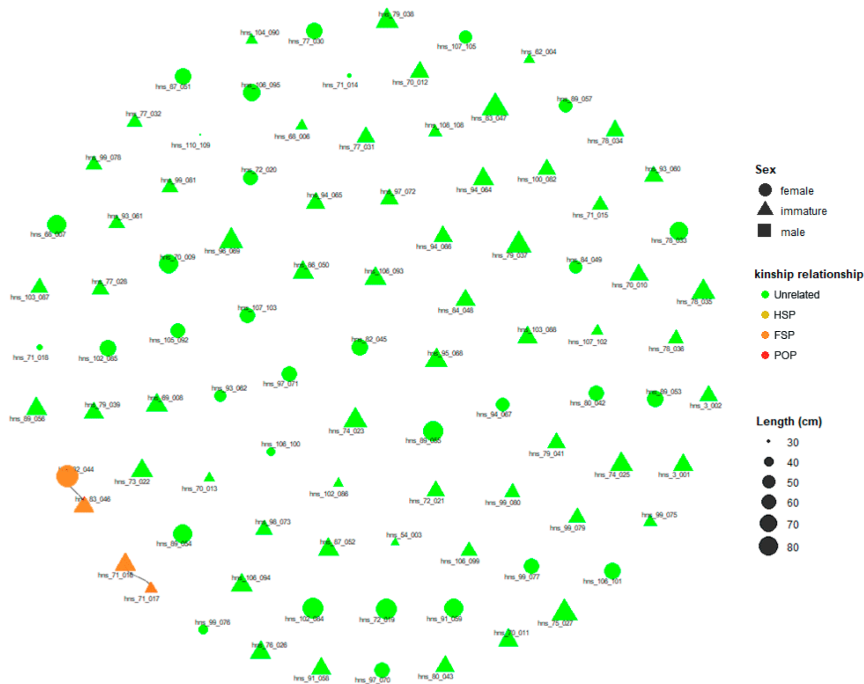



**Figure 17. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, triangle –male) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), red – parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). Two kinship pairs (HSPs) were found among the 94 specimens of hake analysed.**

## CKMR ANALYSIS OF BALEARIC ISLANDS HAKE

The initial goal of the FishGenome project was to collect a total number of 250 hake specimens in the Mediterranean. The Consortium sampled all the fish captured during the MEDITS survey around the Balearic Islands, which amounted for 348 samples of which 288 were included in the RAD libraries. A total of 281 (97.5%) passed the quality criteria and the filtering steps. Our analysis produced 8 605 481 loci of which 10 360 passed sample/population constraints and were polymorphic. A total of 39 940 comparisons were performed between individuals which revealed nine kinship pairs among the samples, two POPs/FSPs and seven HSPs.

The next step was the analysis of the individuals showing high relatedness, to determine, based on their biological data (size, weight, age if known) if they belong to the same or different cohorts. Seven of the kinship pairs detected in our analysis clearly corresponded to the same cohort and should not be considered for CKMR. Thus, three kinship pairs were found among the 281 specimens analysed, one POP/FSP and two HSPs (Figure 18).

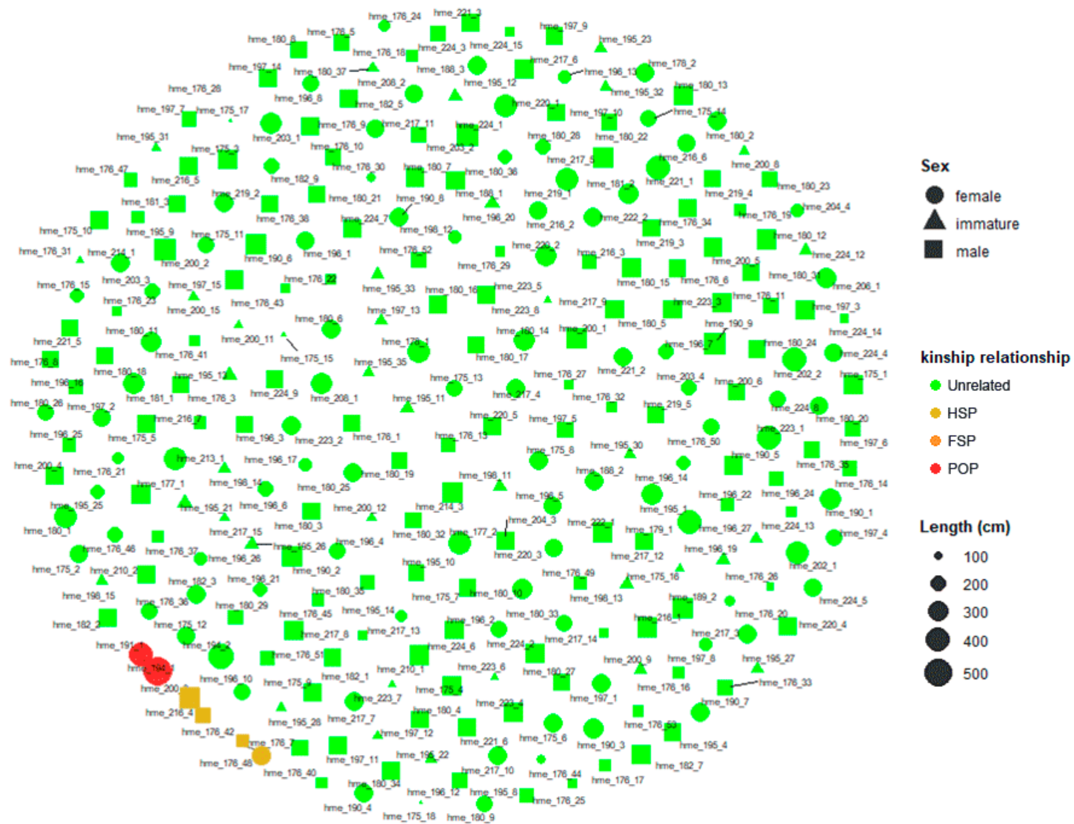



**Figure 18. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, triangle –male) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), red – parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). Three kinship pairs were found among the 281 specimens of hake analysed, one POP and two HSPs.**

CKMR ANALYSIS OF GALICIAN SHELF (NW Atlantic) HAKE

The initial goal of the FishGenome project was to collect a total number of 250 hake specimens in the north-west Atlantic Iberian Peninsula Galician shelf. These specimens were already available in the research group. A total of 144 were included in the RAD libraries and 142 (98%) passed the quality criteria and the filtering steps. Our analysis produced 8 386 443 loci of which 10 084 passed sample/population constraints and were polymorphic. A total of 10 011 comparisons were performed and six kinship pairs (HSP) were found among the specimens. Based on their biological data (size, weight), all but one belongs to the same cohort and should not be considered for CKMR. Thus, one HSP was found among the specimens. The next step consisted on the analysis of the individuals showing high relatedness, to determine, based on their biological data (size, weight, age if known) if they belong to the same or different cohorts. All but one kinship pair (HSP) in our analysis clearly correspond to immature specimens of the same cohort (Figure 19) and should not be considered for CKMR.

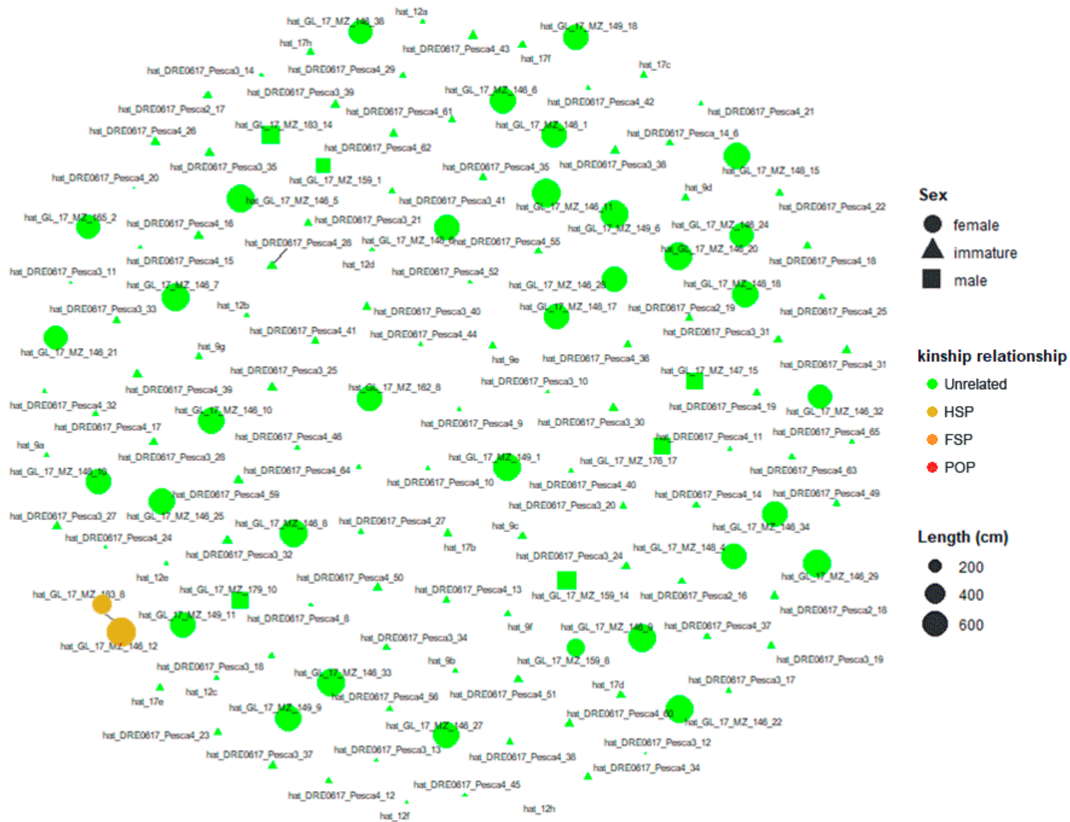



**Figure 19. Kinship network based on the kinship coefficients inferred from the combined application of the methods MLE and MoM. Each geometric figure represents a specimen. The shape of the geometric figure indicates the sex of the specimen (circle – female, triangle –male) while their length is proportional to the size of the geometric figure. The type of kinship is indicated by the colour: green – unrelated (U), red – parent-offspring (POP), orange full-sibling (FSP), and yellow – half-siblings. The average kinship coefficients for a POPs/FSPs is 0.25 while FSPs display, on average 0.125. More distant kinship relationships (cousins, etc.) were not considered in our analysis. As expected in a wild fish population, most individuals are unrelated (green). One kinship pair, a HSP, was found among the 142 specimens of hake analysed.**

### 2.2.4. RAD-SEQ FOR CONNECTIVITY AND STOCK BOUNDARIES

Methodology

The RAD-Seq data produced to test the CKMR methodology was reanalysed to assess its robustness, accuracy and technical power to quantify connectivity and biological stock boundaries, using hake (North Sea, Balearic Islands and Galician shelf) as a case study. The method was not applied to cod or ballan wrasse as only data from a unique population was available and the study of connectivity and stock boundaries requires the exploration of two or more. Starting DNA samples were obtained from fin clips of hake collected at three different locations – North Sea, Galician shelf and Balearic Islands. RAD libraries were prepared in house and sequenced on a HiSeq X platform (Illumina) in 150-bp paired reads. Obtained sequences were processed in Stacks 2.0, following several steps; demultiplexing and cleaning, identification of loci, creation of a catalogue and matching of the individual's sequences, assembly of paired-end contigs, calling of variant sites in the population and individual genotyping. Next, we identified the set of informative SNPs among the complete set of variant sites using the module "populations" in Stacks2. The following filtering was applied: variant sites had to be present in, at least, 80% of individuals or more in each of the three populations and only locus with a maximum observed heterozygosity of 0.70 and a minimum minor allele frequency of 0.05 were processed. First, we used Bayesian clustering and discriminant analysis to visualize differences between populations. We analysed the variance (FST) in neutral and adaptive bi-allelic SNPs to explore (1) population structure and (2) migration rates. Both SNP types are useful to explore patterns of population differentiation. Most SNPs in a given genome are expected to be neutral at the population-level, being characterized by balanced allele frequencies. In contrast, adaptive or outlier SNPs display a reduction in levels of polymorphism with one of the alleles being present in the majority of the individuals at one particular location, while the other allele is rare. This bias indicates that the common allele is advantageous and is, therefore, under selection at this particular location. Thus, outlier SNPs provide a powerful way to unveil local adaptation and migration. To detect loci putatively under divergent selection and identify alleles associated to each geographic region, we used two different outlier tests, Bayescan and OutFlank. Following a conservative strategy, we compared the two sets of potential outliers detected by Bayescan and OutFlank to retain only those common between both approaches and eliminate potentially false positives. This set of common SNPs was used to identify potential migrants and admixed individuals with the software STRUCTURE.

Results

The results in this section have been summarized from section 2 of Deliverable 2.2., *Pilot studies comparative analysis*.

The analysis of the three populations (North Sea, Galician shelf and Balearic Islands) with Stacks2 produced a set of 12 945 bi-allelic SNPs that were analysed using Bayesian clustering and Discriminant analysis of principal components (DAPC) to infer genetic differentiation among them (Figure 20 and Figure 21). All three populations had a differentiated genomic signature but surprisingly, differentiation between the North Sea and the Galician shelf population was significantly stronger than differences between the North Sea and the Balearic Islands population with mean FST values of: 0.014 (Galician shelf-Balearic Islands), 0.011 (Galician shelf-North Sea) and 0.004 (Balearic Islands-North Sea).

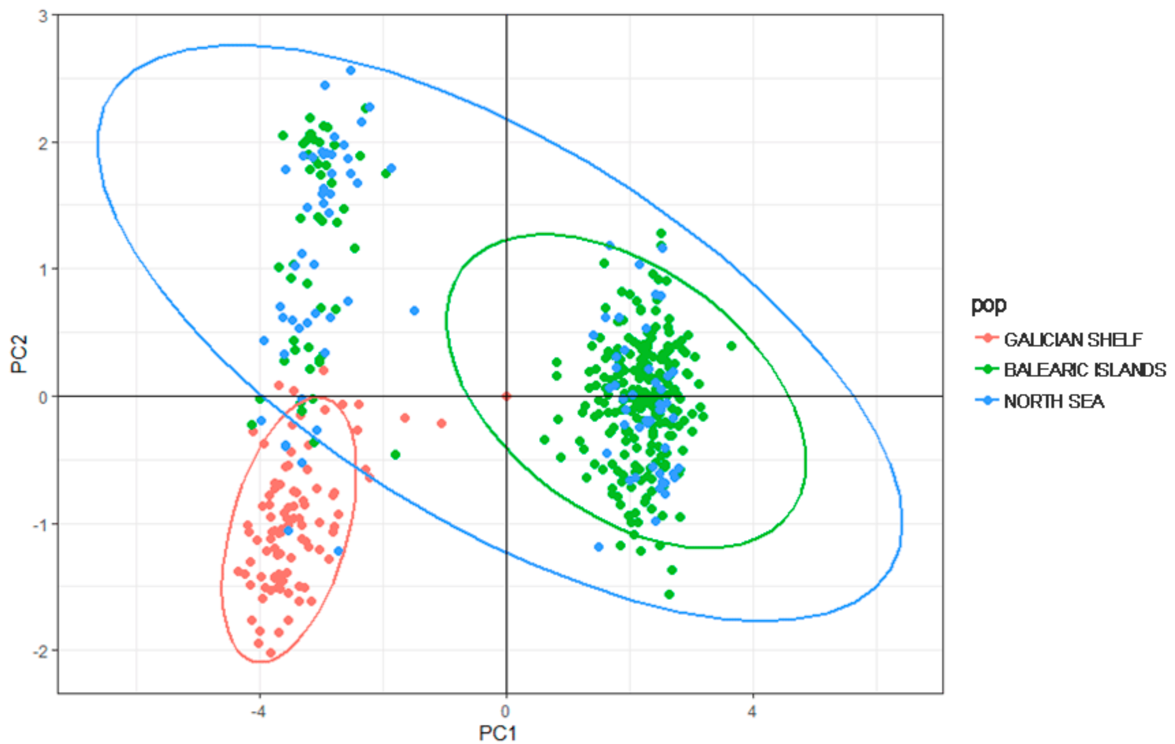



**Figure 20. Population structure of hake populations from the North Sea, the Balearic Islands and the Galician shelf detected by principal component analysis along axes PC1 and PC2 that are used to infer the number of clusters of genetically related individuals, based on 12957 SNP markers. Each dot represents an individual and three different colours are used to differentiate the populations: red, green and blue represent the Galician shelf, Balearic Islands and North Sea populations, respectively. Ellipses illustrate the distribution of individuals within groups.**

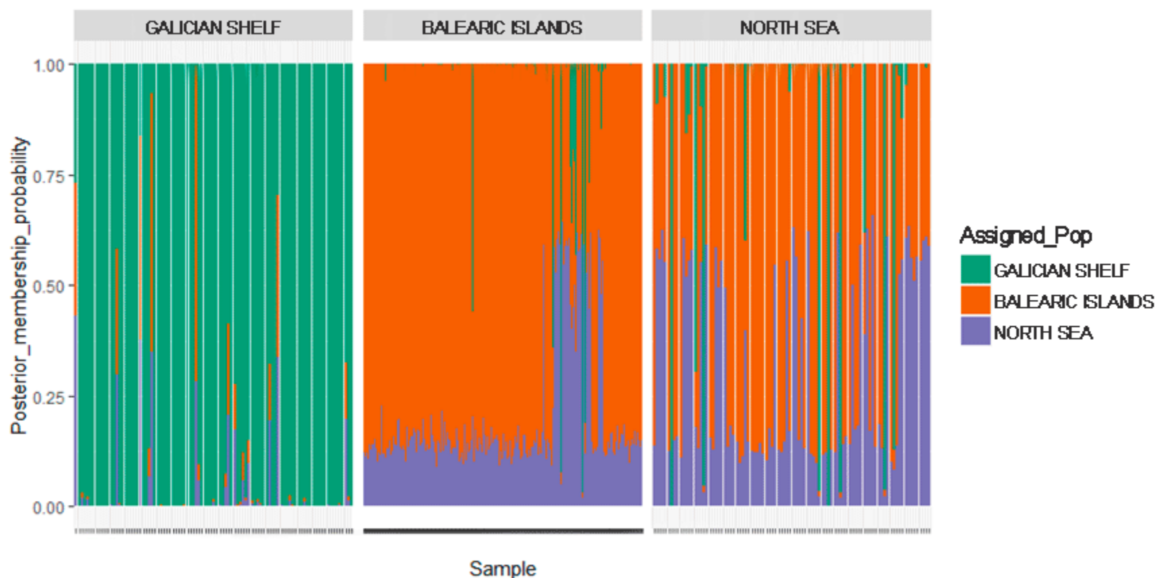



**Figure 21. Plot of posterior probabilities of assigning individual membership to their original populations obtained by Bayesian clustering. Each vertical line represents an individual and the colours refer to the different clusters.**

### 2.2.5. RAD-SEQ FOR STOCK SUBSTRUCTURE

Methodology

The RAD-Seq data produced to test the CKMR methodology was reanalysed to explore its capacity and accuracy to detect fine-scale geographic substructure within three populations of hake (North Sea, Balearic Islands and Galician shelf), one of cod and one of ballan wrasse. Starting DNA samples were obtained from fin clips of hake collected at three different locations – North Sea, Galician shelf and Balearic Islands of cod collected in the North Sea and of ballan wrasse from the Galician shelf, collected at two locations (Vigo and Malpica; ~120 km apart). RAD libraries were prepared in house and sequenced on a HiSeq X platform (Illumina) in 150-bp paired reads. Obtained sequences were processed in Stacks 2.0, following several steps; demultiplexing and cleaning, identification of loci, creation of a catalogue and matching of the individual's sequences, assembly of paired-end contigs, calling of variant sites in the population and individual genotyping. Next, we identified the set of informative SNPs among the complete set of variant sites using the module "populations" in Stacks2. The following filtering was applied: variant sites had to be present in, at least, 80% of individuals or more in each of the three populations and only locus with a maximum observed heterozygosity of 0.70 and a minimum minor allele frequency of 0.05 were processed. First, we used Bayesian clustering and discriminant analysis to visualize differences within populations. We analysed the variance (Fst) in neutral and adaptive bi-allelic SNPs to explore population sub-structure. Both SNP types are useful to explore patterns of differentiation within a given stock. To detect loci putatively under divergent selection and identify alleles associated to genetically "heterogenous" units within each geographic region, we used two different outlier tests, Bayescan and OutFlank. Following a conservative strategy, we compared the two sets of potential outliers detected by Bayescan and OutFlank to retain only those common between both approaches and eliminate potentially false positives.

Results

The results in this section have been summarized from Section 3 of Deliverable 2.2., *Pilot studies comparative analysis*.

NORTH SEA COD SUBSTRUCTURE

The analysis of the North Sea cod with Stacks2 produced a set of 25,571 bi-allelic SNPs that were analysed using Bayesian clustering and Discriminant analysis of principal components (DAPC) to infer genetic differentiation within the stock. The clustering algorithms did not reveal any hidden substructure within the stock (Figure 22).

The search of adaptive SNPs using Bayescan revealed, moreover, no outliers within the stock, reinforcing the lack of sub-structuring within the analysed samples and indicating a discrete population (Figure 23).

NORTH SEA, BALEARIC ISLANDS AND GALICIAN SHELF HAKE POPULATIONS SUBSTRUCTURE

For hake, the three populations (North Sea, Galician shelf and Balearic Islands) were analysed separately with Stacks2, producing, 11 608, 10 084 and 10 360 bi-allelic SNPs, respectively. Bayesian clustering did not infer any genetic sub-structuring within any of the three stocks (figures not shown). Bayescan revealed no outliers within any of the stocks (Figure 24), reinforcing the lack of sub-structuring within the analysed samples and indicating discrete populations in all three locations.

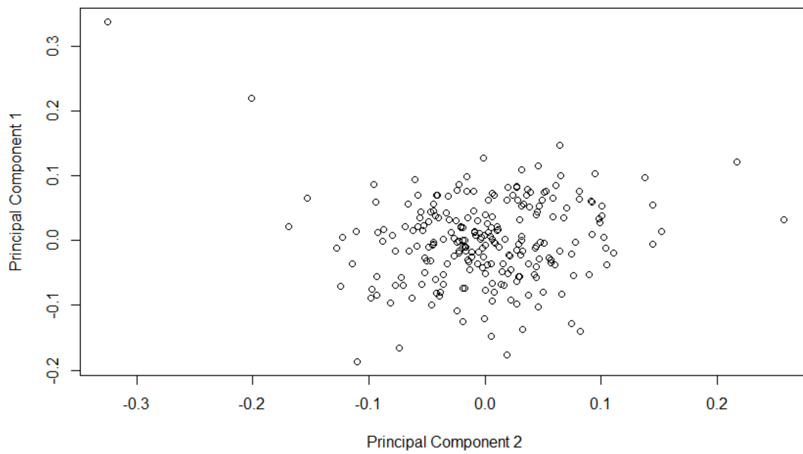



**Figure 22. Population structure of North Sea cod stock detected by principal component analysis along axes PC1 and PC2 that are used to infer the number of clusters of genetically related individuals, based on 25,571 SNP markers. Each dot represents an individual. No distinctive groups are detected indicating a unique genetic stock.**

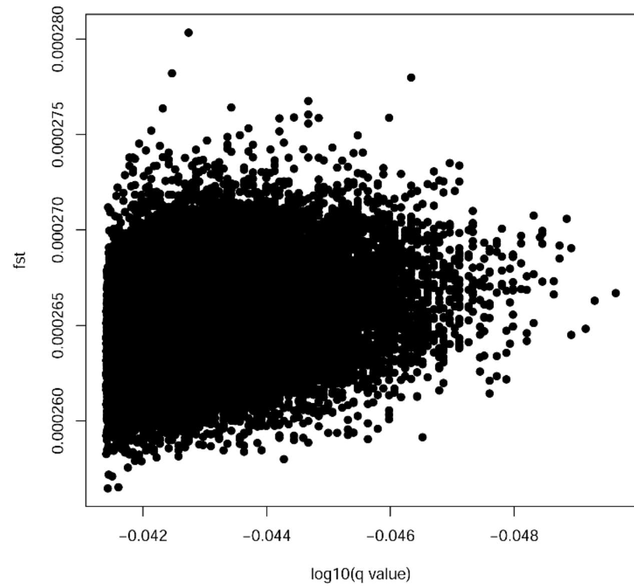



**Figure 23. Bayescan plot of 25,571 SNPs according to FST and log10 (q-value) in genome scan analysis of 235 individuals of cod from the North Sea. Each dot represents a SNP locus and no outliers were identified.**

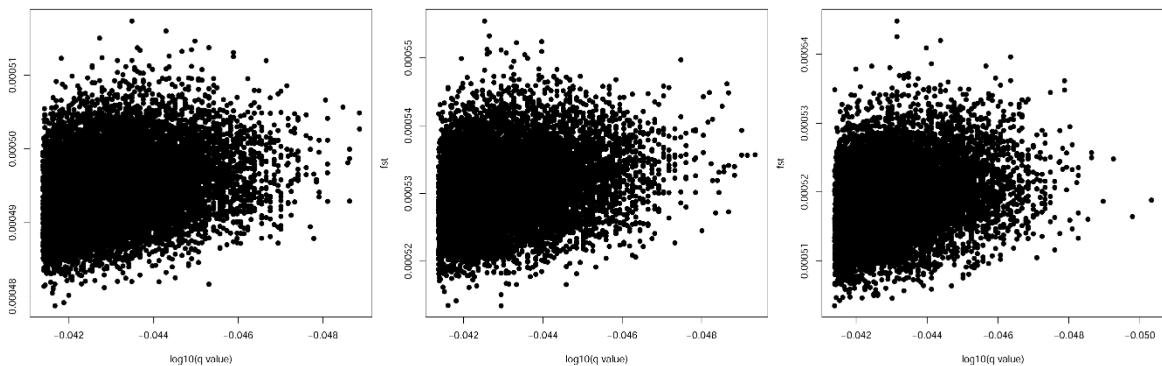


**Figure 24. Bayescan plots of 11,608 (left), 10,084 (center) and 10,360 (right) SNPs according to FST and log10 (q-value) in genome scan analysis of 94, 142 and 281 individuals of hake from the North Sea, the Galician shelf and the Balearic Islands, respectively. Each dot represents a SNP loci and no outliers were identified in any of the three stocks.**

BALLAN WRASSE SUBSTRUCTURE

The analysis of the ballan wrasse with Stacks2 produced a set of 35,096 bi-allelic SNPs that were analysed using Bayesian clustering and Discriminant analysis of principal components (DAPC) using all loci to infer genetic differentiation within the population. The clustering algorithms did not reveal any hidden substructure between both locations sampled (Vigo (Vi) and Malpica (Ma)), however a clear separation within the population was detected at both sites. Segregating individuals corresponded to two different colour phenotypes (plain and spotted) (Figure 25).

Clustering of samples by Hardy–Weinberg equilibrium (HWE) with Structure revealed that values of K = 2 was the most likely configuration according to the cross-entropy criterion. Genetic clusters identified by Structure are congruent with PCA results, separating both color phenotypes, plain and spotted, but not the locations (Vigo and Malpica) (Figure 26).

The search of adaptive SNPs using Bayescan and Outflank revealed, moreover, a large set of SNP loci under selection within the population, when both phenotypes - plain and spotted - were compared. These results reinforce the sub-structure within the analysed samples and clearly indicate two discrete components within the population (Table 5).

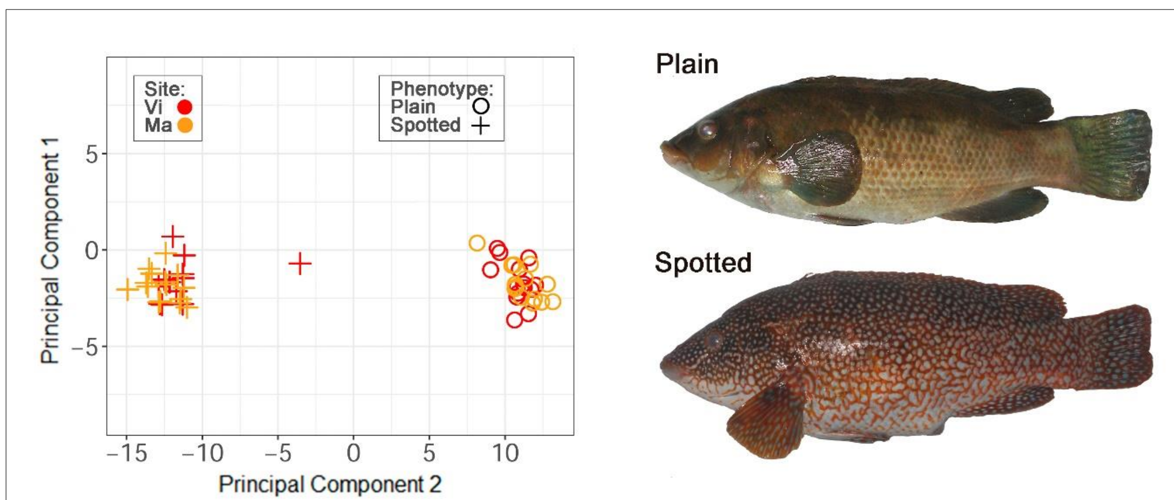



**Figure 25. Population structure of ballan wrasse detected by principal component analysis along axes PC1 and PC2 that are used to infer the number of clusters of genetically related individuals, based on 35,096 SNP markers. Each dot represents an individual. Two distinctive groups are detected indicating two genetic stocks. Samples were collected in two locations, Vigo (indicated Vi) and Malpica (indicated Ma) that are separated by ~120 km. The two differential components correspond to two differential color phenotypes (rigth panel, plain (above) and spotted (below)).**

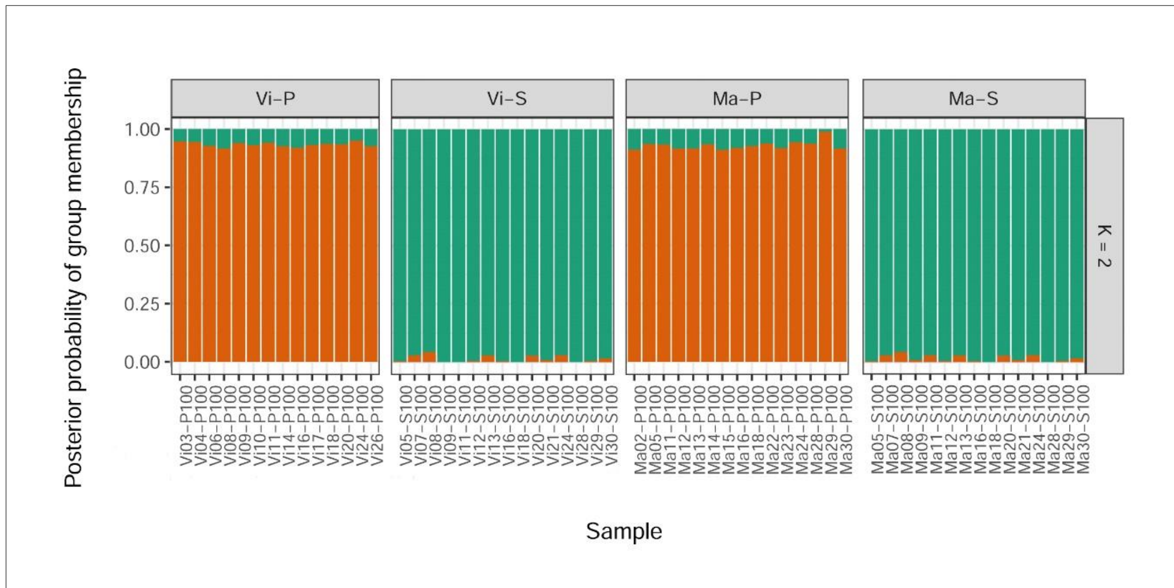



**Figure 26. Individual ancestry coefficients of 60 individuals of ballan wrasse from the Galician shelf for K = 2 are shown in the figure. Each bar represents an individual. Vi indicates Vigo, Ma indiates Malpica, P denotes plain color morphotype and S spotted color morphotype.**

**Table 5. Summary of pairwise comparisons between plain–spotted phenotypes of *Labrus bergylta* across two sites in the Galician shelf. Ma indicates Malpica, Vi indicates Vigo. Pheno indicates phenotype –plain or spotted. Outflank and Bayescan indicate the number of outliers found in the comparison. Differentiation between both phenotypes is indicated by $F_{ST}$ and Dist denotes distance (km) between the sites.**

| Site 1 | Site 2 | Pheno 1 | Pheno 2 | Outflank | Bayescan | $F_{ST}$ | Dist |
|---|---|---|---|---|---|---|---|
| **Malpica (Ma)** | Malpica (Ma) | Spotted | Plain | 280 | 199 | 0.0175 | 0 |
| **Vigo (Vi)** | Malpica (Ma) | Plain | Spotted | 273 | 188 | 0.0167 | 127 |
| **Vigo (Vi)** | Malpica (Ma) | Spotted | Plain | 192 | 139 | 0.0159 | 127 |
| **Vigo (Vi)** | Vigo (Vi) | Spotted | Plain | 184 | 116 | 0.0150 | 0 |

### 2.2.6. RAD-SEQ FOR SEX ASSIGNMENT

#### Methodology

The Pilot study tested the potential of this technique to reveal sex-specific genetic markers, i.e., markers found in one sex but not the other, only in hake. The method was not applied to cod since a sex marker, for this species, was published in 2016 (Star et al., 2016). The ballan wrasse, on the other hand, is a hermaphrodite species and these are characterized by a unique common genome for both sexes.

Starting DNA samples were obtained from fin clips of hake collected at three different locations – North Sea, Balearic Islands and Galician shelf. RAD libraries were prepared in house and sequenced on a HiSeq X platform (Illumina) in 150-bp paired reads. Obtained sequences were processed in Stacks 2.0 using two different pipelines, specifically designed to search for two types of sex markers:

**a)** Regions present exclusively in one of the sexes

**b)** Regions present in both sexes, which contain a SNP that is sex-specific

Both pipelines started with the application of a de novo pipeline consisting in four major stages. First, reads were demultiplexed and cleaned, then loci were identified, a catalogue was created and individuals were matched against this catalogue. Finally, paired-end contigs were assembled and merged; variant sites were called in the population and genotypes in each sample.

The search of both types of markers **a)** and **b)** was conducted in the three populations separately (North Sea, Balearic Islands and Galician shelf) using a balanced number of males and females.

The search of type **a)** markers was performed with the software RADSex (Feron et al., 2021) and the workflow consisted on four steps: 1) the identification of RAD-Seq reads from all individuals, 2) the calculation of the depth of each sequence in each individual, 3) the computation of their distribution between both sexes and 4) the extraction of the markers significantly associated with sex. A minimum read depth threshold of 5 and a p-value of association with sex of 0.05 was applied.

For the search of type **b)** markers, the pipeline consisted on the use of the population's module implemented in Stacks2. A stringent filtering criterion (minimum minor allele frequency $< 0.5$, presence in at least 90% of the individuals and maximum observed heterozygosity $> 0.9$) was applied to identify private alleles (SNPs specific to one sex) in each population.

### Results

The results in this section have been summarized from Section 4 of Deliverable 2.2., *Pilot studies comparative analysis*.

The search of type **a)** markers, corresponding to regions present exclusively in one of the sexes, yielded the following results in each of the three hake populations:

- In the North Sea population, RADSex detected 19,702,186 regions of which 2.6% (520,149) had a minimum depth of 5 reads and a positive association with sex according to a Pearson's chi-squared test of independence. After Bonferroni correction, one of these regions was found to be significantly associated with the male sex ($p < 0.05$). This marker can classify correctly the 91% of the males and79% of the females in the North Sea (Figure 27).
- In the Balearic Islands population, 22,441,872 markers were detected by RADSex of which 540,369 had a minimum depth of 5 reads and positive association with sex according to a Pearson's chi-squared test of independence. After Bonferroni correction none of them was found to be significantly associated with the either sex ($p < 0.05$).
- In the North Atlantic population, 90% of the individuals studied were immature. Thus, sex could not be assigned, preventing their classification and the search of a sex marker in this population.

We tested whether the marker isolated in the North Sea population was also sex-linked in the Balearic Islands population. The search of the region in this population revealed its presence in 23 of 33 M and 23/33 F, indicating that is a specific of the North Sea population.

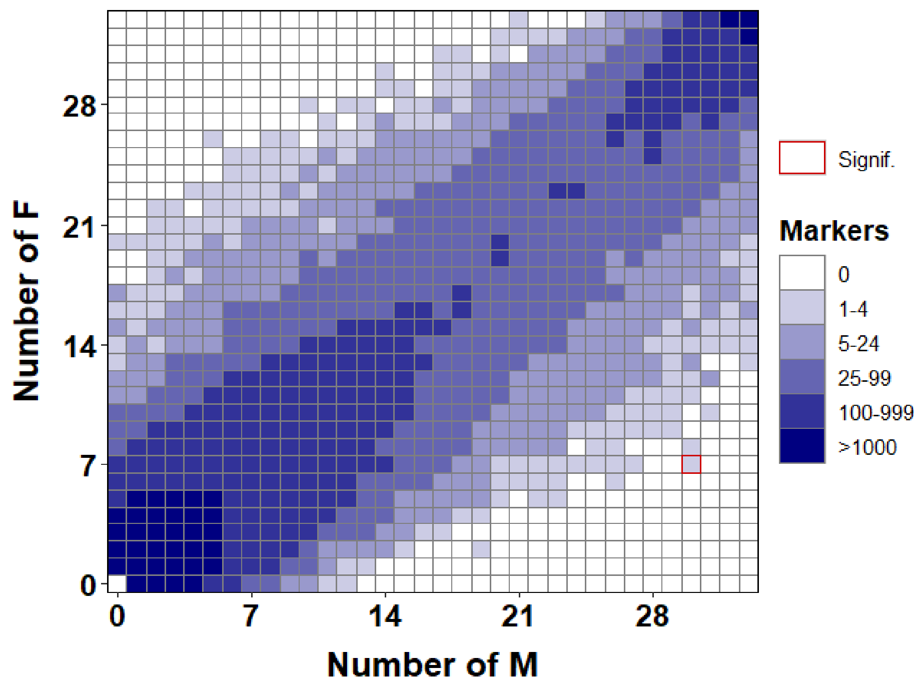



**Figure 27. Heat map of RADSex markers in 33 males and females of the North Sea hake population showing a higher coverage. Positions with significant association with sex are indicated with a red box on the lower right corner and upper left corner for male and female respectively. The intensity of the blue color corresponds to the number of markers.**

**Table 6. Number and characteristics of the private alleles found in the North Sea hake population (number of individuals, variance, standard error, mean frequency of the most frequent allele at each locus in this population)**

| **Variant positions** | | | | | |
|---|---|---|---|---|---|
| **Pop ID** | Private | Num_Indv | Var | StdErr | P |
| **M** | 64 | 57.57277 | 3.17622 | 0.01136 | 0.85066 |
| **F** | 1 | 31.38004 | 1.18599 | 0.00646 | 0.84827 |

The search of type **b)** markers, consisting in regions containing SNP-linked markers produced, in turn, the following results:

- In the North Sea population, our analysis identified 64 private alleles in males and 1 in females (see Table 6)
- In the Balearic Islands population, 0 private alleles were detected, even when the parameter -r (minimum percentage of individuals required to process a locus) was reduced to 0.8, 0.6 and 0.5.
- The use of mostly immature individuals in the Galician shelf population prevented the search in this population

None of the private alleles detected in the North Sea population showed a significant association with sex in the Balearic Islands population (the search could not be conducted in the Galician shelf population due to the immature stage of the majority of the individuals analysed) and no specific private alleles were detected in specimens from the Balearic Islands population.

### 2.2.7. EPIGENETIC AGE DETERMINATION

Methodology

Starting DNA samples were obtained from fin clips of cod collected in the North Sea. We used bisulfite-converted restriction site associated sequencing (bis-RAD-Seq) to build an epigenetic clock. This genome-wide technique consists, briefly, of adding a bisulfite conversion step to the classic RAD-Seq technique, before amplification and sequencing. The analysis of the resulting sequences consisted of several steps, including a quality filtering of the raw reads, PCR duplicate removal, the creation of a bisulfite converted reference genome and the extraction of the methylation calls based on this genome. The calls were subsequently analysed with the package MethylKit that filters, normalizes and unites the data based on coverage. Resulting data was then processed for machine learning for age prediction. The final model was selected as the one minimizing the prediction error and used to evaluate the correlation between actual and predicted age.

Results

The results in this section have been summarized from Section 5 of Deliverable 2.2. *Pilot studies comparative analysis*.

Of the 60 initial samples, 54 samples (90%) passed the quality criteria and had no missing DNA methylation data. In these 54 samples bis-RAD-Seq returned 67,549 CpGs common in all samples. Pre-processing to prepare the data for applying a machine learning model involved several steps. First, CpGs with zero or near-zero variance were filtered. These filtering steps retained a total of 65,841 CpGs. From these, CpGs with methylation significantly correlated with age were selected and retained those with $P < 0.05$ Pearson correlation, which were 12,053 CpGs. Next, the CpGs highly correlated between them ($r > 0.9$, Pearson correlation) were filtered, leaving only 125 CpGs.

To build the machine learning model, we used penalized regression which is a type of linear regression that is penalized for having too many variables and performs well in a situation with a number of variables higher than the number of samples. Three types of penalized regressions were used which differ in the amount of penalty they impose: LASSO (LM), Ridge (RM) and Elastic net (EM). From these methods, we selected LM as the type of regression to build the epigenetic age predictor, since it yielded the lowest number of coefficients, that is, 26 CpGs. Of the 26 CpG positions in the cod genome used to build the epigenetic clock, 8 CpGs showed negative correlation with age while the remaining 18 CpG showed positive correlation.

Thus, penalized regressions (LASSO regression) within a leave-one-out cross-validation context resulted in the construction of a cod epigenetic clock of high accuracy when regressed against chronological age determined by otolith analysis ($r > 0,95$; $P < 0.001$) and capable to estimate age with a precision of ~8 months (Figure 28).

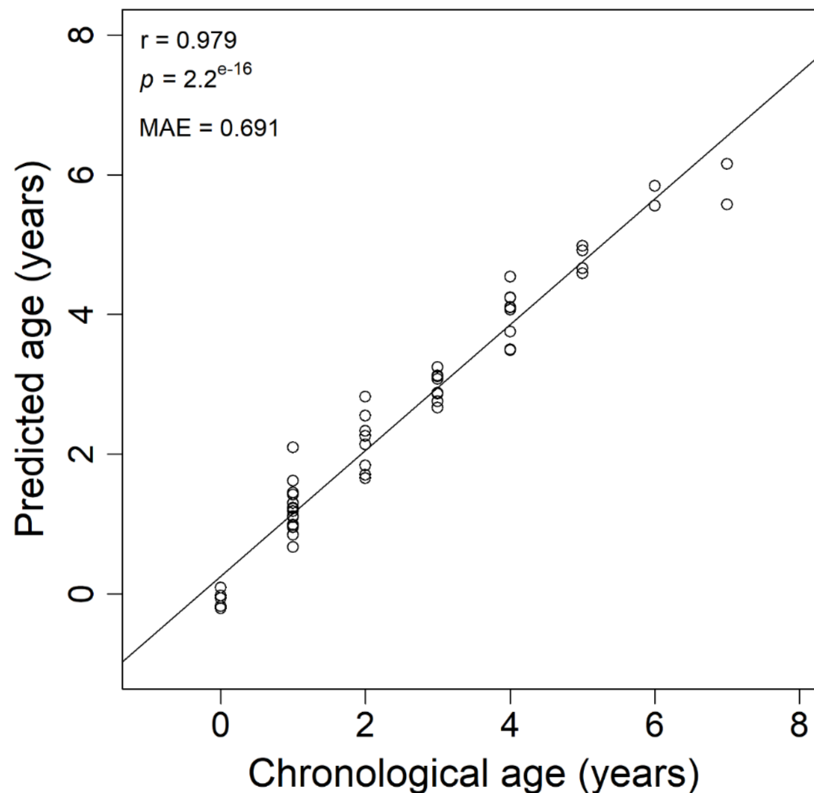



**Figure 28. The Atlantic cod epigenetic clock for age estimation. Number of CpGs retained: 26. Penalized regression: LASSO. Accuracy (Pearson correlation): r = 0.979, P = 2.2. e-16. Precision: leave-one-out-cross-validation (LOOCV). Maximum absolute error (MAE) = 0,691 years ( = 252 days; 8,2 months).**

### 2.2.8. Environmental DNA

#### Methodology

A total of 32 sediment and water samples were collected at 16 stations in the North Sea during a sampling mission on FRV Walther Herwig (mission number WH428) from 8 July to 3 August 2019 conducted by the Thünen-Institute in the framework of the FishGenome Project.

Water and sediment sampling strictly followed the FishGenome "D2.1c_Experimental design and protocols for eDNA". eDNA extraction was performed following the method described there. Briefly, water samples from single Niskin bottles were directly filtered after each CTD cast on 0.45 µm filters. Filters were stored at -20°C until DNA extraction. Accordingly, sediment samples were stored in sterile 100 ml tubes filled with ethanol (99%) at -20°C until extraction within 3 months after collection. Regarding labelling, we refer to the sampling regions as "boxes" (i.e., box A) and the location within the box as "station", with the date of collection.

The results from the eDNA samples collected in the Balearic Islands in the Mediterranean have been excluded from the Final Report due to their inconsistency. The output obtained is contradictory not only with the findings of the North Sea analysis but also with the large body of literature on eDNA research. Although the efficacy of a protocol can differs significantly in response to local environmental conditions, there are two well-established facts. First, eDNA systematically reports the same or higher biodiversity than traditional approaches (e.g.: Stoeckle et al., 2021, Zhou et al., 2022). Second, fish eDNA is found in higher concentrations in sediment than in water due to lower decay

rate of the former (e.g. Turner et al., 2015; Sakata et al. 2020). Traditional trawling around the Balearic Islands detected 40% more species than eDNA from water with only 12% in common whereas no fish could be detected from the analysis of eDNA from sediments. These inconsistencies stem from a combination of insufficient fine-tuning of the protocol and possibly other methodological issues such as the presence of PCR inhibitory compounds in the eDNA samples. Working with eDNA always requires a strict interpretation of the data of the results obtained, since the tiny amounts of material involved in eDNA detection make its analysis prone to methodological (primer specificity, completeness of databases), biological (differential shedding and decay rates) and environmental (transport, influence of temperature, pH,…) biases. After a careful analysis, we have considered that confronting the results obtained in the Balearic Islands and the North Sea are relevant to extract lessons about the difficulties of standardization and implementation of a unique protocol at different laboratories and this is reflected accordingly in the "lessons learnt" section. However, reporting these analyses in the results section would be misleading, as clear methodological issues have been detected.

Two different methodologies were tested:

A metabarcoding approach to estimate biodiversity (qualitative). The optimisation of the 12s rRNA MiFish metabarcoding protocol published by Miya et al. (2015) as well as the subsequent analyses were performed in accordance with the "D2.1c_Experimental design and protocols for eDNA", available in the FishGenome cloud.

A qPCR quantitative approach to estimate the biomass of Atlantic cod (*Gadus morhua*) in the North Sea. To evaluate the performance of eDNA methods in the quantification of cod in the North Sea, we tested different polymerases, three different fluorescence technologies (MGB, TaqMan® and LNA) and different primers/probes covering 8 genes, in addition to different annealing temperatures (Figure 29). Next, we adopted a second strategy based on the concept of analytical sensitivity addressing three analytical values:

The Lower Limit of Blank (LoB) corresponding to the lowest DNA copy number that can be detected with 95 percent confidence above the concentrations of the blank or non-target species.

The Limit of Detection (LoD) corresponding to the number of DNA copies that can be detected with 95 percent confidence, but which cannot be quantified under the experimental conditions.

The Limit of Quantification (LoQ) defined as the smallest number of DNA copies, which can be measured and quantified with a defined precision and accuracy under the experimental conditions by a given method (Armbruster & Pry, 2008). LoQ can never be below LoD.

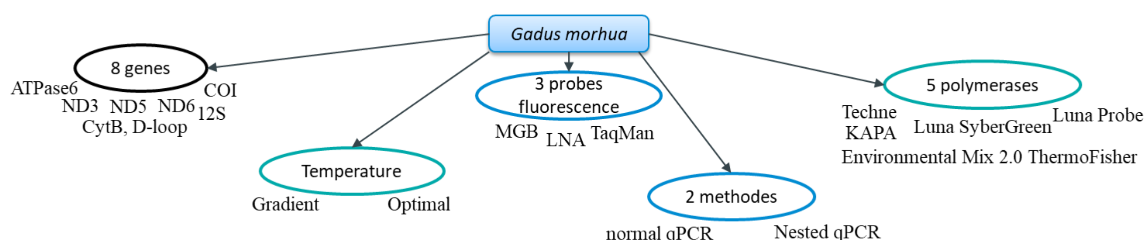



**Figure 29. Overview of the strategy followed to develop a highly specific qPCR test for Atlantic cod (*Gadus morhua*).**

## Results

The results in this section have been summarized from Section 6 of Deliverable 2.2., *Pilot studies comparative analysis*.

- Metabarcoding

Based on the objectives of this part of the project aiming at optimising and evaluating the performance of eDNA methods in revealing marine biodiversity in the North Sea using MiFish primers (Miya et al., 2015), the MiFish protocol was used as a template.

However, the protocol published by Miya et al. (2015) could not be reproduced despite various optimization attempts, because the original manuscript contained serious mistakes. After consultation with the corresponding author, Dr. Masaki Miya, Natural History Museum and Institute, Chiba, Japan, new and corrected primer sequences could be obtained. Dr. Miya also delivered a new protocol.

Once all the problems were detected and each step was fine-tuned, the metabarcoding approach was applied to all the samples collected during the survey. Sediment samples collected at 16 stations produced a total number of 7,208,776 high quality reads (3,604,388 reverse and 3,604,388 forward). From these, 230 fish species were identified (Figure 30). The results obtained from the sediments revealed large differences in read numbers among stations.

For water samples, a total number of 15,982,574 high quality reads (7,991,287 in each direction). From these, 75 fish species were detected. However, following Miya´s protocol, shark species were not detectable, while skates of the genus Raja were found at two stations. Another potential bias detected in the Pilot study was the the absence of 12S rDNA reference sequences for a number of fish species (such as the solenette (*Buglossidium luteum*)) in international databases like NCBI and ENA. This absence can lead to best hits to close species hampering the correct identification.

The percent of reads corresponding to fish was between 80% and 98% per sample, with an average of 90%. The negative controls did not show false positives. In all samples, but three, 0% of the reads corresponded to fish, while in these three 1.5% of the reads with had high similarity with fish sequences.

The capacity of the eDNA to reveal fish biodiversity in the North Sea was then tested by comparing the results obtained from metabarcoding to the trawling data collected during the survey. The overall results showed that eDNA from water samples can reveal the presence or absence of a given fish species with 90% reliability (Figure 30). Moreover, in general, metabarcoding was able to detect a higher number of fish species than trawling, pointing at the detection of eDNA traces of elusive species. The number of species detected by eDNA compared to conventional methods was, at least, three times higher (Figure 31).

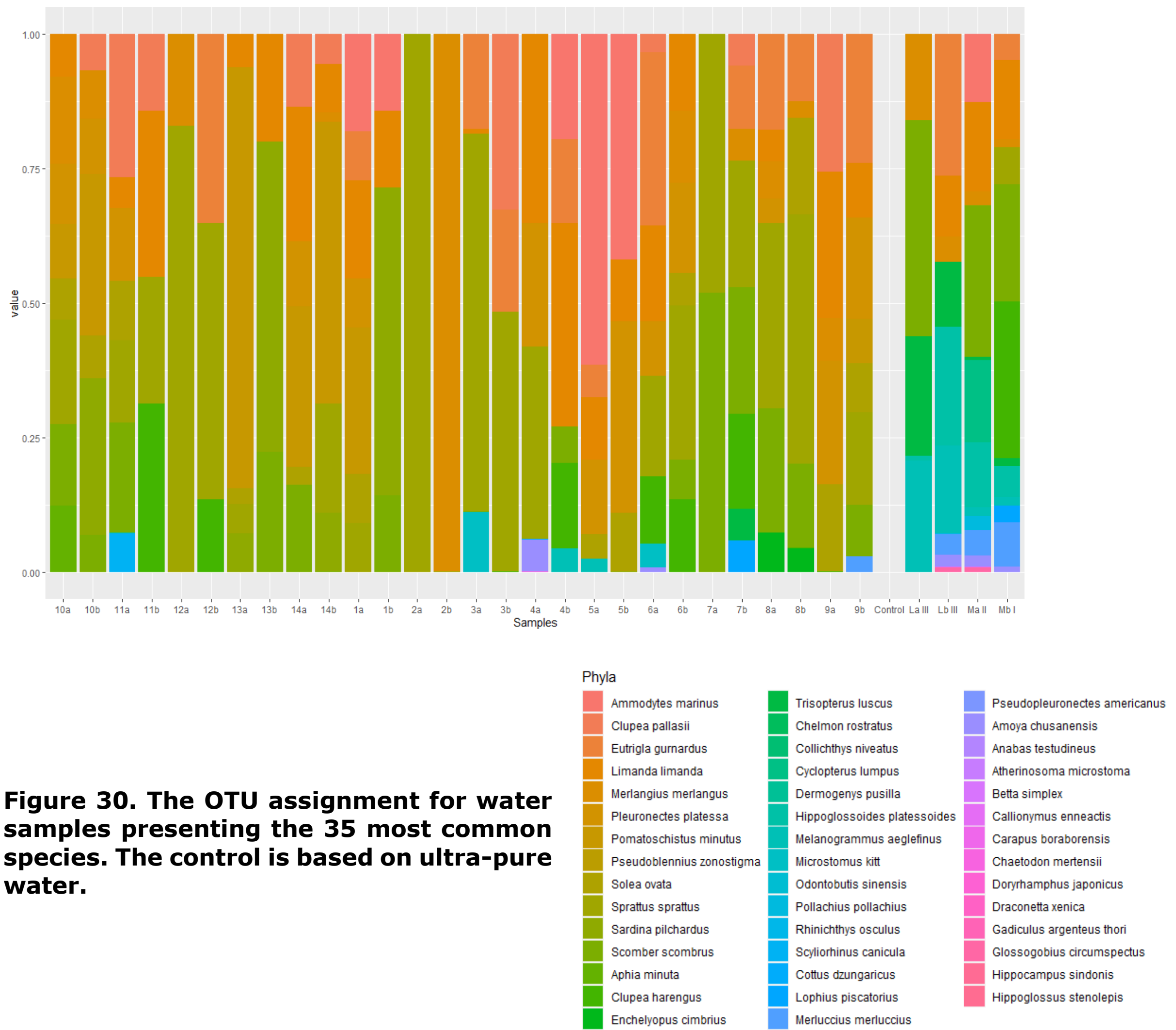



**Figure 30. The OTU assignment for water samples presenting the 35 most common species. The control is based on ultra-pure water.**

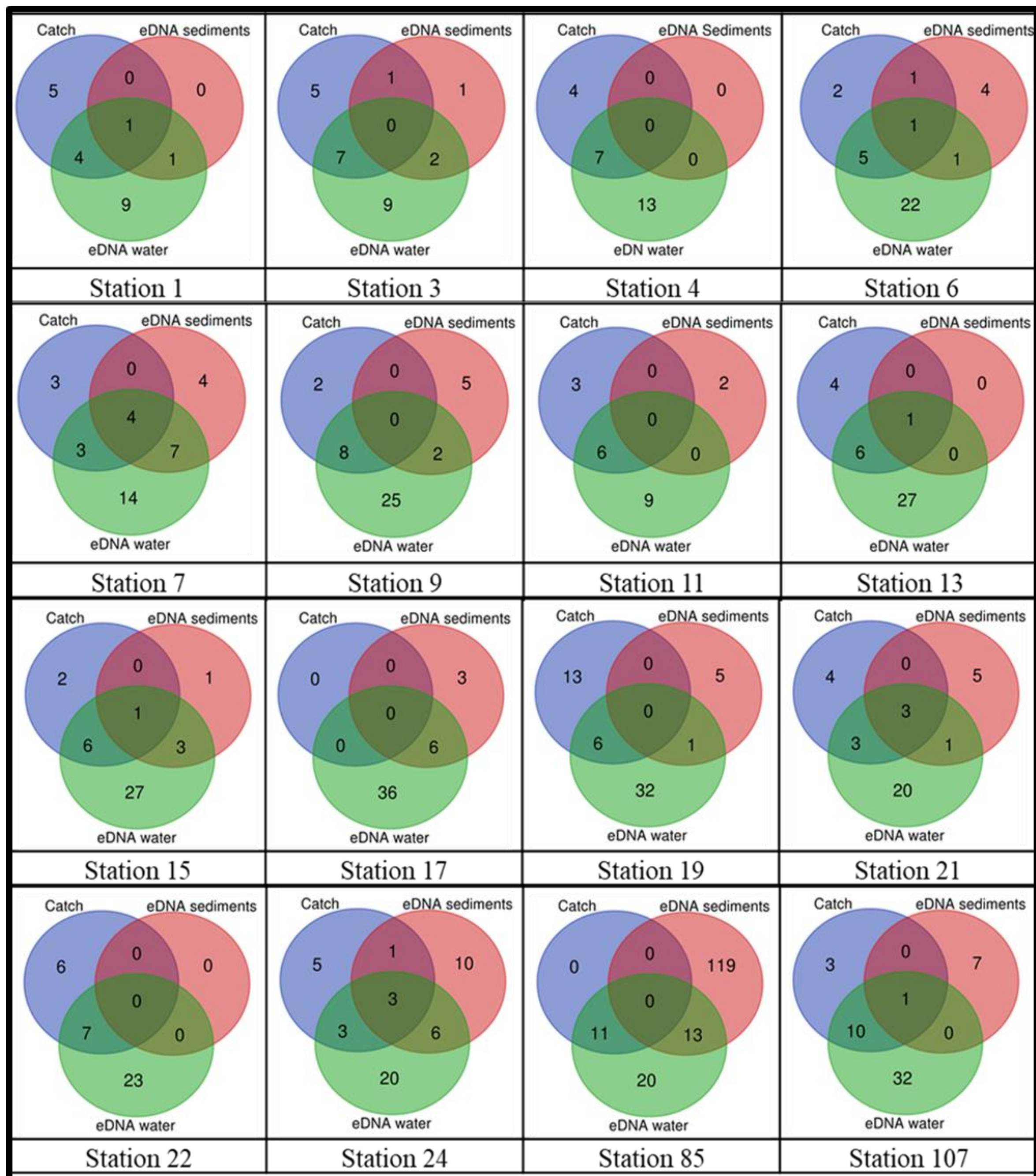



**Figure 31. Venn diagram showing the number of common and distinguishable species in trawl catches versus eDNA from sediment and water samples. 8 stations are presented. Sharks were principally not detected. Numbers for Station 1 are to be explained as: 9 species were only detected via eDNA analysis of water sample, 5 detected only by trawling and 4 species were detected by metabarcoding from water and also trawling, 1 species each was detected in trawling and in both eDNA metabarcoding techniques.**

- Quantitative approach (qPCR)

The first strategy tested, revealed cross-species amplifications of non-target gadid species and even other teleost families in all assays for all eight genes tested.

The second strategy was applied on 9 standard curve replicates, in addition to nine validation water samples providing significantly better results. The limit of detection was

as low as 10 copies per reaction. Therefore, the LoD value of our qPCR assay was 16.40 copies per reaction, equivalent to Ct value 35.4. (Figure 32).

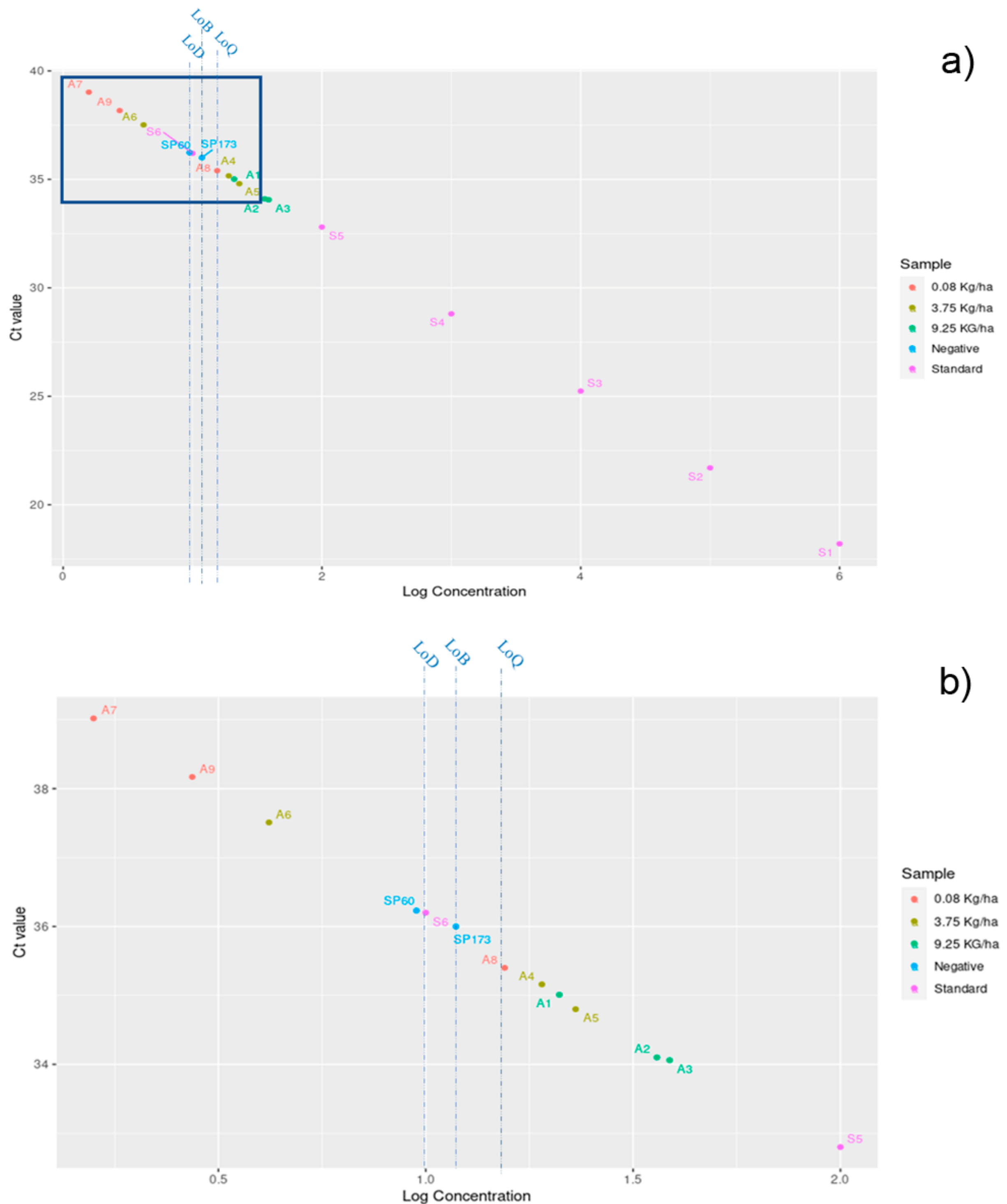



**Figure 32. Panel a) Standard curve of threshold cycle number (Ct values) plotted against the log concentration (copy number). Dark green dots represent 9 replicates for each dilution. In the black square (Figure -a-), the results from different qPCR assays are shown. Panel b) Detail of a) including standard (dark green dots) - the last dilution represented coincides with LoD value - non-target samples (red dots), eDNA water samples in triplicates (blue dots). Dotted lines indicate LoB, LoD and LoQ values. In this assay the standard equation is Ct=-3,57*(concentration)+39,73.**

Based on the standard curve for DNA copy number determination, we applied a 99% regression between copy number and Ct value revealing an efficiency of 90% for the cod qPCR test. The qPCR test was able to show a positive signal in the LoQ range with a confidence of 33% for a sample from a station where only 0.08 kg/ha (corresponds to 1 kg/hour trawling according to the standards of Baltic International Trawl Survey (BITS) program) of cod were caught by bottom trawling. For stations with cod catches larger than or equal to 3.25 kg/ha (45 kg/hour trawling), positive signals were detected in the LoQ range with 100% confidence. Stations with cod catches of 1.25 kg/ha (15 kg/hour trawling) and 1.0 kg/ha (12 kg/hour) revealed averages of 16.5 ± 5 and 8.0 ± 6 DNA copies per litre, respectively. At zero catch stations for cod, a median of 0 DNA copies per litre were observed. However, at 2 stations and for all 3 replicates of each of these stations, a positive signal in the qPCR was detected even in the absence of cod in the corresponding bottom trawls, resulting in a mean equal to 0.09 ± 0.02 DNA copies per litre. All negative controls for eDNA extraction batches (unfiltered membranes) and qPCR no template controls (NTCs) were negative.

Since three negative controls for sample collection (ultra-pure water rinsed from the used Niskin Bottles on board before each station) were found to be positive with very low DNA copies in station 98 in 9 replicates, we subtracted this number of copies from the results of real samples.

In contrast, the eDNA analyses of sediment samples did not reveal any positive signal. All sediment samples turned out to be negative (Figure 33).

To determine a quantitative relationship between the catch per unit effort (CPUE) per hour and the eDNA copy number for *Gadus morhua*, we developed a Matern 5/2 kernel model. The model was set up on the basis of a dataset published by Knudsen et al. (2019). Subsequently, we used our qPCR data for a second validation of the model. The model has a correlation between the predicted and observed data of 98%, with a MAE (Mean Absolute Error) equal to 107 and RMSE (Root Mean Squared Error) equal to 204 (Figure 34).

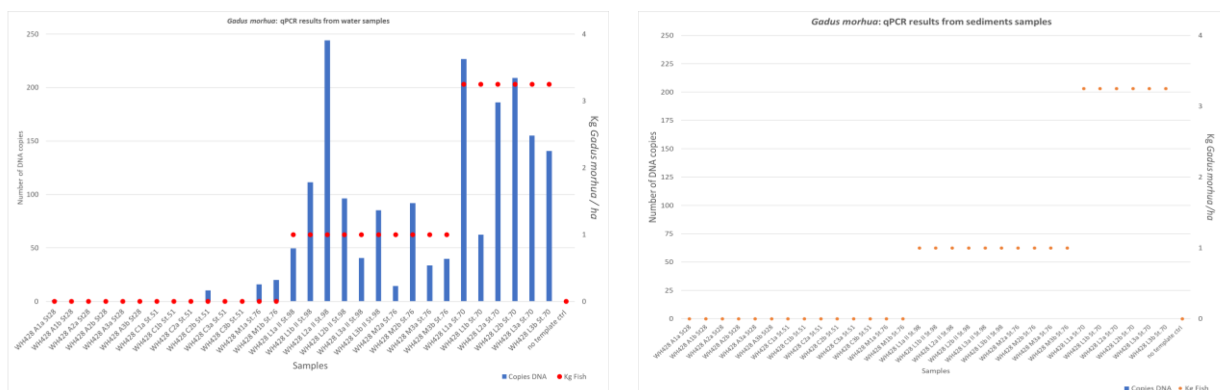



**Figure 33. eDNA sampling results from water (a) and sediment (b) in the North Sea (mission number WH428). In the WH428 cruise (in 2019), we sampled water and sediment from 10 stations in the North Sea, with three different positions (S: start, M: middle and E: end). The sample name is readable as: Cruise number (WH428), followed by the box titled by a letter (e.g., "A1a", where "A" is the box, "1" is the position during trawling, and "a" for the Niskin Bottle replicates.**

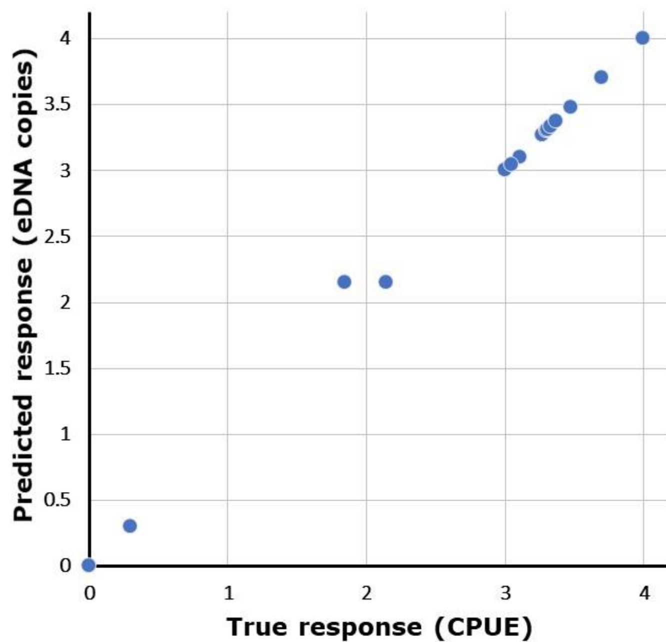



**Figure 34. Comparison between output of Catch results and output of GPR model for eDNA copies. (R = 96%, Error MAE = 0,002). The plot shows the correlation between predicted data by the model (Predicted response) and the real data collected from trawl (True response).**

## 2.3. Lessons learnt

### 2.3.1. CKMR

An ideal Close Kin Mark Recapture sampling design for estimating population size should have an even distribution of both reproductively mature and juvenile groups and both sexes evenly represented. However, regular surveys do not generally allow this type of sample selection, although ratios of the samples analysed by the FishGenome Consortium did not deviate greatly from these ratios. The methodology – RAD-Seq - employed was suitable, as it produced a number of genetic markers an order of magnitude higher than required for a robust kinship analysis for both species. Nonetheless, the number of kinship pairs found in our analysis was clearly insufficient to allow an accurate estimation of population parameters using CKMR.

Considering Bravington et al. (2016a, 2019) and Rodríguez-Ezpeleta et al (2020) recommendations and assuming an equal mix of juveniles and adults, a total of about 50 POPs and about 100 HSPs need to be detected to achieve a CV of less than 20% in the estimation of the parameters of interest with CKMR. All our analysis detected much lower numbers, indicating we would need to increase substantially the number of samples to be analysed to achieve the recommendations (Table 7).

Specifically, sampling sizes of populations from all three locations and for both species should be increased, at least, by 50x-100x, according to our results (and making a rough approximation as the number of kinship pairs detected is very low), assuming an ideal sampling, i.e., an even distribution of both reproductively mature and juvenile groups and even representation of both sexes. A robust CKMR analysis would have required further collection of tissue samples to reach a minimum of 23, 500 individuals for North Sea cod, 4,700 individuals for the North Sea hake, 14,050 individuals for the Balearic Islands hake and 14,200 individuals for the Galician shelf hake (in order to detect 50 POPs and 100 HSPs with a CV of 20%). Although these numbers are still reasonable to be analysed using the proposed technique combined with SNPs chips (RAD-Seq methodology is affordable for analysis of 5,000 individuals or less, but larger sample sizes would require developing SNP chips), the bottleneck here clearly lies on

achieving this number of specimens during current ongoing surveys. Considering that the FishGenome Consortium has analysed all the specimens caught during the regular IBTS and MEDITS surveys of cod and hake, the difficulties to reach required numbers are evident. Thus, **the applicability of CKMR to exploited species with large population sizes would require the collection of samples across several years and an intensification of current sampling surveys**. Multi-year samplings add uncertainty to the estimations, but CKMR could still be a valuable tool as a periodic independent estimate of population parameters to cross-validate population estimates by the regular stock assessments. Alternatively, the sampling scheme could be extended to involve boats/vessels from the commercial fleet in the collection of specimens.

**Table 7. Species (where NS indicates North Sea, BI corresponds to Balearic Islands and GS denotes Galician shelf), N samples (number of samples analysed), N kinship (number of kinship pairs found), N required 50 (number of samples required to achieve 50 POPs), N required 100 (number of samples required to achieve 100 HSPs). For cod NS, hake NS and GS, no POPs were found so we used a rough approximation to calculate the N required. We estimated that the frequency of POPs is approx. half the frequency of HSPs (*indicates that HSPs not provide useful information for CKMR)**

| Species | N samples | N kinship | N required 50 POPs | Increase | N required 100 HSPs |
|---|---|---|---|---|---|
| **Cod NS** | 235 | 1 (1FSP*+1HSP) | 23,500 | 100x | 23,500 |
| **Hake NS** | 94 | 2 (2HSPs) | 4,700 | 50x | 4,700 |
| **Hake BI** | 281 | 3 (1POP+2HSP) | 14,050 | 50x | 14,050 |
| **Hake GS** | 142 | 1 (1HSP) | 14,200 | 100x | 14,200 |

### 2.3.2. RAD-SEQ FOR CONNECTIVITY AND STOCK BOUNDARIES

Fisheries management inherently relies on the spatial extent over which stocks are assessed and accordingly, requires a precise delineation of stock boundaries to be effective. On the other hand, connectivity informs about changes in distribution and migration movements among stocks/populations. Understanding these patterns is also essential for an adequate exploitation of fish stocks, since the degree of connectivity among adjacent populations can affect population persistence, productivity, resilience and response to exploitation. Ignoring stock delineation and population genetic structure in fisheries management may result in local depletion through overexploitation. Although several methodologies have been proven useful for exploring population delimitation and connectivity, they are not used in current fisheries assessments. In practice, a variety of partly conflicting factors are used to delineate these entities, such as biological, geographical, economic, social or even political factors, leading, in many cases, to a mismatch between biological and fisheries management units (Reiss et al. 2009). Moreover, understanding patterns of connectivity and population boundaries in the marine realm is essential for carrying out accurate CKMR studies.

Results obtained by applying the protocol proposed in the FishGenome project in the Pilot studies, provide insight on the robustness, accuracy and technical power of the RAD-Seq methodology, to explore the spatiotemporal variability and quantify connectivity and biological population boundaries, using hake as a case study.

RAD-Seq data provides unprecedented access to survey genome-wide diversity and can generate thousands of single nucleotide polymorphisms (SNPs) markers with direct application to management and conservation. Differences of RAD-Seq-derived SNP frequencies between populations can be analysed to explore population structure and quantify connectivity. Here, we analysed such differences to infer population/stock boundaries by estimating the degree of genetic differentiation between them. Moreover,

we inferred connectivity indirectly by inferring the origin(s) of migrants using assignment methods. We detected a low genetic differentiation between the three populations analysed (Galician shelf, Balearic Islands, North Sea). This could be driven by high gene flow among them that would homogenize allele frequencies. However, considering the large distances among the populations analysed, the most plausible explanation in this case is that large effective population sizes are resulting in low genetic differentiation between them (Hare et al., 2011). This theory is reinforced by a lack of migrants detected among the three populations, possibly prevented by existing oceanographic barriers across the analysed regions that impede gene flow.

Although the results obtained here serve only as a proof of concept due to the large distances among analysed populations, the pipeline developed in the framework of FishGenome shows an enormous potential to establish these parameters accurately. Moreover, it only requires a **small number of individuals per population that can be easily collected during ongoing fisheries research surveys**. Considering that the assessment of stock boundaries and connectivity among adjacent stocks should be urgently adapted in fisheries management to guarantee their conservation, we recommend implementing the method routinely.

### 2.3.3. RAD-Seq for stock substructure

Detecting substructure within stocks is of fundamental importance, as it can have direct consequences on establishing appropriate population management units. The presence of differential patterns of genetic variation within a stock can significantly increase the risk for depletion of genetic resources. Loss of genetic diversity can hamper the ability of a stock to adapt to environmental changes and other pressures, decreasing its resilience. Moreover, substructure provides important insights into the processes of gene flow, genetic drift and selection. Undetected substructure can also lead to biased estimates of a single overall abundance through the CKMR estimate. Thus, it is critical to identify genetically "heterogenous" groups of individuals, in order to define the number of basic units for species that are exploited. Fisheries management should be based on such knowledge.

Although several methodologies have been proven useful for exploring stock sub-structuring, they are not used in current fisheries assessments. RAD-Seq data provides unprecedented access to survey genome-wide diversity and can generate thousands of single nucleotide polymorphisms (SNPs) markers with direct application to management and conservation. Differences of RAD-Seq-derived SNP frequencies within stocks can be analysed to explore population sub-structure. Here, a very strong genomic differentiation was found within the ballan wrasse population. The markedly distinct genomic signatures were strongly correlated with two phenotypic colour phenotype – plain and spotted -, while individuals with the same colour phenotype showed a very low overall genomic differentiation. These results indicate strong reproductive isolation between phenotypes and consequently, the two genomic components should be managed as different sub-stocks, despite being a unique species, *L. bergylta*. Other studies indicate large differences in life-history parameters with spotted individuals reaching larger sizes at age, investing fewer resources in reproduction, and changing sex at larger sizes and older ages than plain individuals. All this evidence indicates that ignoring the existence of two components for the exploitation of this species could lead to unintended overexploitation of one or both sub-populations, potentially leading to the loss of one or both components. For ballan wrasse, genetic markers derived from RAD-Seq have proven to be a powerful tool to unveil genetic differentiation within the local spawning populations and inform fisheries management.

On the other hand, no differential patterns of genetic structuring were detected within any of the stocks analysed of cod and hake, at any of the geographic locations analysed (Galician shelf, Balearic Islands, North Sea). Both marine species are characterized by large numbers of individuals with high dispersal potential, often resulting in none/weak population structure (Bernatchez et al., 2017; Papa et al., 2020). However, the North

Sea cod stock is known to consist of three reproductively isolated sub-populations that are genetically differentiated (Heath et al., 2014). The reason for not detecting them in the present study can be explained, since all the individuals analysed here were collected during the North Sea IBTS survey led by the Thünen-Institut, which targets a limited area, where only one of them is present.

The results obtained here clearly demonstrate the power and accuracy of the RAD-Seq technique to uncover **fine-scale population substructure**. The pipeline developed in the framework of FishGenome is a powerful tool to unveil accurately potential genomic differentiation within stocks. Moreover, obtaining this essential information is simple and only requires a small number of individuals per population that can be easily collected during ongoing fisheries research surveys. Thus, we recommend implementing the method routinely to explore population/stock sub-structuring and ensure a proper consideration of biological management units and fine-tune stock management.

#### 2.3.4. RAD-Seq for sex assignment

Knowing the sex of the individuals is essential to evaluate fisheries stocks. Ongoing surveys, determine the sex by macroscopic inspection/histology of the gonads of adult individuals. However, this method has two main drawbacks. First, sexing of young juveniles is often not possible/uncertain and, second, it requires the sacrifice of the specimen. Genetic sexing of individuals can be performed at any stage of the development/age and only requires a small piece of tissue (such as a fin clip) that can be collected non-invasively. Nonetheless, it is only possible in species with a simple genetic sex chromosomal determination system whereas, in fishes, sex is often determined by the environment or social interactions, as in hermaphrodites (Palmer et al., 2019). In species with sex chromosomes, isolating sex markers is generally an arduous task, since many teleost species have very small sex-determining regions.

RAD-Seq data produces large amounts of sequencing data, which can be easily scanned, to detect differences between both sexes, if a suitable pipeline is available. We tested the pipeline developed within FishGenome to search for a sex marker in hake. The analysis revealed a panel of sex-linked markers in North Sea hake consisting of one sex-specific region restricted to males and sixty-four male heterogametic single nucleotide polymorphism (SNP) loci.

These polymorphisms enable the genetic sexing of hake specimens from this location with a high accuracy of >90% in the case of males and >80% for females. Nonetheless, they are population-specific and, therefore, only informative of sex in this North Sea population. Testing of the markers in the Balearic Islands population did not show a sex-linkage in specimens from this location, indicating that recombination suppression in this region has only evolved in specimens from the North Sea. Annotation of the markers is useful to select further the best candidates, as a "true" sex-linkage would be reflected in a co-location to the same genomic region. However, the poor quality of the hake reference genome has prevented from obtaining this important information.

Still, our results are very useful in providing insights into the potential sex determination system in this species. In hake, the sex determination is unknown but the balanced proportions of males and females in the majority of studied populations, suggest an XX/XY or ZZ/ZW system. For taxa with XX/XY sex-determination, only female individuals can be heterozygous for an SNP located on the X chromosome, with the actual proportion depending on the allele frequency of the SNP. For an SNP located on the Y chromosome, the locus will be missing for all females, and all males will be homozygous. The opposite is true for ZZ/ZW systems. The unique sex-specific region restricted to one sex appeared in males while almost all sex-specific SNPs (64 of 65) were, likewise, found in males but not females, a pattern that agrees with a male heterogametic system, providing strong evidence of an XX/XY sex determination mechanism in hake.

In summary, **no universal sex-specific markers could be isolated for hake but several North Sea population-specific markers were found**, allowing the correct

sex assignment of over 90% of males and 79% of females from this region. These markers, still need to be validated via PCR amplification in a larger number of individuals from the North Sea, as well as tested in specimens from the North Atlantic.

Our strategy is, nonetheless, suitable for the isolation of sex markers in fish, as data from RAD-Seq experiments has allowed the characterization of sex-specific markers in various species, including blennies and rockfishes (Hundt et al., 2019, Fowler and Buonaccorsi, 2016). However, RAD-Seq scans only between 1-10% of a given genome, which might be insufficient for species with very small sex-determination regions. An alternative would be the analysis of whole genome sequence data as for other species, including cod, the isolation of sex-markers was prevented until whole-genome sequencing was performed.

Genetic sexing of individuals offers important advantages since, unlike classical histological sex assignment, it allows sexing of early stages of development (eggs and young juveniles) and only requires a small piece of tissue that can be collected non-invasively. In the framework of FishGenome, obtaining a sex marker for the species of interest would imply an extremely simplified sampling of the specimens, as all information needed for stocks assessment through CKMR could be obtained from a fin clip (age through epigenetics, sex if a sex marker is available).

### 2.3.5. Epigenetic age determination

In fishes, age-class distribution is one of the most relevant parameters in a population, with important influences on biomass distribution, intra-specific interactions and reproductive potential. Thus, accurate age estimation is essential for the monitoring and management of fishery resources. Age estimation in fishes has traditionally relied on the analysis of growth marks in hard structures such as otoliths, but this requires well-trained personnel, is time-consuming, lethal and has low accuracy in some species.

Recently, epigenetic clocks have been developed in animals, mostly mammals, but also in some fish. These clocks build on the fact that aging is associated with changes in DNA methylation in specific cytosine-guanine (CpG) loci. Epigenetic clocks consist of carefully selected loci across the genome that are collectively capable of predicting chronological age with high accuracy and precision. Epigenetic clocks developed so far have relied on targeted approaches, starting with 101-102 CpG loci, which may compromise accuracy and precision, or on genome-wide methods, starting with 105-106 CpG loci, where then >99% of the sequenced loci are not used, implying a waste of resources. Thus, a cost-effective method for the construction of epigenetic clocks to be useful for fisheries is needed.

The epigenetic clock developed here is robust and allows age determination in cod with high accuracy and precision, two requisites for its use in stock assessment. Further, this clock compares very well with other available fish epigenetic clocks (Table 8) in terms of accuracy and precision.

Further developments will include testing this clock in different Atlantic cod populations to make it sure that it can be applied to specimens of different geographic origins. The important aspect is that the technique is now developed and perhaps some fine-tuning of the model will be necessary when more samples are processed.

A major challenge that, nevertheless, needs to be overcome, is to develop a cheap, fast and accurate method to process thousands of samples as an alternative to the traditional otolith-based estimation. We believe that, once set up, it will require the skills routinely in place in any basic molecular biology lab (DNA extraction, sample processing and DNA sequencing), meaning that the method has the potential to be scaled up.

**Table 8. Comparison of the performance of the cod epigenetic clock with other piscine clocks. Abbreviations: yr, yeas; wk, weeks; r, Pearson correlation value**

| | Atlantic cod | Zebrafish | Seabass |
|---|---|---|---|
| **No. samples** | 56 | 67 | 50 |
| **Age range** | 0-7 | 0-60 wk | 0,5-10,5 |
| **CpG initial** | 67549 | 524038 | 48 |
| **CpG final** | 26 | 29 | 28 |
| **Accuracy (r)** | 0,98 | 0,92 | 0,824 |
| **Precision (yr)** | 0,69 yr | 3,7 wk | 2,14 yr |

The epigenetic clock developed here is a molecular method of high accuracy and precision to determine age in cod and represents a suitable and modern alternative to the traditional otolith-based age estimation method. Investing in a bit more effort, the method could be scaled up and deployed for the mass analysis of fish age in fisheries management. However, both methods would need to coexist during a period to ensure consistency of data series and crosscheck the consistency of both methods on large numbers of specimens.

### 2.3.6. Environmental DNA

The Pilot study addressed the evaluation of the performance of eDNA methods based on metabarcoding to reveal fish biodiversity in the North Sea and the Balearic Islands. Methodological problems produced inconsistent results in the last ecosystem that have been excluded from this report. However, this failure serves to highlight the challenges of implementing the eDNA methodology, even when using a standardized protocol, in different environments. An intense fine-tuning tailored to the ecosystem studied is still required, as the characteristics of the samples collected in different regions (e.g. more or less contaminated) can strongly influence the performance of eDNA protocols (Kumar et al., 2022). In the North Sea, our overall results showed that eDNA from water samples can reveal the presence or absence of a given fish species with 90% reliability. This result supports the general ability of eDNA to provide solid qualitative information on fish biodiversity in the North Sea. However, the MiFish 12S rDNA primers clearly have limitations in species detectability (Miya et al., 2015, 2020; Sato et al., 2018) and do not cover some species, such as sharks, for which additional primers are required. A combination of the Miya protocol with other target genes like COI or CytB could help in overcoming this problem. Moreover, public databases such as NCBI and ENA are far from being complete and do not contain sufficient 12S rDNA gene sequences for quite a few common fish species, which makes the detection of these species impossible or leads to incorrect records of closely related species. The combination with COI or CytB could also help with this, as they are better covered in international databases. Thus, the protocol requires further optimization before its application in fisheries management.

The results of eDNA metabarcoding, nonetheless, show a remarkable potential for biodiversity assessment, since they point to a detection of eDNA traces of species that were not found by trawling. The number of these species detected by eDNA compared to conventional methods was, at least, three times higher. This is perfectly in line with previous studies reporting that eDNA metabarcoding performed better than conventional methods in assessing species richness (Afzali et al., 2021; Fujii et al., 2019; Nester et al., 2020; H. Sato et al., 2017; Valsecchi et al., 2020; Yamamoto et al., 2017).

The outputs of eDNA sequencing from water samples were more consistent with the results of traditional methods, than those of eDNA from sediments. Sediments act as

DNA repositories of ancient biodiversity (Coolen & Overmann, 2007; De Schepper et al., 2019; Willerslev et al., 2003) and do not necessarily represent present-day conditions, whereas eDNA in water samples is characterised by its short half-life of less than 2 days, together with its easier accessibility. However, sediment eDNA can be of crucial importance to detect sediment-dwelling species or changes in biodiversity through time.

The evaluation of the eDNA as a tool to estimate diversity through qPCR also revealed very promising results. The pilot study addressed the performance of eDNA for recording North Sea cod stocks from sediment and water samples using the data from trawl catches to validate the results. The specific quantitative PCR test developed for cod (*Gadus morhua*) is able to detect Atlantic cod even at low abundances (up to 0.08 kg/ha). Moreover, eDNA was able to reveal the presence of Atlantic cod in water samples even at stations where the species did not appear in trawl catches. Since the half-life of eDNA in water is only two days (Collins et al., 2018; Lance et al., 2017), we can assume that at these two stations cod was either present at low abundance just outside the trawled area or that a reasonable number of cod passed through the trawled area in the two days prior to sampling. The cod qPCR test developed within the FishGenome project allows the detection of cod at very low levels (up to 0.08 kg/ha), while guaranteeing a degree of specificity up to 100%, as well as an excellent sensitivity of around 95%.

The reason why we could not detect cod DNA in North Sea sediments remains unclear. The negative results could be explained by unfavourable conditions for the preservation of DNA in the sampled sediments. However, since sediments are reported to represent a historical repertoire of species colonising a certain area and to be able to preserve DNA for up to 100,000 years (De Schepper et al., 2019), or even 270 000 years (Coolen & Overmann, 2007), this line of reasoning is weak.

Based on these results, the analysis of water samples for eDNA seems to be an appropriate method to determine a quantitative assessment of a fish stock. However, the method still requires fine-tuning, before an ample use on fisheries assessments. Correlation between DNA copy number and CPUE is still challenging. In our Pilot study, a statistically weak correlation between CPUE and eDNA concentrations was observed, when assessed station by station. Despite the good general correlation (98% after optimization) shown by our model between the observed and the predicted situation, further optimization steps, as well as an integration of more samples and stations, are required for model validation. Summarizing, the results support the ability of eDNA to provide effective and reliable information for fish monitoring in the North Sea, but further validation is needed.

# 3. ANALYSIS AND STRATEGIC PLANNING

The ultimate purpose of this work package was to develop a roadmap for the implementation of genomic-based approaches in fisheries stock assessment, providing precise and accurate information on whether and how the analysed genomic-based approaches could become part of the regular research surveys, describing the steps, the pathway, and the timeline for its progressive implementation in fisheries management. This roadmap is based on the outputs of WPs 1 and 2, i.e. the technical performance and features of the genomic techniques, on the cost-efficiency analysis of these tools and on a SWOT analysis about the conditions for the future use of these techniques in the EU-MAP for data collection and fisheries assessments.

## 3.1. MAIN ACHIEVEMENTS

- ***Cost-benefit analysis.*** The purpose of this analysis was to evaluate the cost-efficiency of HTS methodologies for fisheries stock assessment. That is, to learn if the use of genomic methodologies can yield, at least, equal, or equivalent outputs (in our case data enabling the stocks assessment) at a lower cost than current methodologies used for scientific advice of the Common Fisheries Policy (CFP). We carried out a contextualised cost comparison and identified some potential pathways for efficiency gains, both on the surveys at sea and on the lab, where the sampling processing takes place until the output data is made available for the stock assessment. Despite this, we also faced several constraints related to lack of past references and scarce data for the analysis, leading to several assumptions. The information obtained, combined with findings from the State-of-the-art reviews and the pilot studies, was used to feed a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis for the future use of these techniques in the EU-MAP for data collection and fisheries assessments. As a second step, it was also used as input for the design of realistic implementation scenarios and to plan whether and how best to progressively incorporate these technologies into the fisheries research surveys at sea and in the assessment. The results of the cost-benefit analysis are presented in Deliverable 3.1 "Cost-efficiency of the application of HTS methods on fisheries research surveys and stock assessment".
- ***SWOT analysis.*** This part of our study gathers the findings on factors that may directly or indirectly affect the implementation of HTS techniques in the fisheries assessments. For this analysis, internal factors (strengths and weaknesses) are those related to robustness and accuracy, reliability, versatility, coverage, cost efficiency and added value. They have been considered both jointly and individually for each of the genomic methods analysed, so the specificities of each one is duly considered. External factors (opportunities and threats) were assessed for the whole set of techniques. The SWOT allowed us to identify the advantages and limitations of the techniques, as well the needs and conditions, for the implementation of the genomic tools in stocks assessment, and, hence, designing a realistic roadmap for such implementation. The results of the SWOT analysis are presented in Deliverable 3.2 "SWOT analysis".
- ***2$^{nd}$ FishGenome Workshop "Co-envisaging the actions needed towards the future implementation of genomic tools for fisheries research surveys at sea and stocks assessment".*** A three session-workshop was carried out during October-November 2021 within the framework of the FishGenome project. The overall objective was to discuss with a panel of experts i) the pros and constrains on using advanced genomic techniques in research surveys and stock assessments; and ii) gathering of relevant information for the design of scenarios and of a roadmap for the implementation of these techniques

in data collection and scientific advice. Expert peers' input enriched the analysis and contributed to confirm the overall appropriateness of the approach.

- ***Roadmap.*** The roadmap sets a pathway to progress towards a full-scale implementation of HTS methods into fish stocks assessment and management. A complex, comprehensive but clear roadmap was designed to achieve five defined objectives in the short- mid- and long-terms. Five strategic challenges make up the general action plan, corresponding mostly to each of the five specific objectives. Within each challenge, several specific and interlinked actions are proposed, totalling 45 actions.
- ***Long-term prospects:*** Based on information from two sources - 1) a comprehensive review of the latest genomic literature, and 2) the analysis and evaluation of the knowledge gathered during FishGenome study contract -, we assessed the potential of emerging cutting-edge genomic technologies to further contribute to fisheries management in the long-term for closing some of current technical gaps or needs. For this, we evaluated the potential of existing and emerging genomic tools which are relevant to estimate the necessary parameters for stock assessment.

## 3.2. Work carried out

### 3.2.1. Cost-benefit analysis

#### Methodology

The cost-efficiency analysis of the genomic methods, compared to the currently used methods, required, in first instance, (i) to identify the cost components in each case, (ii) to measure the relative efficiency and (iii) to understand the origin of the differences. Secondly, it required identifying the output. In our case, outputs are understood as the necessary data for delivering the required scientific assessment and advice for the fisheries management. From these two steps, the most relevant relative efficiency variables can be selected and calculated.

A third step refers to quantitative assessment of the potential benefits derived from the use of the new techniques and its economic implications. However, the constraints derived from the scope of the pilot studies made this part of the analysis unfeasible at the time, and a qualitative approach based on experts' criteria was used instead for the overall identification of potential benefits from HTS implementation.

Data on the fisheries research surveys' costs were obtained for 2012 and 2013 for a total of 18 countries. This information was provided to the FishGenome study by the EC after a formal request directed to the EU member states for the aggregated use of these data. Additional secondary data for analysing existing fisheries research surveys (survey areas covered, stocks addressed, etc.) were obtained from different databases, mainly from ICES -DATRAS.

The costs for the implementation of DNA-HTS methods were estimated from the pilot experiences for IBTS-Q3 and MEDITS-GS5. The calculations do not reflect the total time and financial resources consumed during the FishGenome project in its pilot studies (although it was originally considered as a first reference value). Therefore, we did not include, for instance, planning time, development and fine-tuning of protocols, meetings, etc., but an estimation of the implementation requirements for the genomic methods once the protocols have been set up and tested.

The data obtained were subject to the following analysis:

a) For the surveys:

- Relevant cost items of the fisheries research surveys at sea and relative weight of each cost component to determine major pathways for efficiency gains.
- Identification of relative efficiency variables (mainly cost per day at sea, per survey and cost per stock addressed by each survey, and labour/effort per survey)
- Comparison of relative efficiency at different levels of data aggregation (survey, region)

b) For the genomic methods:

- Relevant cost items (for work on board and in the lab) and relative weight of each cost item for the three different techniques.
- Aggregated cost of the three techniques.

This study/approach depends also on several, practical assumptions for the sake of simplification and to overcome some of the main constraints coming from data availability.

The quantitative values estimated should not be viewed as highly accurate and useful for any larger scale implementation progressions, but to extract some conclusions about:

1. An order of magnitude of what it would take to implement these techniques individually and all together.
2. The cost items, which have a major impact on fisheries research surveys at sea and thus, some clues about possible pathways for efficiency improvements.
3. A new set of bottom-trawl research surveys where the use of the three DNA-HTS techniques can be expected to be at least as cost-efficient as for those surveys targeted for the pilot studies in FishGenome.
4. The approach that could be followed to further study the cost-efficiency of the HTS implementation progression.

## Results

### COSTS OF DCF FISHERIES RESEARCH SURVEYS

Fisheries research surveys are not performed in the same way, they differ significantly one to the other. A variety of stock assessment evaluation methods is regularly used. They also differ on the requirements on a priori information about the stock, in biological data and fishing parameters, in the time-range of this data needed to run it with confidence (Osio et al. 2018), in the type of fisheries, spatial parameters, etc. All these aspects are susceptible to condition the efficiency of the surveys and the eventual optimization of their cost.

We have classified the cost components of DCF mandatory trawl surveys into the following categories:

- Staff Costs
- Sea allowances
- Travelling expenses
- Vessel costs
- Consumables and Computing Costs
- Subcontracting costs

The main cost for trawl surveys is the vessel, amounting, on average, to 66,4% of the total cost, reaching a maximum for some surveys of 89,2% of the survey cost.

A significant driver of vessel costs is its cost per day that ranges from 933,7€/day to 23.911,03€/day. The differences are a consequence of the vessel characteristics (size, etc.) and of the accounting practices. Vessels range from small commercial fishing boats to large research vessels being the latter, on average, more expensive.

Three main factors determine the cost of a survey:

- Number of days at sea
- Cost of the vessel
- Number of enrolled scientists and technicians and cost of scientific staff.

The vessel cost per day increases with the duration of the survey (Figure 35), because long surveys usually demands larger vessels. The number and type of scientific staff are determined by the requirements of the surveys and their cost by the salary conditions in each country. To mitigate the impact of salary differences among countries, labour intensity in terms of effort was considered for comparison instead of staff costs.

**Figure 35. Relationship between vessel cost per day and no. of days at sea, and the total cost per survey.**

An additional factor to be considered is the number of species addressed per survey. If stock assessments are used as the main final output of a marine scientific survey or proxy of the outputs, the number of stocks reported per survey could also be a relevant indicator of efficiency (the higher the number of stocks reported, in principle, the higher the efficiency). However, there are many hurdles to reach an accurate assessment from survey data regarding the stocks (it is clear how many stocks each survey reports, but the relevance of this contribution to the assessment is not being considered) and some stocks are assessed through the information provided by different research surveys, so comparison is not always straightforward.

More relevant than the differences in the cost range is that, in general terms, an inverse relationship between the number of stocks assessed and the cost per stock assessed is observed (Figure 36), i.e. the cost by stock assessed is less if several stocks are targeted by the same survey.

The differences stem from the type of fisheries (mixed vs single-species fisheries), the geographical scope of the survey, the distance from port to the sampling area, the

scientific goals, etc., so that cost per stock assessment is a simplification that can provide guidance for this analysis, but it should be put in context.

It is of interest to check if those differences are diluted when aggregating by fishing region. As displayed in Table 9 and Table 10, stock assessments number and total survey cost are heterogeneous, as are the values for the efficiency variables estimated: cost per stock assessed, survey cost per day and cost per stock assessed and day. The cost per stock assessed ranges from 85.432,61€ to 367.928,32€ in the Atlantic Ocean and the Baltic, while in the Mediterranean it ranges from 11.549,88€ to 189.749,41€. On average, stock assessments for the Mediterranean fisheries are significantly cheaper, but again, the comparison is not to be taken alone. Relevant assumptions have been made to simplify the analysis and one needs to put the results that we present in the right context.

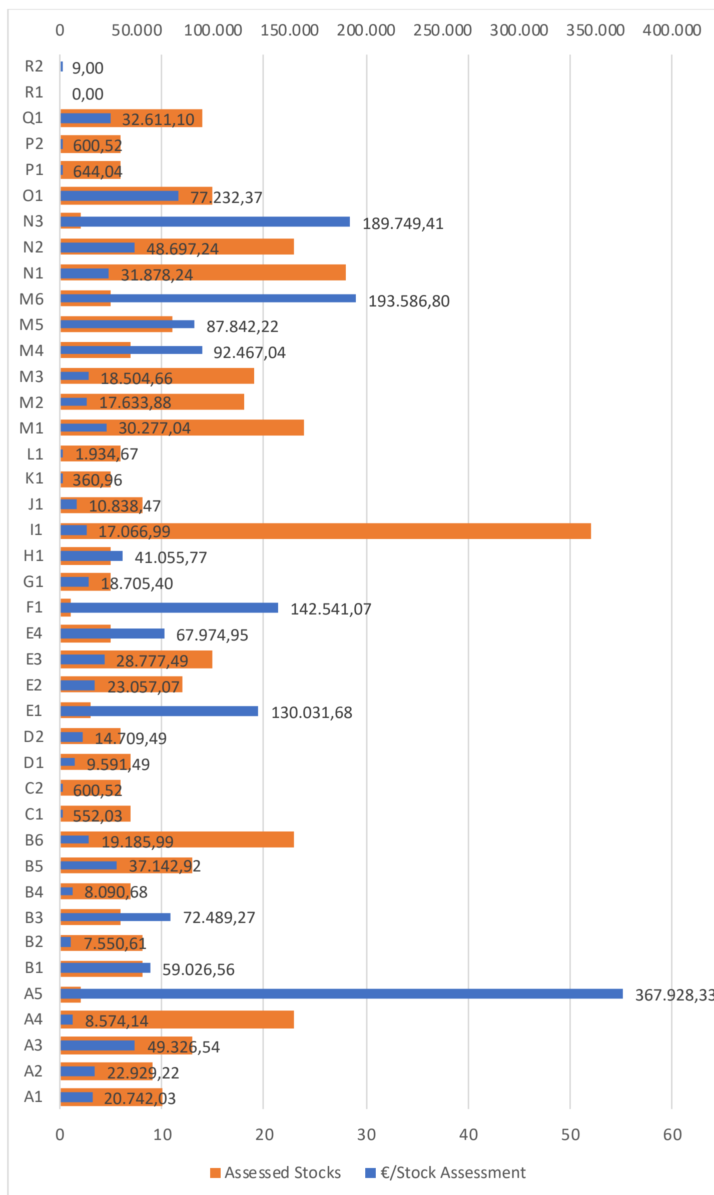



**Figure 36. Efficiency. Cost per stock assessed by country/survey. 2013. Codes in the Y-axis indicates in anonymous manner the 41 surveys conducted by EU countries.**

**Table 9. Atlantic Ocean calculation of the cost-efficiency of independent demersal fishery trawl surveys for stock assessment purposes. Summary table. S.A. Number refers to the number of stocks assessed in the specified subregion. Total cost represents the total cost (in euros) of deploying the fishing survey (taking as reference year 2013). S.A. Cost reports the average cost of a single stock assessment in the corresponding FAO region. Survey Cost / Day represents the survey cost per day. Finally, S.A. Cost / Day reports the average stock assessment cost per fishery-independent survey and per day.**

| Region (FAO) | Surveys | S.A. Number | Total Cost (€) | S.A. Cost (€) | Survey Cost / Day (€) | S.A. Cost/ Day (€) |
|---|---|---|---|---|---|---|
| **27.3** | BITS Q1 & Q4 | 20 | 2.286.691,00 | 114.334,55 | 176.120,21 | 8.806,01 |
| **27.3, 27.4** | IBTS Q1 & Q3 | 39 | 3.883.269,15 | 99.571,00 | 221.137,01 | 5.670,18 |
| **27.6, 27.7, 29.9** | IBTS Q4 | 38 | 3.435.074,32 | 90.396,69 | 98.451,38 | 2.590,83 |
| **NAFO 1** | GGS | 2 | 735.856,65 | 367.928,32 | 20.153,13 | 10.076,57 |
| **NAFO 3** | Flemish Cap & Platuxa | 19 | 1.623.219,63 | 85.432,61 | 33.290,87 | 1.752,15 |
| **Average** | | 23,60 | 2.392.822,15 | 151.532,63 | 109.830,52 | 5.779,15 |

**Table 10. Mediterranean Sea calculation on the cost-efficiency of independent demersal fishery trawl surveys for stock assessment purposes. Summary table. S.A. Number refers to the number of stocks assessed in the specified subregion. Total cost represents the total cost (in euros) of deploying the fishing survey (taking as reference year 2013). S.A. Cost reports the average cost of a single stock assessment in the corresponding marine region. Survey Cost / Day represents the survey cost per day. Finally, S.A. Cost / Day reports the average stock assessment cost per fishery-independent survey day**

| Region (GFCM) | S.A. Number | Total Cost (€) | S.A. Cost (€) | Survey Cost / Day (€) | S.A. Cost / Day (€) |
|---|---|---|---|---|---|
| **1.1 (GSA's 1, 5, 6)** | 20 | 967.934,01 | 48.396,70 | 14.661,08 | 733,05 |
| **1.2 (GSA 7)** | 2 | 379.498,82 | 189.749,41 | 10.083,75 | 5.041,87 |
| **1.3 (GSA's 9, 10, 11)** | 23 | 367.777,07 | 15.990,31 | 8.383,73 | 364,51 |
| **2.2 (GSA's 15, 16, 18, 19)** | 35 | 404.245,88 | 11.549,88 | 16.507,37 | 471,64 |
| **2.1 (GSA 17)** | 15 | 431.631,43 | 28.775,43 | 19.992,84 | 1.332,86 |
| **3.1 (GSA 22)** | 4 | 93.527,00 | 23.381,75 | 2.750,79 | 687,70 |
| **3.2 (GSA 25)** | 1 | 142.541,07 | 142.541,07 | 5.577,24 | 5.577,24 |
| **Average** | 14,29 | 398.165,04 | 65.769,22 | 11.136,69 | 2.029,84 |

COSTS OF HTS METHODOLOGIES' IMPLEMENTATION

No specific surveys have been carried out for the HTS methods' pilot implementation, but samples have been collected during MEDITS GSA5 (by IEO) and IBTS Q3 (by THÜNEN) surveys. So, their costs have been used in this analysis, as the closest case for referencing the economic impact of the HTS methods. The main cost components in these surveys are summarized in Table 11.

**Table 11. Cost components of the MEDITS GSA5 and IBTS Q3 expressed in euro**

| | MEDITS-GSA5 | IBTS-Q3 |
|---|---|---|
| **Staff Costs** | 35.326,72 | 28.840,00 |
| **Sea allowances** | 18.606,25 | nd |
| **Travelling expenses** | 1.275,00 | 300,00 |
| **Vessel costs** | 250.500,00 | 454.433,28 |
| **Durable equipment** | 900,00 | nd |
| **Consumables and Computing Costs** | 2.710,00 | 1.939,46 |
| **Subcontracting costs** | - | 129,84 |
| **TOTAL (€)** | **309.317,97** | **485.642,57** |

As for the costs of implementing HTS methodologies, it should be kept in mind that the vessel costs have been excluded from the calculation. In the short and medium-term, it is not possible to replace current methodologies with HTS methods given that stock assessment accuracy requires a certain length and stability of the data time series. Therefore, the shift towards HTS methodologies would require a period of coexistence. The last statement means that the implementation of HTS methodologies would be an added cost to the currently mandatory fisheries research surveys (according to the EU MAP) and during a coexisting period that is yet to be determined.

The estimated cost for the three HTS methods in the two pilot studies together was 76.933,31€ (IBTS-Q3 and MEDITS GSA5). Assuming, for simplicity of the calculation, that samples from IBTS survey were mostly cod[8] and samples from MEDITS were hake, it is possible to estimate the cost of the HTS methods in the North Sea case study (CKMR and EAD for cod plus eDNA) in 38.230,80€ and the costs of the Mediterranean case study (CKMR and EAD for hake plus eDNA) in 38.702,51€. This would mean a 12,5% increase in MEDITS GSA5 survey and a 7,8% in IBTS Q3, because of the introduction of these techniques. If we assume a linear increase of the cost of CKMR and EAD and no need for additional samples of water of sediments, the cost of applying the three methods to the 5 species surveyed in MEDITS GSA5 would amount to 183.974,75€. Likewise, the cost of applying them to the 13 species surveyed in IBTS Q3 would amount to 468.387,05€. . This result is valid for similar species. Different species, particularly species with bigger genome size, will have higher sequencing costs. As a smaller number of samples could be included in each library, more libraries will be needed. Nevertheless, since this only affects the smallest cost item, the impact of sequencing costs on the total amount will be limited.

Indeed, the main cost incurred in HTS methodologies (Figure 37) is that derived from processing the samples in the laboratory, which includes DNA extraction and other sample preparation, library creation and data processing. This represents 85,4% of the total cost. Furthermore, the cost of specific sampling for HTS methods onboard (tissues

[8] Hake samples were also provided by the IBTS survey, nevertheless given the limited number of samples and the marginal effect on costs, this species has been disregarded.

extraction and preservation onboard and water and sediment samples extraction and preservation onboard) is estimated at 4,2% of the total, while subcontracting (exclusively comprises sequencing) represents 10,3% of the total amount. Therefore, the most relevant cost is not sequencing, but the time employed in processing samples for sequencing and analyzing the sequencing outcomes.

The main cost component of processing samples in the laboratory is the staff costs (68%), the rest being consumables (32%), as shown in Figure 38. The staff is needed to carry out three main tasks: DNA extraction and other sample preparation (23% of the cost), library creation (14%) and data processing (31%) when the sequences are obtained.

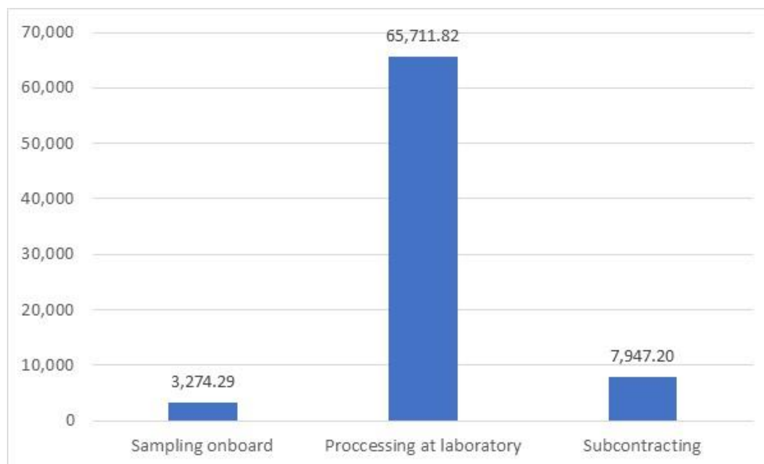



**Figure 37. Aggregated cost per stage for the application of HTS methodologies to hake and cod.**

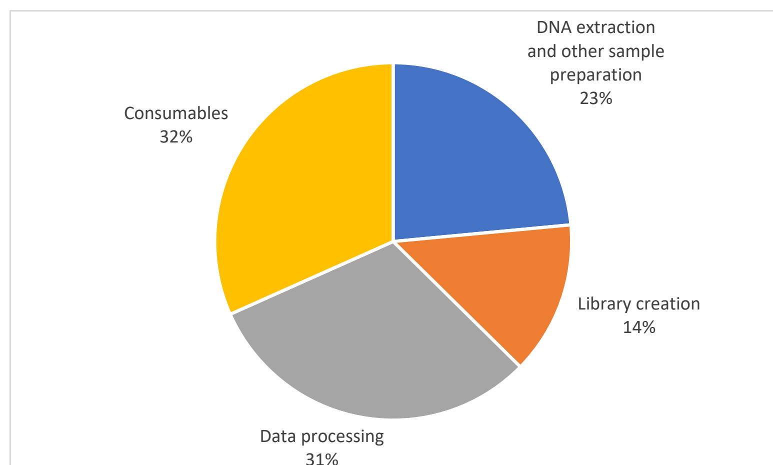



**Figure 38. Disaggregation of the costs of samples processing at the laboratory.**

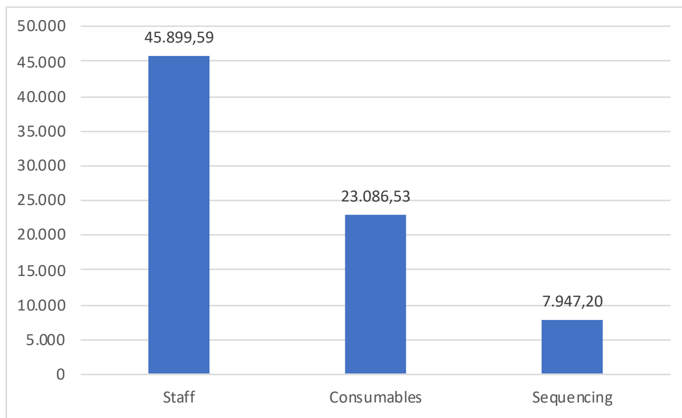



**Figure 39. Aggregated cost per category for the application of HTS methodologies to hake and cod**

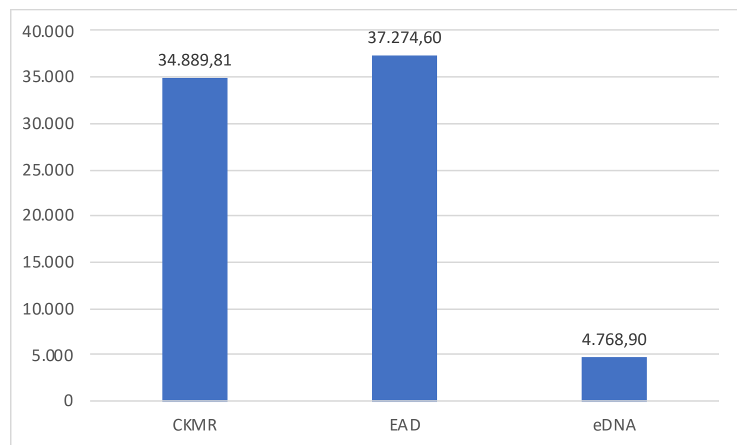



**Figure 40. Aggregated cost per method for the application of HTS methodologies to hake and cod**

Looking at cost items (Figure 39), the staff is the main one (59,7%), followed by consumables (30%) and sequencing through subcontracting (10,3%). Since the staff is the main cost item, there can be expected efficiency gains in the mid- and long-term, through further refined protocols and routinization. Similarly, dedicated trainings, together with the skills developed with practice, can also contribute to increasing labor-time productivity. On the contrary, no dramatic reductions can be expected from sequencing services. Just to illustrate with an example, since September 2001, the cost of sequencing a human-sized genome has dropped from 95,2 million US dollars to only US$689 in August 2020 (Wetterstrand 2021) and the evolution of prices in more recent years seems to stabilize or to decrease slightly.

If each methodology tested in the FishGenome project is considered separately, the costliest at present is EAD, followed by CKMR and, at a great distance, by the eDNA (Figure 40). The higher cost of EAD is mostly explained by/due to the higher cost of the sequencing method required, bis-RAD-Seq. Comparing with RAD-Seq (used, for instance, in CKMR), the former involves more processing time and the use of more expensive consumables. This is the consequence of the comparatively less maturity of EAD, so its cost is liable to decrease in the future, as far as the epigenetic age clock technology is further developed. In other words, these costs are required for the initial development of the epigenetic clock, but once developed, the method can be scaled up to a PCR multiplex, which is cheaper and requires fewer loci, eventually resulting in

savings in processing time and sequencing. On the contrary, the number of fish samples needed for the application of EAD is much lower, in general, than for CKMR.

A second aspect to consider is eventual gains in surveying efforts, that is, eventual reductions of days at sea. Possible sources of efficiency gains could come from eventual adjustments in sampling effort derived from a progressive introduction of the HTS. The use of these techniques could result in the need for a smaller number of samples, the reduction of vessel days, or open new opportunities for sampling on captured fishes onshore. Efficiency gains can also come from optimizing the reuse of samples through biobanking solutions. Techniques such as CKMR, which are quite intensive in sampling needs, could benefit from these alternatives. Nevertheless, so far, there is no information available for estimating a hypothetical effect of the HTS methods studied on sampling efforts, to estimate this scenario quantitatively.

## TOWARDS THE IMPLEMENTATION OF GENETIC HTS METHODS IN FISHERIES RESEARCH SURVEYS.

Cost-efficiency analysis measures the cost ratio between inputs and obtained outputs. Consequently, efficiency can be improved by increasing the amount or value of the outputs at equal or minor cost.

Since HTS methods will have to 'coexist' with current ones, at least during a certain period, it makes sense to share resources to optimize operating efficiency. In this regard, Figure 41 synthesizes the two main indicators used in this report: i) cost per stock assessed, and ii) vessel cost per day by survey. The first one is an index number oriented to output, whereas the second is oriented to inputs. Figure 42 compares average effort invested per stock vs. average vessel cost/day, in each survey. The idea of this part of the analysis is to identify for which surveys other than the ones tested here, the implementation of HTS could be at least as efficient as it has been for the two experimental pilots carried out in FishGenome. The effort invested per stock is the result of dividing the number of working hours by the number of stocks the survey contributes to assess. We have used working hours as an effort measure instead labor costs because of the differences in salaries among countries.

In the coordinate planes represented (Figure 41 and Figure 42) fist quadrant is the one at the upper right of the axes, second is the one at the upper left, third is bottom left and fourth is the bottom right. Horizontal blue dashed lines represent the Stock Assessment Average cost and labor-time per assessed stock, respectively, and the vertical one the average vessel cost per day in both cases.

The third quadrant in the coordinate plane (data under the two averages) reports those surveys that are, in principle, more efficient in terms of cost and in terms of cost per output. If no other factors are considered, this feature places them as good candidates for incorporating sampling for HTS methods.

The fourth quadrant represents surveys (over average cost per stock and under vessel cost per day average) that are efficient in terms of inputs but not in terms of output.

Points in the first quadrant (above the two averages) suggest the surveys with the lowest efficiency and, consequently, the less feasible candidates for additional samplings. For these surveys, the key question is the eventual need for increasing the days at sea. If the total vessel cost does not increase with sampling for HTS methods, the cause of efficiency does not apply. In this area, there is only one country addressing hake and several targeting cod.

Finally, the second quadrant represents surveys with an efficient rate of cost per stock assessed, even with a comparatively high vessel cost per day. In this case, it is likely that additional sampling requires additional vessel days. This would probably translate in a significant cost increase due to the introduction of HTS methods.

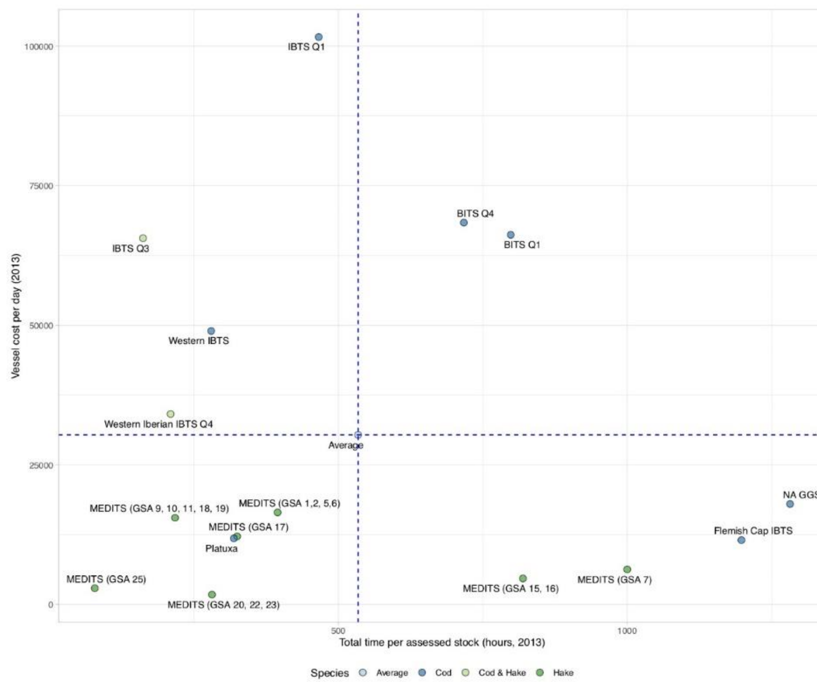



**Figure 41. Cost per assessed Stock vs. vessel cost per day (by survey)**

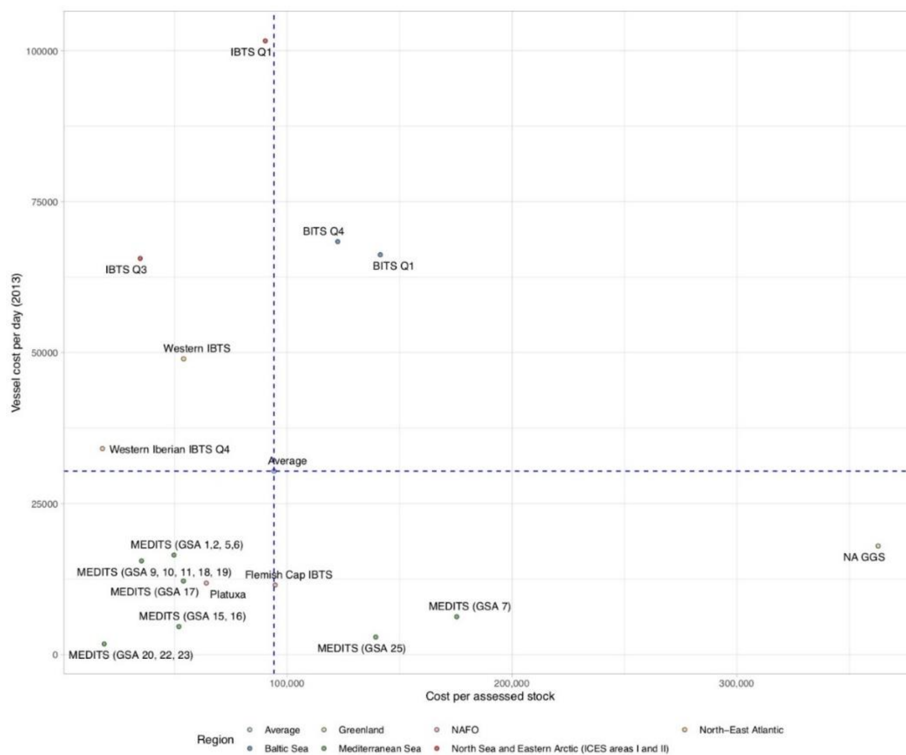



**Figure 42. Average effort per stock reported vs. average vessel cost per day (by survey)**

Additional or complementary criteria should also be considered, even though some of them would need further information and discussion. This particularly applies to the creation of value for money strategy. New value can be created, for instance, by applying HTS methods to those selected stocks where current assessment methods exhibit limitations. Other authors (Ovenden et al. 2015) have pointed out that, given the economic effort, the best candidates for taking up these methods would be high value and long-running research and monitoring programs (as those selected for the FishGenome pilots).

### 3.2.2. SWOT

#### Methodology

The findings on the applicability of the techniques from work carried out in the State-of-the-art reviews (Section 1 of this report), the outputs from pilot studies (section 2), and from the economic analysis performed in the previous task (3.2.1 Cost-benefit analysis) were thoroughly analysed to identify both internal and external factors that have an impact on the potential implementation of HTS techniques for stock assessment. A preliminary SWOT (strengths, weaknesses, opportunities, and threats) analysis was performed to assess those factors boosting and hindering the implementation of the three genomic techniques (CKMR, eDNA and DNAm) in fisheries stock assessment.

A workshop titled "Co-envisaging the actions needed towards the future implementation of genomic tools for fisheries research surveys at sea and stocks assessment", was carried out during October-November 2021, where a panel of experts convened to provide further insight on the outcomes of the SWOT analysis.

The overall objectives of the panel discussions were i) to further understand the pros and constrains on using advanced genomic techniques in research surveys and stock assessments; and ii) to gather relevant information for the design of scenarios and for a roadmap on the implementation of these techniques in data collection and for scientific advice.

**Box 3: SWOT analysis**

Refers to a strategic planning tool that provides relevant information for decision-making through the identification and analysis of internal (strengths and weaknesses) and external (opportunities and threats) factors.

A SWOT analysis is designed to facilitate a realistic, fact-based, data-driven look with a focus on leveraging strengths and opportunities to overcome weaknesses and threats.

The panel of experts shared their expertise and knowledge on surveys at sea, data management (to ensure integrity of long-time series) and provision of scientific advice and stock assessment for fisheries management and Regulation implementation and included:

- Experts from the project on the use of the HTS techniques.
- Experts on survey design, data collection (especially DCF), and stock assessments in ICES and GFCM.
- RCG and relevant ISSGs and other subgroup chairs, experts on the relevant scientific surveys.
- EC officials of different DGs, including the JRC.
- STECF independent experts, as they are directly involved in the evaluation of DCF reporting obligations.

The workshop was held remotely and structured in three individual sessions addressed to the same group of people. These sessions were carried out in a consecutive manner, so that each session built on the outcomes of the previous one, to ensure consistency on debates and conclusions.

All three sessions were designed to address key aspects on:

1. The requirements and pros and cons of the HTS techniques
2. Implications of their implementation on the surveys and data collection and quality (logistic aspects on board, need for biobanking solutions, data accessibility, interoperability and sharing, complementarity or replacement of certain data sets)
3. The fisheries stock assessments and opportunities towards Management Strategy Evaluations (e.g., estimator for Harvest Control Rule)
4. Other relevant needs and considerations from the following perspectives:
    a. Environmental aspects.
    b. Economics and financial aspects (mostly about foreseeable cost implications but also on funding requirements and framework).
    c. Regulatory and governance aspects.

A final SWOT analysis was then produced, setting the basis for the strategic planning that will be embodied in the roadmap for implementation (see 3.2.3).

### Results

The most relevant issues discussed during the **first session** were the following:

- The optimization of research surveys has taken a long time and considerable effort, so any changes derived from technology and scientific progress should be proposed and assessed under a strategic and systematic approach.
- There is a general agreement on the need to set a period of co-existence of both methods, traditional and genomic. Thus, the potential impact of implementing genomic techniques on the cost of the survey was also a concern.
- However, the cost-benefit is expected to become more favourable, providing that some of the traditional techniques can be progressively replaced.
- The strength of genomic techniques increase sharply when combined with other sources of knowledge.
- The relevance of systematizing the estimation of some parameters, such as age for fisheries assessments. Classical ageing methods have proven inaccurate for some species, for which only length distribution can be, at present, used for stock assessment. Even for those species where otoliths provide accurate estimations of age, any method that provides a more mechanistic/measurable process rather than based on individual interpretation is considered a potential improvement and a huge contribution. Other parameters could also benefit in the same way.
- The possibility of using samples from the commercial fleet to fulfil the high sampling requirements of the CKMR technique was a recurring topic. Due to the nature of genomic methods, these samples could be considered as a fishery independent data source in spite of being collected from the commercial fleet. This approach should contribute to improve assessments, especially for data-poor stocks.
- The implementation of tools that allow a better understanding of connectivity is essential in a context of global change, as it would allow to monitor changes in populations over time, on a regular basis.

During the **second session** of the Workshop a selection of key elements and necessary assumptions to define the long-term vision for the implementation of the HTS methods' roadmap was discussed.

**Three different scenarios** of implementation were presented:

1) A conservative option, that would be to keep the status quo, with HTS methods progressing under a science-driven approach.
2) An optimistic ambition scenario, entailing a full-scale and rapid implementation of the HTS techniques, where genomic and traditional methods co-exist and complement each other.

3) An intermediate scenario, with a progressive and steadily implementation of the genomic tools into stock assessment.

After a short discussion, the third scenario – the intermediate – was considered the most realistic and therefore adopted as the "target scenario"; then the discussion focused on this scenario. It involves a mid-scale genomic sampling partially routinized (only implemented at full scale for just some surveys and stocks) and partially implemented in the EU-MAP, with a limited use in stock assessment in the mid-term. However, the assessment would be improved for at least a few stocks. Nevertheless, this scenario envisages future steps towards a full implementation.

In this session, a set of criteria that are likely to affect the steps and timing for the implementation of genomic methods in surveys' protocols and their use in advice were also discussed. In addition, we debated what is needed for a successful implementation. The outcome was that criteria related to stock assessment, technical readiness and sample availability should be prioritized, while economic and environmental criteria were less relevant. As a result, the project team produced a list of prioritized surveys, species and implementation conditions, and some relevant research needs were pointed out, concluding that the roadmap should work under an operational perspective, while it should be science-oriented for the development of the techniques. These results feed the roadmap and are shown in the next section (3.2.2).

The **third and last session** enabled the setting of the main elements needed to build a *plausible implementation scenario* and the possible actions towards the implementation of the genomic techniques in the coming years. This also included the suitable time frame and the necessary means for a successful adoption of the HTS techniques within the framework of fisheries stock assessment. Particularly important was the recommendation of finding the best framework that integrates and builds upon all the relevant actors and existing structures, namely experts of different profiles, advisory and management bodies, to follow-up the implementation of the roadmap actions and to foster success.

The SWOT analysis was revisited, based on discussions and information described above and a final deliverable D3.2 "SWOT analysis" was produced. The comprehensive SWOT integrates detailed assessment of the relevant elements for all the techniques together, but also the specific strengths and weaknesses for the implementation of each of the HTS techniques. They are described in detail below.

**STRENGTHS**

**Global**

- POTENTIAL: Modern DNA HTS methods have the potential to help overcome some of the limitations of traditional methods to assess the state of fish stocks through scientific surveys. Traditional methods are costly, which, coupled with complex logistics, may result in sparse data in space and time, and requires a long-time to treat and analyze the collected data.
- ACCURACY: once fully developed and tuned for fisheries assessment purposes, HTS methods will be able to provide more accurate data than traditional methods on their fields of application.
- RELIABILITY: HTS methods could provide a better solution/alternative in some situations where survey approaches fail and may produce more reliable data than current methods used in research surveys (for example, related to the characteristics of the species, like when otholit reading is not viable for age estimation, or related to constraints of trawl research surveys, like in areas where those cannot be carried out).
- VERSATILITY/COVERAGE: Improvements are expected on age and sex determination, population dynamics, abundance, connectivity and stock

substructure, biodiversity, etc. In some situations, survey time may be reduced if some of the data are estimated by HTS methods
- COST EFFICIENCY: Genomic technologies are progressing extremely fast and the cost for sequencing follows a clearly decreasing trend. However, this services no longer represent a major cost item in the whole implementation process.
- COST EFFICIENCY: Given that staff costs for lab work represent the most important cost item among those strictly derived from the HTS methods' implementation (not from the use of survey vessels), routinisation, standardisation, bioinformatics developments and capacity building are pathways to cost-efficiency.
- COST EFFICIENCY: The pathways for cost efficiency at the survey level have also been identified although it is not straight forward to decide which cost component should be addressed to increase cost-efficiency: vesel cost per day, no. of days at sea and staff-effort required (measured in time units).
- COST-EFFICIENCY pathways enable the identification of CANDIDATE surveys for the implementation of HTS in the future. However cost criteria need to be used with caution and several other criteria need to be considered.

**CKMR**

- CKMR offers an ALTERNATIVE FOR estimating effective population size, abundance and other demographic parameters like mortality, fecundity, recruitment and adult survival rate. It enables the identification of close relatives amongst large sample sizes of wild fish with no need to have a reference genome.
- ACCURACY for adult survival rates (CKMR through POPS and HSPs) provided that information of length/age composition is available. Collecting parameters such as size and age allows reducing assignment errors to extremely low values, if an adequate number of markers is used in the analysis.
- When combined with age data, it requires only a short-term cross-sectional sample of a population, rather than long data series, to produce RELIABLE ESTIMATIONS.
- ADDED VALUE INPUTS: Provides effective spawning stock biomass, that could be used as a very interesting BRP –biological reference point- (vs. abundance = whole adult population-). Good proxy for the precautionnary? approach to biomass.
- ADDED VALUE: Offers a better understanding of stocks' productivity, relative importance of different age classes to total reproductive capacity and a better understanding of the biology of the target species.
- RELIABILITY: Compared to traditional methods, it can reduce uncertainty in current assessments
- ADDED VALUE: It provides assessment independent of classical assessment models

**eDNA**

- Environmental DNA coupled to metabarcoding methodologies is a promising tool for rapid, non-invasive, affordable biodiversity assessent and monitoring, with enormous POTENTIAL and VERSATILITY to inform aquatic conservation and management (allowing detection of elusive/rare and or invasive species, diet analysis - trophic relationship and non-invasive population genetics)
- eDNA extraction protocols are being OPTIMIZED within the frame of the FishGenome project to target three main applications: single species detection, estimation of abundance and biomass of target species, and biodiversity assessments. (eDNA: HTS for biodiversity assessment, and quantitative Polymerase Chain Reaction (qPCR) for the quantification of a target species)
- SENSITIVITY – high detection probability. Higher sensitivity than conventional survey methods. Collection methods typically have sought to identify organisms at low densities and, thus, should be optimized for detection sensitivity
- APPLICABILITY under both multiple- and single-species approach
- AFFORDABILITY: No/low need for sophisticated field equipment. eDNA has proven to be the most affordable technique among those used in FishGenome, but cost efficiency needs to be formulated in terms of the investment needed to obtain the

same output, therefore, in cost terms, there not much use in comparing with the HTS techniques.
- EASY SAMPLING: Allows significant increase of spatial and temporal biological monitoring in aquatic ecosystems, due to the ease of collecting water samples. Can become significantly cheaper and less time-consuming than conventional survey methods.
- ACCESS TO DIFFICULT-TO-REACH ENVIRONNMETNS: Allows study of inaccessible or very difficult to reach environments, such as rocky areas or deep sea.
- ADDED VALUE: Could provide info on the co-occurrence of species and relationships through ecological networks – enhance ecological quantifiable info.
- ADDED VALUE: Can produce biodiversity data at unprecedented scales.

**DNAm**

- Age-class distribution is one of the most important demographic parameters in a wild fish population, with a key role on biomass distribution and reproductive potential. Epigenetics age determination by DNA methylation provides high ACCURACY, PRECISION, REPRODUCIBILITY, SENSITIVITY AND SUFFICIENT RANGE.
- APPLICABILITY: An epigenetic clock should be possible to build for any fish species since clock-like DNA methylation changes in some loci seem to be a conserved feature in all vertebrate genomes. Common conserved loci would enable the develpment ot multi-species clock, which would represent a major advantage.
- APPLICABILITY: Epigenetic clocks should work well in both short and long-lived species. For clock construction, all age classes should be targeted with sufficient number of individuals per class.
- EASY SAMPLING: The tissue of choice for clock development would be the fin clip because it is easy to obtain and preserve and is already used for many genetic studies.
- EXPECTED LOW COST: Advances in techniques of measuring DNA methylation will make it possible to estimate age in large sample sizes at a very low cost.

## WEAKNESSES

**Global**

- MATURITY: HTS methodologies are scientific developments which still need further innovation steps to fit the stock assessment specific needs.
- ACCURACY: The improved accuracy of the HTS methods compared to traditional approaches is yet to be demonstrated in a variety of scenarios.
- COST_EFFICIENCY of HTS techniques has not been totally proven at present; as most cases claiming such an advantage do not refer to stock assessment but to different applications, cost-efficiency cannot be directly inferred from those studies.
- ASSUMPTIONS: The estimation of benefits is not straightforward; some assumptions need to be made for correct estimation. It is difficult to make a quantitative assessment of the benefits, as those may be of a very diverse nature.
- REPLACEMENT POTENTIAL: HTS methods do not provide all the parameters that traditional methods are able to provide (size structure, maturity, abundance-at-age) and comparative analysis suggest that both approaches shall be used complementarily rather than as substitutes.
  Furthermore, fisheries research surveys also provide data for environmental assessments related to oceanographic conditions or the presence of abiotic contaminants that cannot be estimated from HTS methods
- LACK OF STANDARDISATION: in most HTS techniques, one of the drivers is the constant evolution in bioinformatics, that requires a continuous adaptation of the protocols. The learning curve for the massive and routine use of the techniques in the assessment is still unknown.
- CAPACITY BUILDING: Lack of suitable facilities and equipment to perform the genomic analyses and specialized/trained staff to carry it out makes the application

of HTS techniques still highly dependent on capacity investment and/or external services and this is not always an option (e.g. bis-RAD-Seq is not offered on a regular basis as a service)
- COST EFFICIENCY: HTS cost data are not directly comparable out of genome-size similar species. The total cost of the application of epigenetics by DNAm and CKMR highly depends on the species' genome size. For that reason, for different species sequencing costs may not be directly comparable and, subsequently, not generalizable for every species assessed in EU demersal fisheries trawl surveys.
- COST EFFICIENCY: The harmonisation of information of survey costs at the EU level is not always possible, thus hampering the identification of totally comparable cost units and cost-efficiency indicators and being necessary to make several assumptions.

**CKMR**

- MATURITY: needs to be demonstrated and validated by further studies, especially on fish species characterized by large population sizes.
- COST-EFFICIENCY: The high number of samples required to detect sufficient numbers of kinship pairs, together with the necessary understanding of the biology of the species and of its habitat use, constitute major limitations that might prevent the application of CKMR to some populations. Even when the high costs and effort are feasible, important technical challenges have to be considered.
- BACKGROUND NEEDS: solid knowledge of the biology of target species and its population structure is required. Therefore, it cannot be used for species which are poorly studied or with unknown reproductive biology.
  Taking into account patterns of social structure and habitat-use is essential, but both are unknown for a vast number of species
- TECHNICAL REQUIREMENTS: Bioinformatics analysis may not be straightforward. Parentage analysis becomes markedly more challenging when neither parent is known by observation.
- TECHNICAL REQUIREMENTS: No commercial/published software available specifically for CKMR, i.e. that integrates the selection of informative SNP with kinship analysis, when the second is highly dependent on the first:
  - If the selection of the set of SNP markers is insufficient, it might lead to difficulties in establishing kinship o, worse, to an overestimation of kinship..
  - If the selection of the set of SNP markers is inadequate because the SNPs are under a strong selection, they might become useless in the near future
- TECHNICAL REQUIREMENTS: The generation of datasets for genetic identification of kinship requires strict quality control steps for genetic identification of genetic relatedness
- TECHNICAL REQUIREMENTS: Poor DNA quality may cause problems for the population sequencing analyses, together with low sequencing coverage, PCR duplicates, genotyping errors, allele dropout and null alleles (bioinformatics)

**eDNA**

- BACKGROUND NEEDS: Extreme water-volume to biomass ratio, the effects of sea currents and waves on dispersion and dilution of eDNA, and the impact of salinity on the preservation and extraction of eDNA can influence capture and detection. An appropriate use of the technique in the marine realm requires a better understanding of eDNA mainly in four domains: origin, state, transport and fate, as well as taking into account several parameters that characterize this environment. Concentration of eDNA is dependent on biomass, age and feeding activity of the organisms, as well as their physiology, life history and use of space. It is not always feasible to have all this knowledge at hand
- LIMITATIONS: eDNA does not provide information on population structure or fish condition: it is not efficient to estimate age and size structure of the population. Also, it does not detect species' hybridization

- DNA DEGRADATION: The DNA in environmental samples is typically highly degraded into fragments, limiting the scope of eDNA studies, as often only small segments of genetic material remain.
- RELIABILITY. Since eDNA is a sensitive method, there are many potential sources of "error". Some of these errors are associated to collection, laboratory and bioinformatics procedures and include contamination, inhibition, amplification and sequencing errors, computational artifacts and inaccurate taxonomic assignment. From these errors, the most serious is probably the risk of contamination and hence the possibility of false positive results.
- SENSITIVITY: Collection (DNA capture) methods typically have sought to identify organisms at low densities and, thus, should be optimized for detection sensitivity
  It is also possible that protocols available in literature are quite generic, need adaptation and fine-tuning and may not be thoroughly described, hampering replicability.
- STANDARDISATION: Diverse approaches for sampling and interpreting DNA data that result in a variety of protocols (lack of standardization). There is no single universal processing workflow that provides a unified and streamlined manner for satisfactorily treating eDNA data from raw sequences to taxonomic identification and diversity analysis.
  Slight variations in the protocol might affect its robustness. This is emerging as one of the major challenges for implementing the method as a routine assessment: Protocols should be very detailed to allow standardization and reproducibility. Calibration test runs are needed.
- RELIABILITY: Results taken from eDNA metabarcoding data must be cautiously considered, given that some taxa could be present in the final dataset by erroneous assignments, due to contaminations, mistagging, or PCR and sequencing errors (false positives) and some other taxa can remain undetected, due to partial sampling, DNA extraction, PCR amplification, or HTS bias (false negatives).
- BACKGROUND NEEDS: In cases where no sufficient annotation and genetic database information exists, it might only be possible to group sequences by nucleotide similarity, using clustering methods. One solution to overcome this problem is the construction of a private database where the sequences, species labelling and geographic origin are carefully verified; however, it is a costly and laborious task.
- MATURITY: To date, there is no clear consensus on the correlation of the number of molecules estimated by qPCR or ddPCR and the actual abundance or biomass of fish in a given sample.
- BACKGROUND NEEDS: Comparing water samples and sediments, the latter can act as DNA repository of ancient biodiversity and does not necessarily represent present-day conditions, so seawater is preferred.
- TECHNICAL REQUIREMENTS: The most abundant species take over the majority of sequencing reads – deeper sequencing might be necessary to identify the least abundant ones.

**DNAm**

- MATURITY: The technique is still in at an emerging scientific stage. Further research and developing/testing the technique are necessary.
  In contrast to other molecular approaches that can be applied to fisheries management, only a few scientific publications exist on the subject of epigenetic clocks to estimate age in fish.
- TECHNICAL REQUIREMENTS: Need of an independent calibration method for fine-tuning of an epigenetic clock – calibration possible only when the true age of the individuals is known (species where otoliths provide an accurate estimation of age). However, most fish species, unlike mammals or birds, have an undetermined growth pattern. This results indifficulties for validation of epigenetic age estimates.
- REPLACEMENT POTENTIAL: Data to be collected during sampling would need to include weight, length and sexthe ongoing work in surveys should continue.

- PRECISSION below 1 year: The ability to detect age classes between 0 and 1+ may be relevant when addressing highly dynamic populations or short-lived species, and precission is lower in this case.
- TECHNICAL REQUIREMENTS: The sensitivity (minimum value –age- that can be detected) of an epigenetic clock depends on the minimum age of the individuals used to build the clock and, in some fisheries, it might not be easy to capture individuals of all ages.
  Changes in the environment may affect the tick rate of piscine epigenetic clocks and these might represent a biological challenge that needs to be assesed on a case by case basis.
- MATURITY: Bis-RAD-Seq is believed to be best technique for epigenetic age determination. So far, this is not offered on a regular basis by companies dedicated to genomic services.

**OPPORTUNITIES**

**Global**

- There are several initiatives promoting biobaking solutions to realize the principle 'sample once and use the data several times'. This not only can have a positive effect on cost-efficiency, but also enable some flexibility to any implementation scenario.
  The setup of a biobank network for fisheries research surveys could open a new opportunity for regional cooperation in the context of the DCF, with potential benefits for many other policies (MSFD, Research, Biodiversity, etc.)
- Artificial intelligence and big data technologies are expected to have a short term impact on bioinformatics, making the tools more powerful to deliver high quality estimates from the available data inputs.
- An increasing effort in genome sequencing of marine life worldwide is expected, which, combined with open access to scientific data and information, will also speed-up progress and enable valuable reference data. For example, genome reference data for more species, which will impact possitively the potential of the HTS tools studied.
- DNA HTS methods have the potential to provide relevant information to different policy frameworks and regional conventions for the protection of the marine environment: coordination, efficiency and synergies can be further explored.
- Efforts towards strengthening regional cooperation in data collection are yielding sucessful outcomes: Member States (MS) shall agree at marine region level on the data to be collected, based on the identified needs of end users of scientific data ('end-user needs'), including, where appropriate, the species, stocks, regions, variables, methodology and frequency of data collection. The progress and efforts towards regional coordination of the data collection through the Regional Coordination Groups (RCGs) constitute a suitable forum for discussion, agreement and recommendations to MS on the adoption of HTS methods. The new EMFAF will continue providing financial support in this direction.
- Collaboration of the commercial fishing fleets could make viable the collection of large amounts of samples, which cannot replace independent surveys, but supplement the data available and reinforce scientific advice.
- MSFD is under revision and, among its priorities for the future period, it is stated that it shall contribute to the modernisation of monitoring and reporting. This willincrease the opportunity for new technologies and reduce the cost of these activities.
- The EU Green Deal reinforced commitments towards biodiversity restoration and sustainable exploitation of fisheries and the Recovery Package represent an opportunity for the adoption of new measures towards this objectives and for investment that could favour decisions (at EU level, by MS and by the fishing industry) towards the implementation of the DNA HTS methods for fisheries data collection and fisheries assessments.

- The claims for the Ocean Science Decade at international level, the recent launch of the new framework programme for research and innovation, Horizon Europe, and within it the EU Mission Restore our Ocean and Waters, open a unique and timely opportunity for addressing many of the research needs identified for the use of HTS methods within FishGenome and other initiatives (e.g. the H2020 PANDORA project), as well as to take advantage from the progress achieved so far by these and other related projects. It is important to highlight the target established by the EU Mission that 20% of DNA of life in our ocean and waters is to be fully sequenced and publicly available by 2025, 50% by 2030.

## THREATS

### Global

- The existing gap between academic research priorities and the science-policy advice processes represents a challenge when attracting geneticists and bioinformatic experts to the assessment and advice processes – there's a need to set new incentives, visibility and recognition for the work around fisheries assessment and advice.
  Geneticists dedicated to support fisheries management must understand and commit to address issues and objectives that are identified in policy and management. Policy and decision makers must be aware of the opportunities and limits of genetic and genomic approaches. All of this needs a dedicated dialogue framework that is yet to be built.
- The pace of technological change around bioinformatics and genetics represents a big challenge in the fisheries data collection and assessment realm. Despite regional harmonisation of procedures and protocols, long-term series for data, a multi-level governance framework etc. it is yet difficult, in such a complex framework, to encompase and take full advantage of technological innovations. On the other hand, extremely rapid technology advancement in genomics may cause that fisheries related research gets biased towards technology, rather than management and policy. Finding the right balance represents a challenge to any implementation scenario.
- Curation of DNA dabasases for fisheries species is of paramount imporance for an accurate use of the HTS in biodiversity assessments. Marine life DNA databases of generic use include errors and this has lead to a proliferation of specific databasis, which do not solve the overarching problem. Curation and integration efforts need to be encouraged and requires the engagement of the scientific community internationally.
- Disruption caused to routinized procedures, long-term series and data warehousing by the fisheries research surveys and scientific assessment communities may represent a barrier for an ambitious HTS implementation scenario.
- A regular implementaion of the HTS methods (irrespective of the ambition) would need incorporation to the EUMAP. This would imply a mid/long-term process, likely needing a stepwise approach. It should start from a careful selection of the surveys and of the stocks (FishGenome is expected to aid in this process), and ensuring the capacity and resources for the adaptation of the necessary National Work Plans, exploring the feasibility to incorporate this effort to a Regional Work Plan and carefully considering also future adaptation of the reference databases managed by the EC-JRC, STECF and ICES.
- Political instability and increasing trend of energy prices could have an impact on the economic recovery forecasts: post-brexit, post-pandemic, increasing inflation rates affecting all the sectors, disruption on global transportation etc. could have an impact on public finances. As a result, investment towards improving the common framework for fisheries assessments may not be among the priorities. The rise of fuel costs will directly impact the cost of fisheries research surveys.
- Administrative burden and restrictions derived from international regulation frameworks:

    - Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits arising from their utilization.
    - Future International Legally Binding Instrument for access to genetic resources in waters beyond national jurisdictions (UN Convention on the Law of the Sea)
- Increasing political and societal pressure towards the baning of research based on animals use. However, it is unclear whether this represents a threat or an opportunity in the context of the HTS methods implementation, as the techniques are non lethal and e-DNA does not even need fish samples.

### 3.2.3. Roadmap for the implementation of genomic methods

FishGenome roadmap sets a pathway to progress towards a full-scale implementation of HTS methods into fisheries stock assessment and management. It provides precise and accurate information on whether and how the analysed genomic-based approaches could become part of the regular research surveys, describing the steps, pathway, and timeline for a progressive implementation of the genomic methods. Below is a summary of the deliverable D3.3 *Roadmap for the implementation of genomic-based approaches in fish stocks' assessment*.

#### Methodology

To set a realistic and suitable target for the roadmap, a thorough description of the current situation was the first step. To do so, a set of criteria were proposed, completed, and validated by the representative group of experts. As a result, five families of criteria, considered to condition the steps and timing for the implementation of genomic methods in surveys' protocols and their use in advice, have been proposed:

- Technical criteria: readiness, versatility, accuracy, precision and replicability, complexity, value of the information provided on requested parameters;
- Survey logistics: vessel (characteristics and equipment constraints), time availability (by staff on board), noise (in the standard survey procedures), samples availability;
- Stock assessment: robustness (gains in accuracy, precision), disruption of time series, impact on assessment models (and databases, etc.), added value;
- Economic criteria: number of days at sea (vessel cost/ day), staff needs (time and skills), material and infrastructure needs, implementation costs;
- Environmental aspects: animal welfare, carbon footprint, waste production (single-use products).

Additionally, it was necessary to set the parameters that help defining the needs for the implementation of HTS techniques, which have been grouped into four categories:

- Candidate surveys: selection, sampling design;
- Candidate species/approaches: stock biological background knowledge, stock assessment role, commercial/ecological importance, alternative approaches rather than species (e.g. for eDNA no specific species are targeted in the survey);
- Implementation needs: capacity building, infrastructure/ facilities, regulation needs;
- Research needs: refinement of tested methods, standardization and demonstration, other HTS methods, biobanking, future challenges.

All the above criteria and parameters set the basis to define the most plausible scenario for the implementation of HTS methods. The most likely scenario would entail mid-scale, well-designed and partially routinized genomic sampling, that is applied only in some

surveys and/or to some stocks, therefore partially implemented in the EU MAP. The mid-term use in stock assessment would be, thus, limited to only selected ecosystems.

Under this approach, additional costs would have to be borne, with no significant risks, while achieving improvement in those stock assessments where it is applied. Regarding the EU MAP, no modifications are deemed necessary, since it is flexible enough; introduction of new techniques and set-up of pilot studies is foreseen, and it allows for modifications of the National/Regional Work Plans accordingly. Therefore, the improvement in fisheries assessment due to the application of genomic methods will be step-wise rather than disruptive, as initially expected.

The state-of-the-art analyses and the pilot studies, conducted in FishGenome, already identified the difficulties of implementing genomic tools into stock assessment for a large number of stocks in the short-term. The reasons for this were identified in the SWOT analysis and later confirmed during the workshop held with experts linked to the EU Data Collection Framework and stocks assessment.

Based on the above analyses, the most plausible scenario advocates for a **stepwise implementation** with a **progressive adoption/integration** of genomic information in the assessment and advice procedures and envisioning a full implementation in 10 years. This progressive implementation will provide sufficient time and possibilities to further **improve the methodologies** (genomics, bioinformatics, sampling, modelling, assessment) and **increase the capacities** required for the full implementation, while benefiting from the fact that, for some techniques and in some stocks, a short-term implementation is possible. Thus, in this stepwise scenario, it will be possible to **demonstrate the benefits** of the methods, although **stronger coordination** is needed to learn from each step before taking the next leap.

Following the selection of the scenario for the implementation, five specific objectives of the roadmap were defined:

1. Ensure the progressive uptake of genomic information for assessment and advice, throughout several phases, by i) integrating and standardizing HTS methods across surveys; ii) improving robustness of data series and stock assessment; and iii) exploring collaboration with commercial fishing fleets and other opportunistic sampling scenarios.

2. Guarantee the continuous improvement of the methodologies by adjusting and standardizing protocols as required, performing lab intercalibration and incorporating further scientific findings and technology developments (chips, different sampling alternatives, etc, while exploring other genomic methods for the same goal and for other biological parameters.

3. Coordinate the implementation efforts in the context of the EUMAP DCF, involving the Regional Coordination Groups, ICES, STECF, RFMOs, etc. to ensure a successful HTS methods uptake.
4. Develop capacities across Europe by i) improving skills and building engagement to achieve experts' buy-in; and ii) reinforcing infrastructures and laboratories, and collaboration among them, especially for samples' processing, biobanking of samples and data, large-scale sequencing solutions and automation of bioinformatic analysis.
5. Demonstrate the benefits of the methods more systematically, gathering relevant data for addressing cost-efficiency analysis across objectives and proposed action lines and doing so beyond stock assessment.

To achieve the overarching goal of the roadmap, the most relevant strategic areas for activity orientation were identified, defined as strategic pillars. Each pillar requires, the involvement of different stakeholders to a certain extent. Five strategic pillars were defined:

SP1. Genomic techniques
SP2. Survey and logistics
SP3. Scientific advice and stock assessment
SP4. Financial and economic aspects
SP5. Governance and other policies

Finally, the roadmap envisions a full-scale implementation in 10 years, counting from the launching of actions. The timeframe for the roadmap is divided into three periods of varying duration:

- Short-term: years 1 and 2;
- Mid-term: years 3 to 5;
- Long-term: years 6 to 8 and beyond.

This timeframe has been established considering 2023 as starting point and according to the milestones related to the DCF National Work Plans (NWPs) and the upcoming Regional Work Plans (RWPs). For both, the entry into force of the revised NWPs/RWPs is foreseen in 2025 while the expected start date of new NWPs/RWPs is in 2028.

### Results

Based on the scenarios outlined, the objectives and pillars, as well as the timeframe for the implementation, we have defined a general Action plan to address each of the five objectives of this roadmap:

1. Towards a progressive adoption of the genomic information
2. Continuous improvement of the methodologies
3. Fostering a coordinated roadmap
4. Developing capacities for a successful implementation
5. Ensuring value for money

For each objective several specific actions are identified based on the rationale described above, summing a total of 45 actions (described in detail in Deliverable 3.3).

#### 1. Towards a progressive adoption of the genomic information

This strategic challenge aims to ensure the progressive uptake of genomic information for stock assessment and scientific advice. The various genomic techniques have different levels of maturity and readiness, the research surveys differ in capacity to implement the demanding routines, and some species (or more properly, stocks) are more suitable to embrace genomic approaches. Thus, going from the most plausible scenario to the full-scale implementation scenario requires a stepwise approach. Consequently, the actions within this challenge are organised in three phases: short-, mid- and long-term.

The definition of each phase depends entirely on the criteria used to propose candidate stocks and surveys. The criteria to select candidate surveys for the application of genomic methods during the next three years (Phase I), avoiding disruption of the survey protocol, are outlined below:

- Surveys from the EU MAP mandatory list of surveys, whose use in scientific advice has been assessed positively by STECF. Further information can be found in the STEFC report on the Evaluation of Mandatory Surveys under the DCF[9].
- Surveys providing data for stock assessment mostly and, among these, stocks for which data is provided by only one or very few surveys. This will reduce the risk on implementing the genomic methods in a single survey that provides only partial information of a stock; but also copes with the necessity on implementing the genomic methods in too many surveys to cover the whole stock distribution.

[9] https://stecf.jrc.ec.europa.eu/documents/43805/2457962/STECF+19-05+-+Ev+mandatory+surveys+DCF.pdf/758dda47-836a-44c4-bd63-9b4c9edf7d4f

- Surveys that are cost-effective and where the vessels' characteristics facilitate the implementation due to the availability of suitable space and facilities onboard.

Regarding the selection of candidate species, the following recommendations should be considered:

- Cover several stocks of the same species, to test if the techniques work within the same species and that the achieved results are not stock-specific.
- Select species with different characteristics, paying particular attention not only to teleosts, but also to crustaceans, cephalopods and elasmobranchs, among other taxa.
- Consider both data-rich (to facilitate calibration) and data-limited (to provide valuable information) stocks.
- Consider ecosystem indicators/parameters that can be estimated from these methodologies.

In the two subsequent phases (Phase II and Phase III), case studies should be expanded both in terms of, firstly, species/stocks and then, surveys.

Finally, the actions within this challenge, must ensure a proper integration and standardization of HTS methods across surveys, as well as an improvement in robustness of the genomic information, the data series and, more importantly, stock assessment. To this end, 12 specific actions are defined, eight for phase I and four more for the other two phases (Table 12).

**Table 12.- List of proposed actions related to Objective 1, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets].**

| Objective 1: Progressive Adoption | Pillar | Timeframe |
|---|---|---|
| Action 1.1. Phase I: Definition of case studies I | SP5 Governance | Short [1] |
| Action 1.2. Phase I: Sampling and data protocol | SP2 Survey & Logistics | Short [1] |
| Action 1.3. Phase I: Testing results and robustness of ageing data | SP3 Advice & assessment | Short-Mid [2-3] |
| Action 1.4. Phase I: Testing results and robustness of eDNA data for abundance estimation | SP3 Advice & assessment | Mid [3-5] |
| Action 1.5. Phase I: Testing results and robustness of genomic data for stock structure and connectivity | SP3 Advice & assessment | Short-Mid [2-3] |
| Action 1.6. Phase I: Testing results and robustness of CKMR data for abundance estimation | SP3 Advice & assessment | Mid-Long [4-6] |
| Action 1.7 Phase I: Simulate assessment and optimize sampling plans | SP3 Advice & assessment | Mid-Long [4-6] |
| Action 1.8. Phase I: Final analysis on impact on assessment | SP3 Advice & assessment | Mid-Long [3-6] |
| Action 1.9. Phase II: Definition of case studies II | SP5 Governance | Mid [4] |
| Action 1.10. Phase II: Sampling and data protocol | SP2 Survey & Logistics | Mid [4] |
| Action 1.11. Phase III: Full implementation | SP5 Governance | Long [7] |
| Action 1.12. Phase III: Sampling and data protocol | SP2 Survey & Logistics | Long [7] |

Following the definition of cases studies for phase I (Action 1.1), detailed sampling and data protocols must be developed and agreed among the parties involved (Action 1.2) to ensure sufficient quantity and quality of samples for a robust development of genomic approaches. We do not define specific actions regarding the sampling itself, as the sampling must be conducted within ongoing surveys already supported and coordinated by DCF and the RCGs. However, several coordination actions are proposed (see "*Fostering a coordinated roadmap"* below) to ensure that sampling develops accordingly.

Four actions are then defined to conduct the analyses, test results and the robustness of each of the genomic approaches: epigenetics for ageing individuals (Action 1.3), eDNA for abundance estimation (Action 1.4), RAD-Seq for stock structure and

connectivity (Action 1.5) and the use of CKMR for abundance estimation in vulnerable, data-limited, and/or difficult to assess stocks (Action 1.6).

These four actions should provide sufficient information to conduct the two key actions needed for the implementation of genomic approaches into stock assessment. First, conducting simulation exercises with the new genomic information incorporated into assessment models, to evaluate their performance and to optimise sampling plans (Action 1.7). Second, conducting a final analysis on the impact of the new approaches into assessment and advice should ensure that assessment, advice, and management can make use of the new information (Action 1.8).

Phases II and III will replicate the above scheme, but, in their case, only two actions are defined for each phase. First, an expansion of case studies to incorporate additional species and surveys (Actions 1.9 and 1.10). Then, a final boost towards a full implementation in those assessed stocks and surveys conducted within EU MAP, where feasibility has been demonstrated (Actions 1.11 and 1.12).

### 2. A continuous improvement of the methodologies

As mentioned above, the genomic techniques have different levels of maturity and readiness. Moreover, the related technology is still progressing very fast and some refinements and optimizations to the current genomic methods are still required. The actions within this strategic challenge should guarantee the continuous improvement of the genomic methods chosen to be implemented in surveys and used for stock assessment by adjusting and standardizing protocols as required. This includes lab intercalibration and incorporating further scientific findings and technology developments (chips, different sampling alternatives etc.). At the same time, the actions within this initiative will explore the use of other genomic methods towards the same goal, as well as the use of the same methods to estimate other biological parameters of interest in scientific advice and fisheries management. To this end, nine specific actions are proposed (Table 13).

The pilot studies have demonstrated that an important effort is still required to further develop suitable protocols and standardize the genomic methods across laboratories (Action 2.1). All tested methods have shown a very high potential for use in stocks' assessment and scientific advice; however, since they have been developed recently and are continuously evolving, a refinement and optimization of the approaches is needed (Action 2.2). The methodologies intended to estimate stock abundance (eDNA, CKMR) will produce parameters that are not, nowadays, of direct use in age-structured assessment models, because they will not provide abundance-at-age. This is especially the case for eDNA, which does not require the collection of the individuals, therefore preventing the measurement of their size and obtaining the age. However, the estimation of abundance with these methods provides valuable information to build time series of biomass/abundance indexes and/or can be used as Biological Reference Points (BRPs), improving stock-recruitment relationships or similar indicators to be used in stock benchmarking (Action 2.3).

Similarly, epigenetics can estimate individual age, but initially requires the analysis of individuals of known chronological age, as a reference to calibrate the epigenetic clocks. However, it can be applied in species where ageing is problematic, and for which our capacity to obtain age data is limited (Action 2.4). Moreover, epigenetics could generate in the future universal epigenetic clocks that are evolutionarily conserved across a broad range of species. They would constitute an invaluable tool to estimate age in stock assessment, considering that the method is non-lethal and suited to automation; however, it still needs considerable research effort (Action 2.5). CKMR in its current form seems to be applicable only to very small-sized populations; and although its use could be expanded by extending the array of close-kin categories to include more distant relationships, it should be carefully tested, as it could exacerbate errors of assignment (Action 2.6).

The SWOT analysis shows that many of these approaches require a technically demanding sampling (eDNA) or the collection of a vast number of samples (CKMR). They both limit the capacity of the surveys to uptake these methodologies at large scale. We propose two actions related to sampling technologies: (i) to improve the collection of eDNA, by developing “DNA laboratories” that can be deployed subsurface (Action 2.7); and (ii) to develop automatic or semi-automatic tissue collectors to speed-up the process of gathering tissue samples (Action 2.8) and avoid contaminations and traceability errors.

Finally, these techniques will produce a large amount of data that requires specific infrastructure and a well-designed workflow in order to be used in stock assessment, advice and management (Action 2.9).

***Table 13.- List of proposed actions related to Objective 2, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets].***

| Objective 2: Improve methodologies | Pillar | Timeframe |
|---|---|---|
| Action 2.1. Protocolization and standardization of methods across laboratories | SP1 Genomic techniques | Short-Mid [1-3] |
| Action 2.2. Refinement and optimization of current approaches | SP1 Genomic techniques | Short [1-2] |
| Action 2.3. Alternative use of genomic information in advice | SP3 Advice & assessment | Short [1-2] |
| Action 2.4. Studies of epigenetics in age data limited species | SP1 Genomic techniques | Short-Mid [1-3] |
| Action 2.5. Development of a multi-species epigenetic clock | SP1 Genomic techniques | Mid [3-5] |
| Action 2.6. Further development of CKMR capacities | SP1 Genomic techniques | Short-Mid [2-4] |
| Action 2.7. Improving eDNA collection techniques | SP2 Survey & Logistics | Short-Mid [2-3] |
| Action 2.8. Development of new sampling methodologies | SP2 Survey & Logistics | Mid [3-5] |
| Action 2.9. Development of workflows and infrastructure for the management of genomic information | SP4 Financial & Economic | Short-Mid [1-3] |

## 3. Fostering a coordinated roadmap

Attaining a successful integration of the HTS methods into fisheries stock assessment and management requires the engagement of all the relevant disciplines and stakeholders under a sound coordination strategy. Such strategy should rely on the existing initiatives for coordination and cooperation (e.g., Regional Coordination Groups), working groups (ICES, RFMOs) and committees such as STECF, acknowledging the fact that most relevant experts are involved in these already existing groups and networks and that these are the forums where technical discussions on surveys and assessment take place. Thus, a coordination should be conceived as an umbrella or overarching structure for the overall management of the roadmap to facilitate the implementation, the dialogue among parties and ensure that each step in the roadmap is being taken. Moreover, this roadmap includes several activities and actions with a scope and expected benefits beyond fisheries assessments. The overarching approach of the roadmap is seeking to maximise the value for money from its implementation. However, such an approach makes the relevant stakeholders' network even wider, and its efficient coordination becomes a critical challenge. To cope with this challenge six actions have been defined (Table 14).

Coordination and cooperation in the context of fisheries data collection and fisheries assessment is not new. Sound structures with achievements in this context are available. A key driver for the success of the roadmap will be its capacity to harness and underpin the potential of the already existing initiatives, minimising the new structures and avoiding unnecessary overlaps and duplications. This roadmap does not replace research initiatives, but on the contrary should couple, complement and reinforce the research carried out across the EU.

We propose a strategy that links the existing initiatives with the aim to contribute to the roadmap vision (Action 3.1), that clearly identifies the relevant agents and their stakes

towards the application of genomics in the fisheries assessment context (Action 3.2) and that embraces mechanisms for engagement (Action 3.6). Thus, the roadmap network will not be *a built from scratch* initiative. As in any other network initiatives, the links among the stakeholders will become stronger for some of the actions and weaker for some others. The kernel of the coordination efforts will focus on the genomics technologies and their application potential for fisheries stocks assessments (3.3), without disregarding the co-lateral benefits of this principal goal. Due to the complexity of this component, the actions within this strategic challenge will also embrace the definition of a risk management plan (3.5), as well as a plan for the management and sharing of data and information (3.4). The right compromise between the core technical focus on fisheries and genomics and the opportunities to contribute to other objectives - such as a positive externality of action and investment in this field (e.g., environmental, and public health policies, research and innovation, and blue economy entrepreneurship) - should be sought.

***Table 14.- List of proposed actions related to Objective 3, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets].***

| Objective 3: Coordination activities | Pillar | Timeframe |
|---|---|---|
| Action 3.1. Setting up a coordination strategy and building a coordination plan | SP5 Governance | Short [1] |
| Action 3.2. Building the stakeholders network | SP5 Governance | Short [1] |
| Action 3.3. Building the genomic network | SP5 Governance | Short [1-2] |
| Action 3.4. Setting up a data policy and data management plan | SP5 Governance | Short [1] |
| Action 3.5. Setting a risk management and mitigation strategy for the network activities | SP5 Governance | Short [2] |
| Action 3.6. Coordination and management of the network | SP5 Governance | Short-Long [2-6] |

## 4. Developing capacities for a successful implementation

Two key elements are essential for the successful implementation of the HTS methods into fisheries stock assessment and management.

First, implementation of new methodologies requires the reinforcement and development of new infrastructures, always keeping in mind the overarching priority of maximizing the use of existing ones. Current functional capacities do not include laboratories with genomic equipment or personnel trained in the use of these techniques. Thus, implementation in the short-term will require the selection of reference laboratories across Europe (Action 4.1) to receive and process the specimens and perform the genomic-based analyses. The samples (tissues, DNA) and data resulting from analysis need to be stored in biobanks and biorepositories, as these facilities are essential to guarantee their standardization, curation, and maintenance over large periods (4.2). They will play a fundamental role in promoting contemporaneous and retrospective studies and in enabling the creation of long-time series. The process of sequencing itself is normally outsourced to private companies, due to the high cost of maintaining updated sequencing equipment in small laboratories. However, developing a genomic sequencing infrastructure at the EU level would allow centralizing the sequencing needs in a unique location, lowering the costs due to the large volume of samples, while guaranteeing independence and freedom of operation from third country infrastructures (4.4). If readiness of HTS techniques is demonstrated, many of the large vessels will have the capacity to create onboard laboratories for genomic analysis (4.3). DNA laboratories onboard could produce real-time results that would allow researchers to make evidence-based decisions on sampling locations maximizing the information obtained during research surveys, and ultimately, their efficiency.

Second, it is very important to improve skills and build engagement to gain fisheries experts' buy-in. This means the involvement and compromise in the adoption of

genomic techniques from those involved in surveys and samples processing, and experts exploiting the resulting data (stock assessment). There is a critical knowledge gap between most geneticists and scientists involved in fisheries management (fisheries scientists, assessment experts, modellers, and geneticists). The implementation of HTS methods requires close communication and common understanding between scientist and the use of a common language. Moreover, it is essential to engage technical personnel in charge of classical analysis on the use of novel technologies, to avoid job losses and subsequent rejection of the implementation of new techniques. A tool to achieve these objectives is the creation of training networks (4.5 and 4.6) in genomics and bioinformatics methods, to support an array of training activities including online courses, tutorials, webinars, advanced hands-on workshops (4.10 and 4.11) and summer schools, among others. Short-stay exchanges and long-term scientific missions of researchers and technicians could bring new and updated expertise, skills and knowledge to the organizations and to strengthen collaboration links among institutions (4.8 and 4.9). Implementing concerted quality assurance and quality control (QA/QC) measurements is essential to guarantee reproducibility and reliability of the results obtained by HTS techniques (4.7).

***Table 15.- List of proposed actions related to Objective 4, with indication of corresponding strategic pillar and timeframe [indicating when the action should be initiated, years in brackets].***

| Objective 4: Developing capacities | Pillar | Timeframe |
|---|---|---|
| Action 4.1. Selection of reference labs | SP4 Financial & Economic | Short [1-2] |
| Action 4.2. Creation of a biobanking solution to manage samples and data | SP4 Financial & Economic | Short-Mid [1-3] |
| Action 4.3. Preparing vessels for the future | SP2 Survey & Logistics<br>SP4 Financial & Economic | Short-Mid [2-3]<br>Long [6-7] |
| Action 4.4. Building sequencing facilities in Europe | SP4 Financial & Economic | Mid [3-5] |
| Action 4.5. Genomics methods training network | SP4 Financial & Economic | Short [1-2] |
| Action 4.6. Bioinformatics training network | SP4 Financial & Economic | Mid [3-4] |
| Action 4.7. QA&QC in ecological genomics | SP4 Financial & Economic | Mid-Long [4-6] |
| Action 4.8. Capacity building on current technicians | SP4 Financial & Economic | Short-Mid [1-3] |
| Action 4.9. Scientific missions programme | SP4 Financial & Economic | Short-Long [2-6] |
| Action 4.10. Workshop on genomics applied to stock assessment | SP3 Advice & assessment | Short [1-2] |
| Action 4.11. Workshop on survey design and protocols | SP2 Survey & Logistics | Short [1] |

## 5. Ensuring value for money

This strategic challenge aims at assessing and demonstrating the benefits of the methods more systematically, through gathering of relevant data for evaluating cost and investment efficiency across objectives and actions and doing so beyond stock assessment.

According to the project findings, it is quite common for cost-efficiency to be taken for granted, when it refers to the use of genomic technologies. This is mainly because of the rapid technological development in recent years and, consequently, the dramatic drop in sequencing costs. However, applying genomics methods in fisheries assessments is quite innovative and there is scarce information available on the costs and investment needs derived from this specific application. On one hand, it is expected that, gradually, survey design, sample analysis, data management and assessment will need to accommodate genomic technologies (A.5.2). The magnitude of these changes can only be accurately estimated with the progression of roadmap Phases (A.5.3). Thus, to assess cost-efficiency and to decide on the way forward for the same or a better output at a lower cost, a systematic economic data gathering, and financial analysis of new processes is needed. The systematic assessment of costs will facilitate the identification of pathways for efficiency improvements through cost reduction, such as exploring the possibility of use of different sample sources and data (A.5.4). Efficiency is not only achieved reducing costs but also increasing the value of the yields, i.e. the outcomes on implementing the genomic technologies.

Finding the adequate proxies for the benefit analysis is always a major challenge is this context. However, there are opportunities to better understand such benefits, at least from a qualitative perspective, and this is something to be further studied. As an example, broadening the policy context where fisheries genomic data can be of interest (e.g. MSFD) is one of the fields for benefit increase with, in principle, no necessary extra costs (A.5.4, A.5.6, A5.7).

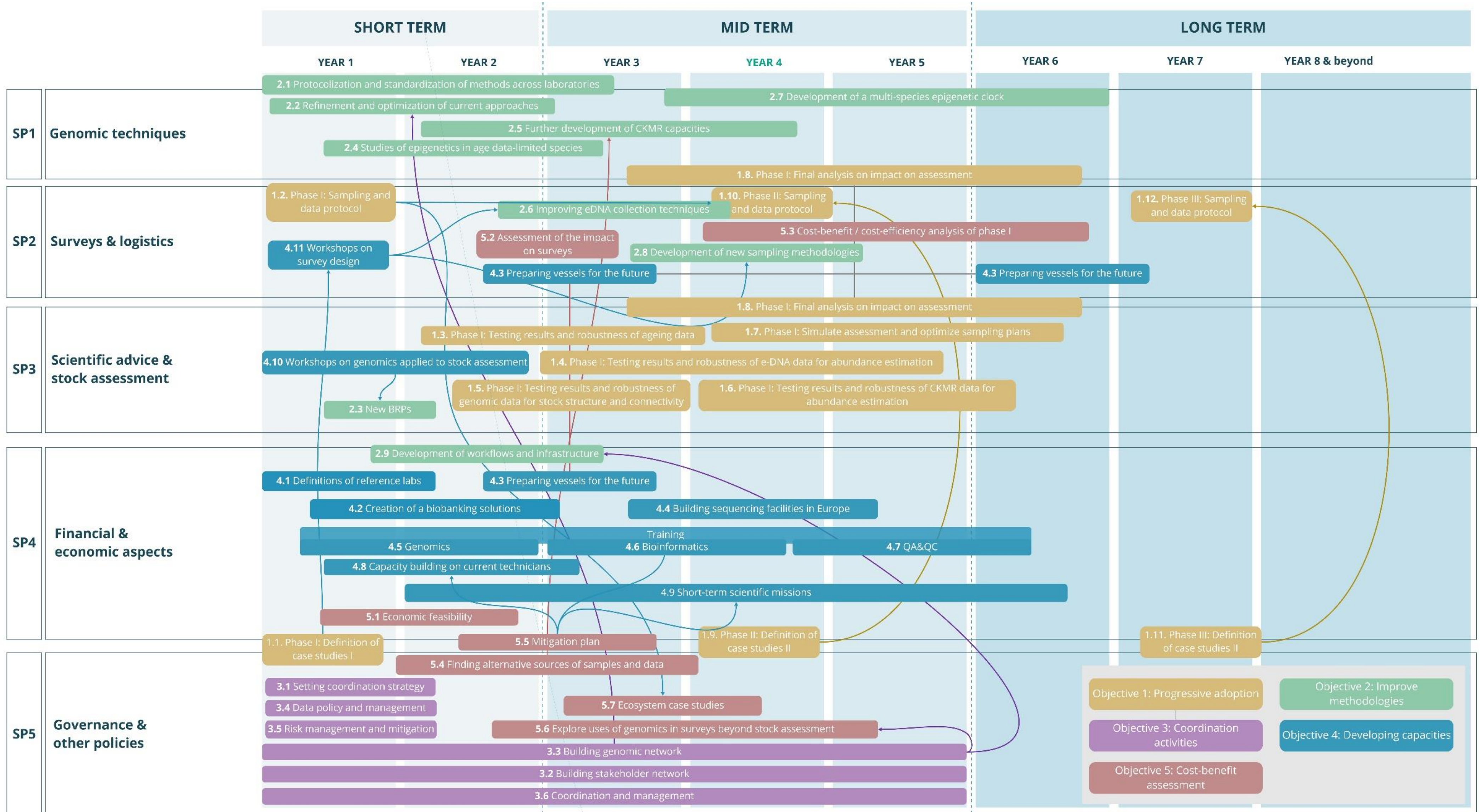
On the other hand, apart from new processes and procedures, the implementation of the roadmap involves the consideration of some tangible and non-tangible investment needs and decision-making. So, not only operational costs are expected to change. In many cases, these needs will be related to the capacity-building requirements identified in this roadmap (infrastructure, coordination structures, addressing training needs and upscaling of skills etc.). Decisions for such investments need to be complemented with financial and strategic-impact assessment to guarantee their long-term sustainability (A.5.1). Such an approach should not only apply to the planning and investment on the new infrastructures (biobank, sequencing facilities, etc.), but also to the investment on the knowledge and skills needed, including staff motivation and incentives, and risk mitigation costs (A5.5).

***Table 16.- List of proposed actions related to Objective 5, with indication of corresponding strategic pillar and timeframe [years in brackets].***

| Objective 5: Cost-benefit assessment | Pillar | Timeframe |
|---|---|---|
| Action 5.1. Economic feasibility studies for dimensioning new initiatives and support infrastructures | SP4 Financial & Economic | Short [1-2] |
| Action 5.2. Assessing the impact on surveys | SP2 Survey & Logistics | Short-Mid [2-3] |
| Action 5.3. Cost-benefit / cost-efficiency analysis on phase I, phase II, phase III | SP2 Survey & Logistics | Mid-Long [4-6] |
| Action 5.4. Finding alternative sources of samples and data | SP5 Governance | Short-Mid [2-3] |
| Action 5.5. Mitigation plan | SP5 Governance | Short-Mid [2-3] |
| Action 5.6. Explore uses of genomics in surveys beyond stock assessment | SP5 Governance | Short-Mid [2-5] |
| Action 5.7. Ecosystem case studies | SP5 Governance | Mid [3-4] |

The overall graphical synthesis of this roadmap is presented below. Many of the actions proposed and mentioned above are interconnected. These connections are represented in the roadmap scheme and further explained, together with each individual action, in Deliverable D3.3 *Roadmap for the implementation of genomic-based approaches in fisheries stocks' assessment*.

**Figure 43. The FishGenome Roadmap at a glance**



### 3.2.4. Long term prospects

The above roadmap focusses on the feasibility of the implementation of the current genomic tools to estimate stock parameters used in stock assessment within the next ten years. However, on one hand, genomics is a discipline that is evolving rapidly, and, on the other hand, the analyses performed have identified gaps and needs that could be solved beyond the ten years' timeframe.

This final task addresses a foresight analysis Of the potential evolution of the genomic tools and their progressive implementation in fisheries assessment, beyond the roadmap period. The foreseen evolution of genomic science, economic conditions, governance framework, etc. may point towards a larger feasibility of genomic methods into fisheries research in the long-term. Following is a summary of deliverable D3.4 *Long term prospects*.

#### Methodology

The analysis of long-term prospects aims at defining the future after the roadmap timeframe in fisheries genomics science. We analysed information from two sources: (i) a comprehensive review of the latest genomic literature, followed by (ii) an analysis and evaluation of the acquired knowledge during FishGenome. The latter includes the development of pilot studies to test the use of novel genomic techniques, experts' advice and opinions gathered during two workshops with the participation of European stakeholders, geneticists, fisheries scientists, and fisheries managers (management and policy).

The long-term prospects analysis provides key insight for better informed decisions regarding future applications of genomics into fisheries management. The document is focused on those tools with greater potential to deliver significant advances in fisheries management in the long-term. It also highlights their main advantages and possible barriers for their implementation in stock assessments and scientific advice. We analysed emerging ways in which research using genomic techniques can provide a better advice to fisheries managers to maintain productive and sustainable fisheries.

The report is organized in seven themes relevant to fisheries, each of which encapsulates one or several genomic sub-disciplines, focused on specific types of biological or management questions:

1) The identification of stock structure: The analysis of the distribution of genetic variation can be used to divide the range of harvested species into demographically independent regions suitable for independent management, to infer connectivity between stocks and to determine fine-scale structure of fish stocks.

2) Determination of key population parameters: There is an urgent need of new methodologies to allow more precise estimations of key population parameters, as a mean to improve fisheries assessment. Several genomic tools can address a comprehensive spectrum of needs and applications relevant to fisheries:

2.1. *Effective population size*: Estimates of genetic effective population size can be used to estimate changes in abundance through time.

2.2. *Abundance*: Novel genetic mark-recapture methods can directly estimate this critical population parameter in exploited or bycatch species.

2.3. *Size structure*: Size in fishes is largely dependent on environmental factors; however, growth has also a genetic basis and fishing mortality is inducing adaptive evolution that genomic tools can reveal.

2.4. *Biological age*: Preliminary research suggests that DNA methylation has potential to estimate age, an essential parameter in fisheries assessments. This would be particularly relevant for species that cannot be aged by conventional methods.

2.5. *Sex and maturity*: The development of genetic tools to identify sex in fish species with sex-determination systems will increase as reference genomic resources grow. On the contrary, the development of simple genomic tools to determine the maturation status of fishes is unlikely, since genomics can only partially explain the variability in this trait.

2.6. *Offspring abundance*: Knowledge of ichthyoplankton dynamics is essential to guide management of fish stocks. Mounting evidence shows the taxonomic precision and reliability of DNA metabarcoding to estimate abundance of eggs and fish larvae.

2.7. *Total and natural mortality*: Novel genetic mark-recapture methods CKMR can be useful in estimating total adult mortality rate if the adequate kinship relationships are analysed.

3) Ecosystem-based approach to fisheries management: Adopted by obligation under the CFP, this approach contributes to improve fisheries management by considering the entire ecosystem of the species being managed, e.g., by accounting for predator-prey interactions, human influence, the fish habitat, as well as other important factors.

3.1. *Food webs*: DNA-based food web analysis of feeding habits (based on plankton and water collected in the feeding area) and stomach content DNA analysis can provide detailed information on predator-prey relationships, trophic linkages, and seasonal shifts.

3.2. *Environmental monitoring*: eDNA has proven very useful to characterize biodiversity and to detect rare and invasive species. The potential of this method to monitor environmental shifts is promising, although various technical issues need to be overcome before wide large-scale application for fisheries and ecosystem assessments.

4) Fisheries control and surveillance: Even the most accurate modelling of exploitation is bound to fail in the absence of regulation, enforcement, and surveillance. Given appropriate reference material, genomic analysis can provide a suite of tools to guarantee the rapid identification of species or the assignment to the population of origin, among other capabilities.

4.1. *Species recognition*: Many aspects of fisheries management rely on the accurate identification of both harvested and non-harvested organisms. DNA-based species recognition (DNA barcoding) is a rapid, universal, and highly accurate tool to assist in the recognition of species.

4.2. *Determination of geographic origin*: Current genomic methods can trace relationships between tissue samples representing species, populations, family groups and individuals providing a tool to keep track of the origin of fishery products along the supply chain.

5) Disease detection in fisheries: Diagnosis and monitoring of diseases in exploited populations is urgently needed, due to global warming. A disease can reduce population size, cause adverse health issues in consumers, or spread in the environment. Genomic tests can be applied to both the exploited species and to the surrounding environment for disease surveillance and for understanding disease epidemiology.

6) Fisheries- and climate-induced changes in exploited populations: Monitoring the role of fisheries on increasing the frequency of undesirable traits in exploited species and identifying adaptation to specific environmental conditions is essential to maintain sustainable fisheries. Novel genomic tools have the capacity to address both issues. Climate change is affecting fish and their habitats. Warmer temperatures are already influencing productivity, abundance, migratory patterns and mortality rates of stocks. Genomic tools can be applied to monitor some of these changes.

6.1. *Fishing-induced selection*: Fishing can play a role in evolutionary change and often results in regime shifts towards early maturation, reduced growth and higher mortality rates. Losses of genetic diversity due to fishing can decrease stock resilience. It is, thus, important to monitor such changes. Emerging genomic methods can measure fishery-induced changes in exploited fish stocks.

6.2. *Monitoring evolutionary responses to climate change*: Climate change can also influence marine and aquatic organisms through novel and strong selective forces. Detecting and forecasting their responses to the changing environment is essential to identify species at risk and adjust management strategies accordingly. Several genomic tools can aid in the monitoring of such evolutionary responses.

7) Biotechnology applied to fisheries: While conventional genomic tools will most likely play a pivotal role in the improvement of fisheries assessments, relying solely on the protection of fish stocks might not be sufficient to address the severity of the mounting threats affecting the ocean (such as habitat loss, overexploitation, pollution, invasive species, increase of epidemic diseases and climate change, among others). Fish stocking is a common practice for rebuilding fish stocks, but highly controversial as it can affect negatively essential evolutionary and ecological processes. The latest genomic techniques have the potential to scale-up the benefits of this practice and manage mistakes. It may be one of the few options to protect and restore affected ecosystems and enhance fish stocks resilience.

7.1. *Selection or genetic rebuilding*: Selection and stock rebuilding entails the release of organisms, selected based on their genomic characteristics, into wild stocks that have been depleted by overfishing or other environmental threats, with the aim of accelerating recovery or enabling recovery.

7.2. *Genetic enhancement of fish stocks*: Stock enhancement is the practice of releasing genetically engineered organisms into the natural habitat of the same species, with the aim of improving resilience or other favourable characteristics of the stock to ultimately increase abundance or harvest beyond the level supported prior the intervention.

## Results

Below, we summarise the main findings for each of the themes proposed:

1) The **identification of stock structure** is a central concept for stock management and requires knowledge on three parameters: the level of (i) genetic difference within and (ii) between adjacent stocks and the (iii) degree of connectivity (or lack of) between them. Novel genomic techniques such as RAD-Seq or whole genome re-sequencing (WGS) offer a rapid and cost-effective way to delineate stocks. However, these parameters are not currently estimated, despite documented changes in the distribution of a large number of fish species due to climate change, a trend that is only expected to increase in the near future. Advancing our understanding of functional genomics is key to expanding the information obtained about stocks, including the added capacity of predicting their future trajectories. However, their identification requires accumulating knowledge on local adaptation, adaptive response to global change and evolutionary consequences of selective harvesting in a diverse array of species/geographic locations to decipher the genetic basis of these processes. Advances in this field will largely depend on the creation of biobanks and biorepositories to archive tissues/DNA/RNA and store data over large periods. In the long-term, genomic analyses should be integrated with demographic and hydrodynamic modelling pursuing multidisciplinary assessments of stock structure and connectivity. Complementary analysis such as hydrodynamic simulations, micro-chemical analysis, fatty acid analysis, and Geographic Information Systems, coupled to demographic-genetic computer simulations will likely play a role in future assessments.

2) **Determination of key population parameters** is often neither accurate nor free of limitations with current methods, leading to a continuous active search for new methodologies for the estimation of:

2.1. Genetic monitoring of the *Effective population size* is used to track changes in abundance, which can be inferred from changes in genetic diversity (e.g. expected heterozygosity and allelic diversity), allele frequencies and contemporary effective population size (cNe). This is a rapidly developing area of theoretical and applied research, but the future lies in the dual application of genetic and demographic estimates and in close comparison between CNe and estimates of abundance derived from genetic close-kin mark-recapture (see following section 2.2) that will benefit both methodologies (Waples & Feutry, 2021).

2.2. *Abundance*: The Close Kin Mark Recapture (CKMR) method has the potential to provide accurate estimates of abundance and other key population parameters in marine species (Bravington et al., 2016a,b), however, there may be both biological and financial barriers to its uptake in fisheries assessments in its current form (see section 2.3.1). Although it has the potential to provide valuable baseline or monitoring data for wild fisheries, further developments of this methodology are needed before it can be widely deployed in fisheries assessments. An important step forward in their application would be the integration of estimates of CKMR-derived abundance to generational effective population size, CNe, although this remains a challenging prospect at the moment (Waples & Feutry, 2021).

2.3. *Size structure*: At present, no genomic tools exist to estimate size structure in fishes. However, as high-quality genomic resources become available for exploited species, the deciphering of the genomic mechanism underlying this trait and the identification of genomic regions linked to the variation in size might be possible. Still, it is important to note that growth and size in fishes are largely dependent on environmental factors Boltaña et al. 2017), and thus, genomic tools alone are unlikely to be useful as size predictors in fish.

2.4. *Biological age*: Data to date indicates that DNA methylation-based age-estimation may offer a robust alternative to assess age in fishes, although the method has been tested in very few marine species. Further evidence is needed to validate its use across teleosts. In the long-term, it is essential to test the possibility of generating universal epigenetic clocks that are evolutionary conserved across a broad range of species, as fisheries species represent a large number of diverse taxa. Such clocks would be robustly calibrated even in species that lack a current method to estimate age. Universal clocks would provide a rapid, reader-independent tool to mass-aging fishes in an accurate and non-invasive manner.

2.5. *Sex and maturity*: The lack of high-quality genomic resources has been an important barrier for the development of sex-linked markers to determine sex in species of interest. As more resources become available, the isolation of sex markers would become more and more common, in those species with genetic sex determination systems. Such tools would provide a fast, cheap and easy way to determine sex, irrespective of the developmental stage of the individual and should be easy to implement at large-scale. On the contrary, the development of simple genomic tools with the capacity to determine the maturation status of fishes is unlikely due to its moderate heritability. Attaining sexual maturation in fishes implies a complex process determined by both genetic and environmental cues that is highly variable, with extremes in age and size at maturation being the result of adaptation to maximise fitness and reproductive success. Recent data has shown a genetic predisposition for variation of age at maturity with moderate heritability, thus suggesting a polygenic or complex nature of this trait, similarly to size.

2.6. *Offspring abundance*: Metabarcoding is a rapid, cost-efficient tool that can be used to quantitatively assess ichthyoplankton communities. It requires collection of plankton using small-size mesh nets and it provides accurate quantification of eggs and larval community composition, avoiding the challenge of morphological identification of early developmental stages. The main barrier to its implementation in fisheries assessments lies on the collection of samples, as it requires ad-hoc surveys aligned with the spawning seasons, but this could be alleviated by the development of remote robotic samplers.

2.7. *Total and natural mortality*: The CKMR method can be useful to estimate (adult) selectivity-at-size and natural mortality (Bravington et al., 2017), a famously difficult task in stock assessment Maunder and Piner, 2015. Further developments are needed before this methodology can be deployed in fisheries assessments.

3) **Ecosystem-based approach to fisheries management** accounts for, among other ecosystem factors, the trophic interactions, changes in fish habitats, functional guilds and communities, and human influence (see point 6 below). It is considered a more adaptive and integrative approach to managing fisheries resources and inform decisions. To adopt the principles of ecosystem-based management in fisheries, several emerging applications of high-throughput genomic techniques could prove useful for the analyses of food-webs, trophic linkages, and ecosystem dynamics.

3.1. *Food webs*: Food webs map out networks of predator-prey relationships amongst ecosystem components. DNA-based food web analysis can provide very fine resolution of predator-prey relationships, trophic linkages, and seasonal shifts (Pan et al., 2021). Such genomic analyses can provide much higher taxonomic resolution and capacity for high-throughput than conventional methods, which are mostly based on morphological analysis of stomach contents or on isotopic signatures. They also are less reliant on expert taxonomic knowledge and can identify prey items even when no morphological features are evident. DNA-based food web analysis can be performed using metagenomics to analyse the biodiversity of the feeding habitats (plankton and water samples collected in the feeding area) and the stomach contents, contributing to understand prey selection mechanisms. The large capacity of this technique provides the opportunity to rapidly and exhaustively assess complex diets or environmental assemblages from hundreds of samples, with a high taxonomic and temporal resolution.

Although the technique is promising, it is not free of limitations in its current form. Metagenomics can provide accurate estimates of relative biomass amounts, but the method needs refinement to provide absolute abundances. Nonetheless, relative estimations are still informative in this context, as they enable longitudinal studies of change in food web structure. The analysis relies on the completeness of public databases, but appropriate reference DNA sequences are not yet available for many dietary items or plankton species, particularly for species collected in poorly explored deep-water environments, leading to misidentifications and omissions in the analysis. This problem will be mitigated in the near future as reference genomic databases grow.

3.2. *Environmental monitoring*: eDNA analysis also allows the simultaneous examination of organisms across multiple trophic levels and domains of life. It can provide critical information on the complex biotic interactions related to ecosystem change, including predator-prey relationships (previous section 3.1), trophic linkages, and seasonal shifts. eDNA-based analyses have the potential to provide detailed information on marine ecosystem dynamics and identify sensitive biological indicators that reflects ecosystem changes, regime shifts and inform about potential conservation strategies (Djurhuus, 2020).

It is particularly useful for the characterization of biodiversity – distribution and relative abundance – in Vulnerable Marine Ecosystems (VME), habitats that are

difficult to explore with other methodologies due to their low resilience to mechanical impacts/ disturbance; also, for the early detection of invasive species, allowing the implementation, where possible, of eradication measures before they can become established. This tool has some exceptional advantages as it is a non-invasive method that can be used to monitor species assemblages and their distribution, with little disturbance to both their habitat and to the species itself.

However, important challenges remain for its application in fisheries assessments, as little is known about the rates of decay and dispersion of the eDNA in different habitats and no consensus on best practices for collection and analysis has been reached yet. The complexity of establishing reproducible protocols is one of the main take-home messages from the pilot studies in FishGenome. Moreover, the method is limited by the incompleteness of reference genomic databases and not yet properly calibrated to estimate biomass, an application that would increase the usefulness of the technique exponentially. Nonetheless, mounting data shows promising results in this respect, especially when used together with quantitative PCR.

eDNA is likely to become an essential tool for the monitoring of fisheries and ecosystems, but we see two essential advancements needed to reach full potential and facilitate implementation. First, the development of remote robotic samplers that would avoid disturbing survey procedures, ease operations on board and facilitate standardization. Still, the full potential of the eDNA will only be achieved when "DNA laboratories" can be deployed subsurface. There are already prototypes consisting of Environmental Sample Processors that can collect and filter the water, extract the DNA, perform qPCR analysis, and report the results in real-time (Hansen et al., 2020). Full-development of these devices could represent a game-change for fisheries, as it could allow marine monitoring in real-time to detect temporal and spatial changes in species occurrence and abundance, given that the method can be refined to achieve this last objective. Second, the use of epigenetics together with eDNA. Although not free of challenges, future developments might allow the detection of methylation sites in degraded DNA. This is currently possible when analysing ancient DNA thousands of years old (Llamas et al., 2012), so it should be technically feasible. This would open new avenues of research, including the possibility of inferring the relative age composition in a given population or of assessing remotely the level of environmental stress in the population.

4) **Fisheries surveillance** is essential to contribute to the sustainable harvest of wild fisheries. Even the best exploitation models based on the more accurate and less biased estimates of population parameters might be thwarted in the absence of regulation, enforcement, and inspection. Their successful implementation relies on a way to accurately identify exploited organisms and their products. This can be difficult; for instance, when diagnostic morphological characters are not evident, as in cryptic species or when the specimens have been transformed and processed (fish fillets trunks). The identification of the geographic origin of a product or whether several products represent a single organism is also needed to ensure sustainable practices. To this end, genomic tools are very powerful and, indeed, widely used. They are broadly applied to enforce accurate labelling of seafood, for tracking the fate of individuals in the marketplace, to validate catch records or to determine the population of origin.

4.1. *Species recognition*: Many aspects of fisheries management rely on the accurate identification of both harvested and non-harvested organisms. For example, mapping species distributions, the discovery of cryptic species, recognizing larval stages or the identification of bycatch. Individuals harvested from a fishery or unintentionally caught or affected (e.g., by-catch and threatened, endangered and protected species) need to be identified to species level to maintain accurate catch records and to assist with fisheries enforcement. Species

recognition using DNA is a burgeoning scientific field. It is probably the most rapidly growing area where genetic tools are being taken up for fisheries management. Fisheries species are relatively well represented in international DNA databases because of the ongoing, dedicated program designed to establish this baseline (Ward et al., 2009), but there is still an urgent need of high-resolution genomic data and curation of databases. More and better data will facilitate a growing uptake of barcoding technology in fisheries management and enable more accurate and consistent attributions of catch and bycatch than have been possible in the past. We foresee that the emerging CRISP technology will be an important tool for the identification of species due to its characteristics. It is highly specific, could yield detection results within an hour, does not require DNA extraction, and eliminates as well the need of specialized lab equipment and expertise (Novak et al., 2020). However, the technology is still far from being ready and the economic cost possibly prevents its implementation in the short-term.

4.2. *Determination of geographic origin*: Genomics can trace relationships between tissue samples representing species, populations, family groups and individuals. Samples can be identified as coming from the same, or different, individuals and also assigned to a particular stock. However, the extent to which the method can be used to determine the population of origin of an unknown sample depends on the genetic distinctiveness of the population compared to others. Genetically distinctive populations of wild fisheries are not as frequent or as easily characterized as some freshwater or terrestrial populations, where dispersal barriers are more common and population sizes may be smaller. Novel throughput techniques have an increased resolution power that ameliorates this problem in most cases. Reference data for genetically distinct populations needs to be validated regularly, as temporal changes in gene frequencies and DNA composition can occur continuously due to a variety of processes (drift, selection, immigration with breeding, mutation). This requires funded programs, coordinated across Europe that are not currently in place. In the long term, developments in gene discovery and genetic engineering may lead to the bespoke genetic marking of fisheries product from aquaculture and wild fisheries. As well as having application to product provenance, this activity would be the precursor to property rights over fisheries strains and stocks.

5) **Disease detection in wild fisheries** is increasingly urgent due to the rising temperatures caused by climate change, that are expected to increase the risk of disease emergence in the coming decades and to affect infectious disease patterns. Shifts in species distributions are already occurring globally in response to climate change, potentially bringing pathogens to new areas. Besides, these immigrants do not have immune adaptation to diseases present in the colonized marine environments, boosting up the risk of outbreaks. A disease has the potential to reduce population size, spread through the environment or might even cause adverse health effects in consumers. Current genomic tools have the capacity to diagnose a number of common diseases outbreaks but their implementation, to date, has been restricted to the aquaculture sector. These tools are highly sensitive and specific, while having the potential to quantify the abundance of disease organisms but can only target specific pathogens. The detection of novel pathogens would require the development, evaluation, and testing of novel assays. Fisheries management agencies should consider the implementation and development of a strategy to monitor diseases in wild fisheries and take actions accordingly. Novel genomic methods could shift from detecting the pathogen to detecting changes in expression in genes involved in the fish immune system. Such analyses rely on an understanding of the molecular pathways involved in the immune response. Despite intense research in this field, little is known, especially related to wild marine species. Thus, this strategy would require extensive studies across diverse species exploring diverse pathogen agents. Although the extensive knowledge required possibly prevents the

implementation in the short-term, it is highly promising to detect new diseases before the populations are heavily affected.

6) **Fisheries- and climate-induced changes in exploited populations** can extensively influence the status of fish stocks. Despite that, evolutionary processes have rarely been considered in fisheries management. However, understanding evolutionary processes is likely to become increasingly important, in, at least, two different aspects. Firstly, fishing has the potential to introduce undesirable evolutionary changes to harvested populations. Secondly, climate-driven changes in the marine environment can also have evolutionary impacts in exploited populations. Genomic tools can be applied to wild individuals and ecosystems to monitor some of these changes.

6.1. *Fishing-induced selection*: Levels of genetic diversity dictate the rate at which a species can adapt in response to environmental and fisheries-induced change, and as such provides a measure of evolutionary resilience. Genetic diversity within species and populations is necessary for their long-term survival, as it allows adaptation not only for individuals, but also for populations, species, and entire ecosystems (Frankham, 2005). High-throughput genomic techniques provide the power to generate population-scale genomic data of genetic diversity. However, these estimates are only useful in relative terms or when compared with a baseline. Currently available technologies include methods that allow obtaining and sequencing DNA from ancient samples, nonetheless, due to their high cost, they have been implemented mainly in humans. In the next decades, the cost is expected to plummet as novel methodologies emerge. This would provide the possibility to use them to explore variation, estimate the loss of diversity and infer the impact of climate change and fishing, using comparative analysis of both contemporary and historical specimens. The latter can be obtained from a broad range of sources, including archaeological excavations and museum collections. An alternative method to establish baseline levels of diversity is the use of historical effective size (HNe, see section 2.1), since it has the potential to estimate abundance prior to harvesting. The growing availability of genomic resources can provide direct access to genes selected for by fishing, thus providing a way to monitor changes by screening temporal or spatial collections. The development of mathematical and statistical models for the integration of an index of resilience – based on the levels of genetic diversity - into fisheries science is needed, as this parameter, although currently neglected, is essential to maintain the exploitation of fish stocks at sustainable levels.

6.2. *Monitoring evolutionary responses to climate change*: Like fishing-induced selection, climate change has the potential to introduce novel and strong selective forces on marine and aquatic organisms. Most organisms exhibit some level of adaptation to local conditions and, when confronted with a changing environment, can respond in several different ways: i) decline and/or become extinct; ii) move to more suitable environments; and iii) adapt to the new conditions. If an organism colonizes a new area, it might be compelled to share habitat with a native "sister" species and end up mating with it (hybridization). Alternatively, only certain phenotypes of a species might have the capacity to move to more suitable areas, causing the separation of the species (speciation and radiation). Climate change is likely to increase drastically the likelihood of speciation and hybridization events, therefore, significantly affecting the distribution, dispersion and abundance of marine and aquatic organisms (Avaria-Llautureo et al., 2021, Chunco et al., 2014). The response of the organisms is intimately related to the genetic diversity available to confront environmental change and emphasizes the importance of maintaining large and regionally representative populations. Detecting and forecasting the responses of the organisms to the changing environment is essential to identify species at risk and adjust management strategies accordingly. Understanding and identifying adaptive traits is likely to be of increasing importance for monitoring and predicting the effects of climate change on fisheries

species. High-throughput genomic sequencing tools will be a key resource for a better understanding of evolutionary processes. They will be most effective when used in combination with experiments in captivity and when applied to historical sample collections.

7) **Biotechnology applied to fisheries** might be one of the few options to protect and restore affected ecosystems and enhance fish stocks resilience, given the severity of the mounting threats affecting the ocean (habitat loss, overexploitation, pollution, invasive species, increase of epidemic diseases and climate change, among others). Direct intervention using genomic-based tools for rebuilding fish stocks is a widespread activity that began in the latter part of the nineteenth century and today, encompasses over 200 species. Nonetheless, a careful assessment of genetic and ecological risks is lagging behind implementation in many countries, putting fisheries at risk. Genetic management is vital to minimize risks to the genetic integrity of wild stocks and maximize post-release fitness and enhancement effectiveness. Genomic assessment tools can be used today in an effective and timely manner to improve emerging and established programs. Such tools, include an array of genomic techniques, which range from technology that is ready to be applied to other that still require more extensive technological developments before it can be implemented. Here, we differentiate between genetic rebuilding and enhancement, the former being a genomic–guided technology ready to be applied, while the second implies genomic engineering and requires more extensive developments before it can be implemented.

7.1. *Selection or genetic rebuilding*: Genomic technologies started to be used to inform repopulation management practices after the realization that re-stocking introductions could create severe problems for the native populations (outbreeding depression and decreases of genomic diversity, among others). Genetic rebuilding of fish stocks has been done so far using mainly two different approaches: 1) genomic-guided re-stocking using hatchery-reared specimens; 2) genomic-guided translocation of individuals. The former, has been widely applied in the marine environment to combat sharp declines in abundance of numerous wild stocks of Pacific salmonids, mainly anadromous salmon and trout. The second, consisted in the translocation of specimens, guided by genomic insight, from one environment to another, with the aim of increasing the genetic diversity of native stocks. The selection of suitable source stock is key for any re-stocking program (Houde et al. 2015) but until now, the reduced fitness of hatchery-born fish has not provided specimens as efficient as wild fish in acclimation. Similarly, translocation of individuals has proven unsuccessful, often due to the little knowledge on the genetic variation underlying adaptation in different environments. The development of high-quality reference genomes will be pivotal in increasing our knowledge of the mechanisms underlying local and climate adaptation. It will also help to identify potential functional fitness-related variation, enabling direct assessments of potential for acclimation capacity of candidate source stocks to facilitate adaptation (Cauwelier et al. 2018). We predict that programmes of genomic guided re-stocking and translocations will be expanded to new regions and species. Also, genomic-guided advanced reproductive technologies will gain importance, as they can be used to select wild-founders with desired genomic characteristics, reproduce them in captivity and release specimens with a unique combination of genomic features that increase their fitness. Novel genomic tools can aid not only the re-stocking practices but also the estimation of migration rates, dispersal, and reproduction levels of re-stocked individuals, thus informing about the effectiveness and impacts of the repopulation events.

7.2. *Genetic enhancement of fish stocks*: A step further of genomic-guided restocking is the use of genetically modified specimens. Novel gene-editing technologies give the opportunity to overcome some of the problems of the genetic rebuilding by producing enhanced individuals, with improved fitness compared to wild specimens. Gene editing is a group of technologies that enable scientists to

change an organism's DNA, i.e., add or remove fragments, or modify particular locations in the genome. Developments in gene discovery and genetic engineering (CRISPR) are already being applied to fisheries products from aquaculture in America and Asia. While controversial, the potential benefits of genetic engineering are enormous, as it can significantly improve the efficiency of land-based aquaculture, providing sustainable food, that can be easily tracked and recognized, and therefore, reducing overharvesting of wild stocks. But also, it can be an invaluable tool for rebuilding of wild fish stocks to levels that produce maximum sustainable yield. This is more likely in the context of climate change as severe depletions of fish stocks might occur. While the genetic insight provided by all other tools enumerated in this document have the potential to greatly improve the management of fish stocks, it is very relevant to consider that they may be insufficient to face the severity of the many anthropogenic threats. Genomic intervention might be needed, and genetic engineering has the capacity to produce disease-resistant or fast-growing specimens, for instance, that could speed the rebuild of depleted stocks and ecosystems. Nonetheless, many aspects would need to be considered before genetically engineered fishes are used for re-stocking. First, an ethical framework for biotechnological interventions in wild ecosystems is needed, to guarantee effective and responsible practice. Issues such as whether this activity would be the precursor to property rights over fisheries strains and stocks would need to be resolved by improving legislative aspects. Moreover, carefully assessing the opportunities and risks of repopulating the waters with genetically enhanced individuals is essential, as are the consequences of releasing into the ocean engineered specimens that outcompete native fish, which could cause more harm than non-intervention. Before considering implementation, further research on the ecological effects of genetically engineered fish is needed, as specimens with a higher adaptive capacity could grow faster, cause changes in predator-prey relationships or in ecological niches, among other ecological impacts. Thus, it is essential to promote research in this field and reinforce data-collection.

## 3.3. LESSONS LEARNT

In the short and medium-term, it is not possible to replace current methodologies with HTS methods, given that stock assessment accuracy requires a certain length and stability of the data time series. Therefore, a period of co-implementation will be needed. Thus, it is necessary to identify cost-efficiency pathways and indicators.

Based on the results of the FishGenome project, the cost of the implementation of the 3 genetic HTS methods for two demersal species, hake, and cod, with current technology, has been estimated in 76.933,31€. In relative terms, to incorporate the HTS methods for a single species would mean a 12,5% increase in MEDITS GSA5 survey and a 7,8% in IBTS Q1. Considering both current and HTS methods cost structure, three main factors were found to contribute towards efficiency gains: days at sea, vessel cost (and characteristics) and staff effort.

Potential reductions in the number of days at sea can be gradually expected because of progressive incorporation of the new techniques and thus, this would reduce the costs derived from the use of the research vessels. This represents a promising pathway for efficiency gains. Such gains might derive not just from any potential reduction in the number of samples needed, but from possibilities for their reuse, mainly due to biobanking solutions, or to the use of samples collected onshore, from commercial fleets. The latter may also have a significant impact on cost reductions and enable the applicability of some of the techniques, such as CKMR, even for species where the number of samples required is too high. Similarly, once the epigenetic age-clock is developed for cod, age determination for this species would need less samples, cheaper sequencing methods could be applied, and gains could even come from a multi-species

epigenetic-clock as a future research outcome. Finally, the use of stations for collecting water samples could also reduce the cost linked to eDNA.

Since staff is the main cost component of HTS methods, efficiency gains can be expected in the long-term through the refinement of protocols and routinization. Along the same lines, effort in training and increasing skills can also contribute to enhance staff productivity for the new activities.

Finally, the classification of surveys by efficiency ratios has provided not only identification of those cases where HTS methods could be incorporated more efficiently, but primarily answers to the key questions on the identification of potentially viable candidates, at the survey level. It does not mean that scientific objectives should be subordinate to economic performance, but suggests potential pathways for a balanced approach.

The **proposed roadmap** is very complex due the number of actions required, and the high number and diverse typology of actors required to be involved in each of the proposed actions. However, thanks to the cost-benefit analyses, the SWOT and especially the interaction with experts within the framework of the 2nd workshop, a clear, realistic and prioritised list of five families of criteria were defined, which facilitate significantly the implementation of this proposal. The precedent work was essential to define the needs for the implementation, the most plausible scenario and hence the specific objectives of the roadmap. A careful definition of the timeframe for the implementation along with a precise definition of the criteria to select case studies make the proposed roadmap feasible and with high chances for success.

However, reaching a successful integration of the HTS methods into fisheries stock assessment and management requires engaging all the relevant disciplines and stakeholders under a sound and strong coordination strategy. The genomic techniques have different levels of maturity and readiness, the research surveys differ in capacity to implement the demanding routines and some species (or more properly, stocks) are more suitable to embrace genomic approaches. Thus, achieving the full-scale implementation scenario from the most plausible scenario is only possible through a stepwise approach, organised in phases. Although we have made here a proposal for case studies at phase I, there are some other alternatives that may be considered at the onset of the implementation. Given the readiness of the genomic tools and that survey experts and stock assessment experts are very much inclined to incorporate these techniques and results into survey procedures and assessment models, it is advisable to initiate the implementation as soon as possible after the end of FishGenome project.

Initiating the implementation as soon as possible is crucial also because genomics is a discipline that is evolving rapidly and there is a constant need of adapting new achievements into the planned roadmap. There is a risk of the roadmap getting outdated if a number of actions proposed related to the improvements of the methodologies and with the adoption of new capacities are not initiated in due time. Yet, we acknowledge that the main difficulties for a successful implementation lie in two aspects that require important funding support: i) a quick development of some of the genomic techniques that still requires refinement and standardization, as well a full demonstration of the robustness of the resulting data to be used in assessment; and ii) an increased capacity in infrastructures and skills across the research institutes involved in fisheries and ecological genomics. Obviously, this funding demand must be accompanied by a constant surveillance of the implementation outcomes and of an analysis of the benefits derived from the implementation.

Genomic techniques have the potential to answer questions that have a biological basis and a number of methods, at present, can already provide essential information, unobtainable (or too costly) by other methods, to enhance the management of fish stocks and to inform decisions. **Cutting-edge genomic techniques** can produce estimates of population parameters, without relying on the process of stock assessment

modelling and provide support for Ecosystem Based Fisheries Management and should be incorporated, accordingly, in a progressive manner for the evaluation of fish stocks. However, many targeted fish species are new to genetic studies, which means that they have little (if any) genomic resources (whole genomes, sequence data). Poor genetic resources limit the usefulness of some tools and imply significant start-up cost for research projects. There is still an urgent need of high-resolution genomic data and curation of databases for many species of interest, but also for other taxonomic groups that are key for the dynamics of the stocks, such as the zooplankton organisms that affect fish recruitment or the pathogens that affect the health of wild stocks. An exponential growth of genomic databases is expected within the next decade, but there should be a dedicated program to address the specific needs in fisheries. The creation of biobanks and data repositories for the long-term preservation of biological specimens (fish tissues) and associated genomic data is also essential, not only valuable for real-time monitoring but fundamental for deferred analysis and evaluation. They enable to extend research into the past, to document changes in stocks over time and space and verify in the future past analytical results. This is particularly relevant, as the capacity of genomic tools to address fisheries management issues is diverse and continuously evolving, but many of them still have limitations in their application at their current stage of development. The routine implementation of genomic tools in fisheries assessments will rely, to a large extent, on the development of remote robotic samplers that ease sample collection and processing, facilitate standardization, and allow real-time reporting of the results. Automatic sample processors alleviate the large costs associated to ad-hoc surveys and permit to increase temporal and spatial scales of studies, allowing a better monitoring of fish stocks across their distribution to detect temporal and spatial changes in occurrence and abundance. This is fundamental for detecting and mitigate the effects of climate change and fishing on fisheries resources. In fact, the severity of anthropogenic threats on marine fish stocks might deplete fish stocks and ecosystems to an extent that the use of biotechnological solutions may be required. Genomic-guided restocking using enhanced individuals, with improved fitness compared to wild specimens might become one of the few tools with capacity to rebuild wild fish stocks to levels that produce maximum sustainable yields. Although the use of such techniques to enhance wild stocks might be controversial, the growing challenges and threats faced by fish stocks in a changing environment, together with the increasing demand for food by a growing human population, demands considering every effective tool available.

## 4. REFERENCES


Afzali, S. F., Bourdages, H., Laporte, M., Mérot, C., Normandeau, E., Audet, C., & Bernatchez, L. (2021). Comparing environmental metabarcoding and trawling survey of demersal fish communities in the Gulf of St. Lawrence, Canada. Environmental DNA, 3(1), 22–42. https://doi.org/10.1002/edn3.111

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Anastasiadi, D. (2016). Intrinsic and environmental influences on DNA methylation and gene expression in fish (Ph.D. Thesis, Universitat Pompeu Fabra). Retrieved from http://www.tdx.cat/handle/10803/511360

Anastasiadi, D., & Piferrer, F. (2020). A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. Molecular Ecology Resources, 20(2), 387–397. https://doi.org/10.1111/1755-0998.13111

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Armbruster, D. A., & Pry, T. (2008). Limit of blank, limit of detection and limit of quantitation. The Clinical Biochemist. Reviews, 29 Suppl 1(Suppl 1):S49-52. PMID: 18852857

Avaria-Llautureo, J., Venditti, C., Rivadeneira, M. M., Inostroza-Michael, O., Rivera, R. J., Hernández, C. E., & Canales-Aguirre, C. B. (2021). Historical warming consistently decreased size, dispersal and speciation rate of fish. Nature Climate Change, 11(9), 787–793. https://doi.org/10.1038/s41558-021-01123-5

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., & Lodge, D. M. (2014). Environmental Conditions Influence eDNA Persistence in Aquatic Systems. Environmental Science & Technology, 48(3), 1819–1827. https://doi.org/10.1021/es404734p

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics, 17(1), 1–17. https://doi.org/10.1007/s10592-015-0775-4

Barros-Silva, D., Marques, C., Henrique, R., & Jerónimo, C. (2018). Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. Genes, 9(9), 429. https://doi.org/10.3390/genes9090429

Bérénos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. Molecular Ecology, 23(14), 3434–3451. https://doi.org/10.1111/mec.12827

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H., Therkildsen, N. O., & Withler, R. E. (2017). Harnessing the Power of Genomics to Secure the Future of Seafood. Trends in Ecology & Evolution, 32(9), 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology & Evolution, 29(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Boltaña, S., Sanhueza, N., Aguilar, A., Gallardo-Escarate, C., Arriagada, G., Valdes, J. A., Soto, D., & Quiñones, R. A. (2017). Influences of thermal environment on fish growth. Ecology and Evolution, 7(17), 6814–6825. https://doi.org/10.1002/ece3.3239

Bravington, M. V., Eveson, J. P., Grewe, P. M. & Davies, C. R. (2015). SBT Close-Kin Mark-Recapture: options for the medium term. Contract Rep. CCSBT-ESC/1509/19

Bravington, M. V., Grewe, P. M., & Davies, C. R. (2016a). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. Nature Communications, 7(1), 13162. https://doi.org/10.1038/ncomms13162

Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016b). Close-Kin Mark-Recapture. Statistical Science, 31(2). https://doi.org/10.1214/16-STS552

Bravington, M.V., Eveson, J.P., Grewe, P.M., Davies, C.R. (2017). SBT Close-Kin Mark-Recapture with Parent-Offspring and Half-Sibling Pairs: update on genotyping, kin-finding and model development. CSIRO CCSBT-ESC/1709/12

Bravington, M., Feutry, P., Pillans, R.D., Hillary, R., Johnson, G., Saunders, T., Gunesekera, R., Bax, N.J. and Kyne, P.M. (2019). Close-Kin Mark- Recapture population size estimate of Glyphis garricki in the Northern Territory. Report to the National Environmental Science Program, Marine Biodiversity Hub. CSIRO Oceans and Atmosphere, Hobart.

Bruce, P., & Bruce, A. (2017). Practical Statistics for Data Scientists: 50 Essential Concepts. O'Reilly Media, Inc. ISBN: 9781491952962

Campbell, E. O., Brunet B .M. T., Dupuis J. R., Sperling F. A. H. (2018). Would an RRS by any other name sound as RAD?. Methods Ecol Evol, 9(9): 1920-1927. https://doi.org/10.1111/2041-210X.13038

Casey, J., Jardim, E., & Martinsohn, J. T. (2016). The role of genetics in fisheries management under the E.U. common fisheries policy. Journal of Fish Biology, 89(6), 2755–2767. https://doi.org/10.1111/jfb.13151

Cauwelier, E., Gilbey, J., Sampayo, J., Stradmeyer, L., & Middlemas, S. J. (2018). Identification of a single genomic region associated with seasonal river return timing in adult Scottish Atlantic salmon (*Salmo salar* ), using a genome-wide association study. Canadian Journal of Fisheries and Aquatic Sciences, 75(9), 1427–1435. https://doi.org/10.1139/cjfas-2017-0293

Choi, Y., Wijsman, E. M., & Weir, B. S. (2009). Case-control association testing in the presence of unknown relationships. Genetic Epidemiology, 33(8), 668–678. https://doi.org/10.1002/gepi.20418

Chunco, A. J. (2014). Hybridization in a warmer world. Ecology and Evolution, 4(10), 2019–2031. https://doi.org/10.1002/ece3.1052

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., & Genner, M. J. (2018). Persistence of environmental DNA in marine systems. Communications Biology, 1(1), 185. https://doi.org/10.1038/s42003-018-0192-6

Coolen, M. J. L., & Overmann, J. (2007). 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. Environmental Microbiology, 9(1), 238–249. https://doi.org/10.1111/j.1462-2920.2006.01134.x

Creer, S., & Seymour, M. (2017). Marine ecology: Genetics from a drop in the ocean. Nature Ecology & Evolution, 1(1), 0037. https://doi.org/10.1038/s41559-016-0037

De Schepper, S., Ray, J. L., Skaar, K. S., Sadatzki, H., Ijaz, U. Z., Stein, R., & Larsen, A. (2019). The potential of sedimentary ancient DNA for reconstructing past sea ice evolution. The ISME Journal, 13(10), 2566–2577. https://doi.org/10.1038/s41396-019-0457-1

Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. Frontiers in Zoology, 3(1), 11. https://doi.org/10.1186/1742-9994-3-11

Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation, 183, 53–63. https://doi.org/10.1016/j.biocon.2014.11.018

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology, 26(21), 5872–5895. https://doi.org/10.1111/mec.14350

Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., Walz, K. R., Andruszkiewicz, E. A., Olesin, E., Hubbard, K., Montes, E., Otis, D., Muller-Karger, F. E., Chavez, F. P., Boehm, A. B., & Breitbart, M. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. Nature Communications, 11(1), 254. https://doi.org/10.1038/s41467-019-14105-1

Dortel, E., Massiot-Granier, F., Rivot, E., Million, J., Hallier, J.-P., Morize, E., Munaron, J.-M., Bousquet, N., & Chassot, E. (2013). Accounting for Age Uncertainty in Growth Modeling, the Case Study of Yellowfin Tuna (*Thunnus albacares*) of the Indian Ocean. PLoS ONE, 8(4), e60886. https://doi.org/10.1371/journal.pone.0060886

Eichmiller, J. J., Miller, L. M., & Sorensen, P. W. (2016). Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. Molecular Ecology Resources, 16(1), 56–68. https://doi.org/10.1111/1755-0998.12421

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2012). SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing (pp. 157–178). https://doi.org/10.1007/978-1-61779-228-1_9

Evans, N. T., & Lamberti, G. A. (2018). Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. Fisheries Research, 197, 60–66. https://doi.org/10.1016/j.fishres.2017.09.013

Feron, R., Pan, Q., Wen, M., Imarazene, B., Jouanno, E., Anderson, J., Herpin, A., Journot, L., Parrinello, H., Klopp, C., Kottler, V. A., Roco, A. S., Du, K., Kneitz, S., Adolfi, M., Wilson, C. A., McCluskey, B., Amores, A., Desvignes, T., … Guiguen, Y. (2021). RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. Molecular Ecology Resources, 21(5), 1715–1731. https://doi.org/10.1111/1755-0998.13360

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. Biology Letters, 4(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources, 16(3), 604–607. https://doi.org/10.1111/1755-0998.12508

Fowler, B. L. S., & Buonaccorsi, V. P. (2016). Genomic characterization of sex-identification markers in Sebastes carnatus and Sebastes chrysomelas rockfishes. Molecular Ecology, 25(10), 2165–2175. https://doi.org/10.1111/mec.13594

Fraija-Fernández, N., Bouquieaux, M., Rey, A., Mendibil, I., Cotano, U., Irigoien, X., Santos, M., & Rodríguez-Ezpeleta, N. (2020). Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. Ecology and Evolution, 10(14), 7560–7584. https://doi.org/10.1002/ece3.6482

Frankham, R. (2005). Genetics and extinction. Biological Conservation, 126(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Fraser, D. J., Jones, M. W., McParland, T. L., & Hutchings, J. A. (2007). Loss of historical immigration and the unsuccessful rehabilitation of extirpated salmon populations. Conservation Genetics, 8(3), 527–546. https://doi.org/10.1007/s10592-006-9188-8

Fujii, Y., Rémy, E., Zuo, H., Oke, P., Halliwell, G., Gasparin, F., Benkiran, M., Loose, N., Cummings, J., Xie, J., Xue, Y., Masuda, S., Smith, G. C., Balmaseda, M., Germineaud, C., Lea, D. J., Larnicol, G., Bertino, L., Bonaduce, A., … Usui, N. (2019). Observing System Evaluation Based on Ocean Data Assimilation and Prediction Systems: On-Going Challenges and a Future Vision for Designing and Supporting Ocean Observational Networks. Frontiers in Marine Science, 6. https://doi.org/10.3389/fmars.2019.00417

Gay, L., Siol, M., & Ronfort, J. (2013). Pedigree-Free Estimates of Heritability in the Wild: Promising Prospects for Selfing Populations. PLoS ONE, 8(6), e66983. https://doi.org/10.1371/journal.pone.0066983

Goldberg, C. S., Strickler, K. M., & Pilliod, D. S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biological Conservation, 183, 1–3. https://doi.org/10.1016/j.biocon.2014.11.040

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution, 7(11), 1299–1307. https://doi.org/10.1111/2041-210X.12595

Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., Izeta, A., de Haan, G., Geiger, H., & Wagner, W. (2018). Epigenetic age-predictor for mice based on three CpG sites. ELife, 7. https://doi.org/10.7554/eLife.37462

Hansen, B. K., Jacobsen, M. W., Middelboe, A. L., Preston, C. M., Marin, R., Bekkevold, D., Knudsen, S. W., Møller, P. R., & Nielsen, E. E. (2020). Remote, autonomous real-time monitoring of environmental DNA from commercial fish. Scientific Reports, 10(1), 13272. https://doi.org/10.1038/s41598-020-70206-8

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., Ruegg, K., & Palstra, F. (2011). Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. Conservation Biology, 25(3), 438–449. https://doi.org/10.1111/j.1523-1739.2010.01637.x

Harrison, A., & Parle-McDermott, A. (2011). DNA Methylation: A Timeline of Methods and Applications. Frontiers in Genetics, 2. https://doi.org/10.3389/fgene.2011.00074

Harrison, J. B., Sunday, J. M., & Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings of the Royal Society B: Biological Sciences, 286(1915), 20191409. https://doi.org/10.1098/rspb.2019.1409

Heath, M. R., Culling, M. A., Crozier, W. W., Fox, C. J., Gurney, W. S. C., Hutchinson, W. F., Nielsen, E. E., O'Sullivan, M., Preedy, K. F., Righton, D. A., Speirs, D. C., Taylor, M. I., Wright, P. J., & Carvalho, G. R. (2014). Combination of genetics and spatial modelling highlights the sensitivity of cod (*Gadus morhua*) population diversity in the North Sea to distributions of fishing. ICES Journal of Marine Science, 71(4), 794–807. https://doi.org/10.1093/icesjms/fst185

Heimbrand, Y., Limburg, K. E., Hüssy, K., Casini, M., Sjöberg, R., Palmén Bratt, A., Levinsky, S., Karpushevskaia, A., Radtke, K., & Öhlund, J. (2020). Seeking the true time: Exploring otolith chemistry as an age-determination tool. Journal of Fish Biology, 97(2), 552–565. https://doi.org/10.1111/jfb.14422

Herder, J., Valentini, A., Bellemain, E., Dejean, T., Van Delft, J., Thompsen, F., Taberlet, P. (2014) Environmental DNA. A review of the possible applications for the detection of (invasive) species. Netherlands Food and Consumer Product Safety Authority. Stichting RAVON, Nijmegen. Report 2013-104. DOI: 10.13140/RG.2.1.4002.1208

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., Diez, J., Sanchez-Mut, J. V., Setien, F., Carmona, F. J., Puca, A. A., Sayols, S., Pujana, M. A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A. F., Fraga, M. F., Heath, S. C., … Esteller, M. (2012). Distinct DNA methylomes of newborns and centenarians. Proceedings of the National Academy of Sciences, 109(26), 10522–10527. https://doi.org/10.1073/pnas.1120658109

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biology, 14(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Houde, A. L. S., Garner, S. R., & Neff, B. D. (2015). Restoring species through reintroductions: strategies for source population selection. Restoration Ecology, 23(6), 746–753. https://doi.org/10.1111/rec.12280

Hundt, P., Liddle, E., Nielsen, S., Pinto, B., & Gamble, T. (2019). Sex chromosomes and sex-specific molecular markers in Indo-Pacific combtooth blennies (Blenniidae, Istiblennius). Marine Ecology Progress Series, 627, 195–200. https://doi.org/10.3354/meps13082

ICES (2021). Workshop on Tools and Development of Stock Assessment Models using a4a and Stock Synthesis (WKTADSA). ICES Scientific Reports. 3:33. 197 pp. https://doi.org/10.17895/ices.pub.8004

Ito, H., Udono, T., Hirata, S., & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. Scientific Reports, 8(1), 9998. https://doi.org/10.1038/s41598-018-28318-9

Jung, M., & Pfeifer, G. P. (2015). Aging and DNA methylation. BMC Biology, 13(1), 7. https://doi.org/10.1186/s12915-015-0118-4

Kassambara, A. (2018). Machine Learning Essentials: Practical Guide in R. STHDA Eds. 209 pp. ISBN : 1986406857

Knudsen, E. E., Steward, C. R., MacDonald, D. D., Williams, J. E., & Reiser, D. W. (Eds.). (2020). Sustainable Fisheries Management. CRC Press. https://doi.org/10.1201/9780429104411

Kopps, A. M., Kang, J., Sherwin, W. B. & Palsbøll, P. J. (2015). How Well Do Molecular and Pedigree Relatedness Correspond, in Populations with Diverse Mating Systems, and Various Types and Quantities of Molecular and Demographic Data? Genes|Genomes|Genetics 5, 1815–1826. https://doi.org/10.1534/g3.115.019323

Kumar, G., Farrell, E., Reaume, A. M., Eble, J. A., & Gaither, M. R. (2022). One size does not fit all: Tuning eDNA protocols for high- and low-turbidity water sampling. Environmental DNA, 4(1), 167–180. https://doi.org/10.1002/edn3.235

Lance, R., Klymus, K., Richter, C., Guan, X., Farrington, H., Carr, M., Thompson, N., Chapman, D., & Baerwaldt, K. (2017). Experimental observations on the decay of environmental DNA from bighead and silver carps. Management of Biological Invasions, 8(3), 343–359. https://doi.org/10.3391/mbi.2017.8.3.08

Llamas, B., Holland, M. L., Chen, K., Cropley, J. E., Cooper, A., & Suter, C. M. (2012). High-Resolution Analysis of Cytosine Methylation in Ancient DNA. PLoS ONE, 7(1), e30226. https://doi.org/10.1371/journal.pone.0030226

Lodge, D. M., Turner, C. R., Jerde, C. L., Barnes, M. A., Chadderton, L., Egan, S. P., Feder, J. L., Mahon, A. R., & Pfrender, M. E. (2012). Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. Molecular Ecology, 21(11), 2555–2558. https://doi.org/10.1111/j.1365-294X.2012.05600.x

Maunder, M. N., & Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science, 72(1), 7–18. https://doi.org/10.1093/icesjms/fsu015

Maunder M. N., Lennert-Cody C. E., Aires-da-Silva A. M., Xu H. (2021). Considerations for conducting close kin mark recapture of stocks managed by the IATTC. Document SAC-12-14

Mayne, B., Espinoza, T., Roberts, D., Butler, G. L., Brooks, S., Korbie, D., & Jarman, S. (2021). Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod. Molecular Ecology Resources, 21(7), 2324–2332. https://doi.org/10.1111/1755-0998.13440

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, 17(2), 240–248. https://doi.org/10.1101/gr.5681207

Milligan, B. G. (2003). Maximum-Likelihood Estimation of Relatedness. Genetics, 163(3), 1153–1167. https://doi.org/10.1093/genetics/163.3.1153

Mills, L. S., Citta, J. J., Lair, K. P., Schwartz, M. K., & Tallmon, D. A. (2000). Estimating Animal Abundance Using Noninvasive DNA Sampling: Promise and Pitfalls. Ecological Applications, 10(1), 283. https://doi.org/10.2307/2641002

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science, 2(7), 150088. https://doi.org/10.1098/rsos.150088

Miya, M., Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. Fisheries Science, 86(6), 939–970. https://doi.org/10.1007/s12562-020-01461-x

Nester, G. M., De Brauwer, M., Koziol, A., West, K. M., DiBattista, J. D., White, N. E., Power, M., Heydenrych, M. J., Harvey, E., & Bunce, M. (2020). Development and evaluation of fish eDNA metabarcoding assays facilitate the detection of cryptic seahorse taxa (family: Syngnathidae). Environmental DNA, 2(4), 614–626. https://doi.org/10.1002/edn3.93

Nielsen, R., Mattila, D. K., Clapham, P. J., & Palsbøll, P. J. (2001). Statistical Approaches to Paternity Analysis in Natural Populations and Applications to the North Atlantic Humpback Whale. Genetics, 157(4), 1673–1682. https://doi.org/10.1093/genetics/157.4.1673

Novak, B. J., Fraser, D., & Maloney, T. H. (2020). Transforming Ocean Conservation: Applying the Genetic Rescue Toolkit. Genes, 11(2), 209. https://doi.org/10.3390/genes11020209

Osio, Giacomo C, Maurizio Gibin, Alessandro Mannini, A Villamor, and A Orio. (2018). The Mediterranean and Black Sea STECF Stock Assessment Database. Luxembourg. https://doi.org/10.2760/559579

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, evolutionary and molecular genetics in the management of wild fisheries. Fish and Fisheries, 16(1), 125–159. https://doi.org/10.1111/faf.12052

Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. Molecular Ecology, 28(21), 4709–4724. https://doi.org/10.1111/mec.15245

Pan, W., Qin, C., Zuo, T., Yu, G., Zhu, W., Ma, H., & Xi, S. (2021). Is Metagenomic Analysis an Effective Way to Analyze Fish Feeding Habits? A Case of the Yellowfin Sea Bream *Acanthopagrus latus* (Houttuyn) in Daya Bay. Frontiers in Marine Science, 8. https://doi.org/10.3389/fmars.2021.634651

Paoli-Iseppi, R., Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A., & Jarman, S. N. (2019). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. Molecular Ecology Resources, 19(2), 411–425. https://doi.org/10.1111/1755-0998.12981

Papa, Y., Oosting, T., Valenza-Troubat, N., Wellenreuther, M., & Ritchie, P. A. (2021). Genetic stock structure of New Zealand fish and the use of genomics in fisheries management: an overview and outlook. New Zealand Journal of Zoology, 48(1), 1–31. https://doi.org/10.1080/03014223.2020.1788612

Pardo, S. A., Cooper, A. B., & Dulvy, N. K. (2013). Avoiding fishy growth curves. Methods in Ecology and Evolution, 4(4), 353–360. https://doi.org/10.1111/2041-210x.12020

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. Molecular Ecology Resources, 14(5), 976–987. https://doi.org/10.1111/1755-0998.12247

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-

Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Rauschert, S., Raubenheimer, K., Melton, P. E., & Huang, R. C. (2020). Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. Clinical Epigenetics, 12(1), 51. https://doi.org/10.1186/s13148-020-00842-4

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. Journal of Applied Ecology, 51(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries, 10(4), 361–395. https://doi.org/10.1111/j.1467-2979.2008.00324.x

Rivera-Colón A. G., Catchen J. (2022). Population Genomics Analysis with RAD, Reprised: Stacks 2. Methods in Molecular Biology, 2498, 99-149. doi: 10.1007/978-1-0716-2313-8_7

Rodriguez-Ezpeleta, N., Patterson, T.A., Pereda, I., Grande, M., Davies, C.R., Lezama-Ochoa, N., Zudaire, I. (2020). Feasibility study on applying close-kin mark-recapture abundance estimates to Indian Ocean Tuna Commission shark species. Final report to IOTC. IOTC-2020-WPEB16-13

Rodríguez-Rodríguez, G., Ballesteros, H. M., Sánchez-Llamas, E., Bande, R., & Otero, R. F. (2022). The state of the art in cost-benefit of HTS methods for stock assessment: An overview. Frontiers in Marine Science, 9. https://doi.org/10.3389/fmars.2022.1005534

Ruzzante, D. E., McCracken, G. R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., Mills Flemming, J., & Skaug, H. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. Methods in Ecology and Evolution, 10(9), 1445–1453. https://doi.org/10.1111/2041-210X.13243

Sakata, M. K., Yamamoto, S., Gotoh, R. O., Miya, M., Yamanaka, H., & Minamoto, T. (2020). Sedimentary eDNA provides different information on timescale and fish species composition compared with aqueous eDNA. Environmental DNA, 2(4), 505–518. https://doi.org/10.1002/edn3.75

Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., & Boehm, A. B. (2016). Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. Environmental Science & Technology, 50(19), 10456–10464. https://doi.org/10.1021/acs.est.6b03114

Sato, H., Sogo, Y., Doi, H., & Yamanaka, H. (2017). Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports, 7(1), 14860. https://doi.org/10.1038/s41598-017-14978-6

Sato, Y., Miya, M., Fukunaga, T., Sado, T., & Iwasaki, W. (2018). MitoFish and MiFish Pipeline: A Mitochondrial Genome Database of Fish with an Analysis Pipeline for Environmental DNA Metabarcoding. Molecular Biology and Evolution, 35(6), 1553–1555. https://doi.org/10.1093/molbev/msy074

Skaug, H. J. (2001). Allele-Sharing Methods for Estimation of Population Size. Biometrics, 57(3), 750–756. https://doi.org/10.1111/j.0006-341X.2001.00750.x

Stamatopoulos, C. (2002). Sample-based fishery surveys : A techical handbook. FAO Fisheries Technical Paper. No 425. Rome, FAO, 132 pp. ISBN : 92-5-104699-9

Star, B., Tørresen, O. K., Nederbragt, A. J., Jakobsen, K. S., Pampoulie, C., & Jentoft, S. (2016). Genomic characterization of the Atlantic cod sex-locus. Scientific Reports, 6(1), 31235. https://doi.org/10.1038/srep31235

STECF (2019). Scientific, Technical and Economic Committee for Fisheries (STECF) – Preparation for the evaluation of the list of mandatory research surveys at sea (STECF-19-05). Publications Office of the European Union, Luxembourg, 2019, ISBN : 978-92-76-09516-3, doi:10.2760/2860, JRC117485

Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z., Dunton, K. J., Hinks, G., & VanMorter, S. M. (2021). Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA. ICES Journal of Marine Science, 78(1), 293–304. https://doi.org/10.1093/icesjms/fsaa225

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. Molecular Ecology, 21(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Thompson, M. J., VonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. Aging, 9(3), 1055–1068. https://doi.org/10.18632/aging.101211

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology, 21(11), 2565–2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation, 183, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. PLOS ONE, 11(11), e0165252. https://doi.org/10.1371/journal.pone.0165252

Turner, C. R., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods in Ecology and Evolution, 5(7), 676–684. https://doi.org/10.1111/2041-210X.12206

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. Biological Conservation, 183, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

Valsecchi, E., Bylemans, J., Goodman, S. J., Lombardi, R., Carr, I., Castellano, L., Galimberti, A., & Galli, P. (2020). Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. Environmental DNA, 2(4), 460–476. https://doi.org/10.1002/edn3.72

Waples, R. S., & Feutry, P. (2022). Close-kin methods to estimate census size and effective population size. Fish and Fisheries, 23(2), 273–293. https://doi.org/10.1111/faf.12615

Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology, 74(2), 329–356. https://doi.org/10.1111/j.1095-8649.2008.02080.x

Wetterstrand, K.A. 2021. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at : www.genome.gov/sequencingcostsdata

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., & Cooper, A. (2003). Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. Science, 300(5620), 791–795. https://doi.org/10.1126/science.1084114

Wright, P. G. R., Hamilton, P. B., Schofield, H., Glover, A., Damant, C., Davidson-Watts, I., & Mathews, F. (2018). Genetic structure and diversity of a rare woodland bat, Myotis bechsteinii: comparison of continental Europe and Britain. Conservation Genetics, 19(4), 777–787. https://doi.org/10.1007/s10592-018-1053-z

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., & Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. Scientific Reports, 7(1), 40368. https://doi.org/10.1038/srep40368

Zhang, Q., Vallerga, C. L., Walker, R. M., Lin, T., Henders, A. K., Montgomery, G. W., He, J., Fan, D., Fowdar, J., Kennedy, M., Pitcher, T., Pearson, J., Halliday, G., Kwok, J. B., Hickie, I., Lewis, S., Anderson, T., Silburn, P. A., Mellick, G. D., … Visscher, P. M. (2019). Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Medicine, 11(1), 54. https://doi.org/10.1186/s13073-019-0667-1

Zhou, S., Fan, C., Xia, H., Zhang, J., Yang, W., Ji, D., Wang, L., Chen, L., & Liu, N. (2022). Combined Use of eDNA Metabarcoding and Bottom Trawling for the Assessment of Fish Biodiversity in the Zhoushan Sea. Frontiers in Marine Science, 8. https://doi.org/10.3389/fmars.2021.809703

## 5. GLOSSARY

**Abundance (absolute, relative and abundance index):** It is a measure of the number of individuals in a population. In fisheries science often referred to a given stock. Absolute abundance is the total number of individuals in the stock and it is an essential parameter in fisheries stock assessments. However, the estimation of absolute abundance is difficult and requires extensive data collection, such as a precise estimate of density at sampling sites and a probability-based array of those sites. Sampling methods often provide abundance indexes or relative abundance. However, absolute abundance can be estimated from population dynamic models (often used in stock assessment), or from other methods as those based in genomics.

**Accuracy:** Closeness with the true value of the magnitude being measured. Depends upon sample techniques of analysis. A measurement can be said to be accurate if their average is close to the true value of the quantity being measured. See also Precision, Bias, and Reproducibility.

**Age class:** A group of individuals of a certain fish population that have the same age. The age group is expressed as an integral number of years. For example, the fish in a population may be separated into classes such as age-0, age-1, age-2 and age-3+ (The plus group contains all fish of a certain age and older). A fish can be classified into the appropriate age class once its age has been established (e.g., by reading the number of rings of the otolith). The percentage of different age classes among the components of a population greatly affects the reproductive possibilities, and thus its evolutionary development.

**Allele:** Each of the forms that the same gene can have or can express within a given population. In most animals, an individual inherits two alleles, one from each parent, for any given genomic location where such variation exists. If the two alleles are the same, the individual is homozygous for that allele. If the alleles are different, the individual is heterozygous.

**Benthos:** Organisms living in or on the sediment on the seafloor. The organisms that live on hard substrates above the sediment are called epibenthos and the organisms that live in the sediment are called infauna.

**Bias:** Statistical bias is anything that leads to a systematic difference between the true parameters of a population and the statistics used to estimate those parameters. See also Accuracy, Precision, and Reproducibility.

**Biobank:** A facility collects, catalogs, stores, and distributes samples of biological material and the data associated with such material. Also called biorepository.

**Biodiversity:** The variability among living organisms and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems.

**Bioinformatics:** Is a computational approach that develops methods and software tools to analyze and interpret biological data, in particular when the data sets are large and complex. .

**Biological monitoring:** Method for observing and assessing the status and changes of species in ecosystems. Usually, biological indicators (e.g., individual species or communities) are used to document and understand changes in in the ecosystems, especially changes caused by the activities of an expanding human population. Indicator species effectively indicate the condition of the environment because of their moderate tolerance to environmental variability. In this project, biomonitoring using eDNA is proposed, which requires less effort, as the water column is sampled instead of each species individually. In this project, biomonitoring using eDNA is proposed, which requires less effort, as the water column is sampled instead of each species individually.

**Biological Reference Point:** Quantitative indicator against which fish biomass, fishing mortality rate, or other stock properties can be compared to determine stock status and provide management advice. Some examples of BRP are a) the Biomass limit reference point (Blim), which is the biomass limit below which a stock is considered to have reduced reproductive capacity; b) the Fishing mortality limit reference point

(Flim) which is fishing mortality rate (F) which leads the spawning stock biomass to Blim and c) the Rate of Fishing mortality consistent with achieving Maximum Sustainable Yield (FMSY).

**Biomarkers:** A portmanteau of "biological markers", refers to biological molecules used as indicators of a biological state, which can be measured and evaluated objectively. In genetics, a biomarker is a measurable DNA and/or RNA characteristic that indicates a biological process by reflecting the expression, function or regulation of a genomic region.

**Biomass:** The total quantity or weight of living organisms in an area or volume at a given time. It can refer to community biomass (all species in the community) or species biomass (mass of one or more species). In fisheries, it usually refers to the total quantity or weight of a stock or a component of a stock (e.g., spawning stock biomass, which is the total weight of the reproductively mature individuals in a stock). .

**Capital and income breeders:** In reproductive strategies, and to ensure an optimal egg production, some species need to store energy well before the breeding season, because this takes place during an unfavourable time of the year in terms of food availability and temperature, among other factors. The energy stored can be mobilized later for reproduction, so storage constitutes the primary energy source for the cost of egg production. These are capital breeders. Some other species, on the contrary, breed under more favourable conditions, take benefit of a relatively more productive environment and use recently acquired energy as income for egg production, and there is no need for storing energy. These are income breeders.

**Carbon footprint:** Environmental indicator that reflects the amount of carbon dioxide and other greenhouse gases generated by human activity. It is expressed as a weight of $CO_2$ emissions produced in tonnes.

**Close-Kin Mark-Recapture:** A direct method for estimating the size of a fish population. Is based on the principle that an individual's genotype can be considered a "recapture" of the genotypes of each of its parents and analyses the number and pattern of close-kin pairs in a mark–recapture (MR) framework. Assuming the sampling of offspring and parents to be independent of each other, the number of kinship pairs genetically identified in samples can be used to estimate absolute abundance and other demographic parameters, including mortality and selectivity.

**Cohort:** Group of individuals of the same species that are born in a defined period of time and in the same place, who grow and survive at similar rates. It is considered the basic reference group for demographic studies and analyses of survival, growth, reproduction and mortality. In fisheries, this term refers to all the individuals of a fish population born in the same spawning season.

**Cold-blooded vertebrates:** See Thermoregulation.

**Coverage:** Number of times a portion of the genome is sequenced in a sequencing reaction. Also expressed as depth coverage. The higher the number of times a position is "read", the higher the confidence to the base call at that particular position. In other words, sequencing error reads are statistically irrelevant when they are outnumbered by correct reads.

**CRISPR:** Pronounced "crisper", stands for Clustered Regularly Interspaced Short Palindromic Repeats, which are the hallmark of a bacterial defence system that forms the basis for CRISPR-Cas9 genome editing technology. It can be used to target specific stretches of genetic code and to edit DNA at precise locations, such as genes, permanently modifying them.

**Demersal:** A term describing organisms that live on or just above the ocean floor. In general, they show little movement, and can be strictly benthic, living on the seabed, or benthopelagic, floating in the water column just above the seabed.

**Digital-droplet Polymerase Chain Reaction:** It is a modified form of real-time PCR, also known as the third generation of quantitative PCR. It works by segmenting samples using water-in-oil emulsions to create droplets from which their genetic material can be identified and quantified. In contrast to qPCR, this method does not require a calibration curve to quantify the DNA. The generated data are collected

from a binary signal and can be used directly to count the number of target eDNA molecules in the sample. ddPCR performs better than qPCR in terms of precision and reproducibility, is less susceptible to inhibitors and more suitable for the quantification of DNA.

**DNA amplification:** The production of multiple copies of a sequence of DNA.

**DNA isolation:** It consists of separating the DNA present in the nucleus of the cell from the other cellular components. Briefly, the procedure consists of breaking the cell membranes for extraction, degradation of the histone-like proteins bound to the DNA, and subsequent precipitation and purification of the DNA.

**DNA methylation:** DNA methylation is a epigenetic mechanism that occurs by the addition of a methyl (CH3) group to DNA. The most widely characterized DNA methylation process is the addition of the methyl group at the 5-carbon of the cytosine ring, also informally known as the "fifth base" of DNA. These methyl groups inhibit DNA transcription and are inherited.

**Ecosystem:** System formed by a set of living organisms, the physical environment in which they live (habitat) and the relationships between them. These elements interact as a functional unit.

**Environmental DNA:** Environmental DNA or eDNA for short, is DNA that is released from an organism into the environment in the form of secreted faeces, mucous, and gametes; shed skin and hair; and carcasses. The term can also be used to refer to the array of molecular methods used to detect this type of genetic material.

**Epigenetic:** Epigenetics can be defined as the study of changes in gene function that heritable and that do not entail a change in DNA sequence. *Epi-* means on or above in Greek, and *epigenetic* describes modifications that attach to DNA altering its physical structure to regulate the activity (expression) of the genes and non-coding regions. Epigenetics includes modifications to histone proteins, noncoding RNAs, and DNA methylation.

**Epigenetic age:** Is the measure of the biological age of an individual using epigenetic changes that are correlated with its chronological age.

**Epigenetic clock:** A molecular tool used to determine the biological age of an individual based on a set of epigenetic changes highly correlated with the chronological age of the individual.

**Fecundity:** Number of gametes that an individual is capable of laying during a given period of time (life, annual, seasonal). Female fecundity (ovules or eggs) It is one of the most useful indicators for assessing the reproductive potential of a species.

**Fisheries management:** The integrated process of information gathering, analysis, planning, consultation, decision-making, allocation of resources and formulation and implementation, with enforcement as necessary, of regulations or rules which govern fisheries activities in order to ensure the continued productivity of the resources and the accomplishment of other fisheries objectives.

**Fluorescence:** Fluorescence is used in biology as a non-destructive way of analysing biological molecules, even at low concentrations, by means of the molecule's intrinsic fluorescence, or by attaching it with a fluorophore.

**FST:** Measure of genetic population sub-division or structure, based on the variance in allele frequencies. Generally, FST < 0.03 indicates little genetic sub-division whereas FST > 0.15 indicate large genetic sub-division.

**Gene expression:** Fundamental life process by which the information encoded within a gene is used to make a functional gene product, either RNA molecules that code for proteins or non-coding RNA molecules that serve other functions of the organism.

**Genome:** The complete set of genetic information in an organism. It provides all of the information the organism requires to function.

**Genomic:** Pertaining to genomics, which is concerned with the study of the structure, function, evolution and mapping of genomes.

**Genotyping:** Is the process of determining the DNA sequence, called a genotype, at specific positions within the genome of an individual. Sequence variations can be used as markers.

**Gonochorism:** It is a sexual system where there are only two sexes and each individual organism is either male or female. See also Hermaphroditism.

**Grey literature:** It is a type of scientific information that is not formally published. It is often generated by researchers and practitioners, but is not controlled by commercial or academic publications. Examples of grey literature can be memoirs, manuscripts, research reports, workbooks or technical reports. The fact that it is not formally published can create problems of accessibility.

**Growth:** Process by which a fish experiments an increase in length and weight, while changing its morphology and may depend on several factors, such as food availability, water temperature and light. The growth of fish is **indeterminate**, meaning that they continuously increase in size throughout their lives. Many fish species also exhibit sex-**dimorphic** growth, with one sex growing larger than the other.

**Half-Sibling-Pairs:** A pair of organisms that share a single parent. They may have the same father but different mothers or they may have the same mother but different fathers.

**Haul:** Fishing operation in which the trawl is deployed and retrieved once.

**Heritability:** Is a measure to estimate the proportion of phenotypic variation that can be attributed to inherited genetic factors. It is a population-specific value, not a parameter attributable to all organisms or species. Heritability estimates range from 0 to 1 and are often expressed as a percentage. A number close to 1 may be indicative of a highly heritable trait within a population.

**Hermaphroditism:** It is a sexual system where each individual has both kinds of reproductive organs and can produce both gametes associated with male and female sexes.

**Heterozygosity:** Indicates the presence of two different alleles at a particular gene locus. Heterozygosity is of great interest for the study of genetic variation in natural populations, providing valuable information on the structure and history of a population. In general terms, high heterozygosity means lots of genetic variability, and low heterozygosity means little genetic variability.

**Heterogamety:** Denotes the sex that possesses dissimilar sex chromosomes. For example, in mammals, males are heterogametic (XY) and females homogametic (XX). In fish, a vast array of sex chromosomal systems has been described, including the XX/XY (male heterogamety) or WZ/ZZ type (female heterogamety).

**Hybridization:** In reproductive biology, hybridization refers to the process of producing offspring by mating of two parents from different varieties or species. In fish, hybrids are mostly produced by crosses between two species of the same genus, and the result may give fertile, infertile or semi-fertile progeny, affecting one or both sexes.

**Ichthyoplankton:** This term refers to the eggs and larvae of fish. They belong to the plankton assemblages during their development. Ichthyoplankton abundance informs about the relative population size for the spawning stock.

**Inhibition:** Involves interference with, or restraint of, the activities of biologic molecules or complexes involved in a process. Different substances can produce, for example, inhibition of DNA replication, transcription or amplification.

**Interspecific:** Biological interaction that occurs between individuals of different species.

**Intraspecific:** Biological interaction that occurs between individuals of the same species.

**Invasive species:** An invasive species is an organism that is not indigenous, or native, to a particular area. These species may be introduced naturally, accidentally or intentionally into an environment that is not their own, managing to adapt and colonise it, displacing native species. Species invasion is one of the main causes of biodiversity loss in the world, impacting also on the health and economy of the affected areas.

**Life-history strategies:** Life history refers to the pattern of survival and reproduction events during the life of an organism. Life history traits include maximum body size, longevity, growth rate, age and size at maturity, fecundity, timing of reproduction,

parental care among others. A life history strategy is a collection of life history traits that are well-adapted for their role and environment for a given species or population. The optimal life history strategy is different for each species or population and depends on its characteristics, environment and other constraints.

**Local adaptation:** Refers to the concept that individuals of local populations tend to have a higher mean fitness in their native environment than in other environments. This phenomenon results from the interaction between multiple evolutionary forces (e.g., genetic drift, migration, mutation, and selection).

**Locus:** A locus (plural loci) is the specific physical location of a gene or other DNA sequence on a chromosome, like a genetic street address.

**Mean absolute error:** Metric used to evaluate the performance of regression models. The mean absolute error is defined as the average of the all absolute differences between true and predicted values. Considering two sets of data, one calculated and the other observed, for the same phenomenon, the mean absolute error is used to quantify the accuracy of a prediction technique by comparing the predicted versus the observed values. A mean absolute error of zero would indicate a perfect model without any error in its predictions.

**Metabarcoding:** Entails the analysis of a pool of genetic material using high throughput sequencing (HTS) to determine the sequence information of short regions of one or a few genes (called DNA barcodes), which can then be linked to a DNA barcode database. Metabarcoding does not focus on one specific organism, but instead aims to determine taxonomic species composition within a complex sample.

**Metagenomics:** Similar to metabarcoding, as it uses HTS to characterize complex samples but involves whole community genome sequencing, which provides a more thorough discernment into community diversity and function, thus enabling not only taxonomic identification, but also functional characterization of the environment.

**Metapopulations:** A group of populations that are spatially separated but yet interact at some level. In fisheries, metapopulations are important as they allow gene flow between populations.

**Method of moments:** In statistics, the method of moments (MoM) is a method of estimation of population parameters. It is an alternative to the method of maximum likelihood.

**Microarray:** Solid supports that have molecules such as DNA, RNA, or proteins attached in highly organized arrays and are normally used to evaluate expression of large numbers of genes or proteins at one time.

**Microsatellites:** Short segment of DNA, usually one to six or more base pairs in length that is repeated multiple times in tandem. The number of repeats, not the repeated sequence, creates different alleles. They are widely distributed throughout the genome in eukaryotes and often used as markers for kinship and genetic diversity studies.

**Mitochondrial DNA:** Genetic material found in mitochondria, cell organelles responsible for energy production. The mitochondrial genome consists of a closed, double-stranded circular DNA, which replicates autonomously within the organelle. Mitochondrial DNA, unlike nuclear DNA, is inherited from the mother, because during fertilisation it is the egg that provides the mitochondria.

**Multiplexing:** Entails pooling DNA fragments from different samples and sequence them all together. Individual "barcode" sequences are added to each DNA fragment during HTS library preparation so that each fragment can be assigned to the sample of origin. Multiplexing exponentially increases sample throughput.

**Nagoya Protocol:** It is an international agreement that aims to contribute to the conservation and sustainable use of biological diversity. Its main objective is to regulate the use of genetic resources and the fair and equitable sharing of the benefits arising from their use. For more information, visit https: //www.cbd.int/abs/.

**Nucleotide:** Is the basic building block of nucleic acids (RNA and DNA) and consist of a five-carbon sugar molecule, a phosphate group and a nitrogenous base (adenine (A), cytosine (C), guanine (G) and thymine (T); in RNA, the base uracil (U) takes the

place of thymine). The chains of these nucleotides that encode the information content in RNA and DNA.

**Maturity ogive:** Proportion of mature individuals at age or length. It allows to define the size or age at which 50% of the individuals of the stock are mature. It is a basic parameter in fisheries analysis and stock assessment. The most accurate method for determining a maturity ogive is by histological examination of the gonads, to distinguish between individuals that have never spawned in their lives and those that have spawned.

**Operational Taxonomic Units:** In metagenomics, OTU is an operational definition used to group closely related individuals based on DNA sequence similarity. OTUs have been the most commonly used units of diversity, especially when analysing taxonomic marker gene sequence datasets.

**Otoliths:** Calcium carbonate structures located in the inner ear of teleost fish (bony skeleton fish), with functions related to balance and hearing. The size and shape of otoliths vary by species. Otoliths are used by fisheries scientists to estimate age of the fish, by counting the number of growth rings present in the otolith. The age data gathered from otoliths allow scientists to estimate several key parameters of a fish population, such as growth rates, maximum age and age at maturity. .

**Outbreeding depression:** In biology, is the reduction of fitness resulting from crosses between two genetically distant populations. For example, if individuals from two populations that are each adapted to their natal environments hybridize, their offspring will contain a mixture of alleles that may not be well suited to either environment.

**Overfishing:** Phenomenon that occurs when fish and other marine species are caught faster than their ability to replenish the population. Causes include poor fisheries management, illegal and unregulated fishing, indiscriminate fishing and the use of unselective gear. The main consequences of overfishing are the loss of biodiversity and the imbalance of marine ecosystems.

**Panmictic population:** A population of individuals of both sexes that mate entirely at random. This assumes that there are no mating restrictions, neither genetic nor behavioural, upon the population and that therefore all recombination is possible.

**Parent-Offspring-Pairs:** A pair of organisms that are closely related and share half of their DNA that is inherited by the offspring (O) from the parent (P). CKMR estimates POPs and other close kin to understand population parameters such as abundance, population size, fecundity by size and selectivity and mortality.

**Pedigree:** Genealogy. Study and monitoring of the ancestry and descent of an organism, as well as its characteristics and relationships between them.

**Physiology:** The science that studies the functioning of living beings, from the cellular level to the level of the individual. Its objectives are: the description, analysis and classification of the phenomena presented in the isolated organism; the assignment of each function to its appropriate organ; the study of the conditions and mechanisms that determine each function; and the study of the regulation and coordination of the different functions.

**Polymerase Chain Reaction:** Polymerase chain reaction (abbreviated PCR) is a laboratory technique for rapidly producing (amplifying) millions to billions of copies of a specific segment of DNA. It is based on the natural processes a cell uses to replicate a new DNA strand and consists of repeated cycles of: (1) denaturation, (2) annealing, and (3) elongation. It is used to obtain sufficient amounts of DNA to perform molecular biology procedures.

**Population:** Group of individuals that have similar demographic or genetic characteristics and thus will respond uniquely and independently to fishing.

**Population dynamic model:** Mathematical model that studies the composition of a population of the same species and its variation over time, as well as the biological and environmental processes that drive these changes. A dynamic model includes time as a dependent variable. The study of demographic parameters such as survival, reproductive success, distribution or dispersion (emigration-immigration) allows us

to know and identify the future of a certain species as well as to detect those points where it is convenient to act in order to ensure the maintenance of the population.

**Population genetics:** Study of genetic variation within and among populations and the evolutionary factors that explain this variation. Seeks to understand how and why the frequencies of alleles and genotypes change over time influenced by factors such as natural selection, genetic drift, mutations and gene flow.

**Precision:** it refers to how close measurements of the same item are to each other. See also Accuracy, Bias, and Reproducibility.

**Primers:** Is a short nucleic acid sequence that provides a starting point for DNA synthesis. In the PCR method, a pair of primers is used to hybridize with the sample DNA and define the region of the DNA that will be amplified. Primers are also referred to as oligonucleotides. Blocking primers are modified primers that preferentially bind to DNA which amplification is to be suppressed and are used to enhance PCR amplification of rare sequences in mixed samples.

**Quantitative Polymerase Chain Reaction:** Is a major development of PCR technology that enables reliable detection and measurement of products generated during each cycle of the PCR process. Abbreviated as qPCR, adds two elements to the standard PCR process; a fluorescent dye and a fluorometer to estimate absolute amounts of DNA present.

**Restriction-site associated DNA sequencing:** Restriction site-associated DNA sequencing (RAD-Seq) is a powerful method for SNP detection in genomes, designed to identify polymorphic variants adjacent to restriction enzyme digestion sites. Samples at reduced complexity across target genomes, identifying and scoring thousands of genetic markers, randomly distributed and does not require previous genomic information.

**Recruitment:** This term can refer to two separate processes in fisheries science: i) the process by which young fish that survive the egg, larva and juvenile stages enter the exploitable phase of a fish stock for the first time and are susceptible to be captured by a particular fishing gear. ii) process by which very young, small fish often undergoing a mass mortality survive to become slightly older, larger fish at different life stage with a much reduced mortality mostly density-independent mortality. .

**Reference labs:** Laboratory that is nationally or supranationally accredited for its ability and its accuracy. They arise from the recognised need to promote uniform practices and reliability of methods of analysis, tests and diagnosis. Governments need to control according to standards, like the International Standards Organization, and the standard and the human resources to test for that standard are warehoused in reference laboratories.

**Reference genome:** A dataset that aims to model and represent the DNA sequence, or genome, of an idealised individual within a species. Reference genomes are assembled by scientists to serve as representative DNA sequences of different species that can then be analysed and compared.

**Regression analysis, __ models:** Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It uses different mathematical functions or regression models to describe such relationships.

**Reproducibility:** Reproducibility is the closeness of agreement between the results of measurements performed under different measurement conditions: different operators, different equipment, or different laboratories.

**Research survey (at sea):** Activities involving the monitoring of fish stocks and/or marine biological resources and the ecosystem, carried out on a vessel dedicated to such scientific research and designated for this task by a Member State (as defined by EU MAP (COM Delegated Decision (EU) 2021/1167).

**Restriction enzymes:** Also known as restriction endonucleases, are enzymes that cleave DNA molecules at specific sites. They are used to cut the DNA at sequence-specific sites to produce fragments with a known sequence at each end. The use of restriction enzymes is critical to certain laboratory methods, including RAD-Seq.

**Restriction-site:** A sequence in the DNA that can be recognized and cut by a specific restriction enzyme.

**Robustness:** It is a measure of the capacity of a method to remain unaffected by small, deliberate variations in method parameters and provides an indication of reliability during normal usage. In other words, is the degree of reproducibility of test results obtained by the analysis of the same samples under a variety of normal test conditions, such as different laboratories, different analysts, different instruments, different reagent lots, different elapsed assay times, different assay temperatures, different days, etc.

**Selectivity:** In fisheries, this term refers to the distribution of fishing mortality in line with the age composition of fish. It is determined by the fishing gear type and spatial and seasonal distributions of fishing and by fish growth and migration.

**Sensitivity:** The ability of an analytical method to be able to detect (Qualitative Analysis) or determine (Quantitative Analysis) small concentrations of analyte in the sample, with acceptable accuracy and precision. Refers to the minimum value that can be detected.

**Sequencing:** Refers to the general laboratory technique for determining the exact sequence of nucleotides, or bases, in a DNA or RNA molecule. Methods for DNA sequencing have evolved over the years, but, in general, they consist of breaking the DNA strands into small fragments and applying a technique to determine the order of the nitrogenous bases of the DNA nucleotides. Sequencing information has traditionally been elucidated using a low throughput technique called Sanger sequencing, until **high throughput sequencing** (HTS) technologies were developed. These are capable of sequencing multiple DNA molecules in parallel, enabling hundreds of millions of DNA molecules to be sequenced at a time. **Reduced Representation Bisulfite Sequencing** (RRBS) is a high-throughput sequencing technique used to study DNA methylation on a genome-wide scale at single-nucleotide resolution.

**Single Nucleotide Polymorphism:** A single nucleotide variation in a DNA sequence, in which one nitrogenous base (A, T, G, C) is replaced by another. At the population level, the variation has to occur in more than 1% of the population to be considered a SNP.

**Standard:** A term used to indicate that something serves as a pattern, model or point of reference for measuring or evaluating things of the same kind

**Stock:** Demographically cohesive group of individuals of one species exploited in a specific area.

**Survival:** Percentage of individuals remaining alive: number of individuals alive at the end of the census period, divided by the number of individuals alive at the beginning. Indicates the probability to live on.

**Sustainability:** Term that refers to the ability to persist over the long term. Sustainable fisheries are those that exploit fish resources at a sustainable rate, adapting to the reproductive rate of fish to ensure that fish populations do not decline over time because of fishing practices. It is also essential that the structure, productivity and diversity of marine ecosystems and habitats are respected, minimising impacts on other species, mainly protected or threatened species.

**SWOT analysis:** Strategic planning tool that provides relevant information for decision-making through the identification and analysis of internal (strengths and weaknesses) and external (opportunities and threats) factors.

**Target species:** In fisheries, species of commercial interest to which the fishing effort of a fleet is directed at in a fishery.

**Taxonomy:** In biology, taxonomy is the scientific study of naming, defining and classifying groups of biological organisms based on shared characteristics, as well as the bases, principles, methods and rules or laws that regulate this classification

**Teleost:** The largest infraclass in the class Actinopterygii, the ray-finned fishes, represent more than one-half of the total number of living vertebrate species with over 26,000 species described.

**Thermoregulation:** Ability of an organism to keep its internal temperature stable within certain limits, even when the ambient temperature is very different. Endotherms, also called warm-blooded animals, use internally generated heat to maintain their body temperature, which remains constant regardless of the outside temperature. Ectothermic, or cold-blooded, animals depend on external heat sources, varying their body temperature with the temperature of the surrounding environment.

**Trawling:** Trawling is one of the most common methods of fishing. Trawl nets are designed to be towed by a boat through the water column (mid-water trawl) or along the sea floor (bottom trawl). There are two types of fishing gears used in bottom trawl surveys: otter trawl and beam trawl. The **otter trawl** is the most widely used bottom-fishing gear. As it is dragged forward, a pair of flat plates called otter boards—one on each side of the trawl net and weighing several tons—spreads horizontally to keep the mouth of the trawl open; at the same time, a long rope with steel weights keeps the mouth open along its bottom edge. Meanwhile, **beam trawl** is a type of trawl where the mouth of the net is held open by a wooden or metal beam. In relation to mid-water trawls, the most common fishing method is **pelagic trawling**, where the net is extended horizontally by pelagic trawl doors. By altering the speed of the vessel or the length of the trawl cable, the position of the trawl in the water column can be changed to suit the desired depth.

**Uncertainty:** The lack of certainty, a state of limited knowledge where it is impossible to exactly describe the existing state, a future outcome, or more than one possible outcome. In fisheries management it is a term used to describe those situations in which the risks (e.g. of overexploitation) are unknown.

**Warm-blooded vertebrates:** See Thermoregulation.

## 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.

HZ-04-23-476-EN-N

Doi 10.2926/803359
ISBN 978-92-95231-54-2

Publications Office
of the European Union