frontiers | Research Topics

Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments.

Edited by
Fran Saborido-Rey, Francesc Piferrer, Reinhold Hanel and Laura Casas

Published in
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-3109-9
DOI 10.3389/978-2-8325-3109-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments.

Topic editors

Fran Saborido-Rey — Spanish National Research Council (CSIC), Spain
Francesc Piferrer — Institute of Marine Sciences, Spanish National Research Council (CSIC), Spain
Reinhold Hanel — Thünen Institute of Fisheries Ecology, Germany
Laura Casas — Institute of Marine Research, Spanish National Research Council (CSIC), Spain

Citation

Saborido-Rey, F., Piferrer, F., Hanel, R., Casas, L., eds. (2023). *Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments.* Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3109-9

Table of contents

05 Editorial: Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments
Laura Casas, Reinhold Hanel, Francesc Piferrer and Fran Saborido-Rey

08 A cold-water fish striving in a warming ocean: Insights from whole-genome sequencing of the Greenland halibut in the Northwest Atlantic
A-L. Ferchaud, E. Normandeau, C. Babin, K. Præbel, Rasmus Hedeholm, C. Audet, J. Morgan, M. Treble, W. Walkusz, P. Sirois and L. Bernatchez

26 The state of the art in cost-benefit of HTS methods for stock assessment: An overview
Gonzalo Rodríguez-Rodríguez, Hugo M. Ballesteros, Eduardo Sánchez-Llamas, Roberto Bande and Rosa Fernández Otero

40 Lack of panmixia of Bothnian Bay vendace - Implications for fisheries management
María-Eugenia López, Mikaela Bergenius Nord, Olavi Kaljuste, Lovisa Wennerström, Zeynep Hekim, Joni Tiainen and Anti Vasemägi

53 Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments
Sergio Ramírez-Amaro, Marta Bassitta, Antònia Picornell, Cori Ramon and Bàrbara Terrasa

74 Towards a unified eco-evolutionary framework for fisheries management: Coupling advances in next-generation sequencing with species distribution modelling
Miguel Baltazar-Soares, André R. A. Lima, Gonçalo Silva and Elie Gaget

83 Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology
Francesc Piferrer and Dafni Anastasiadi

99 Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology
Dafni Anastasiadi and Francesc Piferrer

112 Atlantic cod (*Gadus morhua*) assessment approaches in the North and Baltic Sea: A comparison of environmental DNA analysis versus bottom trawl sampling
Yassine Kasmi, Tina Blancke, Erik Eschbach, Benita Möckel, Laura Casas, Matthias Bernreuther, Pedro Nogueira, Gertrud Delfs, Samira Kadhim, Timo Meißner, Marcellus Rödiger, Amina Eladdadi, Christoph Stransky and Reinhold Hanel

126 Genetic markers associated with divergent selection against the parasite *Marteilia cochillia* in common cockle (*Cerastoderma edule*) using transcriptomics and population genomics data
M. Pampín, A. Casanova, C. Fernández, A. Blanco, M. Hermida, M. Vera, B. G. Pardo, R. M. Coimbra, A. Cao, D. Iglesias, M.J. Carballal, A. Villalba and P. Martínez

140 State-of-the-art of data analyses in environmental DNA approaches towards its applicability to sustainable fisheries management
Natalia Petit-Marty, Laura Casas and Fran Saborido-Rey

156 SNP-based discrimination of pink salmon stocks of the Sea of Okhotsk basin: resolution of the approach and possible ways to increase it
Daria A. Zelenina, Alexey A. Sergeev, Anna I. Kositsina, Valeria A. Soshnina and Nina Yu. Shpigalskaya

167 The potential use of genomic methods in bottom trawl surveys to improve stock assessments in Europe
Rebeca Rodríguez-Mendoza and Fran Saborido-Rey

180 A review of genomics methods and bioinformatics tools for the analysis of close-kin mark-recapture
Laura Casas and Fran Saborido-Rey

195 A review of an emerging tool to estimate population parameters: the close-kin mark-recapture method
Laura Casas and Fran Saborido-Rey

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 13 July 2023
DOI 10.3389/fmars.2023.1238133

OPEN ACCESS

EDITED AND REVIEWED BY
Stephen J Newman,
Department of Primary Industries and Regional Development of Western Australia (DPIRD), Australia

*CORRESPONDENCE
Laura Casas
lauracasas@iim.csic.es

RECEIVED 10 June 2023
ACCEPTED 19 June 2023
PUBLISHED 13 July 2023

CITATION
Casas L, Hanel R, Piferrer F and Saborido-Rey F (2023) Editorial: Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments.
Front. Mar. Sci. 10:1238133.
doi: 10.3389/fmars.2023.1238133

Editorial: Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments

Laura Casas[1]*, Reinhold Hanel[2], Francesc Piferrer[3] and Fran Saborido-Rey[1]

[1]Institute of Marine Research, Spanish National Research Council (CSIC), Vigo, Spain, [2]Thünen Institute of Fisheries Ecology, Bremerhaven, Germany, [3]Institut de Ciències del Mar, Spanish National Research Council (CSIC), Barcelona, Spain

KEYWORDS

fisheries, genomics, stock assessment, high throughput sequencing, research surveys

Editorial on the Research Topic

Prospects and challenges for the implementation of HTS genetic methods in fisheries research surveys and stock assessments

The status of exploited fish stocks requires regular assessment to ensure sustainable fishing practices. These assessments rely on the collection of data from multiple sources, which include fishery-dependent data — catches, landings and biological information — as well as fishery independent data obtained from research surveys (Stomatopoulos, 2002). These traditional methods provide valuable and systematized information regarding exploited fish populations, marine biodiversity, and their environment, and constitute an essential piece in stock assessment and scientific advice. However, they also have important limitations such as high economic costs of research surveys coupled with complex logistics. As a result, fisheries data have a generally limited coverage in space and time and are time consuming in their analysis (Stomatopoulos, 2002), leading to often biased and imprecise estimations (Hilborn and Walters, 2013; Pennino et al., 2016). This can have a great impact on the quality of the scientific advice provided to management bodies and, hence, ultimately on fishing activities. Moreover, traditional methods lack the capacity to provide information on crucial parameters for stock management, such as the delineation of stock boundaries or the connectivity, among others, highlighting the crucial need for innovative assessment tools to aid fisheries management (Valenzuela-Quiñonez, 2016). In addition, most exploited species worldwide lack the data needed to assess their status, despite the increasing trend of overfished stocks in the last four decades (FAO, 2022).

Advancements of high throughput molecular technologies have unlocked the power of genomics to complement traditional methods, reducing uncertainty, improving cost-efficiency of fish stock assessments, and opening the possibility for expanding the range of assessed species. Novel technologies provide the statistical power and resolution required

to solve key issues in fisheries management but despite their potential to improve fisheries advice, the integration of genomic-informed methods into fisheries management practices is still very limited and patchy (Bernatchez et al., 2017; Benestan, 2019). The starting point of this Research Topic has resided in the recognition that there is a pressing need to overcome these barriers (Ovenden et al., 2015; Bernatchez et al., 2017) and bridge the gap between the two fields to secure the sustainability of exploited fish stocks.

This editorial summarizes the contributions to the Frontiers Research Topic "*Prospects and Challenges for the Implementation of HTS Genetic Methods in Fisheries Research Surveys and Stock Assessments*", established under the Marine Fisheries, Aquaculture and Living Resources section in the Frontiers in Marine Science journal. This Research Topic aimed at exploring recent developments in the field of genomics applied to fisheries. Here, we compile a set of 14 articles structured in two categories, with nine contributions devoted to reviewing the latest developments in methodologies with potential use in fisheries science, whereas the rest represent practical applications in this field. Below, we summarize each of the articles included in this Research Topic and add thoughts on the potential future directions of fisheries genomics.

Rodríguez-Mendoza and Saborido-Rey carry out a review of the bottom trawl research surveys in the EU to determine how genomic techniques can be used to improve survey data, considering the needs of current and future stock assessment in Europe. This information is key for the implementation of novel genomic methods in fisheries research surveys, as this requires a deep understanding on how the surveys are conducted, which data is collected onboard and how these data are used in stock assessment, but also in ecosystem assessment and other purposes. This key step needs to be taken prior to a large-scale implementation of genomics methods in surveys.

Ramírez-Amaro et al. provide an overview of the environmental DNA methodology and explore the capacity of several approaches based on this emerging tool to inform Ecosystem-Based Fisheries Management. They analyze the main aspects affecting eDNA behavior in the marine environment and provide a detailed compilation of eDNA applications in fisheries management. Petit-Marty et al. highlight the lack of consistency across data analysis in eDNA studies and provide a review of the essential steps of eDNA data processing and of the bioinformatics tools to produce sound, reproducible, and comparable results. This article provides essential guidance for a comprehensive application of eDNA-based approaches in fisheries management.

Piferrer and Anastasiadi present a detailed review of piscine epigenetic clocks, an emerging genomic technology for age estimation, a crucial parameter for fisheries management. The authors provide guidelines for clock development, detailing the steps and considerations required to produce accurate, precise, and reproducible clocks that can contribute to better fisheries management practices. This article is complemented by a second review focused on the computational steps and tools required for the construction of epigenetic clocks in fish and for age prediction (Anastasiadi and Piferrer).

Casas and Saborido-Rey provide an overview of the Close-kin mark-recapture (CKMR) method, an emerging methodology grounded in genomics, to estimate abundance and other demographic parameters (e.g., population trend, survival rates, connectivity) that are essential in fisheries assessments. The authors assess the readiness, viability, and maturity of the method in a fisheries framework, evaluate technical considerations and requirements for a successful implementation and provide advice for planning a CKMR study. A second review provides guidance to address the genomics and bioinformatics steps required to analyse CKMR data, using a simple terminology to reach potential users with no previous expertise in genomics (Casas and Saborido-Rey).

Rodríguez-Rodríguez et al. assess the cost-effectiveness of the three techniques above (eDNA, epigenetic clocks, and CKMR) applied on marine ecosystems and fisheries and for stock assessment purposes. The authors provide a comparison of the information and cost provided by surveys and novel methodologies, highlighting the striking shortage of systematic cost analysis studies of genomic techniques despite the general assumption of their cost-effectiveness and efficiency in terms of effort and time.

Baltazar-Soares et al. present a Perspective article addressing one of the potential limitations of the implementation of genomics methods in fisheries assessments. Although the establishment of modern genomic approaches and subsequent large-scale genomic datasets can clearly enhance our understanding of stock spatial distributions, it may also compromise the temporal depth analysis and consequently the forecasting ability of stock distribution. To resolve this, the authors propose an interesting approach integrating genomic information on temporal projections of species distributions computed by Species Distribution Models (SDMs) that accounts for habitat selection given the current evolutionary potential of the stock.

Kasmi et al. present a practical application of a real-time PCR-based eDNA approach to assess quantitatively the abundance of cod in the North and Baltic Seas. The authors comparatively analyse the concentrations of eDNA with bottom trawl fisheries catches, showing a significant correlation (95%) between eDNA and cod biomass. This study highlights the potential of this non-invasive tool to assess abundance of fish stocks.

Ferchaud et al. provide an excellent example of the utility of novel genomic techniques to inform fisheries management. The authors use whole-genome data of over 1300 individuals of Greenland Halibut, *Reinhardtius hippoglossoides*, sampled across the Northwest Atlantic to infer geographic population structure and local adaptation. The analysis revealed a panmictic population across the Northwest Atlantic with the exception of the Gulf of Saint Lawrence that presents significant genetic differentiation, mainly attributed to environmental variables, suggesting that this stock might be particularly vulnerable to environmental changes.

Pampín et al. use genomics (2b-RADseq) and transcriptomics to identify single nucleotide polymorphisms (SNP) markers associated with resilience to the protozoan parasite *Marteilia cochillia* that infects bivalves causing a disease known as marteiliosis. Populations of the common cockle (*Cerastoderma edule*) in Galicia (Northwest Spain) constitute a valuable socioeconomic resource for coastal communities but have been decimated by this parasite. The authors produce a robust genotyping SNP tool set that can be applied in marker-assisted

selection programs for restoring affected cockle natural bed ecosystems and recovering their production.

López et al. use RAD-seq to investigate the population structure of vendace (*Coregonus albula*), a small salmonid fish, in the Bothnian Bay, the northernmost part of the Baltic Sea. Their analysis rejects the hypothesis of panmixia as a clear genetic differentiation is found at one of the studied locations. Additionally, they reveal a weak structuring between samples from the Swedish and the Finnish coasts. The study demonstrates the power of RAD-sequencing to detect low but significant genetic structuring relevant for fisheries management.

Lastly, Zelenina et al. demonstrate the usefulness of a panel of SNP markers developed from a RAD-sequencing project to study the intraspecific polymorphism of the complex stock structure of pink salmon, *Oncorhynchus gorbuscha*, in the Okhotsk basin. The complexity is the result of the existence of two allochronous lineages. The developed methodology provides an efficient tool to reliably differentiate regional stocks and determine the proportion of fish from the main reproduction areas in these mixed stocks.

We hope that this Research Topic can contribute to the perception of the potential of new emerging genomic technologies to improve scientific advice and enhance fisheries assessments. The application of genomics tools in fisheries management should not be delayed further, but demands an improved communication and reciprocal training among geneticists, fishery biologists and managers. In addition, the uptake of genomic information for assessment and advice should be progressive, starting with an implementation in parallel with traditional assessment to ensure the preservation of data series and a proper standardization of novel methodologies.

Author contributions

The idea of the Research Topic was developed during a Project meeting under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]". LC and FS-R wrote the first draft of the editorial article and all authors contributed equally to the discussion-conclusions and in writing the final manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to express their gratitude to Gary Carvalho, Laszlo Orban and Allan Tucker for their advice and fundamental inputs that led to the organization of this Research Topic.

Conflict of interest

The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods EMFF/2018/015" signed with the European Climate, Infrastructure and Environment Executive Agency CINEA and funded by the European Union. The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Benestan, L. (2019). *Population genomics applied to fishery management and conservation* (Cham: Springer), 399–421. doi: 10.1007/13836_2019_66

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32, 665–680. doi: 10.1016/j.tree.2017.06.010

FAO. (2022). *The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation*. (Rome, FAO). doi: 10.4060/cc0461en

Hilborn, R., and Walters, C. J. (2013). *Quantitative fisheries stock assessment. 3rd ed.* Eds. C. J. Walters and R. Hilborn (Dordrecht: Springer Science & Business Media).

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., and 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 Fish.* 16, 125–159. doi: 10.1111/faf.12052

Pennino, M. G., Conesa, D., López-Quílez, A., Muñoz, F., Fernández, A., and Bellido, J. M. (2016). Fishery-dependent and -independent data lead to consistent estimations of essential habitats. *ICES J. Mar. Sci.* 73, 2302–2310. doi: 10.1093/icesjms/fsw062

Stomatopoulos, C. (2002). *Sample-based fishery surveys: A technical handbook*. FAO Fisheries Technical Paper. No. 425. (Rome, FAO). 2002. 132. Available at: https://www.fao.org/3/Y2790E/Y2790E00.htm.

Valenzuela-Quiñonez, F. (2016). How fisheries management can benefit from genomics? *Brief. Funct. Genomics* 15, 352–357. doi: 10.1093/bfgp/elw006

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fmars.2022.992504

OPEN ACCESS

EDITED BY
Fran Saborido-Rey,
Spanish National Research Council (CSIC), Spain

REVIEWED BY
Pablo Saenz-Agudelo,
Austral University of Chile, Chile
Natalia Petit-Marty,
Institute of Marine Research, Spanish National Research Council (CSIC), Spain

*CORRESPONDENCE
A–L. Ferchaud
annelaureferchaud@gmail.com

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 12 July 2022
ACCEPTED 05 August 2022
PUBLISHED 02 September 2022

CITATION
Ferchaud A-L, Normandeau E, Babin C, Præbel K, Hedeholm R, Audet C, Morgan J, Treble M, Walkusz W, Sirois P and Bernatchez L (2022) A cold-water fish striving in a warming ocean: Insights from whole-genome sequencing of the Greenland halibut in the Northwest Atlantic. *Front. Mar. Sci.* 9:992504. doi: 10.3389/fmars.2022.992504

A cold-water fish striving in a warming ocean: Insights from whole-genome sequencing of the Greenland halibut in the Northwest Atlantic

A-L. Ferchaud [1]*, E. Normandeau [1], C. Babin [1], K. Præbel [2], Rasmus Hedeholm [3], C. Audet [4], J. Morgan [5], M. Treble [6], W. Walkusz [6], P. Sirois [7] and L. Bernatchez [1]

[1]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada, [2]Norwegian College of Fishery Science, UiT the Arctic University of Norway, Tromsø, Norway, [3]Sustainable Fisheries Greenland, Skørping, Denmark, [4]Institut des sciences de la mer de Rimouski, Université du Québec à Rimouski, Rimouski, QC, Canada, [5]Fisheries and Oceans Canada, St. John's, NF, Canada, [6]Fisheries and Oceans Canada, Winnipeg Department, Arctic Aquatic Research Division, Winnipeg, AB, Canada, [7]Chaire de recherche sur les espèces aquatiques exploitées, Département des sciences fondamentales, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

Characterizing the extent of genetic differentiation among individuals and its distribution across the genome is increasingly important to inform both conservation and management of exploited species. The Greenland Halibut is one of the main demersal fish species to be commercially exploited in Eastern Canada, and accurate information on geographic population structure and local adaptation is required to ensure the long-term presence of this species. We generated high-quality whole-genome sequencing data for 1,297 Greenland Halibut sampled across 32 locations throughout the Northwest Atlantic (from Arctic Canadian and Greenlandic coasts to the Gulf of St Lawrence). Population genetic structure was analyzed, revealing an absence of population differentiation between Canada and west Greenland but significant genetic differentiation between the Gulf of Saint Lawrence and the remainder of the Northwest Atlantic. Except for Gulf of Saint Lawrence, Greenland Halibut thus appear to be panmictic throughout the Northwest Atlantic. Environmental Association Analyses revealed that the environment explained up to 51 % might be replaced by 51% of the differentiation observed between the two stocks, with both ocean-bottom and surface variables (*e.g.*, temperature and oxygen) involved in the observed genomic differentiation. Altogether, these results indicate that phenotypic differences previously observed between the Gulf of Saint Lawrence and the Northwest Atlantic likely resulted from functional adaptive divergence to their respective environmental conditions. Using coalescent simulations, we also assessed how high levels of migration between the two stocks would allow Greenland Halibut to potentially escape unfavorable environmental conditions in the Gulf of Saint Lawrence. In addition to supporting the management of this important

exploited species, this work highlights the utility of using comprehensive genomic datasets to characterize the effects of climate change across a wider range of species.

KEYWORDS

Greenland halibut, Gulf of Saint Lawrence, Northwest Atlantic, whole-genome sequencing, environmental association

Introduction

Defining biologically meaningful units with the aim of sustaining biodiversity is one of the major goals of populations management and conservation biology (Moritz, 1994; Allendorf et al., 2012). In particular, the detection of genetic structure provides a crucial tool to identify such units and to assess the degree of connectivity among populations (Bernatchez et al., 2017). Neglecting to consider population structure may increase risks of overexploitation or mis-management (Waples, 1998), especially in a world governed by anthropogenically driven climate change where species and populations are under constant pressure to adapt to new environmental conditions (Smith and Bernatchez, 2008; Sheffers et al., 2016). If environmental changes are persistent, phenotypic plasticity, migration and adaptation can help a species avoid extinction. Species living in marine ecosystems are traditionally considered to be highly connected due to their large population sizes and the limited numbers of effective barriers to gene flow found in these ecosystems (Nielsen et al., 2009). Because of these conditions, combined with a complex life-cycle, high fecundity and the potential for long-distance migration and dispersal, the use of neutral genetic markers in marine organisms has often not been powerful enough to separate populations into divergent groups (Gagnaire et al., 2015). However, with the advent of new and affordable high-throughput sequencing, genomic analyses of population structure in marine species have increasingly revealed diverse and complex signatures of population differentiation (Lamichhaney et al., 2012; Benestan et al., 2015; Bradbury et al., 2015; Lamichhaney et al., 2017, Van Wyngaarden et al., 2018, Xuereb et al., 2018). These patterns range from genome-wide polygenic variation associated with subtle, coordinated shifts in allele frequency at multiple loci (*e.g.* Gagnaire et al., 2012; Pavey et al., 2015; Babin et al., 2017) to localized genomic regions housing structural variants or genes of large effect within otherwise undifferentiated genomes (Kess et al., 2019; Matschiner et al., 2022). Across studies, genomic differentiation in marine species has been found to be associated with behavioral traits (Prince et al., 2017; Kess et al., 2019), spawning time (Lamichhaney et al., 2017) and environmental variation (Bradbury et al., 2010; Lamichhaney et al., 2012; Benestan et al., 2016; Stanley et al., 2018; Kess et al., 2021). These observations support the hypothesis that genetic differentiation in marine species often underlies adaptive differences that delineate significant ecological diversity. Most genomic studies investigating the association between genotypes and environment (GEA, Genotype Environment Association) were first based on reduced-representation sequencing of natural populations (*e.g.* RAD-Seq, RNA-Seq, targeted sequencing, SNP-chips) and allowed the identification of important candidate genes or loci involved in climate adaptation. In contrast, whole-genome sequencing (WGS) delivers data on the entire genome of individuals and provides information on the spatial pattern of variation along chromosomes (when a reference genome is available) (Fuentes-Pardo and Ruzzante, 2017). If GEAs are combined with WGS data, a comprehensive set of loci is used for correlation to environmental variables and, depending on population structure and the strength of differentiation along the environmental gradient, minor effect loci can also be uncovered (De Villemereuil et al., 2014; Bernatchez, 2016). Moreover, GEAs do not require prior knowledge of specific phenotypic traits; they are therefore less labor-intensive and more affordable, even for genomes of intermediate sizes, due to decreasing costs of sequencing technologies.

Greenland Halibut, *Reinhardtius hippoglossoides* (Walbaum, 1792), is a highly-migratory (Vihtakari et al., 2022), cold-water, long-lived, slow-growing, and economically valuable groundfish (DFO, 2021) with a circumpolar distribution throughout the Northern Hemisphere (Vihtakari et al., 2021). It supports several commercial fisheries throughout the Arctic and the North Atlantic oceans as well as in the Estuary and Gulf of Saint Lawrence (EGSL), Canada (Bowering and Brodie, 1995; Bowering and Nedreaas, 2000; Treble et al., 2008; Delaney et al., 2012). Greenland Halibut are known to spawn during winter months (Gundersen et al., 2010). Following emergence, larvae drift along in the upper layers of the water column for a few months and then settle in nursery areas until adult form is reached (Sohn et al., 2010). Knowledge of nursery areas is limited and only a few have been confirmed in the North Atlantic: Disko Bay (Greenland: Stenberg et al., 2016) and the Svalbard archipelago (Albert and Vollen, 2015). In the EGSL, there are two known nurseries described by Ait Youcef et al. (2013), with the main nursery being located in the estuary and the secondary one in the northeast region of the Anticosti Island. In addition to long-distance migration, this groundfish is also

known to undertake vertical movements and can make use of the pelagic environment for up to one fourth of an individual's lifetime (Boje et al., 2014). Both the important larval dispersal and migratory capacity of the species make it difficult to define stock limits and assess connectivity and population dynamics (Jørgensen, 1997; Stenberg et al., 2016; Barkley et al., 2018; Vihtakari et al., 2022, Bassi et al., submitted). Given the growing importance of this fishable resource, several studies have aimed at documenting the population structure of Greenland Halibut in the North Atlantic. However, results have been ambiguous and sometimes contradictory. Studies based on morphological features (Bowering, 1983; Bowering, 1988; Riget et al., 1992), growth and maturity rates (Templeman, 1973; Morgan and Bowering, 1997; Morgan et al., 2003), tag-recovery data (Boje, 2002), physiological patterns and the abundance and prevalence of parasites (Khan et al., 1982; Arthur and Albert, 1993) revealed some differences between sampled localities. In particular, Greenland Halibut from EGSL has been proposed to be a single population distinct from other stocks in the Atlantic based on prevalence of blood parasites (Khan et al., 1982; Arthur and Albert, 1993). Throughout the North Atlantic, previous genetic studies based on either allozymes, microsatellites or few numbers of single nucleotide polymorphism (SNP) have shown some, albeit weak, transatlantic differentiation (Fairbairn, 1981; Riget et al., 1992; Knutsen et al., 2007; Westgaard et al., 2017). Based on microsatellite markers, Pomilla et al. (2008) identified two genetically distinguishable stocks throughout the Northwest Atlantic with a geographic intermingling and a possible admixing suggesting that only a single unit for management through the Northwest Atlantic would be reasonable until more information is available. Similarly, Roy et al. (2014) concluded that a single, panmictic population is present in the Northwest Atlantic. Vis et al. (1997) detected a weak genetic distinctiveness of the EGSL population when compared with the rest of the Atlantic, but this difference was not significant. A recent study based on genotyping-by-sequencing (GBS) approach confirmed the significant distinctiveness of the EGSL and revealed a fluctuating contribution from the Newfoundland – Labrador area to the EGSL stock over the years (Carrier et al., 2020).

In the absence of clear population structure, the Northwest Atlantic Fisheries Organization (NAFO) divided the Northwest Atlantic into separate management units on the basis of limited empirical evidence: (1) Baffin Island – West Greenland (NAFO Subarea 0 + 1 (offshore)), (2) Cumberland Sound and West Greenland fjords (NAFO Divisions 0B and 1B to F), (3) Labrador – eastern Newfoundland (NAFO Subarea 2 + Divisions 3KLMNO), and (4) the Gulf of Saint Lawrence (NAFO Subarea 4) (Vihtakari et al., 2022). While recent stock assessments suggest that the Baffin Island – West Greenland stock is at low risk, with a biomass index above the average throughout the time series 1997 to 2017 (Northwest Atlantic Fisheries Organization 2020a), a recent slight declining trend in exploitable biomass has been observed in the Labrador – eastern Newfoundland stock (Northwest Atlantic Fisheries Organization 2020b), and a decrease of more than 60 % in exploitable biomass has been observed since 2004 in the Gulf of Saint Lawrence stock (DFO 2019) which corresponds to the southern limit of the species' distribution.

In this study, the aim was to characterize the genome-wide variation of Greenland Halibut using high-quality whole-genome sequencing data from 1,297 fish sampled from 32 locations across the Northwest Atlantic, from Arctic Canadian and Greenlandic coasts to the Gulf of St Lawrence. Specifically, (i) population genetic structure throughout the Northwest Atlantic was assessed, (ii) connectivity between distinguishable stocks was estimated, (iii) use of genotype-environment association analyses to identify potential candidate genes or loci involved in climate adaptation was made and (iv) environmental association with the genome-wide differentiation observed between stocks was corroborated.

Materials and method

Sampling

With the collaboration of Fisheries and Oceans Canada (DFO) and volunteer fishermen, an extensive sampling was conducted in the Northwest Atlantic Ocean to collect Greenland Halibut specimens throughout its geographical range. A total of 1,366 fish were sampled across 32 locations from the Gulf of Saint Lawrence to Nunavut (in the Canadian Artic) and along the Greenland Coast during summer and fall 2016 and 2017 across 32 localities (mean N = 43, ranging from 17 to 76, Figure 1, Table 1). Fish were caught using benthic trawls as part of the annual multispecies surveys conducted by the Department of Fisheries Organization. Fish were lethally sampled and a part of the pelvic fin was collected and stored in 95 % ethanol (EtOH) until DNA extraction.

DNA extraction, libraries and sequencing

Genomic DNA was extracted from a fin-clip using a salt-extraction protocol (Aljanabi and Martinez 1997) with an RNase A treatment (Qiagen). DNA quality of each extract was evaluated with Nanodrop 2000 (ThermoFisher scientific) and migration on a 1 % agarose gel electrophoresis. Following Therkildsen and Palumbi (2016), DNA fragments shorter than 1kb were removed by treating each extract with Axygen magnetic beads in a 0.4:1 ratio and eluting the DNA in 10mM Tris-Cl, pH 8.5. We measured DNA concentrations with the Accuclear ultra high sensitivity dsDNA quantification kit (Biotium) and normalized all samples at a concentration of 5ng/µL. Then, sample DNA extracts were randomized, distributed in plates (96 -well see

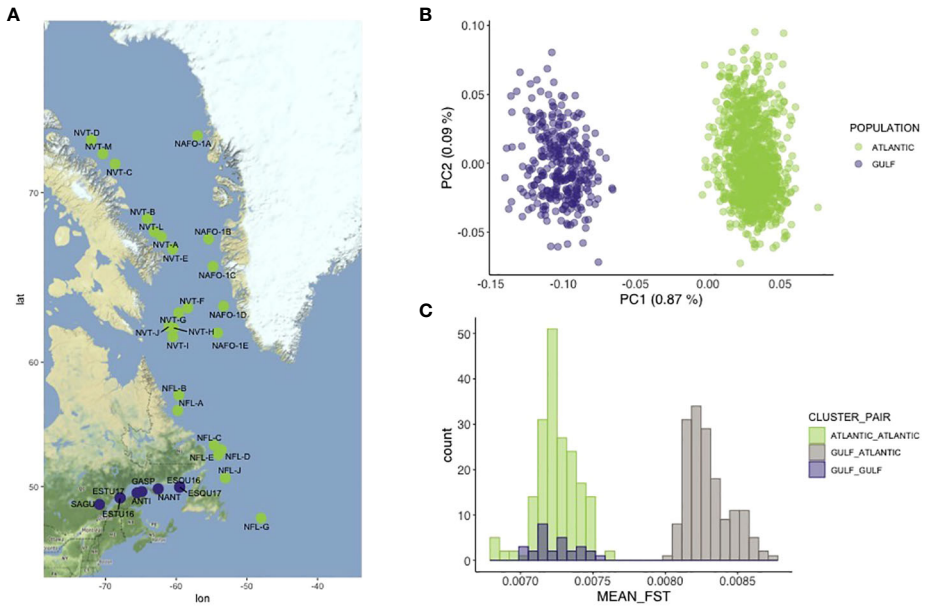


FIGURE 1
(A) Geographic locations of the 32 sampling sites studied across the Northwest Atlantic. Locations are colored according to PCA results discriminating locations form the Gulf of Saint Lawrence (purple) and locations from the rest of the Northwest Atlantic (green). **(B)** Two first axis of the principal component analysis obtained without the 18 detected migrants (see main text for detail). **(C)** Distribution of the global mean pairwise Fst estimated across all sampling localities.

details about randomization and number of plates used in Supplementary Material) and re-normalized at 2ng/µL.

Whole-genome high-quality libraries were prepared for each sample according to the protocol described in Baym et al. (2015) and Therkildsen and Palumbi (2016). Briefly, a tagmentation reaction using enzyme from the Nextera DNA sample preparation kit (Illumina), which simultaneously fragments the DNA and incorporates partial adapters, was carried out in a 2.5µl volume with approximately 2 ng of input DNA. Then, we used a two-step PCR procedure with a total of 12 cycles (8 + 4) to add the remaining Illumina adapter sequence with dual index barcodes and to amplify the libraries. The PCR was conducted with the KAPA Library Amplification Kit and custom primers derived from Nextera XT barcode sets A, B, C and D (total of 384 possible combinations, see Supplementary Material S1). Amplification products were purified from primers and size-selected with a two-steps Axygen magnetic beads cleaning protocol, first with a ratio 0.5:1, keeping the supernatant (medium and short DNA fragments), second with a ratio 0.75:1, keeping the beads (medium fragments). Final concentrations of the libraries were quantified with the Accuclear ultra high sensitivity dsDNA quantification kit (Biotium) and fragment size distribution were estimated with an Agilent BioAnalyzer for a subset of 10 to 20 samples per plate. Finally, equimolar amounts of 93 to 95 sample libraries were combined into 19 separate pools for sequencing on 19 lanes of paired-end 150bp reads on an Illumina HiSeq 4000 at the Norwegian Sequencing Center at the University of Oslo. Several samples were used in different pools and sequencing lanes in order to test for a putative sequencing lane effect in genetic variation (performing a PCA analysis as described in Ferchaud et al., 2022). No sequencing lane effect was detected and only one replicate for each sample (the one with the highest number of reads) was kept for the subsequent analysis. Given our multiplexing (up to 95 individuals per lane) and genome size (~ 600 Mb), we targeted a low sequencing coverage around 1.26 X in average. Individual samples containing too low or too high number of reads (top 2.5 % and bottom 2.5 %) were discarded from the dataset and subsequent analysis were conducted on 1,297 individuals (mean N per location = 41, ranging from 16 to 73, Table 1).

Sequencing filtering and processing

Raw reads were trimmed, filtered for quality, mapped to the reference genome (Ferchaud et al., 2022), cleaned for duplicates reads and mapping quality and then re-aligned using the pipeline available at https://github.com/enormandeau/wgs_sample_preparation, inspired by (Therkildsen and Palumbi, 2016) and fully described in Ferchaud et al. (2022). Given that more than 96 % of the assembly is comprised into the anchored

TABLE 1 32 locations analyzed in this study.

LOCALITY	Number of individuals		POPULATION	LATITUDE	LONGITUDE
	sampled	analyzed			
NAFO-1A	29	28 (28)	ATLANTIC	72.570958	-57.001872
NAFO-1B	30	28 (28)	ATLANTIC	67.630134	-55.419842
NAFO-1C	25	25 (25)	ATLANTIC	66.111617	-54.804607
NAFO-1D	43	38 (38)	ATLANTIC	63.729044	-53.310467
NAFO-1E	76	73 (73)	ATLANTIC	62.001372	-54.145092
NFL-A	45	44 (44)	ATLANTIC	56.416944	-59.768056
NFL-B	17	16 (16)	ATLANTIC	57.601389	-59.634167
NFL-C	45	43 (43)	ATLANTIC	53.617222	-54.668889
NFL-D	45	42 (42)	ATLANTIC	53.218056	-53.600556
NFL-E	47	46 (44)	ATLANTIC	52.8	-54.04
NFL-G	48	46 (45)	ATLANTIC	47.04	-48
NFL-J	49	49 (46)	ATLANTIC	50.78	-53.06
NVT-A	40	40 (40)	ATLANTIC	67.751667	-62.083333
NVT-B	41	38 (38)	ATLANTIC	68.685	-64.107778
NVT-C	42	40 (40)	ATLANTIC	71.335	-68.651667
NVT-D	41	39 (39)	ATLANTIC	72.389722	-72.003611
NVT-E	38	38 (38)	ATLANTIC	67.076944	-60.518333
NVT-F	45	44 (44)	ATLANTIC	63.621667	-58.346111
NVT-G	47	45 (45)	ATLANTIC	63.311389	-59.649722
NVT-H	47	43 (43)	ATLANTIC	62.335833	-60.36
NVT-I	44	42 (42)	ATLANTIC	61.751667	-60.471667
NVT-J	46	43 (43)	ATLANTIC	62.421111	-61.034167
NVT-L	48	46 (46)	ATLANTIC	68.008868	-63.107778
NVT-M	47	47 (46)	ATLANTIC	71.789076	-70.37409
ANTI	40	37 (37)	GULF	49.416667	-65.55
ESQU16	42	38 (35)	GULF	49.9485	-59.516
ESQU17	40	35 (31)	GULF	49.9485	-59.516
ESTU16	50	37 (37)	GULF	48.9548	-67.9288
ESTU17	38	45 (43)	GULF	48.9548	-67.9288
GASP	46	45 (45)	GULF	49.533333	-64.85
NANT	43	39 (37)	GULF	49.7807	-62.5422
SAGU	42	38 (38)	GULF	48.339722	-70.849722

Number of individuals sampled and analyzed (number analyzed in parentheses with migrants removed, see Methods).

24 chromosomes (Ferchaud et al., 2022), raw reads were mapped against a reduced version of the Greenland Halibut reference genome excluding the unassembled scaffolds. Moreover, the sex-linked chromosomes recently identified in this species (chr 10 and chr 21, Ferchaud et al., 2022) were also removed in order to avoid unexpected biases in structuration due to sex-linked markers in further analysis (Benestan et al., 2017). Individuals presenting too low coverage (more than 2sd below the mean) were excluded from the analysis. For low-coverage whole-genome sequencing (lcWGS) data, the recommended practice is to avoid basing downstream analysis on called genotypes (Nielsen et al., 2011) and to use a probabilistic approach based on genotype likelihoods instead. Several models for computing genotype-likelihood-based on read data have been implemented in the program ANGSD (Korneliussen et al., 2014) and it is currently the most widely used and versatile software package for the analysis of lcWGS (Lou et al., 2021). Therefore, ANGSD v0.931 was used for most of our subsequent analyses according to the pipeline documentation available at https://github.com/clairemerot/angsd_pipeline. For all analyses, input reads were filtered to remove reads with a samtools flag above 255 (not primary, failure and duplicate reads, tag -remove_bads = 1), with mapping quality below 30 (-minMapQ 30) and to remove bases with quality below 20 (-minQ 20). We also filtered in order to keep only SNPs covered by at least one read in at least 50 % of individuals (-minInd) and removed SNPs in putative repeated regions allowing a maximum depth of 3 times the number of individuals (-setMaxDepth). Finally, for most of the subsequent

analyses (unless mentioned otherwise), we kept SNPs with minor allele frequency above 5 %.

We first ran ANGSD to estimate genotype likelihoods (GL) with the GATK model (-doGlf 2 -GL 2 -doCounts 1), the spectrum of allele frequency (-doSaf 1) and the minor allele frequency (-doMaf 1) options. The major allele was based on the genotype likelihood and was the most frequent allele across all samples (-doMajorMinor 1). From this first analysis, we generated (i) a beagle file with GL estimates and (ii) a list of variants passing those filters and their respective major and minor alleles that were used for most subsequent analyses. The R program (R Core Team 2020) was employed for graphic output in subsequent analyses, *via* the package ggplot (Wickham, 2016).

Clustering analysis

Genome-wide variation across samples was explored using PCAangsd (Meisner and Albrechtsen, 2018) on the genotype likelihoods. This program extracts a covariance matrix that is then decomposed into principal component analysis (PCA) with R, using a scaling 2 transformation adding an eigenvalues correction, to obtain the individual PC scores (Legendre and Legendre 1998). Two clusters were identified by this analysis (on PC1, see results section). 18 of the 1,297 individuals were not correctly assigned to their geographic cluster and were considered migrants between the Gulf and the Northwest Atlantic. These migrants were confirmed using another clustering identification approach. To do so, NGSadmix (Skotte et al., 2013) was run with K = 2 (the K numbers revealed by the PCA above) to estimate admixture proportions across individuals. Once confirmed by the two approaches (PCAangsd and NGSadmix), migrants were removed before conducting a new PCA analysis and subsequent analyses.

Genome-wide pairwise F_{ST} comparisons was estimated in two rounds. First, in order to avoid bias due to lower numbers of samples, only locations with sample sizes above 30 were kept (28 locations out of 32, see Table 1) and pairwise F_{ST} estimations were performed (378 comparisons in total). A bimodal distribution of the mean global F_{ST} values from those 378 comparisons confirmed the existence of the two populations revealed by the PCA (Figure 1C). The presence of Isolation-By-Distance (IBD) was tested using a correlation between genomic differentiation estimates ($F_{ST}/(1- F_{ST})$) and log (geographic distance) (Rousset, 1997) across the sampled locations. Secondly, estimation of genome-wide F_{ST} was conducted between these two identified populations (Gulf *versus* Northwest Atlantic) in order to detect any genomic regions implied in this differentiation. To achieve this, individuals were pooled into two groups accordingly to the results of the PCA and, in order to avoid bias due to sample size differences between the two groups, the largest group (Northwest Atlantic) was randomly sub-sampled to 303 individuals, the number of samples available for the Gulf. To perform genome-wide F_{ST} estimations, allele-frequency spectrum (-doSaf 1) and minor allele frequencies were calculated for each locality (or population in the second round) with the previous list of variant positions (-sites) and their polarisation as major or minor alleles (-doMajorMinor 3). Then, genome-wide F_{ST} was estimated using the realSFS function in ANGSD between localities and then summarized across sliding-windows of 25Kb with a step of 5Kb.

Given that recent studies have documented the increasing importance of structural variants in genome diversity, and notably chromosomal inversions (Wellenreuther and Bernatchez, 2018), including in marine species (*e.g.* Berg et al., 2016; Cayuela et al., 2020), we scanned the genome for putative inversions or non-recombining haploblocks. To do so, we first ran localPCAs (Li and Ralph, 2019) with PCAangsd on genotype likelihoods in non-overlapping windows of 200 SNPs in each chromosome to extract local covariance matrices and obtained local PCAs of genomic variation (as detailed above). We then used the R package Lostruct (Li and Ralph, 2019) which measures the similarity between local PCA (PC1 and PC2 for each 200 SNP window) using Euclidean distances. Similarity was mapped using multidimensional scaling (MDS) of up to 20 axes. Clusters of outlier windows (presenting similar PCA patterns) were defined along each MDS axis as those with values beyond 4 standard deviations from the mean, following (Huang et al., 2020). Adjacent clusters with less than 20 windows between them were pooled, and clusters with less than 5 windows were not considered. Different window sizes (from100 to 1000), different subsets of PCs (1 to 3 PCs) and different thresholds yielded consistent results. A typical signature of a polymorphic inversion is three groups of individuals appearing on a PCA: the two homokaryotypes for the alternative arrangements and, as an intermediate group, the heterokaryotypes. All clusters of outlier windows were thus examined either by a PCA as single blocks, or divided into several blocks when discontinuous.

Environmental association/adaptation

Recent values (2000 - 2014) for 14 environmental variables were extracted as monthly averages from Bio-Oracle (Tyberghein et al., 2012), one of the most commonly used global environment datasets for marine species (data used were updated on September 3th 2021). Mean surface layer values were obtained for the following variables: temperature, salinity, current velocity, ice thickness, chlorophyll, dissolved oxygen, nitrate, phosphate, phytoplankton and silicate. Mean benthic layers were also collected for variables showing variation at maximum depth across locations (*i.e.*, current velocity, dissolved oxygen, temperature and salinity). Layers were imported into ArcGIS and values at each geographic coordinate corresponding to Greenland Halibut sampling

locations were extracted using the "Extract Values to Points" tool implemented in the program.

Two regression approaches were used to investigate GEA, utilizing matrices of dependent and independent (explanatory) variables. The explanatory matrix contained environmental variables and the dependent matrix contained genotypic data (here a matrix of minor allele frequencies by sampling locations). From the 28 locations containing more than 30 individuals, we excluded two locations that were sampled across two consecutive years at the same location (ESTU17 and ESQU17) as well as the SAGU locality, for which environmental data was not available in the Bio-Oracle database. We thus aimed to identify SNPs associated with both surface and benthic variables among Greenland Halibut from different sampling locations. First, Redundancy Analysis (RDA) was conducted as a multi-locus GEA method to detect loci under selection (Forester et al., 2018). The function "rda" was used to compute the RDA on the model for each collected environmental variable independently (see Laporte et al., 2016; Le Luyer et al., 2017 and Ferchaud et al., 2020 for examples of similar methodology). An analysis of variance (ANOVA; 1,000 permutations) was then performed to assess the global significance of the RDAs, and the percentage of variance explained (PVE) was computed with the function "RsquareAdj". When not mentioned, R functions were part of the VEGAN package (Oksanen et al., 2020). Only significant environmental variables were retained (Table S2) and correlation between them was tested with a Pearson correlation test. Correlated variables (with a Pearson coefficient > 0.7) were reduced by performing a PCA and retaining significant PCs following the Kraiser-Guttman and Broken Stick model (Borcard et al., 2011) (See Supplementary Material S3). Final redundancy analyses were then performed on each group of correlated variables and resumed by a PCA. SNPs linked to environmental variables were then defined following instructions from the online tutorial proposed by Brenna Forester (Forester et al., 2018; https://popgen.nescent.org/2018-03-27_RDA_GEA.html). We defined outliers as loci with loadings outside of 3 standard deviations from the mean (two-tailed p-value = 0.0027). The distribution of those outliers throughout the genome was explored by plotting the absolute value of the SNPs loadings on a Manhattan plot using gglplot2 library in R. For our second GEA method, we used Latent Factor Mixed Models (LFMMs), as implemented in the lfmm2 R package (Cayle et al., 2019) on the environmental variables defined as significant by the RDA analyses. The number of latent factors was set equal to K =2, corresponding to the elbow value in the PCA scree plots as evaluated from a PCA on a LD-pruned dataset. False discovery rate was assessed following the recommendations of François et al., (2016), using a Benjamini-Hochberg correction. The log10 of the p-value of SNPs was plotted on a Manhattan plot using the gglplot2 library on R. Finally, in order to assess how environmental conditions may modulate the extant of genetic differentiation between Greenland Halibut from the Gulf of Saint Lawrence and the Atlantic, a linear regression was performed between the number of SNPs harboring Fst values above 0.05 and the numbers of outliers detected both by RDA and LFMM (fdr= 0.01) approaches across the chromosomes, using the GLM function implemented in R with a binomial variance and logit link.

Migration estimations

The pattern of migration between the two observed populations was assessed using fastsimcoal2 (Excoffier et al., 2021). Fastsimcoal2 is a continuous time coalescence-based genetic simulation program that enables the estimation of demographic parameters under flexible scenarios from the site frequency spectrum (SFS) under a maximum-likelihood approach. Site frequency spectrum is defined as the sampling distribution of allele frequency at any random polymorphic locus in the genome and is commonly used for inferring population genetic parameters (Chen et al., 2007; Chen, 2012). SFS-based methods have received considerable great interest since the emergence of next generation sequencing, particularly due to computing time of these methods being independent of the length of the genome being analyzed (Gutenkunst et al., 2009; Chen, 2012). Moreover, several approaches of fitting the SFS using exact derivations or approximations have been developed (Gutenkunst et al., 2009; Excoffier et al., 2013, Liu and Fu, 2020) and it has been shown that the expected SFS could be robustly estimated using coalescent simulations (Excoffier et al., 2013). The statistics summarized in the SFS can also be considered for multiple populations and corresponds to the joint distribution of allele frequencies across di-allelic variants (joint SFS). SFS and joint SFS can be predicted from low-coverage data notably through the use of the genotype likelihood approach implemented in ANSGD (Nielsen et al., 2011).

Site allele frequency was first estimated for each of the two populations (Gulf of Saint Lawrence and the Atlantic) using -dosaf 1 in ANGSD with the GATK model (-GL 2), and the command -anc referring to the Greenland Halibut reference genome (Ferchaud et al., 2022). No minor allele frequency filter was applied since variants present in low frequency could be highly informative on demographic history. We kept only SNPs covered by at least one read in at least 80 % of individuals (-minInd) and removed SNPs in putative repeated regions, allowing a maximum depth of 3 times the number of individuals (-setMaxDepth) to filter putative genotyping error and paralogous. The realSFS function was then performed to get an unfolded joint SFS. This joint SFS was projected in the δaδi program (version 1.6.3 Gutenkunst et al., 2009) to remove missing data before being converted to a fastsimcoal2 format using a custom script. Estimating demographic parameters is highly time consuming for large sample sizes, therefore each population was randomly reduced to 100 individuals. The resulted observed joint SFS was then fed into fastsimcoals2 while an expected joint SFS was estimated under a model of

two populations (with constant effective size) connected by gene flow. While recent studies have pointed out that failure to include complex parameters in extensive simulations can lead to biases in parameter estimations (*e.g.* Momigliano et al., 2021), the objective of this analysis was to get relative migration rates between the two groups rather than to get a comprehensive picture of the demographic history and we decided to use a simpler model. Effective size of the two populations (NPOP1 and NPOP2 uniformly distributed between 100 and 100 000) and migration rates between them (MIG21 and MIG12) were the parameters of the model. This model was replicated 100 times, simulating 200,000 coalescents (-n 200000), and containing at least 10 observed SFS entry counts (-C 10) over 40 optimization cycles, to estimate demographic parameters by maximum composite likelihood (-M) using the –nosingleton option to exclude genotyping error. Point estimates of the different demographic parameters were selected from the runs with the highest maximum composite likelihood.

Finally, we calculated interval confidences of parameter estimates from 10 parametric bootstrapped join SFS, and re-estimating parameters each time. The outputs of the 10 independent fastsimcoal runs from the 10 bootstrapped Join SFS were individually examined before being pooled together. For each run, the difference between the maximum likelihood estimated from the model and the maximum likelihood observed from the join SFS was calculated and only runs expressing the lowest difference (*i.e.*, in the plateau with the lowest values) were retained. Estimated parameters of the retained runs were pooled together over the 10 independent runs to get confidence interval.

Results

After mapping the reads of 1,366 Greenland Halibut samples to the reference genome, cleaning for quality and processing for SNP identification with genotype likelihoods, we identified 5,347,751 single nucleotide polymorphisms (SNPs) with a minor allelic frequency (MAF) above 5 % across the 1,297 individuals with adequate coverage (see Material and Method section).

Differentiation between Gulf of Saint Lawrence and Northwest Atlantic

Genome-wide variation analyzed by a global PCA displayed two clusters, with PC1 explaining 0.86 % of the total genetic variation (Figure S1A). Those two clusters were respectively composed of the individuals from the Gulf and those inhabiting the rest of the Northwest Atlantic. Other PC axes did not reveal additional clusters. In particular, this suggests that there is no genetic differentiation between fish from the Canadian and Greenlandic waters, suggesting that they form a single panmictic population (Figure S1A). However, 18 individuals that were not correctly assigned to their respective geographic clusters were detected. It is suspected that these individuals could be migrants between the Gulf and the Northwest Atlantic (Figure S1B). When the PCA was re-run without the migrants, it displayed the same two distinct groups composed of individuals from the Gulf on one side and individuals from the rest of the Northwest Atlantic on the other side, with PC1 explaining 0.87 % of the genetic variation (Figure 1B).

Pairwise genome-wide F_{ST} estimations performed across sampling localities corroborated this clustering result. The distribution of mean estimates revealed two non-overlapping modes of F_{ST} values corresponding to (i) mean pairwise F_{ST} estimates among sampling locations within each group and (ii) mean estimates among sampling locations between the two groups (Figure 1C). Pairwise genome-wide F_{ST} between Gulf and Northwest Atlantic exhibited higher and non-overlapping values (median = 0.0083, range: 0.0080-0.0087) than estimates within the Gulf (median = 0.0072, range 0.0070-0.0075) or estimates within Atlantic (median = 0.0072, range: 0.0067-0.0075). No evidence of isolation by distance was detected, either between Gulf and Northwest Atlantic (r^2 adj = 0.006369, p-value = 0.1575) or within the Gulf (r^2 adj = 0.03719, p-value = 0.1648), although a very weak but not significant relationship within the Northwest Atlantic cluster was found between geographical distance and genetic differentiation (r^2 adj = 0.009861, p-value = 0.09122, Figure S2).

Genome-wide F_{ST} estimations between the Gulf and the Atlantic revealed low levels of genetic differentiation along the genome (mean F_{ST} = 0.0018, range:0.00023-0.32). This mean global F_{ST} between the two groups could appear low compared to the mean estimates among sampling localities between the two groups (median Fst = 0.0083). In Supplementary Material S2, we show that this effect is attributed to the sample sizes used: 30 *versus* 30 individuals when estimating pairwise localities compared to 303 *versus* 303 individuals when estimating F_{ST} between the two groups. However, several peaks of differentiation were found along the genome (Figure 2). In particular, three sliding-window peaks exhibited F_{ST} values above 0.05, respectively in Chr03 ($maxF_{ST}$ = 0.25), Chr05 ($maxF_{ST}$ = 0.054) and in Chr23 ($maxF_{ST}$ = 0.063). Figure S3 displays the SNP F_{ST} values for 200 Kbp windows surrounding these peaks as well as other regions revealing highest Fst values (Fst_{Chr03} = 0.30, Fst_{Chr05} = 0.18, Fst_{Chr07} = 0.17, Fst_{Chr12} = 0.18, Fst_{Chr15} = 0.15 and Fst_{Chr23} = 0.32).

We retrieved 25,000 pb flanking regions upstream and downstream of the genomic location of the three SNPs exhibiting high Fst, based on the reference genome, to search for coding regions. The search was conducted using the database Nucleotide collection (nr/nt) using Megablast on the NCBI Blast + platform (Camacho et al., 2009). In Chr03, a match with a transcript of the green-sensitive opsin gene was found (99.51 %

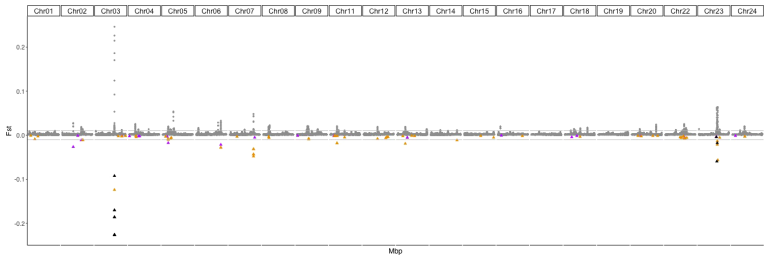


FIGURE 2
Genome-wide Fst differentiation estimated between fish from the Gulf of Saint Lawrence and the Northwest Atlantic in sliding windows of 25 Kb within each chromosome (sex-linked chromosomes, Chr 10 and Chr21 (Ferchaud et al., 2022) have been removed. Horizontal grey lines indicate Fst values of 0.01 and -0.01. Colored triangles correspond to outliers found in association with environmental variables by the RDA and LFMM (fdr = 0.0001) approaches. For clarity, Fst values of those SNPs have been reported in negative values as triangles. Purple triangles correspond to sliding windows containing SNPs found in association with bottom dissolved oxygen, those in yellow correspond to the sliding windows presenting SNPs associated with temperature-related variables. Black triangles identify sliding windows containing SNPs putatively in association with both the bottom dissolved oxygen and temperature-related variables and non-outliers in dark grey dots.

and 99.67 % of identity with Atlantic Halibut (*Hippoglossus hippoglossus*) and Pacific Halibut (*Hippoglossus stenolepsis*) transcripts respectively), in Chr05 the best hit was with the RIC8 guanine nucleotide exchange factor B transcript (94.99 % of identity with Atlantic and Pacific Halibut transcripts), in Chr07 the putative translation initiation factor eIF4E (81.75 % identity with the turbot, *Scophtalmus maximus*), the zinc finger protein 365 transcript (93.69 % of identity with Pacific Halibut) and the guanylate cyclase activator 1d (guca1d, 92.91 % of identity with Atlantic Halibut) were respectively the best hit for Chr12 and Chr15 and the transcript of the 24-dehydrocholesterol reductase (dhcr24) matched with the highest peak of differentiation in Chr23 (with 92.65 % of identity with the Japanese flounder (*Paralichthys olivaceus*) transcript).

Lostruct analysis detected five outlier genomic windows along the genome, respectively on Chr05, Chr06, Chr13, Chr19 and Chr23 displaying three distinct clusters in each that could represent the signature of an inversion (Figure S4). However, there was no variation in the geographic distribution of the different karyotypes for any of those putative structural variants, suggesting that there are not involved in the differentiation between fish from the Gulf and those from the rest of the Northwest Atlantic.

Genotype-environment association

Out of 14 environmental variables tested, seven were found to be significantly associated with genomic variation in Greenland Halibut using the RDA; two variables at maximum depth (bottom temperature and bottom dissolved oxygen) and five environmental variables at the surface of the sea (surface temperature, surface salinity, surface chlorophyll, surface phytoplankton and surface nitrate) Table S2. Pairwise Pearson correlation tests conducted among those seven variables revealed that six of them were highly correlated ($r > 0.7$, Supplementary Material S3). Only bottom dissolved oxygen was not highly correlated to the other variables (mean coefficient correlation with other variables = 0.45). Variation brought by the 6 related variables, hereafter denoted as "temperature-related" variables, was resumed by the first axis of a PCA (explaining more than 77% of the variation, see Supplementary Material S3) and relative loadings of each locality were then used to perform the final RDA. An RDA performed with the bottom dissolved oxygen was also used to detect outliers specifically associated with this variable. A total of 2,097 SNPs was detected as putative temperature-related outliers, whereas 1,851 SNPs were detected as putative oxygen-related outliers. For both variables, putative outliers were spread all along the genome (Figure S5). SNPs in association with environmental variables were also detected by LFMM. While considering a very conservative false discovery rate of 0.0001 in order to exclude putative false positives, a total of 95 SNPs was detected to be associated with temperature-related variables and 41 were associated to the bottom dissolved oxygen. All of those putative oxygen-related outlier SNPs were also detected by the RDA approach, while 94/95 of the putative temperature-related were in common with RDA outliers. With a less conservative false discovery rate (fdr = 0.01), 537 SNPS were detected as putatively associated with temperature-related variables and 282 SNPs putatively associated with the bottom dissolved oxygen and more than 92 % of them (respectively, 497/537 and 262/282) were confirmed by the RDA. As for the RDA analysis, outliers detected by LFMM were widely distributed throughout the genome but they were more concentrated in Chr 3 and Chr 23 (Figure 3). Moreover, while most of the outliers detected by both GEA analyses were unique to a single environmental variable, 15 of them were shared between the temperature and the bottom dissolved oxygen. All of those 15 putative outliers were located within the two most highly

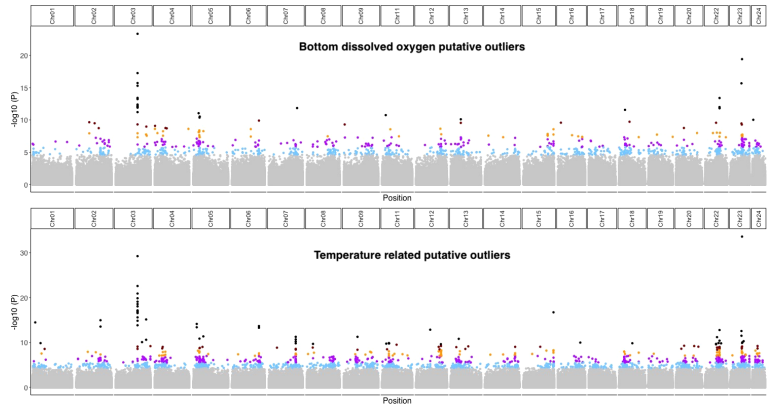


FIGURE 3
Genome-wide Environmental Association detected by LFMM 2 with bottom dissolved oxygen (upper panel) and temperatures related variables (lower panel). The Manhattan plot show the -log10 of the p-value provided by LFMM2 for each SNP. Dots are colored according to False Discovery Rate (black:< 0.00001, brown:<0.0001, orange:< 0.001, purple:< 0.01, light blue:< 0.05 and grey: no outliers).

differentiated genomic regions between the Gulf and the Northwest Atlantic (Chr 03: 17,709,400 bp and Chr23: 9,606,635 bp, Figure 3). 80 % of them (12 out of 15) were located within a 21,000 bp region of Chr03 (Chr03: 17,709,400 bp, Figure 3) and the remaining SNPs (3 out of 15) were located in Chr23 (Chr23: 9,606,635 bp). Finally, a significant positive correlation was found across the chromosomes between the number of outliers associated with environment variables and the number of SNPs with high Fst values (Figure 4, r^2_{adj} = 0.51, p-value = 1.27 10^{-4}).

Migration between the Gulf and the Northwest Atlantic

Migration rate (proportion of migrants per generation) from the Northwest Atlantic to the Gulf (m_{Atl_Gulf} =0.0022 [0.0020; 0.0024]) was significantly higher than the migration rate from the Gulf to the Northwest Atlantic (m_{Gulf_Atl}=0.0017 [0.0015; 0.0019], t= -2.629, p-value = 0.008). According to the following estimated effective size, N_{Gulf} = 47,517 [45,083; 49,951] and N_{Atl} = 57,757 [55,135; 60,378], significantly more migrants per generation were estimated to move from the Northwest Atlantic to the Gulf (63 [54; 72]) than in the other direction (41 [34; 48]). Empirical numbers of migrants detected by clustering analyses also report an asymmetric gene flow between the two populations. Seven migrants were observed in the Northwest Atlantic (sample size =983), whereas 7 migrants were detected in the Gulf (sample size = 314), leading to a higher proportion of migrants detected from Northwest Atlantic to the Gulf (10/314 = 0.03), than from the Gulf to the Northwest Atlantic (7/983 = 0.007).

Discussion

In this study, we conducted the first genome-wide investigation of genomic variation in Greenland Halibut, sequencing the whole genome of 1,297 individuals distributed throughout the Northwest Atlantic. This high genomic resolution allowed us to statistically confirm the previously supposed specificity of the Gulf stock and to determine that the divergence is mainly driven by environmental variables such as sea temperature, salinity and oxygen levels. As a result, the Gulf stock could be particularly vulnerable to the environmental changes that are expected in the context of a warming climate. Altogether, our results revealed a pattern of spatial selection in a marine fish species with high dispersal potential that will help support the management of this important exploited species.

Significant divergence between two stocks

Our analysis of population structure identified a weak but significant divergence between Greenland Halibut from the Gulf and those from the Northwest Atlantic. The small proportion of genomic variation explained (0.87 %), and low differentiation observed between the two stocks (mean F_{ST} = 0.0018), are both consistent with the weak structuration identified in other highly dispersed marine species (Jiménez-Mena et al., 2020; Le Moan et al., 2021) and often referred to as 'cryptic structuration' (Benestan et al., 2015; Le Moan et al., 2021). For example, Kess et al. (2021) identified a similar structure in Atlantic Halibut (*Hippoglossus hippoglossus*), based on 86,000 SNPs defined by restriction-site associated DNA sequencing, and estimated a lower genomic

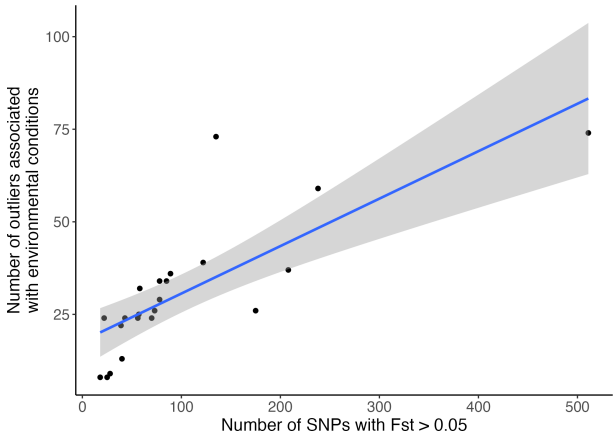


FIGURE 4
Relationship between the number of SNPs with Fst estimates higher than 0.05 and the number of SNPs detected in putative association with both bottom dissolved oxygen and temperature-related variables for each chromosome (adjusted r-squared = 0.51, p-value = 1.26e-04).

differentiation between the Gulf and adjacent areas on the continental shelf than we found for the Greenland Halibut (Atlantic Halibut; meanF_{ST} = 0.0005, minF_{ST} = 0, maxF_{ST} = 0.068 *versus* Greenland Halibut; meanF_{ST} = 0.0018, minF_{ST} = 0.00023, maxF_{ST} =0.32). This difference between the two species could be attributed to intrinsic biological features (notably, Atlantic Halibut spawn in nearshore waters along the west coast of Newfoundland (*e.g.* Belle-Isle Strait) that is geographically intermediate between the Northwest Atlantic and the Gulf (Le Bris et al., 2018), while the main Greenland Halibut spawning area is presumed to be in the Davis Strait (Gundersen et al., 2010). Alternatively, the two different genomic approaches to cover the genomic variation could explain the difference in observed Fst values between the two studies. The whole-genome screening we used enabled finding all the SNPs that were differentiated, including in the very restricted genomic regions explaining most of the differentiation between the two groups (see below for the discussion about those differentiation peaks). It is likely that a restricted approach like the one used for the Atlantic Halibut would miss such narrow regions, potentially leading to a lower average differentiation. Moreover, the high sample size used in our Fst estimations (303 individuals for each population) reduces the stochasticity that comes with to low sampling sizes and this could also lead to more power to find the differentiation peaks (Supplementary Material S2). Previous studies conducted on Greenland Halibut were based either on allozymes (Fairbairn, 1981; Riget et al., 1992), 12 microsatellites (Roy et al., 2014), or 96 SNPs (Westgaard et al., 2017). They reported an absence of structure and even claimed panmixia throughout the Northwest Atlantic (Roy et al., 2014). However, the reversal in relative frequency of the two most abundant mitochondrial haplotypes highlighted by Vis et al., (1997) did hint at a genetic difference between the Gulf of Saint Lawrence and the Atlantic. Recently, Carrier et al. (2020) used a Genotyping-by-Sequencing approach to reveal significant genetic differences between Greenland Halibut living in the Gulf and two offshore locations in Newfoundland. In this study, we have elucidated the genomic structure for Greenland Halibut throughout the Northwest Atlantic by covering the entire genomic variation across 1,297 individuals. In addition to confirming the significant specificity of the Gulf, we also observed an absence of distinction between Greenland Halibut from the Canadian and Greenland waters. This result suggests extensive mixing along the slope of the entire North Atlantic continental shelf where Greenland Halibut occurs. In addition to analyses of spatial distribution of settling larvae showing that larvae from geographical separate spawning areas could intermix (Stenberg et al., 2016), tagging studies indicate that mixing occurs between stock areas at juvenile and adult life-stages (Vihtakari et al., 2022). Our study fills gaps related to lack of knowledge about stock structure in Greenland Halibut and allows for the first time to conclude that Greenland Halibut occurring from NAFO Subareas 0 to 3 could be managed as a single unit.

Moreover, the absence of Isolation-By-Distance observed throughout the Gulf and the Northwest Atlantic, confirms the panmictic situation of Greenland Halibut in the Northwest Atlantic and is consistent with what we know about Greenland Halibut migratory behavior, in that the spatial scale within each stock is within the dispersal capabilities of this species. Indeed, Greenland Halibut of the Northwest Atlantic are highly migratory; fish tagged and released in Davis Strait, Baffin Bay, and the fjords of southwestern and eastern Greenland have moved south and been caught on the northern slopes of the Grand Bank of Newfoundland and as far east as Denmark Strait (between Greenland and Iceland) (Boje, 2002; Vihtakari et al., 2022).

Environment association

Despite high effective population sizes and connectivity between the two clusters, our results revealed a significant genetic differentiation between the Gulf and the Northwest Atlantic. This observation suggests that natural selection, in addition to genetic drift, might have driven contemporary Greenland Halibut population structure in the Northwest Atlantic. Our results revealed that environmental variables explain 51 % of the differentiation between the two populations. Specifically, we found that both temperature-related variables (bottom temperature, surface temperature, surface salinity, surface nitrate and primary productivity, estimated with phytoplankton and chlorophyll) and bottom dissolved oxygen represent key selective agents that appear to drive adaptive divergence between the Gulf and the Northwest Atlantic. Marine landscape genomic studies combining traditional landscape approaches with large genomic datasets have identified significant associations between environmental variables and genetic structure in other marine species such as Atlantic Herring (*Clupea harengus*) (Limborg et al., 2012), Atlantic cod (*Gadhus morhua*) (Berg et al., 2015), eastern oyster (*Crassostrea virginica*), purple sea urchin (*Strongylocentrotus purpuratus*) (Pespeni et al., 2013), sea scallop (*Placepecten magellanicus*) (Van Wyngaarden et al., 2018), sea cucumber (*Parastichopus* californicus) (Xuereb et al., 2018), European hake (*Merluccius merluccius*) (Milano et al., 2014), American Lobster (*Homarus americanus*) (Benestan et al., 2015; Dorant et al., 2020) and flatfish species such as Sole (*Solea solea*) (Diopere et al., 2018) and Atlantic Halibut (*Hippoglossus hippoglossus*) (Kess et al., 2021). In the Northwest Atlantic, based on a climate-associated multispecies cryptic cline, Stanley et al. (2018) revealed a biogeographic break along with a steep climatic gradient driven by temperature.

Strong selection acting on individual locus

Outliers in association with environmental variables have been discovered all along the genome and it is noteworthy that every high peak of differentiation between the two populations is associated with environmental variation (Figure 2 and Figure 4). Each of the two highest peaks (located in Chr03 (Fst = 0.30) and Chr23 (Fst = 0.32)) harbors SNP outliers associated with both the temperature-related variables and the bottom dissolved oxygen. The striking differences in allele frequencies between the two identified populations suggests that these restricted genomic regions as well as the other highly differentiated regions (like the ones found in Chr5, 7, 12 and 15) might be under strong selective pressure (Figure 2 and Figure S3). For these regions, and particularly for Chr12 (Figure S3), a single SNP is responsible for the large changes in allele frequencies among locations from the Gulf and those within the Northwest Atlantic, suggesting that an individual locus is experiencing strong selection.

Are Greenland Halibut adapting to climate change?

Our Genotype-Environment Association analysis revealed that both surface and bottom environmental variables are associated with genomic variation in Greenland Halibut. Although Greenland Halibut is primarily considered a deep-water fish, it uses both pelagic and bottom environments (Boje et al., 2014). For example, Vollen and Albert (2008) and Albert et al. (2011) documented extensive vertical activity of Greenland Halibut along the continental slope in Europe. It was found that individuals spend up to one fourth of their lives in the pelagic environment. This time is divided between the pelagic larvae phase and what is thought to be foraging on pelagic prey (Nygaard & Boje, 2011), thus exposing the fish to every layer of the water column and their associated environmental conditions. Warming ocean temperatures, both at the surface and in the deep waters, were at record levels in the Gulf of Saint Lawrence in 2020 (Galbraith et al., 2021). Climate data revealed that overall warming was detected in the Gulf in 2021. The observed surface warming falls in line with global warming data (about one degree per 100 years) but in the deep layers an increase of one-and-a-half Celsius degrees was measured over a period of 12 years (Galbraith et al., 2021), clearly outpacing global climate models. The last year for which below-normal temperatures at 300 m were observed was 2009. Since then, these temperatures have risen steadily. Given that deep water from the continental shelf flowing and diffusing into the Gulf of Saint Lawrence is also very warm, the situation is likely to continue (Neto et al., 2021). That flow is a mix of the deep-reaching Gulf Stream current moving north and the Labrador current moving south (Claret et al., 2018; Neto et al., 2021). Lately, the Gulf Stream has become dominant over the Labrador current (Neto et al., 2021), and while we don't know if this is a permanent, cooling events are not expected in this area (Galbraith et al., 2021). With the Gulf Stream (oxygen-poor tropical and sub-tropical water) becoming more important than the Labrador current (well-oxygenated sub-arctic water) important environmental changes could be coming to the Gulf ecosystem. Oxygen levels in the Gulf of Saint Lawrence are rapidly decreasing, even outpacing the global and the North Atlantic basin average deoxygenation rates (Claret et al., 2018). Our study revealed that genomic variation in Greenland Halibut is associated with dissolved oxygen at deep layers of the water column. Interestingly, we found that the highest differentiation expressed between the Gulf and the Northwest Atlantic, located in Chr23, matches with the Delta (24)-sterol reductase gene known to be involved in the response to oxidative stress.

Greenland Halibut inhabiting waters of the Saint Lawrence system might already suffer from hypoxia and allele frequency shift observed at this genomic region might be seen as an adaptive response to those conditions. Youcef et al. (2013) found a strong association between high Greenland Halibut densities and low dissolved oxygen concentration in the Gulf, suggesting that this species could already be highly tolerant to hypoxic conditions. They report that the negative effects of hypoxia, if present, could be compensated by other factors such as food availability and/or refuge from predation. However, Dupont-Prinet et al. (2013) revealed that juvenile Greenland Halibut are less tolerant to hypoxia than adults, notably due to a slower digestion process. The authors concluded that juveniles from the Saint Lawrence system were living at the edge of their metabolic capacities and that growth and distribution could be affected if further declines in dissolved oxygen occurred. The recent observed reduction in oxygen availability could make Greenland Halibut less tolerant to hypoxia. In addition to warm and poorly oxygenated waters, Greenland Halibut also has to cope with the recent nitrate increase in the Gulf of Saint Lawrence (Blais et al., 2021; Lavoie et al., 2021). Our analysis indeed revealed a genome-wide association with nitrate in surface waters, suggesting that Greenland Halibut may be responding to this observed increase. Nitrate occurs naturally at low concentrations in waterways but human activities, such as extensive use of agricultural fertilizers, the consumption of fossil fuels and increasing urban pressures, have increased nitrate concentrations to records levels in many places (Camargo et al., 2005; Claret et al., 2018), including in the Saint Lawrence system, where nitrate inventories in the surface water in recent years have been slightly above normal (Blais et al., 2021). Excess nitrate can trigger eutrophication events, where the growth of aquatic plants and cyanobacteria is spurred, subsequently stripping the water column of oxygen. Ultimately, fish exposed to nitrate increases become more susceptible to hypoxia (Gomez-Isaza et al., 2021). Finally, the genomic association found with productivity (chlorophyll and phytoplankton) may be mainly attributed to the high productivity, notably the phytoplankton spring bloom, occurring throughout the Saint Lawrence system (Lavoie et al., 2021).

The genotype-environment associations highlighted by our results and particularly the association between environmental variables and specific narrow genomic regions suggest that natural selection is acting and that Greenland Halibut are undergoing adaptation to their changing environment. For example, environmental association with the Green sensitive opsin gene located in the second highest peak of differentiation between the Gulf and the Northwest Atlantic (Chr03) was revealed by our study. This gene has already been shown to be associated with somatic growth in the Barfin Flounder (*Verasper moseri*) (Kasagi et al., 2015). In addition, on the third highest peak of differentiation (Chr05) we found an association with the Guanine nucleotide exchange factor B (RIC8), known to be involved in gastrulation and in embryonic development (Tonissoo et al., 2010). Moreover, other putative candidate genes involved in the differentiation between the Gulf and the Atlantic support this scenario of ongoing adaption to environmental change. In Chr07, the transcript of the translation initiation factor eIF4E, known to regulate stress homeostasis and modulate host invasion (Batool et al., 2021), seems under selection. The detected Zinc finger protein 365 in Chr12 has been recently defined as a new maternal LPS-binding protein that defends zebrafish embryos against gram-negative bacterial infections (Du et al., 2018) and, in Chr15, the detected guanylate cyclase activator 1d has been shown to be involved into the regulation of the photoreceptor GC in the teleost's retina (Imanishi et al., 2004). Altogether, our results indicate that previous phenotypic differences observed between the Gulf and the Northwest Atlantic in terms of parasitism (Arthur and Albert, 1993), growth rate (Templeman, 1973), fecundity (Morgan et al., 2003) and physiology (Khan et al., 1982) may all be part of the signature of a functional adaptive divergence in Greenland Halibut that is mainly driven by the stronger climate changes observed in the Gulf of Saint Lawrence than elsewhere is the North Atlantic.

Another coping mechanism – escape from the Gulf of Saint Lawrence

If regional warming and deoxygenation continue in the Gulf of Saint Lawrence, one can anticipate that adaptation would not be sufficient for Greenland Halibut to cope with this climate change (Pershing et al., 2015). Greenland Halibut is not only a species with a narrow optimum temperature, but it is the Gulf's species which prefers the coolest temperatures (1°C to 5°C) along with the Northern shrimp (*Pandalus borealis*) (Brennan et al., 2016). In a 50-year prediction for the decade 2060-2070, Chabot et al. (2013) concluded that Greenland Halibut could potentially disappear from the Gulf with warmer waters negatively affecting productivity, abundance and distribution. The Greenland Halibut is also in competition for its food with the redfish (*Sebastes fasciatus* or *S. mentella*), for which warming waters are beneficial (DFO, 2021). Recently, it has been reported that Greenland Halibut cohorts that were abundant in 2012-2013 grew slower and took more time to reach their commercial size than was expected in 2018-2019. It is difficult to determine if this was the consequence of warming and poorly oxygenated waters, competition with redfish or other factors, but indices of abundance and biomass of fish above 40 cm are decreasing (DFO, 2021). Another point brought by our study is the connectivity detected between the Gulf and the Northwest Atlantic. The asymmetrical connectivity detected by our

analysis is congruent with the main nursery located in the Saint Lawrence Estuary (Ait Youcef et al., 2013 and Bassi et al. submitted) and GBS results revealing the contribution of Newfoundland stocks to this nursery (Carrier et al., 2020). It has been hypothesized that fish from the Atlantic could potentially drift during their prolonged pelagic larval phase and enter into the Saint Lawrence system *via* Belle-Isle strait, where there is a cold-water current oriented toward the Gulf coming from Labrador (Carrier et al., 2020). The proportion of larvae entering the Gulf would thus depend on the strength of this current. Considering the ongoing ocean circulation shift mentioned above (Claret et al., 2018), we can hypothesize that the asymmetrical migration detected by our study could potentially reverse and that emigration from the Gulf would become more important than immigration from the Atlantic. This could be seen as a passive response to the current dynamic, but it could be also an active strategy of Greenland Halibut to cope with environmental conditions becoming unfavorable in the Gulf of Saint Lawrence.

Conclusion

The high genomic resolution used in our study allowed to confirm the previously supposed specificity of the Gulf of Saint Laurent and the very high connectivity within the remainder of the Northwest Atlantic, suggesting that except for the Gulf of Saint Lawrence, Greenland Halibut appear to be panmictic throughout the Northwest Atlantic. Environment association analyses revealed that divergence between the two stocks is mainly attributed to environmental variables such as sea temperature and dissolved oxygen and indicated that phenotypic differences previously observed between the Gulf of Saint Lawrence and the Northwest Atlantic likely resulted from functional adaptive divergence to their respective environmental conditions. Altogether our results suggest that the Gulf stock could be particularly vulnerable to the environmental changes that are expected in the context of a warming climate. Moreover, the high levels of migration assessed between the two stocks would allow Greenland Halibut to potentially escape unfavorable environmental conditions in the Gulf of Saint Lawrence. In addition to supporting the management of this important exploited species, this work highlights the utility of using comprehensive genomic datasets to characterize the effects of climate change across a wider range of species.

Data availability statement

The data presented in the study are deposited in the SRA NCBI repository, accession numbers are provided in Supplementary Material S4.

Ethics statement

Ethical review and approval were not required for the animal study because samples were collected during commercial fisheries.

Author contributions

LB and CA conceived the study. KP, JM, MT, RH and WW designed the sample collection. ALF and CB performed the DNA extraction and lab manipulation. ALF and EN analyzed the data. ALF led the writing. All authors contributed to the article and approved the submitted version.

Funding

This project was funded by a National Sciences and Engineering Research Council of Canada (NSERC strategic partnership program) grant to LB, CA, PS, and Réjean Tremblay, as well as by the Canadian Research Chair in Genomics and Conservation of Aquatic Resources, and Ressources Aquatiques Québec (RAQ).

Acknowledgments

We thank all biologists and technicians of Fisheries and Oceans Canada and the Greenland Institute of Natural Resources for collecting the samples. We also thank Cécilia Hernandez for help gathering all the samples in the lab and Justine Létourneau for help in DNA extraction. Brian Boyle, Claire Mérot and Nina Therkildsen provided key advice about sequencing and libraries preparation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2022.992504/full#supplementary-material

SUPPLEMENTARY FIGURE 1
(A). The four first axes of the principal component analysis performed with all the 1,297 Greenland Halibut individuals. **(B).** Assignment of each of the 1,297 Greenland Halibut to the two clusters using NGSadmix.

SUPPLEMENTARY FIGURE 2
Isolation-By-Distance among the Greenland Halibut localities sampled throughout the Northwest Atlantic. Relationship between the logarithm of the geographic distance and the corrected genetic differentiation for all the pairwise sampled localities.

SUPPLEMENTARY FIGURE 3
SNP Fst values between the Gulf of Saint Lawrence and the Atlantic in the 100 Kbp flanking regions around the highest Fst peaks found in the genome.

SUPPLEMENTARY FIGURE 4
Local PCA conducted for each of the outlier window detected by the Lostruct analysis. Despite the detection of three clusters within these windows, no geographical structuration is distinguishable. Therefore, those putative inversions are unlikely to be involved in the Gulf-Northwest Atlantic differentiation.

SUPPLEMENTARY FIGURE 5
Genome-wide RDA absolute loading values for bottom dissolved oxygen and temperature-related variables found to be significantly associated with allele frequencies. Black dots correspond to SNPs that are significantly associated with the environmental variables, while grey dots indicate non-outliers SNPs.

SUPPLEMENTARY TABLE 1
Mean pairwise Fst estimates among each locality containing more than 30 individuals.

SUPPLEMENTARY TABLE 2
Results of the Redundancy of Analysis.

References

Ait Youcef, W., Lambert, Y., and Audet, C. (2013). Spatial distribution of Greenland halibut *Reinhardtius hippoglossoides* in relation to abundance and hypoxia in the estuary and gulf of St. *Lawrence Fish. Oceanogr.* 22, 41–60. doi: 10.1111/fog.12004

Albert, O. T., Lambert, Y., Vollen, T., Freitas, C., and Heggenbakken, L. (2011). "Distinguishing pelagic and demersal swimming of deepwater flatfish by recording of body angles," in *Advances in fish tagging and marking technology*. Eds. J. McKenzie, B. Parsons, A. Seitz, R. K. Kopf, M. Mesa and Q. Phelps (Bethesda, MD: American Fisheries Society), 507–528.

Albert, O. T., and Vollen, T. (2015). A major nursery area around the Svalbard archipelago provides recruits for the stocks in both Greenland halibut management areas in the northeast Atlantic. *ICES J. Mar. Sci.* 72, 872–879. doi: 10.1093/icesjms/fsu191

Aljanabi, S. M., and Martinez, I. (1997). Universal and rapid salt-extraction of high-quality genomic DNA for PCR-based techniques. *Nucleid Acids Res.* 25 (22), 4692–4693. doi: 10.1093/nar/25.22.4692

Allendorf, F. W., Luikart, G. H., and Aitken, S. N. (2012). *Conservation and the genetics of populations (2nd Ed)* (West Sussex: U. K. Wiley-Blackwell).

Arthur, J., and Albert, E. (1993). Use of parasites for separating stocks of Greenland halibut (*Reinhardtius hippoglossoides*) in the Canadian Northwest Atlantic. *Can. J. Fish. Aquat. Sci.* 50, 2175–2181. doi: 10.1139/f93-243

Babin, C., Gagnaire, P.-A., Pavey, S. A., and Bernatchez, L. (2017). RAD-seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American eel (*Anguilla rostrata*). *Genome Biol. Evol.* 9 (11), 2974–2986. doi: 10.1093/gbe/evx226

Barkley, A. N., Fisk, A. T., Hedges, K. J., Treble, M. A., and Hussey, N. E. (2018). Transient movements of a deep-water flatfish in coastal waters: Implications of inshore-offshore connectivity for fisheries management. *J. Appl. Ecol.* 55, 1071–1081. doi: 10.1111/1365-2664.13079

Bassi, L., Tremblay, R., Robert, D., Bernatchez, L., Ferchaud, A.-L., and Sirois, P. *Submitted Connectivity and natal sources of Greenland halibut in the gulf of st. Lawrence inferred from otolith chemistry* (Canadian Journal of Fisheries Aquatic Science).

Batool, W., Shabbir, A., Lin, L., Chen, X., An, Q., He, X., et al. (2021). Translation initiation factor eIF4E positively modulates conidiogenesis, appressorium formation, host invasion and stress homeostasis in the filamentous fungi *Magnaporthe oryzae*. *Front. Plant Sci.* 12, 646343. doi: 10.3389/fpls.2021.646343

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., Kishony, R., et al. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS One*, 1–15. doi: 10.1371/journal.pone.0128036

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., Bernatchez, L., et al. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Mol. Ecol.* 24 (13), 3299–3315. doi: 10.1111/mec.13245

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., et al. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol. Ecol.* 25, 5073–5092. doi: 10.1111/mec.13811

Benestan, L., Moore, J.-S., Sutherland, B. J.G., Le Luyer, J., Maaroufi, H., Clément, R., et al. (2017). Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Mol. Ecol.* 26, 6767–6783. doi: 10.1111/mec.14217

Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., et al. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* l.). *Genome Biol. Evol.* 7 (6), 1644–1663. doi: 10.1093/gbe/evv093

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M.I., Knutsen, H., et al. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci. Rep.* 6, 23246. doi: 10.1038/srep23246

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *J. Fish Biol.* 89, 2519–2556. doi: 10.1111/jfb.13145

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32 (9), 665–680. doi: 10.1016/j.tree.2017.06.010

Blais, M., Galbraith, P. S., Plourde, S., Devred, E., Clay, S., Lehoux, C., et al. (2021). *Chemical and biological oceanographic conditions in the estuary and gulf of saint Lawrence during 2020* (Canadian Science Advisory Secretariat (CSAS). Research Document 2021/060: Fisheries and Ocean Canada).

Boje, J. (2002). Intermingling and seasonal migrations of Greenland halibut (*Reinhardtius hippoglossoides*) populations determined from tagging studies. *Fish Bull.* 100, 414–422. http://hdl.handle.net/1834/31074

Boje, J., Neuenfeldt, S., Sparrevohn, C. R., Eigaard, O., and Behrens, J. W. (2014). Seasonal migration, vertical activity, and winter temperature experience of Greenland halibut *Reinhardtius hippoglossoides* in West Greenland waters. *Mar. Ecol. Prog. Ser.* 508, 211–222. doi: 10.3354/meps10874

Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R. Springer Science & Business Media*.

Bowering, W. R. (1983). Age, growth, and sexual maturity of Greenland halibut, *Reinhardtius hippoglossoides* (Walbaum), in the Canadian Northwest Atlantic. *Fish. Bull.* 81, 599–611.

Bowering, W. R. (1988). An analysis of morphometric characters of Greenland halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic using a multivariate analysis of covariance. *Can. J. Fish. Aquat. Sci.* 45 (4), 580–585. doi: 10.1139/f88-070

Bowering, W., and Brodie, W. (1995). "Greenland Halibut (Reinhardtius hippoglossoides). a review of the dynamics of its distribution and fisheries off eastern Canada and Greenland," in *Deep-water fisheries of the north Atlantic oceanic slope*. Ed. A. G. Hopper (Dordrecht, The Netherlands.: Kluwer Academic Publishers), 113–160. doi: 10.1007/978-94-015-8414-2_4

Bowering, W. R., and Nedreaas, K. H. (2000). A comparison of Greenland halibut (*Reinhardtius hippoglossoides* (Walbaum)) fisheries and distribution in the Northwest and northeast Atlantic. *Sarsia* 85, 61–76. doi: 10.1080/00364827.2000.10414555

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., et al. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic ocean in response to temperature. *Proc. R. Soc. B: Biol. Sci.* 277 (1701), 3725–3734. doi: 10.1098/rspb.2010.0985

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., et al. (2015). Transatlantic secondary contact in Atlantic salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Mol. Ecol.* 24 (20), 5130–5144. doi: 10.1111/mec.13395

Brennan, C. E., Blanchard, H., and Fennel, K. (2016). Putting temperature and oxygen thresholds of marine animals in context of environmental change: A regional perspective for the scotian shelf and gulf of st. Lawrence. *PloS One*, 0167411. doi: 10.1371/journal.pone.0167411

Camacho, J. A., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10 (421), 1–9. doi: 10.1186/1471-2105-10-421

Camargo, J. A., Alonso, A., and Salamanca, A. (2005). Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates. *Chemosphere* 58, 1255–1267. doi: 10.1016/j.chemosphere.2004.10.044

Carrier, E., Ferchaud, A.-L., Normandeau, É., Sirois, P., and Bernatchez, L. (2020). Estimating the contribution of greenland halibut (*Reinhardtius hippoglossoides*) stocks to nurseries by means of genotyping-by-sequencing: Sex and time matter. *Evol. Appl.* 13, 2155–2167. doi: 10.1111/eva.12979

Cayle, K., Basile, J., Johanna, L., and François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* 36 (4), 852–860. doi: 10.1093/molbev/msz008

Cayuela, H., Rougemont, Q., Laporte, M., Mérot, C., Normandeau, É., Dorant, Y., et al. (2020). Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Mol. Ecol.* 29, 2379–2398. doi: 10.1111/mec.15499

Chabot, D., Guénette, S., and Stortini, C. (2013). "A review of the physiological susceptibility of commercial species of fish and crustaceans of the northwest Atlantic to changes in water temperature, dissolved oxygen, pH and salinity," in *Climate change impacts, vulnerabilities and opportunities analysis of the marine Atlantic basin. vol. 3012*, vol. p . Eds. N. Shackell, B. Greenan, P. Pepin, D. Chabot and A. Warburton, 83–167.

Chen, H. (2012). The joint allele frequency spectrum of multiple populations: A coalescent theory approach. *Theor. Population Biol.* 81, 179–195. doi: 10.1016/j.tpb.2011.11.004

Chen, H., Green, R. E., Pääbo, S., and Slatkin, M.. (2007). The joint allele-frequency spectrum in closely related species. *Genetics* 177, 387–398. doi: 10.1534/genetics.107.070730

Claret, M., Galbraith, E. D., Palter, J. B., Bianchi, D., Fennel, K., Gilbert, D., et al. (2018). Rapid coastal deoxygenation due to ocean circulation shift in the northwest Atlantic. *Nat. Climate Change* 8, 868–872. doi: 10.1038/s41558-018-0263-1

Delaney, A., Jacobsen, R. B., and Hendriksen, K. (2012). "Greenland Halibut in upernavik: a preliminary study of the importance of the stock for the fishing populace," in *Aalborg: Innovative fisheries management* (Aalborg University).

De Villemereuil, P., Frichot, E., Bazin, E., François, O., and Gagiotti, O. E. (2014). Genome scan methods against more complex models: when and how much should we trust them? *Mol. Ecol.* 23, 2006–2019. doi: 10.1111/mec.12705

DFO (2019). Assessment of the gulf of St. Lawrence (4RST) Greenland Halibut stock in 2018). *DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2019/023*

DFO (2021). *Proceedings of the regional advisory meeting on the assessment of the gulf of St* Vol. 23-24 (DFO Can. Sci. Advis. Sec. Proceed. Ser. 2021/017: Lawrence (4RST) Greenland halibut), 2021.

Diopere, E., Vandamme, S. G., Hablützel, P. I., Cariani, A., Van Houdt, J., Rijnsdorp, A., et al. (2018). Original article seascape genetics of a flatfish reveals local selection under high levels of gene flow. *ICES J. Mar. Sci.* 75 (2), 675–689. doi: 10.1093/icesjms/fsx160

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., et al. (2020). Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species. *Mol. Ecol.* 29 (24), 4765–4782. doi: 10.1111/mec.15565

Du, X., Zhou, Y., Song, L., Wang, X., and Zhang, S.. (2018). Zinc finger protein 365 is a new maternal LPS-binding protein that defends zebrafish embryos against gram-negative bacterial infections. *FASEB J.* 32, 979–994. doi: 10.1096/fj.201700694R

Dupont-Prinet, A., Vagner, M., and Chabot, D. (2013). Impact of hypoxia on the metabolism of Greenland halibut (*Reinhardtius hippoglossoides*). *Can. J. Fish. Aquat. Sci.* 70, 461–469. doi: 10.1139/cjfas-2012-0327

Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PloS Genet.* 9 (10), e1003905. doi: 10.1371/journal.pgen.1003905

Excoffier, L., Marchi, N., Marques, D. A., Mattey-Doret, R., Gouy, A., and Sousa, V. C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* 37 (24), 4882–4885. doi: 10.1093/bioinformatics/btab468

Fairbairn, D. J. (1981). Biochemical genetic analysis of population differentiation in Greenland halibut (*Reinhardtius hippoglossoides*) from the Northwest Atlantic, gulf of St Lawrence, and Bering Sea. *Can. J. Fish. Aquat. Sci.* 38, 669–677. doi: 10.1139/f81-090

Ferchaud, A.-L., Leitwein, M., Laporte, M., Boivin-Delisle, D., Bougas, B., and Hernandez, C. (2020). Adaptative and maladpative genetic diversity in small populations: Insights from the brook charr (*Salvelinus fontinalis*) case study. *Mol. Ecol.* 29, 3429–3445. doi: 10.1111/mec.15566

Ferchaud, A.-L., Mérot, C., Normandeau, E., Ragoussis, J., Babin, C., and Djambazian, H. (2022). Chromosome-level assembly reveals a putative y-autosomal fusion in the sex determination system of the Greenland halibut (*Reinhardtius hippoglossoides*). *G3: GenesGenomesGenetics* 12 (1), doi: 10.1093/g3journal/jkab376

Forester, B. R., Lasky, J. R., Wagner, H. H., and Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* 27, 2215–2233. doi: 10.1111/mec.14584

François, O., Martins, H., Caye, K., and Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology* 25, 454–469.

Fuentes-Pardo, A. P., and Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* 26, 5369–5406. doi: 10.1111/mec.14264

Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., et al. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol. Appl.* 8, 769–786. doi: 10.1111/eva.12288

Gagnaire, P.-A., Normandeau, É., and Bernatchez, L. (2012). Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. *Mol. Biol. Evol.* 29 (10), 2909–2919. doi: 10.1093/molbev/mss076

Galbraith, P. S., Chassé, J., Shaw, J.-L., Dumas, J., Caverhil, C., Lefaivre, D., et al. (2021). *Physical oceanographic conditions in the gulf of St* (Canadian Science Advisory Secretariat (CSAS). Research Document 2021/045: Lawrence during 2020).

Gomez-Isaza, D. F., Cramp, R. I., and Franklin, C. E. (2021). Exposure to nitrate increases susceptibility to hypoxia in fish. *Physiol. Biochem. Zool.* 94 (2), 124–142. doi: 10.1086/713252

Gundersen, A. C., Stenberg, C., Fossen, I., Lyberth, B., Boje, J., and Jørgensen, O. A. (2010). Sexual maturity cycle and spawning of Greenland halibut *Reinhardtius hippoglossoides* in the Davis strait. *J. Fish Biol.* 77, 211–226. doi: 10.1111/j.1095-8649.2010.02671.x

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PloS Genet.* 5, e1000695. doi: 10.1371/journal.pgen.1000695

Huang, K, Andrew, RL, Owens, GL, Ostevik, KL, and Rieseberg, LH. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol. Ecol.* 29, 2535–2549.

Imanishi, Y., Yang, L., Sokal, I., Filipek, S., Palczewski, K., and Baehr, W. (2004). Diversity of guanylate cyclase-activating proteins (GCAPs) in teleost fish: Characterization of three novel GCAPs (GCAP4, GCAP5, GCAP7) from zebrafish (*Danio rerio*) and prediction of eight GCAPs (GCAP1-8) in pufferfish (*Fugu rubripes*). *J. Mol. Evol.* 59 (2), 204–217. doi: 10.1007/s00239-004-2614-y

Jørgensen, O. A. (1997). Movement patterns of Greenland halibut, *Reinhardtius hippoglossoides* (Walbaum), at West Greenland, as inferred from trawl survey distribution and size data. *J. Northw. Atl. Fish. Sci.* 21, 23–37. doi: 10.2960/J.v21.a2

Jimenez-Mena, B., Le Moan, A., Christensen, A., van Deurs, M., Mosegaard, H., Hemmer-Hansen, J., et al. (2020). Weak genetic structure despite strong genomic signal in lesser sandeel in the north Sea. *Evol. Appl.* 13, 376–387. doi: 10.1111/eva.12875

Kasagi, S., Mizusawa, K., Murakami, N., Andoh, T., Furufuji, S., Kawamura, S., et al. (2015). Molecular and functional characterization of opsins in barfin flounder (*Verasper moseri*). *Gene* 556 (2), 182–191. doi: 10.1016/j.gene.2014.11.054

Kess, T., Bentzen, P., Lehnert, S. J., Sylvester, E. V.A., Lien, S., Kent, M. P., et al. (2019). A migration-associated supergene reveals loss of biocomplexity in Atlantic cod. *Sci. Adv.* 5, eaav2461. doi: 10.1126/sciadv.aav2461

Kess, T., Einfeldt, A. L., Wringe, B., Lehnert, S. J., Layton, K. K.S., McBride, M. C., et al. (2021). A putative structural variant and environmental variation associated with genomic divergence across the Northwest Atlantic in Atlantic halibut. *ICES J. Mar. Sci.* 78 (7), 2371–2384. doi: 10.1093/icesjms/fsab061

Khan, R. A., Dawe, M., Bowering, R., and Misra, R. K. (1982). Blood protozoa as an aid for separating stocks of Greenland halibut, *Reinhardtius hippoglossoides*, in the northwestern Atlantic. *Can. J. Fish. Aquat. Sci.* 39 (9), 1317–1322. doi: 10.1139/f82-176

Knutsen, H., Jorde, P. E., Albert, O. T., Hoelzel, A. R., and Stenseth, N. C. (2007). Population genetic structure in the north Atlantic Greenland halibut (*Reinhardtius hippoglossoides*): influenced by oceanic current systems? *Can. J. Fish. Aquat. Sci.* 64, 857–866. doi: 10.1139/f07-070

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15(356).

Lamichhaney, S., Martinez Barrio, A., Rafati, N., Sundström, G., Rubin, C.-J., Gilbert, E. R., et al. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *PNAS* 109 (47), 19345–19350. doi: 10.1073/pnas.1216128109

Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., et al. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the north Atlantic ocean. *Proc. Natl. Acad. Sci.* 114, E3452–E3461. doi: 10.1073/pnas.1617728114

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzent, M., Budzinski, H., et al. (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in north Atlantic eels. *Mol. Ecol.* 25 (1), 219–237. doi: 10.1111/mec.13466

Lavoie, D., Lambert, N., Starr, M., Chassé, J., Riche, O., Le Clainche, Y., et al. (2021). The gulf of st. Lawrence biogeochemical model: A modelling tool for fisheries and ocean management. *Front. Mar. Sci.* 8, 732269. doi: 10.3389/fmars.2021.7322

Le Bris, A., Fisher, J., Murphy, H. M., Galbraith, P. S., Castonguay, M., Loher, T., et al. (2018). Original article migration patterns and putative spawning habitats of Atlantic halibut (*Hippoglossus hippoglossus*) in the gulf of st. Lawrence revealed by geolocation of pop-up satellite archival tags. *ICES J. Mar. Sci.* 75 (1), 135–147. doi: 10.1093/icesjms/fsx098

Le Luyer, J., Laporte, M., Beacham, T. D., Kaukinen, K. H., Withler, R. E., Leong, J. S., et al. (2017). Parallel epigenetic modifications induced by hatchery rearing in a pacific salmon. *Proc. Natl. Acad. Sci.* 114 (49), 12964–12969. doi: 10.1073/pnas.1711229114

Le Moan, A., Bekkevold, D., and Hemmer-Hansen, J. (2021). Evolution at two time frames: ancient structural variants involved in post-glacial divergence of the European plaice (*Pleuronectes platessa*). *Heredity* 126, 668–683. doi: 10.1038/s41437-020-00389-3

Legendre, P., and Legendre, L. (1998). *Numerical ecology. 2nd English* Edition, Elsevier, Amsterdam.

Limborg, M., Helyar, S. J., De Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R., et al. (2012). Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol. Ecol.* 21, 3686–3703. doi: 10.1111/j.1365-294X.2012.05639.x

Li, H., and Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics* 211, 289–304. doi: 10.1534/genetics.118.301747

Liu, X., and Fu, Y.-X. (2020). Stairway plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 21 (280), 1–9. doi: 10.1186/s13059-020-02196-9

Lou, R. N., Jacobs, A., Wilder, A. P., and Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol. Ecol.* 30, 5966–5993. doi: 10.1111/mec.16077

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., et al. (2022). Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* 6, 469–481. doi: 10.1038/s41559-022-01661-x

Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210, 719–731. doi: 10.1534/genetics.118.301336

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., et al. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Mol. Ecol.* 23, 118–135. doi: 10.1111/mec.12568

Momigliano, P., Florin, A.-B., and Merilä, J. (2021). Biases in demographic modeling affect our understanding of recent divergence. *Mol. Biol. Evol.* 38 (7), 2967–2985. doi: 10.1093/molbev/msab047

Morgan, M. J., and Bowering, W. R. (1997). Temporal and geographic variation in maturity at length and age of Greenland halibut (*Reinhardtius hippoglossoides*) from the Canadian north-west Atlantic with implications for fisheries management. *ICES J. Mar. Sci.* 54, 875–885. doi: 10.1006/jmsc.1996.0200

Morgan, M. J., Bowering, W. R., Gundersen, A. C., Hoines, Å., Morin, B., Smirnov, O., et al. (2003). A comparison of the maturation of Greenland halibut (*Reinhardtius hippoglossoides*) from populations throughout the north Atlantic. *J. Northw. Atl. Fish. Sci.* 31, 99–112. doi: 10.2960/J.v31.a7

Moritz, C. (1994). Defining "Evolutionarily significant units". *Conserv. Trends Ecol. Evol.* 9 (190), pp. 373–375. doi: 10.1016/0169-5347(94)90057-4

Neto, A. G., Langan, J. A., and Palter, J. B. (2021). Changes in the gulf stream preceded rapid warming of the Northwest Atlantic shelf. *Commun. Earth Environ.* 2 (74), 1–10. doi: 10.1038/s43247-021-00143-5

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., and Bekkevold, D. (2009). Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol. Ecol.* 18, 3128–3150. doi: 10.1111/j.1365-294X.2009.04272.x

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S.. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986

Northwest Atlantic Fisheries Organization (2020a) *Greenland Halibut in subarea 0+1 (offshore)*. Available at: https://www.nafo.int/Portals/0/PDFs/Advice/2020/ghl_off.pdf.

Northwest Atlantic Fisheries Organization (2020b) *Updated SCAA base case assessment for Greenland halibut*. Available at: https://www.nafo.int/Portals/0/PDFs/sc/2020/scr20-030REV.pdf.

Nygaard, R., and Boje, J. (2011). *Standardized logbooks from the inshore longline fishery on Greenland halibut in the inshore part of div. 1A* (Dartmouth: NAFO SCR 11/43 Serial No. N5928. Northwest Atlantic Fisheries Organization).

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). *Vegan: Community Ecology Package*. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Pavey, S. A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C., et al. (2015). RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Curr. Biol.* 25, 1666–1671. doi: 10.1016/j.cub.2015.04.062

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., et al. (2015). Slow adaptation in the face of rapid warming leads to collapse of the gulf of Maine cod fishery. *Sci. Res.* 350 (6262), 809–812. doi: 10.1126/science.aac98

Pespeni, M. H., Sanford, E., Gaylord, B., Hill, T. M., Hosfelt, J. D., Jaris, H. K., et al. (2013). Evolutionary change during experimental ocean acidification. *PNAS* 110 (17), 6937–6942. doi: 10.1073/pnas.1220673110

Pomilla, C., Treble, M. A., Postma, L. D., Lindsay, M. M., and Reist, J. D. (2008). Initial genetic evidence of population structure of Greenland halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic. *J. Northw. Atl. Fish. Sci.* 40, 1–15. doi: 10.2960/J.v40.m637

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., et al. (2017). The evolutionary basis of premature migration in pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv.* 3, e1603198. doi: 10.1126/sciadv.1603198

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riget, F., Boje, J., and Simonsen, V. (1992). Analysis of meristic characters and genetic differentiation in Greenland halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic. *J. Northw. Atl. Fish. Sci.* 12, 7–14. doi: 10.2960/J.v12.a1

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from f-statistics under isolation by distance. *Genetics* 145, 1219–1228. doi: 10.1093/genetics/145.4.1219

Roy, D., Hardie, D. C., Treble, M. A., Reist, J. D., and Ruzzante, D. E. (2014). Evidence supporting panmixia in Greenland halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic. *Can. J. Fish. Aquat. Sci.* 71 (5), 763–774. doi: 10.1139/cjfas-2014-0004

Sheffers, B. R., De Meester, L., Bridge, T. C.L., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. C., et al. (2016). The broad footprint of climate change from genes to biomes to people. *Science* 354 (6313), aaf7671. doi: 10.1126/science.aaf7671

Skotte, L., Korneliussen, T. S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195, 693–702. doi: 10.1534/genetics.113.154138

Smith, T. B., and Bernatchez, L. (2008). Evolutionary change in human-altered environments. *Mol. Ecol.* 17, 1–8. doi: 10.1111/j.1365-294X.2007.03607.x

Sohn, D., Ciannelli, L., and Duffy-Anderson, J. T. (2010). Distribution and drift pathways of Greenland halibut (*Reinhardtius hippoglossoides*) during early life stages in the eastern Bering Sea and Aleutian islands. *Fish. Oceanogr.* 19 (5), 339–353. doi: 10.1111/j.1365-2419.2010.00549.x

Stanley, R. R. E., Di Bacco, C., Lowen, B, Beiko, R. G., Jeffery, N. W., Van Wyngaarden, M., et al. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science* 4 (3), sciadv.aaq0929. doi: 10.1126/sciadv.aaq0929

Stenberg, C., Ribergaard, M. H., Boje, J., and Sundby, S. (2016). *Larval drift and settling of Greenland halibut (R. hippoglossoides walbaum) in Northwest Atlantic with special focus on greenlandic waters* (Danish Meteorological Institute) DMI Report, 16–21.

Templeman, W. (1973). *Distribution and abundance of the Greenland halibut, reinhardtius hippoglossoides (Walbaum)* Vol. 10 (the Northwest Atlantic. ICNAF) Res. Bull., 84–98.

Therkildsen, N. O., and Palumbi, S. R. (2016). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Mol. Ecol. Resour.* 17 (2), 194–208. doi: 10.1111/1755-0998.12593

Tonissoo, T., Lulla, S., Meier, R., Saare, M., Ruisu, K., Pooga, M., et al. (2010). Nucleotide exchange factor RIC-8 is indispensable in mammalian early development. *Dev. Dynamics* 239, 3404–3415. doi: 10.1002/dvdy.22480

Treble, M. A., Campana, S. E., Wastle, R. J., Jones, C. M., and Boje, J. (2008). Growth analysis and age validation of a deepwater Arctic fish, the Greenland halibut (*Reinhardtius hippoglossoides*). *Can. J. Fish. Aquat. Sci.* 65, 1047–1059. doi: 10.1139/F08-030

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., De Clerck, O., et al. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecol. Biogeogr.* 21, 272–281. doi: 10.1111/j.1466-8238.2011.00656.x

Van Wyngaarden, M., Snelgrove, P. V. R., Dibacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Zhan, L., et al. (2018). Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*. *Ecol. Evol.* 8, 2824–2841. doi: 10.1002/ece3.3846

Vihtakari, M., Elvarsson, B. Þ., Treble, M., Nogueira, A., Hedges, K., Hussey, N. E., et al. (2022). Migration patterns of Greenland halibut in the north Atlantic revealed by a compiled mark–recapture dataset. *ICES J. Mar. Sci.* 79 (6), 1902–1917. doi: 10.1093/icesjms/fsac127

Vihtakari, M., Hordoir, R., Treble, M., Bryan, M. D., Elvarsson, B., Nogueira, A., et al. (2021). Pan-Arctic suitable habitat model for Greenland halibut. *ICES J. Mar. Sci.* 78 (4), 1340–1356. doi: 10.1093/icesjms/fsab007

Vis, M. L., Carr, S. M., and Davidson, W. S. (1997). Greenland Halibut (*Reinhardtius hippoglossoides*) in the north Atlantic are genetically homogeneous. *Can. J. Fish. Aquat. Sci.* 54, 1813–1821. doi: 10.1139/f97-088

Vollen, T., and Albert, O. T. (2008). Pelagic behavior of adult Greenland halibut (*Reinhardtius hippoglossoides*). *Fish Bull.* 106, 457–470.

Walbaum, C. (1792). Pisces: Rejectee as a generic name and placed on the official index. *Bull. of Zool. Nomencl.* 23, 41–45.

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Heredity* 89 (5), 438–450. doi: 10.1093/jhered/89.5.438

Wellenreuther, M., and Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33 (6), 427–440. doi: 10.1016/j.tree.2018.04.002

Westgaard, J.-I., Saha, A., Kent, M., Hansen, H. H., Knutsen, H., Hauser, L., et al. (2017). Genetic population structure in Greenland halibut (*Reinhardtius hippoglossoides*) and its relevance to fishery management. *Can. J. Fish. Aquat. Sci.* 74 (4), 475–485. doi: 10.1139/cjfas-2015-0430

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (NewYork, NY: Springer-Verlag).

Xuereb, A., Normandeau, É., Daigle, R. M., Curtis, J. M.R., Bernatchez, L., and Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Mol. Ecol.* 27, 2347–2364. doi: 10.1111/mec.14589

Youcef, W. A., Lambert, Y., and Audet, C. (2013). Spatial distribution of Greenland halibut reinhardtius hippoglossoides in relation to abundance and hypoxia in the estuary and gulf of st. Lawrence. *Fish. Oceanogr.* 22 (1), 41–60. doi: 10.1111/fog.12004

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 18 October 2022
DOI 10.3389/fmars.2022.1005534

OPEN ACCESS

EDITED BY
Fran Saborido-Rey,
Centro Superior de Investigaciones Científicas (CSIC), Spain

REVIEWED BY
Monica Gambino,
NISEA Società cooperativa, Italy
Siquan Tian,
Shanghai Ocean University, China

*CORRESPONDENCE
Gonzalo Rodríguez-Rodríguez
gonzalo.rodriguez@usc.es

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 28 July 2022
ACCEPTED 21 September 2022
PUBLISHED 18 October 2022

CITATION
Rodríguez-Rodríguez G, Ballesteros HM, Sánchez-Llamas E, Bande R and Otero RF (2022) The state of the art in cost-benefit of HTS methods for stock assessment: An overview.
Front. Mar. Sci. 9:1005534.
doi: 10.3389/fmars.2022.1005534

The state of the art in cost-benefit of HTS methods for stock assessment: An overview

Gonzalo Rodríguez-Rodríguez[1]*, Hugo M. Ballesteros[2], Eduardo Sánchez-Llamas[2], Roberto Bande[3] and Rosa Fernández Otero[4]

[1]IDEGA and Department of Applied Economics, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, [2]Department of Applied Economics, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, [3]IDEGA and Department of Economics, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, [4]Centro Tecnológico del Mar (CETMAR), Vigo, Spain

Over the past two decades, enormous progresses have been made in high-throughput sequencing (HTS) method development. This fact unveiled the potential usefulness of HTS methods in a wide range of fields such as fishery assessment and management, for which their application has been extensively discussed. As a consequence of the rapid development, sequencing costs have continuously declined, leading to a general claim that HTS methods are cost-efficient compared with traditional ones. Within this context, the underlying research objective is to assess the cost-effectiveness of genomic techniques through a review of the state of the art (SoA) on three HTS methods: i) environmental DNA (eDNA); ii) epigenetics method for age determination through DNA methylation (DNAm), and; iii) close-kin mark-recapture (CKMR) applied on marine ecosystems and fisheries and for stock assessment purposes. The SoA review of the literature on HTS methods was performed through the snow-balling systematic reviewing approach. The analysis has considered the set of processes and variables necessary to perform the stock assessment and compared the capacity of current and HTS methods for providing the required data. Research reveals that HTS methods constitute a promising tool for fishery research and, particularly, for improving scientific advice. Nevertheless, up to now, only one research, on a non-commercial species, has been conducted on the application of HTS methods for stock assessment purposes. Although some partial data are present in the literature, no systematic analysis on costs has been found. This paper suggests that the future research agenda should attempt to straddle both the scenarios for the transition process, considering complementary implementation and substitution possibilities and their cost-efficiency. Clarifying these questions is likely to pave the way for the effective and step-wise implementation of these methods in fishery management; thus, further research is recommended to encompass the transition process.

KEYWORDS

fisheries, stock assessment, HTS methods, close-kin mark-recapture, eDNA, DNAm

Highlights

- HTS methods have been claimed to be cost-efficient; nevertheless, very few publications have systematically and accurately addressed the issue.
- Most cases that claim cost-efficiency are not referred to stock assessments but to other objectives such as biodiversity observation or traceability of fishes, and, hence, cost-efficiency in stock assessment cannot be directly inferred from those.
- As a matter of fact, most of the research on HTS methods that claims to be cost-effective is eDNA metabarcoding in non-marine habitats such as rivers, lakes, or ponds.
- Even those very few cases referred to stock assessment, these works are focusing on species that differ from most of the commercially exploited species.
- In terms of information outputs, traditional surveys provide a broader scope of variables needed for stock assessment, whereas 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 the fact that fishing surveys responsible authorities regularly collect cost-related data as a systematic endorse system, there is a clear shortage 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 reviews, both for currently used and new methodologies. Therefore, further specific research on cost-efficiency is encouraged.
- HTS methods can provide additional valuable information outputs for managing not only the fisheries but the marine ecosystems.

1 Introduction

Society places multiple pressures on marine ecosystems, threatening their capacity to keep providing the multiple services and benefits that they are yielding (Costello et al., 2012). Consequently, improving the understanding of the marine resources is key to manage them (Thomsen et al., 2012). Responding to such challenges will require not only diverse types of knowledge (Rodríguez-Rodríguez et al., 2021) but cost-effective monitoring tools that enable the collection of accurate data to assess the health status of large marine areas (Borja et al., 2016).

Furthermore, fishery management under the Common Fisheries Policy (CFP) aims to ensure that fishing and aquaculture are environmentally, economically, and socially sustainable and that they provide a source of healthy food for the European Union (EU). Since the inception of the first CFP in 1983, the primary fishery management instrument has been the setting of annual TACs (Total Allowable Catches), which are assigned to different EU member states (Casey et al., 2016). In the Mediterranean, fishery management has tend to rely mostly in input control and technical measures (Carpi et al., 2017). In any case, both the CFP and the ecological, social, and economic sustainability of the European stocks currently rely on a robust and accurate scientific advice (Hoydal, 2007; Carpi et al., 2017). In other words, accurate data are a requirement for fulfilling the aims of the CFP by setting up accurate and robust management measures according to the stock status (Thomsen et al., 2012; Thomsen and Willerslev, 2015; Jerde et al., 2019). TACs are set annually (or every 2 years for demersal stocks) by the EU council of fishery ministers taking as basis scientific advice on stock status from advisory bodies [such as the International Council for the Exploration of the Sea (ICES) or the Scientific, Technical and Economic Committee for Fisheries (STECF)] (Daw and Gray, 2005; European Union, 2013). Within this context, some authors—such Chen (2003); Chen et al. (2003); Cope and Punt (2011)—assessed and highlighted the benefits of employing the most accurate available data in fishery stock assessment for reducing subjective uncertainties in determining current fishery status (Chen et al., 2003).

Regarding the marine species research, the data are largely surveyed using selective and invasive methods, which are mostly limited to commercially exploited species and restricted to particular areas. To ensure consistency, information of marine species could be derived from two main sources, namely, fishery-independent data (scientific surveys at sea) and fishery-dependent data (commercial catches analysis and sampling by observers on board) (Pennino et al., 2016). Both sources of information have complementary information that can be jointly used for fishery evaluation and management purposes (Pennino et al., 2016).

Fishery-dependent data are significantly cheaper to obtain, given the fact that the information can be captured in the process of fishing (Dennis et al., 2015). However, negative factors associated to commercial fleet catches data collection such as hyperstability [understood, as defined by de Mitcheson and Erisman (2012), as the phenomenon in which an observed index of stock abundance remains stable while the stock abundance is declining], spatial variability of fishing effort, variable fishing capacity, or erroneous data collection stated the necessity of including fishery-independent information into stock assessment data collection framework. Inherent features of fishery-independent data such as the employment of fishery scientists, the use of a specific fishing vessel and gears or the samples analysis make it invariably more expensive to attain per

unit data unit than fishery-dependent data. Despite this fact, many authors—such as Caddy and Cochrane (2001); Punt et al. (2002), or Dennis et al. (2015)—highlighted the need for fishery monitoring systems that are both robust to the inherent uncertainty associated to the stock assessment process and cost-effective in the relationship between the productivity of the information obtained and its cost.

While it is true that fishery-independent data contain critical information for stock assessment, it is also true that marine scientific surveys are costly, in which cost-efficiency and eventual alternatives open debates (Dichmont et al., 2017; Zimmermann and Enberg, 2017). Along the years, sampling mechanisms have been kept unchanged for both sources of information to maintain homogeneous estimates of stock populations (Stamatopoulos, 2002). At the same time, technical innovations and development of genetic-based methods have supposed a revolution in stock assessment field. Concretely, high-throughput sequencing (HTS) methodologies have been signaled as a breakthrough, able to overcome the traditional methods of data collection and, consequently, to better support institutions and managers in the fishery management (Ovenden et al., 2015; Bravington et al., 2016; Casey et al., 2016; Deiner et al., 2017; Martinsohn et al., 2019; Friedman et al., 2022). Also known as next-generation sequencing (NGS), HTS refers to technologies that sequence DNA and RNA in a rapid and increasingly accessible manner (Nkrumah-Elie et al., 2018). Different advantages with respect to traditional methods have been pointed out in the scientific literature: simplicity, higher precision and accuracy, non-invasiveness, or less time-consuming have been the most highlighted features of HTS methods (Bourlat et al., 2013; Thomsen et al., 2016; Mauvisseau et al., 2017; Hansen et al., 2018). In general, it is assumed that genomic techniques are cost-effective and also efficient in terms of effort and time (Bourlat et al., 2013; Rees et al., 2014; Smart et al., 2016; Gillet et al., 2018; Hering et al., 2018; Lugg et al., 2018; Waples et al., 2018). Genetic analyses have much to offer fishery managers, especially in the provision of tools enabling unequivocal specimen identification and assessment of stock structure (Ward, 2000).

Nevertheless, because of the lack of available information regarding cost-benefit on the implementation of genetic techniques, the very limited published research related to genetic methods for stock assessment purposes (Kolody and Bravington, 2019), although there is an increasing number of papers using genetic studies to present stock structures of different species (Bravington and Grewe, 2007; Pita et al., 2016; Papa et al., 2021) and the high operational costs of implementing those methods in fishery-dependent and fishery-independent data acquisition, makes it necessary to check the evidence supporting such statements. Questions such as the scope for application in terms of gears, species, environments, or fishing areas; the links with policy objectives and the stages covered during the assessment process; the limitations taken into consideration or the benefits of the implementation of HTS methods for stock assessments are key for evaluating the performance of the genetic methods. It constitutes the main objective of this review. In general terms, this implies reviewing what has been stated on the cost-efficiency of HTS methods, which are the current limitations of the available knowledge and the controversies about the suitability for their cost-efficient use in stock assessments.

For that reason, the main purpose of this report is to review the state of the art (SoA) on the cost-benefit/cost-effectiveness of the application of HTS methods for providing biological data for stock assessments. On the basis of the FishGenome project requirements, this review will focus on three relevant HTS methodologies for stock assessments data acquisition process: 1) environmental DNA (eDNA), 2) close-kin mark-recapture (CKMR), and 3) epigenetic age determination based on DNA methylation (DNAm), and it will go from general to the specificities, with particular focus to bottom-trawl fishery-independent surveys and to some of the most representative demersal species subject to TACs, like cod, hake, and wrasse, because cost-efficiency may vary significantly between species (highly migratory, demersal, pelagic, etc.).

Ultimately, the FishGenome study is intended to help getting better and broader scientific knowledge to support future decisions such as an upgrade in the design of the surveys within the European CFP Data Collection Framework (DCF). Therefore, the focus can be put not only on the technology but also on how the current genomic technologies can efficiently contribute to policy and management needs, reducing the gap between science and policy (Casey et al., 2016).

2 Methodology

The SoA review of the existing literature on HTS methods was performed through the snow-balling systematic reviewing approach. 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 20 June 2019 using the following core concepts and terms: i) NGS; ii) epigenetic age determination method (DNAm) (NGS_1); iii) eDNA studies (NGS_2); iv) CKMR studies (NGS_3); v) cost-effectiveness; vi) NGS_1, NGS_2, and NGS_3 combined with "cod", "hake", and "wrasse"; vii) NGS_1, NGS_2, and NGS_3 combined with "North Sea", "North-West Iberian Peninsula", "Balearic Islands", and "Mediterranean"; viii) NGS_1, NGS_2, and NGS_3 combined with "Trawl", "Trawlers", and "Demersal"; ix) fishery research surveys/traditional surveys; and, finally; x) fish stock assessments.

2.1 SoA browsing results and literature review limitations

Table 1 shows the results of the key terms searched on the Thomson Reuters' Web of Knowledge. It should be highlighted

TABLE 1 Combination of terms and search sequence in Thomson Reuters' Web of Knowledge.

NGS	Search	Papers	Reviewed
eDNA	eDNA + Fisheries + Marine	9	8
	eDNA + Fisheries + Marine + Cod	2	2
	eDNA + Trawling	3	3
	eDNA + Cod	4	2
	eDNA + Hake	0	0
	eDNA + Wrasse	0	0
	eDNA + Cost-effectiveness	5	3
	eDNA + Fisheries + Marine + Costs	1	1
	eDNA + Surveys	55	18
	eDNA + North Sea + North-West Iberian Peninsula + Balearic Islands +Mediterranean	0	0
Next Generation	NGS + Fisheries + Marine	20	3
Sequencing	NGS + Fisheries + Marine + Cod	3	0
	NGS + Fisheries + Marine + Hake	2	0
	NGS + Fisheries + Marine + Wrasse	0	0
Close-kin	CKRM	7	4
Mark-recapture	CKRM + Fisheries	3	3
	CKMR + Demersal	0	0
	CKMR + Trawling	0	0
	CKMR + Cod	0	0
	CKMR + Hake	0	0
	CKMR + Wrasse	0	0
	CKMR + Cost-effectiveness/CKMR + Fisheries + Marine+ Costs	0	0
	CKMR+ surveys	1	0
	CKMR+ North Sea + North-West Iberian Peninsula + Balearic Islands +Mediterranean	0	0
DNA	DNAm + Fisheries	6	3
Methylation	DNAm + Marine	1	0
	DNAm + Demersal	0	0
	DNAm + Trawling	0	0
	DNAm + Cod	12	3
	DNAm + Hake	0	0
	DNAm + Wrasse	1	1
	DNAm + Cost-effectiveness + Fisheries + Marine/DNAm + Fisheries + Marine+ Costs	0	0
	DNAm + Surveys	4	0

that there are limited pieces of research on some of the core searching topics. 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 analyzed 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 as hake, cod, and wrasse.

In this way, Jerde et al. (2019) performed a similar literature review, searching in Google Scholar and Web of Science to collect published papers using the metabarcoding approach to estimate fish biodiversity. The authors used tags such as "environmental DNA", "metabarcoding", and "fish", finding n = 46 works on freshwater systems and just n = 7 on marine habitats. The underlying fact is that the literature on freshwater systems is currently much more advanced than focused on the marine ones (Hering et al., 2018). In this regard, the novelty of the use of eDNA metabarcoding on seawater samples to account for marine fish biodiversity is such as recent that it was unprecedented until 2012 (Thomsen et al., 2016).

As for the results of the search, no publications have been found on cost-benefit of the application of HTS methods (CKMR; eDNA and DNAm) in stock assessments or even marine fisheries. It should be noticed that cost-benefit analysis (CBA) is not limited just to monetary values and, ideally, involves more variables of environmental and societal nature, including its costs and benefits (Bateman et al., 2003;

Sartori et al., 2014; Martinsohn et al., 2019). However, at least the identification of certain costs of HTS methods was possible through the literature review. In the same way, no published research in cost-efficiency–related research was found. Thus, it was approached using the available material, which included research on non-marine systems and gray literature (ICES, 2014; ICES, 2015; ICES, 2018; IEO, 2018).

Finally, it also can be highlighted that there is still insufficient research when the contexts of application of the HTS methods are marine or coastal environments. In this regard, the CKMR method on coastal systems is limited to a few papers, like the ones signed by Bravignton and his team (Bravington et al., 2014; Bravington et al., 2016; Waples et al., 2018) who focus their research on just one species (Blue Tuna). In the same context, the cases of study based on DNAm method are quite limited (Table 1). On the contrary, e-DNA methods are better documented.

3 Traditional marine evaluation surveys vs. HTS methods

3.1 Efficient, for what?

Before any evaluation or review, it is necessary to identify and keep in mind for what the methods evaluated are intended to be efficient. In this case, they should serve for carrying out stock assessments. The purpose of a stock assessment is to provide support for decision-making by (1) describing alternative possible states of nature, (2) determining the consequences of taking different management actions under different states of nature, and (3) calculating the probability of different states of nature (Hilborn, 2003). In the case of European waters, stock assessments are the base for TAC allocation that is a key pillar of the CFP. Each EU member state receives a fixed proportion of whatever TAC is agreed for each fish stock (Casey et al., 2016).

Current methods of stock assessment tend to use all available information in a unified framework and may simultaneously include surveys, catch per unit effort (CPUE), age-distributions, length distributions, and tagging (Hilborn, 2003).

Therefore, a key question in terms of efficiency is whether all of these methods are able to provide the information required for a stock assessment. To the best of our knowledge, no comparisons have been done between the results of traditional versus HTS methods. Thus, a first attempt was needed for this review. In that sense, Table 2 shows the type of information obtained with the traditional fishery surveys and by HTS methodologies.

Table 3 shows the parameters obtained from traditional methods versus the alternative provided by HTS methodologies. From a quantitative point of view, the first direct observation is that traditional methods are currently providing all the parameters needed for a stock assessment, either directly as a part of the survey (species, weight and size, number, age, sex, maturity, fertility, abundance, trophism, etc.) or indirectly through the models, as is the case of the stock status and its distance to management target reference points. On the other hand, HTS methodologies do not provide all the required variables and parameters. For instance, they do not provide information on weight and size. The CKRM method is able to provide key parameters for stock assessment: stock status and biodiversity, whereas eDNA can accurately determine the species and provide information on diversity, and EAD (epigenetics for age determination) supplies accurate data on age and sex.

At this point, it should be noticed that the interest in HTS methodologies stems, on a broad level, from its potential to provide unique understandings of ecological processes in marine environments and supports more precise approaches for ecosystem-based management (Ovenden et al., 2015) and going down to each methodology for the potential to provide efficiently accurate (specific) data. This is the case for CKMR, which is expected to widen the scope of population-level inference relative to currently used monitoring programs (Conn et al., 2020). In the same vein, epigenetic clocks have proven themselves to be accurate (Simpson and Chandra, 2021), with recent studies revealing new examples of DNAm age association in several new species increasing the potential for developing DNAm age biomarkers for a broad range of wild animals (De Paoli-Iseppi et al., 2017). eDNA sampling can be a highly sensitive method for detecting aquatic taxa (Smart et al., 2016); however, its cost-efficiency has been scarcely studied.

Toward the end, currently used methodologies provide a broader scope of variables, whereas HTS methodologies focus on the improvement of certain key variables. At the same time, currently used methodology ensemble is the outcome of a long process of adaptation to the goals and needs of stock assessments, whereas HTS methodologies are scientific developments dealing with their innovation path for fitting with the stock assessment specific needs, which may pose a path-dependent problem.

3.2 About the information outputs of the HTS methods

Once the broad picture of the information outputs provided by the two big groups of methodologies regarding stock assessment has been identified, it is necessary to go deeper into the properties of the HTS methods. The more extensive body of literature in this field is devoted to eDNA. Regarding this method, Deiner et al. (2017) carried out a literature review on eDNA metabarcoding on animals and plants, observing that environmental metabarcoding of DNA can, in some cases, complement and even improve the results of conventional

TABLE 2 Information provided by the traditional fishery surveys, compared with the potential information provided for the combination of the following HTS methods: CKMR + eDNA + epigenetic age determination.

Method	Specie	Weight and Size	Species abundance	Age	Sex	Maturity	Fertility	Stock	Abundance	Biodiversity	Trophism	Additional Information (marine litter, pictures, etc.)
Traditional Survey	✓	✓	✓	✓	✓	✓	✓	?	✓	?	✓	✓
CKMR + eDNA + DNAm	✓	✗	✗	✓	✓	✗	✗	✓	✓	✓	✗	✗

Columns in table refer to different parameters and variables obtained during the traditional fishery surveys and by HTS methods in a fishery-independent survey. This information is as follows: i) species as the identification of existent marine species in the survey area; ii) weight and size as stock weight and size structure; iii) species abundance as the amount of species in a given target area; iv) age as stock age structure; v) sex as target stocks sex proportion; vi) fertility makes reference to the target stock reproductive ability; vii) stock as different stocks identification in a given area; viii) abundance as the target stocks abundance and biomass estimation; ix) biodiversity as the variety of species in the survey area; and x) trophism as information on the structure of fish community and its trophic interactions.

TABLE 3 Information provided by traditional fishery surveys and by HTS methods.

Method	Specie	Weight and Size	Species abundance	Age	Sex	Maturity	Fertility	Stock	Abundance	Biodiversity	Trophism	Additional Information (marine litter, pictures, etc.)
Traditional Survey	✓	✓	✓	✓	✓	✓	✓	?	✓	?	✓	✓
CKMR	✗	✗	✗	✗	?	✗	✗	✓	✓	✗	✗	✗
eDNA	✓	✗	✗	✗	✗	✗	✗	✗	?	✓	✗	✗
Epigenetic Age Determination	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗

Theoretically, the method could offer information on this issue but partially or subjected to other additional processes. Columns in table refer to different parameters and variables obtained during the traditional fishery surveys and by HTS methods in a fishery-independent survey. This information is as follows: i) species as the identification of existent marine species in the survey area; ii) weight and size as stock weight and size structure; iii) species abundance as the amount of species in a given target area; iv) age as stock age structure; v) sex as target stocks sex proportion; vi) fertility makes reference to the target stock reproductive ability; vii) stock as different stocks identification in a given area; viii) abundance as the target stocks abundance and biomass estimation; ix) biodiversity as the variety of species in the survey area; and x) trophism as information on the structure of fish community and its trophic interactions.

10.3389/fmars.2022.1005534

methods by identifying different species, sampling greater diversity, and increasing the resolution of taxonomic identifications. This literature review included n = 21 studies in different ecological systems, and only n = 3 related to fish species in marine environments. The results showed that the number of marine species detected by eDNA metabarcoding could be complementary, similar, or even more significant.

A recent study presented by Yamamoto et al. (2017) showed that the level of identification is similar between the traditional evaluation methodology (long-term observation) and the genomic method (eDNA, surface, water column, and ocean sediment), implying that each technique identified more or less the same number of individuals, but each method detected some specific species that the other could not reveal. In particular, they demonstrated that eDNA metabarcoding is a more time-efficient method for examining a whole fish community than a visual census, having a very high detection performance among the HTS methods. This and the next one are two of the cases where it is particularly relevant to bear in mind what we are comparing when assessing the efficiency.

In the same vein, Port et al. (2016) compared another technique, scuba-diving, with eDNA (water column) being the latter much more effective, identifying a higher number of species than the traditional assessment technique. The most consistent study with the purpose of this review is the one carried out by Thomsen et al. (2016). They compared traditional techniques (trawl catch data) with eDNA samples (bottom sediment and water column) obtaining similar family richness. Whereas the eDNA identified species that do not frequently enter the nets, the trawling technique detected other species that were not recognized, at the species level, by the eDNA analysis.

Interestingly, the only assessment based on the CKMR methodology not only provides key parameters needed for stock assessment (abundance) but also requires traditional measures (weight, sex, etc.), suggesting the complementarity between both methods. In fact, different studies suggest the complementarity between traditional and HTS methodologies, because they seem to offer a broader picture of the state of the oceans and their resources (Deiner et al., 2017; Evans et al., 2017; Gillet et al., 2018; Stat et al., 2019).

3.3 Unraveling the value of a sample: Are the HTS methods cheaper than the traditional evaluation methods?

3.3.1 Value of bottom trawl surveys

The identification of costs and economic information related to the bottom trawl surveys has been unsuccessful. Although the literature on currently used methods of evaluation has been explored, with a particular interest in bottom trawl surveys, the consulted manuals do not present economic data regarding costs, salaries, hours of work, or any similar variables. In this sense, the manuals on bottom trawl surveys were reviewed using keywords such as expenses, costs, or outputs. Those terms appeared only on two occasions (ICES, 2014; IEO, 2018), revealing the difficulty of accessing economic information on evaluation projects. Manuals were useful to identify general survey practices, on-board processes, materials, and even the observation personnel necessary for the data collection but do not obtain the data required to establish the bases of cost analysis. Therefore, one of the revealed aspects along the literature review process was the lack of (published) cost studies on this topic.

Dennis et al. (2015) compared the cost-benefit ratio of fishery-independent versus fishery-dependent methods in the small-scale Torres Strait lobster fishery (between Australia and Papua New Guinea). Although the sampling method, randomly allocated stations, is not directly comparable with bottom trawl surveys, the paper yield some interesting insights. The authors based their analysis on the premise that fishery-independent surveys' higher cost would be economically justified when the profit attained due to an additional catch allocation estimated by the fishery model at least matched the survey cost and assuming that the fishery-independent outputs were incorporated by managers to set the TAC. Results show a positive net present value on the long term. Beyond the specificities of this case, the paper demonstrates, on the one hand, the contribution of CBA for decision-making and, on the other, the intrinsic relevance of accurate data.

3.3.2 Value of HTS methods

Only n = 3 works were found, in which the value of an HTS method sample was defined, answering the following question: How much does it cost to generate the information using a genomic technique? Three for eDNA and three for CKMR (but being part of the same project). In addition, one paper has been found on DNA analysis that, although it does not fall exactly within the scope of this review, may be representative of the costs of DNA sequencing.

The group of papers based on eDNA methodology share certain similarities and results (Table 4). All of them are on fish species but in freshwater systems. They were compared with currently used methodologies but only at the survey stage (they do not compare the total cost from sampling to modeling). These works suggest that the cost-efficiency of genomic techniques is based on the reduction of effort and work time in observation campaigns. Therefore, if conducting a biodiversity analysis, then these techniques may be adequate and cost-effective, but they do not provide enough information for a stock assessment of commercial fisheries. At the same time, the sampling process differs significantly in rivers, lakes, ponds, and oceans.

Indeed, these works suggest that the cost-effectiveness of genomic techniques is based on the reduction of effort and work time in observation campaigns. This factor, although reasonable

TABLE 4 Cases of study where the costs of getting information with traditional evaluation methods are compared.

Reference	Traditional Survey	HTS	Samples	Sample Value	Effort/Time	Are HTS much cheaper?
(Evans et al., 2017) Freshwater Namekagon River (EEUU)	• Triple-pass electrofishing • Single-pass electrofishing • Presence-absence electrofishing	eDNA	42	$ 16,14	• eDNA approach required (6.8 person/h) • Triple-pass electrofishing (90 person/h) • Single-pass electrofishing (30 person/h) • Presence-absence electrofishing (20 person/)	• (YES) Triple-pass electrofishing. (e-DNA 67% Cheaper) (YES) Single-pass electrofishing • (E-DNA ± equal) (NO) Presence-absence electrofishing • (e-DNA 33% more expensive) COST: eDNA 42 Samples • A = Materials = Cost of selecting dPCR samples was $ 4.02 per sample + cost of DNA extraction at $ 8.49/sample = $ 525 • B = Labor (6.8 h/person * 22.5 $/h) = $ 153 • Total Cost 42 e-DNA samples = $ 678. • eDNA sample = $ 16.14
(Qu & Stewart, 2019) Freshwater/ Yangtze River/ China	• Traditional Surveys, Capture and Visual Monitoring	eDNA • CPCR • qPCR	45	eDNA (CPCR): €211 eDNA (qPCR): €25.2 Annual value for Visual and Capture monitoring	E-DNA (CPCR) sampling, E-DNA (qPCR): sampling, Visual Monitoring (7 days, 1 boat, 3 personnel) Capture Monitoring (10 days, 15 boats, 40 personnel)	• E-DNA (CPCR) sampling, including labor, filtering water collections, extractions, amplifications, and sequencing. Total cost: 9,531.90. • eDNA (qPCR): sampling, including labor, filtering water collections, extractions, amplifications, and sequencing. Total cost 1,134.07 euros• Visual Monitoring (7 days, 1 boat, 3 personnel). Total cost: 4,466.59 euros/year• Capture Monitoring (10 days, 15 boats, 40 personnel). Total cost: 41,874.29 euros/year.
(Stein et al., 2014) Freshwater /EEUU	• Bioassessment	eDNA • Sanger Sequency	¿?		Substantially less: not necessary to sort specimens, clip tissues, and place extracts into individual wells on a plate	• Potentially, fish traditional method = $850 (Sorting $350 + Taxonomic ID $400) • ADN barcoding using Sanger sequencing includes sorting and, when required, clipping tissue samples. Sanger = $2,900 (Sorting $400 + Taxonomic ID $2,500) • ADN barcoding NGSr = $500–1,000 (Sorting $0 + Taxonomic ID $500–1,000)

from the economic point of view, is not the only one that influences the decreasing cost of the use of genomic techniques. In this sense, possible economies of scale associated with the processes, the number of samples needed, or how many of them are processed will influence the costs of obtaining the information (Smart et al., 2016; Lugg et al., 2018).

On the reduction of effort, it was argued (Stein et al., 2014) that "Next-Generation Sequencing costs are substantially less than those associated with Sanger sequencing because it is not necessary to sort specimens, clip tissues, and place extracts into individual wells on a plate". In this sense, it was observed that, for the identification of fish/invertebrates/algae in freshwater systems:

- Fish traditional method = $850 (Sorting $350 + Taxonomic ID $400).
- ADN barcoding using Sanger sequencing includes sorting and, when required, clipping tissue samples. Sanger = $2,900 (Sorting $400 + Taxonomic ID $2500).
- ADN barcoding using NGS = $500–1,000 (Sorting $0 + Taxonomic ID $500–1,000)

Qu and Stewart (2019) conducted a study on the Yangtze River freshwater system (China) where they compared the costs of the traditional surveys (capture and visual monitoring) with two eDNA protocols (cPCR and qPCR) to identify the status of a specific aquatic mammal (*Neophocaena asiaeorientalis*). Table 5 shows the results of their research, indicating that the technical eDNAs were cheaper than the traditional evaluation methodologies.

In that sense, the value of eDNA (cPCR) taking into consideration literature review about sampling, including labor, filtering water collections, extractions, amplification, and sequencing, had a total cost of €594.20 (for 45 samples).

Regarding the currently used evaluation methods, the difference was the effort invested in the information gathering. Visual monitoring method demanded more time and resources (boat and personnel) implying at less 7 days of works, rent of a boat, and paid three experts who addressed the evaluation processes. The total cost of the operation if it was to be implemented monthly would be €1,116.65. If the same study was replicated seasonally (3X), the estimated value would be €3,349.95; whereas if it was carried out 12 times a year, it would amount to €13,399.78 per year.

The method of capture monitoring was much more expensive, because it demanded 10 days of work, 15 boats, and 40 personnel doing the monitoring work. This operation had a total cost: €41,874.29 year.

In this case, as expressed by Qu and Stewart (2019), "Visual surveying on a monthly basis thus costs 1.88× that of eDNA

TABLE 5 Cost comparison eDNA vs. traditional evaluation methods in a freshwaters system (Qu and Stewart, 2019).

Survey method	Details	Cost
cPCR	eDNA collection labor	€69.09
	Filter papers + consumables	€25.12
	Extraction QIAGEN DNEasy blood and tissue kit	€237.43
	Amplification	€56.53
	Confirmation (visualization)	€94.22
	PCR labor	€67.84
	Total	€594.20
qPCR	eDNA collection labor	€69.09
	Filter papers + consumables	€25.12
	Extraction MOBIO DNEasy PowerWater kit	€516.59
	Amplification and quantification	€32.12
	qPCR labor	€150.75
	Total	€793.66
Visual monitoring	X = 7 days, 1 boat, 3 personnel	
	Per month (1X)	€1,116.65
	Per season (3X)	€3,349.95
	Per year (12X)	€13,399.78
Capture monitoring	X = 10 days, 15 boats, 40 personnel (EFFORT)	
	Per year (1X)	€41,874.29

collections utilizing cPCR (species detection) at the same temporal schedule. If, however, eDNA sampling using cPCR occurred only once per season (3X), then visual surveys would approximate 5.64× more expensive. Similarly, visual surveys compared to eDNA sampling utilizing qPCR would equate to 1.41× on a monthly sampling schedule and 4.22× on a seasonal sampling schedule".

Evans et al. (2017) illustrated how the same genomic technique can be more expensive or cheaper, depending on the traditional method with which it is compared. As a novelty, in this paper, it emphasized the identification of the workforce, the number of employees, salaries, and work time. Precisely, the reduction of effort is the fundamental factor that lowers the costs (Evans et al., 2017).

The closest research to the purpose of this review is the group of publications stemming from the research in the estimation of the spawning biomass of bluefin tuna using close-kin genetic markers (Bravington et al., 2014; Bravington et al., 2016; Waples et al., 2018), because the objectives of their project were i) to provide a fishery-independent estimate of the number of adult Southern bluefin tuna and ii) to provide direct estimates of age-specific fecundity and a better definition of spawning stock biomass (Bravington et al., 2014). The three publications linked to this research suggested that CKMR is a cost-effective method. The sources of cost reduction (comparing with currently used methods) are as follows (Bravington et al., 2014): i) sample sizes are likely to be lower; ii) possibility of re-using samples reducing the cost of sampling in future; iii) progressive reduction of genotyping cost; and iv) no ship or aircraft time.

Nevertheless, no systematic or specific data about cost is presented. Interestingly, they suggest a limitation for the application of this methodology to most of the marine fish species: They are too abundant to make the method cost-effective, so that it will be necessary a further reduction of genotyping costs (Bravington et al., 2014).

In this project, all fish sampled for genetics had their length measured and were sexed by checking for residual female gonads, as a part of the regular catch sampling program. A portion of the fish genotyped form part of the otolith collection set and therefore will be of known age (Bravington et al., 2014), which, as a matter of fact, is suggesting complementarity between traditional and HTS methods.

To the best of our knowledge, one of the most detailed papers about costs of the application of DNA analysis in fisheries is the one by Martinsohn et al. (2019). This paper considered both total cost of monitoring and approximate costs associated with laboratory setup in a monitoring and forensic context based on a number of previous fishery and aquaculture compliance investigations. Results suggest that the application of such methodologies is i) affordable because the costs of sequencing have been dropping over the last years and ii) economically justifiable given that in all cases examined in their study, and analytical costs (including administrative costs) were lower than the value of confiscated catches, illegal imports, and associated

fines. Nevertheless, in this case, DNA analysis is providing information that no other method can provide, and, furthermore, the application of this technology responds to quite different policy and management objectives (fishery control, enforcement, and traceability) to those that we are considering.

4 Discussion

A critical requirement of stock assessment is the availability of data as precise as possible. Precision is based on an optimal trade-off between bias (approximation error) and variance (errors in estimating parameter values from the limited data available) as the errors of prediction are influenced by both (Dennis et al., 2015). Up till now, the key way for efficiently increasing accuracy was to combine dependent and independent data; nevertheless, with the significant reduction of sequencing cost, HTS methodologies have become a promising way to provide accurate data efficiently and to reduce the error of estimated TACs. Under the precautionary approach applied to fishery management framework, increased precision would reduce the risk of severe and irreversible damage to the fishery resources and the environment while maximizing economical profits and more steady management scenarios.

Fishery stock assessment usually requires a large number of historical information sets to characterize various fishery aspects (Chen et al., 2016). As stated by some authors, time-series length reflects the completeness of information collection for targeted variables, which is highly correlated with fishery economic/social/ecological importance (Chen et al., 2003; Rotherham et al., 2007; Chen et al., 2016). The continuity in fishery-dependent and (specially) in fishery-independent data collection programs and the gradual implementation of HTS methods applied for stock assessment purposes would suppose a revolution in how decision-making process will change in relation to the data (Hilborn, 2003). This revolution, understood in the sense of evolution, would be focused on providing the best possible technical support toward the stock assessment process development and, subsequently, an improved fishery management for policy making. These improvements would be mainly focused on the following: i) gathering and integration of new stock information (i.e., stock identification and new stock borders); ii) reducing uncertainty on stock assessment estimation process (improvements on community age structure information and reproductive parameters); iii) significant improvement on assessed species number due the reduction of on-sea data gathering techniques time and cost (i.e., data gathering from commercial landings for abundance estimation); vi) data quantity increase and, subsequently, increase on reference points calculation number; and v) inclusion of environmental information in the stock evolution (through the eDNA data incorporation in stock assessment process and policy-making process).

Under the presented context, HTS methods have been claimed to be a breakthrough in marine science (Ovenden et al., 2015; Bravington et al., 2016; Casey et al., 2016; Martinsohn et al., 2019; Friedman et al., 2022), raising the interest in its application in stock assessments. Hence, to check its suitability in terms of cost-efficiency, an SoA review was carried out. Several gaps appear that prevent to support the cost-efficiency by now.

4.1 Lack of research related on the applications of HTS techniques for stock assessment

First of all, only one project addresses the application of HTS methods to stock assessments (Bravington et al., 2014; Bravington et al., 2016; Waples et al., 2018). This research deals with an application of the CKMR method to estimate the spawning biomass of the Southern bluefin tuna, proving to be a suitable and also a cost-efficient method. Nevertheless, although certain possible sources of costs reductions are identified, i) no systematic cost analysis has been presented and, in addition, ii) characteristics of these species differ from most commercial species (lower number of individuals, no need for catching individuals for the assessment, etc.). It means that, for other commercial fisheries or different target species, a higher number of samples are needed (as the number of individuals is higher), increasing the cost of sampling (boat and crew time). Therefore, even with the same purpose, results are not directly transferable. Furthermore, recent contributions (Friedman et al., 2022) based on expert's advice also pointed out that cost-effective studies on the use of genetic technologies should be conducted.

4.2 Current application of new genetic techniques

The remaining literature dealing with the application of new genetic methodologies to fisheries is oriented to objectives that are different from stock assessments. Results in terms of efficiency or efficacy of any method depend on the purposes that they are used for.

In other words, eDNA seems to be efficient when compared with direct observations, scuba observations, etc. Nevertheless, for the moment, there is no evidence about their efficiency for stock assessments when compared with currently used methodologies. Stock identification remains one of the most confusing but relevant challenges in fishery science (Cadrin et al., 2014). Within this context, understanding intraspecific stock subdivisions remains a challenge in fishery science. Despite

this, molecular genetic techniques such as eDNA appear to be a robust tool in conservation biology for identifying key aspects such as reproductive isolation between stocks, permitting delineation of management units, and allowing assessment of conservation priorities from an evolutionary perspective (Begg and Waldman, 1999). In particular, coastal and demersal species represent the main target on the application of eDNA techniques because of the species characteristics.

Epigenetic (DNAm) age determination seems to be a very accurate method to obtain sex and age, but it is a small part of the information required. Despite the short amount of information provided by DNAm for stock assessment purposes, it is key to estimate the fishery current status (population age structure, reproductive analysis, stock recruitment relationships, etc.). This information is generally difficult to obtain by means of traditional techniques (i.e., otoliths analysis), or, in other cases, measures are inaccurate. In addition to the benefits on the application of epigenetics on stock assessment processes, from an ecological point of view, these techniques will therefore improve our understanding of the mechanisms underlying natural variation in ecologically important traits and will provide insights into the mechanisms that allow organisms to respond to the environment (Bossdorf et al., 2008). In addition, in general terms, epigenetic processes may increase the evolutionary potential of organisms in response to abiotic stress and other environmental challenges, which could potentially be highly relevant in the context of global environmental change (Bossdorf et al., 2008). Therefore, understanding epigenetics in fishery populations could constitute a key tool in adaptive management toward the mitigation of climate change negative implications.

Despite the lack of variety on research studies about the application of CKMR methods for stock assessment purposes, it seems to be a promising HTS method in fishery research area. As stated by some authors, CKMR can be used effectively and cheaply as a mid/long-term monitoring tool for stock assessment. Some features—such as i) the ability of performing long-term forecasts; ii) the independent estimation of selectivity, fecundity, and mortality; iii) the capacity to estimate accurately stock abundance; and iv) accurate estimation of stock-recruitment relationship—place CKMR as a promising tool for stock assessment. Despite this, in marine context, most fish species are simply too abundant to have made the method cost-effective, although this may change as genotyping costs continue to drop (Bravington and Grewe, 2007).

In general terms, optimal levels of spatial and temporal replication according to fishery characteristics, target species, or the benefits of increasing sample sizes, etc., should be taken into account on deploying CBA (Underwood, 1996). This fact implies the inclusion of an inherent stage of the surveys that will determine appropriate compromises between survey precision and the collected information amount (Bravington and Grewe, 2007).

4.3 Surveys

Despite the fact that there remains a need to measure the relative importance of different types of fishery data for stock assessment in allocating sampling effort to ensure the optimal data collection (Chen et al., 2003), the inclusion of HTS techniques in stock assessment data collection frameworks and, subsequently, in management processes would suppose a revolution in the fishery-dependent and fishery-independent sampling.

A common practice while obtaining biological information about a fishery resource during a fishery-dependent sampling is to complement catch and effort data with biological observations of caught individuals. Traditional mechanisms for measuring sex, maturity, or age are highly invasive and generally require to purchase a wide number of individuals. HTS techniques put over the table a non-invasive mechanism for obtaining wide amount of information from a single fin-clip, eliminating (to a great extent) the need to engage trawl or acoustic surveys for determining stock size. Related literature on this topic is unclear, but some manuals recommend a minimal sample size of 50~100 individuals to obtain relevant biologic information (Martinsohn et al., 2015). Performance and benefits of implementing HTS methods on biological fishery-dependent information sampling are highly determined by target species' commercial value.

Related with fishery-independent surveys, many authors such as Caddy and Cochrane (2001) or Dennis et al. (2015) highlighted the need and benefits for fishery monitoring systems that are both robust to inherent uncertainty and cost-effective. Despite this, the general intuition is that new HTS methods could reduce costs and improve efficiency and precision of fishery information. In addition, despite the further work needed, genetic response to environmental changes such as climate change information could be obtained by means of HTS methods. It would offer wide amount of information to the application of a climate-based adaptive management in fishery research area (Frost et al., 2012).

4.4 On the benefits of including genetic information on stock assessment

Literature revision on the benefits of including HTS information on fishery data for stock assessment highlighted the value of the new techniques. Those could be summarized in i) precise stock identification; ii) accurate sex, maturity, and age determination; and iii) reduction of the associated uncertainty of parameter estimation and, subsequently, improvement of fishery model outcomes for stock assessment precision, which could lead to a more accurate management measures.

Stock identification is a key aspect in modern fishery stock assessment. However, considering the importance of identifying the target species stock structure, there is a scarcity of assessments that actually include stock identification requirements (Begg et al., 1999). Understanding the genetic variation between stocks provides a reliable source of information for management purposes (Begg and Waldman, 1999).

Many authors addressed the fact that a large amount of biological knowledge is potentially useful in stock-assessment and management context. Sex, age, and maturity parameter estimation involves a wide number of processes and methods for sampling. In addition, sometimes, the big number of samples and the difficulties to obtain precise observations (i.e., otoliths sampling) increase the process uncertainty (observer error). In addition, differing by species and region, this kind of information could not be sampled (e.g., otoliths observation in tropical species and/or small-pelagic individuals cannot be performed). The inclusion of HTS techniques and, concretely, the use of DNAm methods could suppose a revolution in the ability to estimate precise fishery parameters and, subsequently, improve the scientific advice—the latter through reducing the sources of uncertainty in fishery stock assessment and, therefore, potentially improving management.

4.5 HTS method limitations

In addition, just a few works expose the limitations of such novel techniques, putting in the spotlight the need of delving into the study of the HTS methods to advise them as substitute measures to traditional methodologies (Bravington et al., 2014; Bravington et al., 2016; Yamamoto et al., 2016; Evans et al., 2017; Yamamoto et al., 2017; Harper et al., 2018; Qu and Stewart, 2019). Moreover, there are very few works that go beyond analyzing the HTS methods from a critical point of view questioning their alleged cost-efficiency in a broader analytical frameworks (Stein et al., 2014; Hansen et al., 2018; Jerde et al., 2019).

4.6 Further applications of HTS methods

Finally, it is worthy to point out that HTS can provide additional information that could be of interest for marine research beyond stock assessment, contributing, for instance, to move toward ecosystem-based fishery management. HTS-based methods such as CKMR in the estimation of biomass and species abundance, epigenetic analysis as an alternative to traditional ageing techniques, genotypic analysis of marine species, and eDNA sampling could enable a more adaptive management, contributing to mitigate negative climate change implications.

5 Conclusions

Once the literature is revised, the HTS method seems to be a promising methodology for marine science and, particularly, for stock assessment. Nevertheless, no complete evidence of their cost-efficiency/cost-benefit for marine fish stock assessments has been provided up to now. Almost all available analysis focuses on specific components or activities, but there is a lack of full-stream assessments.

An eventual cost-efficient analysis comparing currently used versus HTS methodologies should take into account, on the one hand, the information provided (and its suitability for stock assessment) and, on the other, the cost of gathering such information considering the sampling process, data collection, and the processing of the information until the assessment is done. In addition, different commercial species should be considered as costs can vary significantly depending on the abundance, the habitats, etc.

As it is not possible to assess the efficiency without taking into consideration the outputs (both intermediate and final), it would be convenient to identify the amount and quality of information generated by the genomic techniques individually and compare them with the data obtained in traditional sampling, to confirm whether the use of HTS methods is more efficient.

Finally, the future research agenda should attempt to straddle both the scenarios for the transition process, considering complementary implementation and substitution possibilities and their cost-efficiency. A key requirement in this regard is that fishery stock assessment usually requires a large number of historical information sets. Clarifying these questions is likely to pave the way for the effective and step-wise implementation of these methods in fishery management; thus, further research is recommended to encompass the transition process.

Author contributions

GR-R provided the original idea. HB, ES-L and GR-R designed the manuscript content and outline. GR-R, HB, ES-L, RB and RO contributed to writing and re- viewing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

HB's work has been funded by a Postdoctoral Research Contract by the Xunta de Galicia, Project ED481D 2022/009. The information and views set out in this paper are based on

scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bateman, I. J., Lovett, A. A., and Brainard, J. S. (2003). *Applied environmental economics: A GIS approach to cost-benefit analysis, applied environmental economics: A GIS approach to cost-benefit analysis* (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511493461

Begg, G. A., Friedland, K. D., and Pearce, J. B. (1999). Stock identification and its role in stock assessment and fisheries management: an overview. *Fish. Res.* 43, 1–8. doi: 10.1016/S0165-7836(99)00062-4

Begg, G. A., and Waldman, J. R. (1999). An holistic approach to fish stock identification. *Fish. Res.* 43, 35–44. doi: 10.1016/S0165-7836(99)00065-X

Borja, A., Elliott, M., Snelgrove, P. V. R., Austen, M. C., Berg, T., Cochrane, S., et al. (2016). Bridging the gap between policy and science in assessing the health status of marine ecosystems. *Front. Mar. Sci.* 3 doi: 10.3389/fmars.2016.00175

Bossdorf, O., Richards, C. L., and Pigliucci, M. (2008). Epigenetics for ecologists. *Ecol. Lett.* 11, 106–115. doi: 10.1111/j.1461-0248.2007.01130.x

Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., et al. (2013). Genomics in marine monitoring: New opportunities for assessing marine health status. *Mar. pollut. Bull.* 74, 19–31 doi: 10.1016/j.marpolbul.2013.05.042

Bravington, M., and Grewe, P. (2007). "A method for estimating the absolute spawning stock size of SBT, using close-kin genetics," in *CCSBT 8th meeting of the stock assessment group and the 12th meeting of the extended scientific committee* (CSIRO, Hobart, Australia). https://doi.org/FRDC Report 2007/034.

Bravington, M. V., Grewe, P. M., and Davies, C. R. (2014). Fishery-independent estimate of spawning biomass of southern bluefin tuna through identification of close-kin using genetic markers. *FRDC Rep.* CSIRO. https://doi.org/FRDC Report 2007/034

Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016). Close-kin mark-recapture. *Stat. Sci.* 31, 259–274 doi: 10.1214/16-sts552

Caddy, J. F., and Cochrane, K. L. (2001). A review of fisheries management past and present and some future perspectives for the third millennium. *Ocean Coast. Manag* 44, 653–682. doi: 10.1016/S0964-5691(01)00074-6

Cadrin, S. X., Karr, L. A., and Mariani, S. (2014). *Chapter one - stock identification methods: An overview*. Eds. S. X. Cadrin, L. A. Kerr and , S.B.T.-S.I.M Mariani (San Diego: Academic Press), 1–5. doi: 10.1016/B978-0-12-397003-9.00001-1

Carpi, P., Scarcella, G., and Cardinale, M. (2017). The saga of the management of fisheries in the Adriatic Sea: History, flaws, difficulties, and successes toward the application of the common fisheries policy in the Mediterranean. *Front. Mar. Sci.* 4. doi: 10.3389/fmars.2017.00423

Casey, J., Jardim, E., and Martinsohn, J. T. H. (2016). The role of genetics in fisheries management under the E.U. common fisheries policy. *J. Fish. Biol.* 89, 2755–2767 doi: 10.1111/jfb.13151

Chen, Y. (2003). Quality of fisheries data and uncertainty in stock assessment. *Sci. Mar.* 67, 75–87. doi: 10.3989/scimar.2003.67s175

Chen, Y., Chen, L., and Stergiou, K. I. (2003). Impacts of data quantity on fisheries stock assessment. *Aquat. Sci.* 65, 92–98. doi: 10.1007/s000270300008

Chen, Y., Shan, X., Jin, X., Yang, T., Dai, F., and Yang, D. (2016). A comparative study of spatial interpolation methods for determining fishery resources density in the yellow Sea. *Acta Oceanol. Sin.* 35, 65–72. doi: 10.1007/s13131-016-0966-y

Conn, P. B., Bravington, M. V., Baylis, S., and Ver Hoef, J. M. (2020). Robustness of close-kin mark–recapture estimators to dispersal limitation and spatially varying sampling probabilities. *Ecol. Evol.* 10, 5558–5569. doi: 10.1002/ece3.6296

Cope, J. M., and Punt, A. E. (2011). Reconciling stock assessment and management scales under conditions of spatially varying catch histories. *Fish. Res.* 107, 22–38. doi: 10.1016/j.fishres.2010.10.002

Costello, C., Ovando, D., Hilborn, R., Gaines, S. D., Deschenes, O., and Lester, S. E. (2012). Status and solutions for the world's unassessed fisheries. *Science* 80-,). doi: 10.1126/science.1223389

Daw, T., and Gray, T. (2005). Fisheries science and sustainability in international policy: A study of failure in the European union's common fisheries policy. *Mar. Policy* 29, 189–197. doi: 10.1016/j.marpol.2004.03.003

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., et al. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* 26, 5872–5895. doi: 10.1111/mec.14350

de Mitcheson, Y. S., and Erisman, B. (2012). *Fishery and biological implications of fishing spawning aggregations, and the social and economic importance of aggregating fishes BT - reef fish spawning aggregations: Biology, research and management*. Eds. Y. Sadovy de Mitcheson and ,. P. L. Colin (Springer Netherlands: Dordrecht), 225–284. doi: 10.1007/978-94-007-1980-4_8

Dennis, D., Plagányi, É., Van Putten, I., Hutton, T., and Pascoe, S. (2015). Cost benefit of fishery-independent surveys: Are they worth the money? *Mar. Policy* 58, 108–115. doi: 10.1016/j.marpol.2015.04.016

De Paoli-Iseppi, R., Deagle, B. E., McMahon, C. R., Hindell, M. A., Dickinson, J. L., and Jarman, S. N. (2017). Measuring animal age with DNA methylation: From humans to wild animals. *Front. Genet* 8. doi: 10.3389/fgene.2017.00106

Dichmont, C. M., Fulton, E. A., Gorton, R., Sporcic, M., Little, R. L., Punt, A. E., et al. (2017). From data rich to data-limited harvest strategies-does more data mean better management? *ICES J. Mar. Sci.* 74, 670–686. doi: 10.1093/icesjms/fsw199

European Union (2013). *Regulation (EU) no 1380/2013 of the European parliament and of the council of 11 December 2013 on the common fisheries policy, amending council regulations (EC) no 1954/2003 and (EC) no 1224/2009 and repealing council regulations (EC) no 2371/2002 and (EC, Official Journal of the European Union. European Union)*.

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., et al. (2017). Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Can. J. Fish. Aquat. Sci.* 74, 1362–1374 doi: 10.1139/cjfas-2016-0306

Friedman, K. J., Bartley, D. M., Rodríguez-Ezpeleta, N., Mair, G. C., Ban, N., Beveridge, M., et al. (2022). "Current and future genetic technologies for fisheries and aquaculture: implications for the work of FAO," in *FAO fisheries and aquaculture circular* (Rome: FAO). doi: 10.4060/cc1236en

Frost, M., Baxter, J., Buckley, P., Cox, M., Dye, S., and Harvey, N. (2012). Impacts of climate change on fish, fisheries and aquaculture. *Aquat. Conserv. Mar. Freshw. Ecosyst.*, 22. doi: 10.1002/aqc.2230

Gillet, B., Cottet, M., Destanque, T., Kue, K., Descloux, S., Chanudet, V., et al. (2018). Direct fishing and eDNA metabarcoding for biomonitoring during a 3-year survey significantly improves number of fish detected around a south East Asian reservoir. *PloS One*. 13, 1–25 doi: 10.1371/journal.pone.0208592

Hansen, B. K., Bekkevold, D., Clausen, L. W., and Nielsen, E. E. (2018). The sceptical optimist: Challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish. Fish.* 19, 751–768 doi: 10.1111/faf.12286

Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H. C., Gough, K. C., et al. (2018). Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (Triturus cristatus). *Ecol. Evol.* 8, 6330–6341 doi: 10.1002/ece3.4013

Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., et al. (2018). Implementation options for DNA-based identification into ecological status assessment under the European water framework directive. *Water Res.* 138, 192–205 doi: 10.1016/j.watres.2018.03.003

Hilborn, R. (2003). The state of the art in stock assessment: Where we are and where we are going. *Sci. Mar.* 67, 15–20 doi: 10.3989/scimar.2003.67s115

Hoydal, K. (2007). Viewpoint: the interface between scientific advice and fisheries managementa. *ICES J. Mar. Sci.* 64, 846–850. doi: 10.1093/icesjms/fsm010

ICES (2014). "Manual for the Baltic international trawl surveys (BITS)," in *Series of ICES survey protocols SISP 7 - BITS* (Copenhagen, Denmark). doi: 10.17895/ice.pub/7580

ICES (2015). *Report of the international bottom trawl survey working group (IBTSWG)* (Copenhagen, Denmark). doi: 10.17895/ices.pub.8617

ICES (2018). *Report of the working group on application of genetics in fisheries and aquaculture (WGAGFA)* (Copenhagen, Denmark). doi: 10.17895/ices.pub.8132

IEO (2018). *PLAN DE CAMPAÑA MEDITS_ES_2018 (GSA5)* Madrid.

Jerde, C. L., Wilson, E. A., and Dressler, T. L. (2019). Measuring global fish species richness with eDNA metabarcoding. *Mol. Ecol. Resour.* 19, 19–22. doi: 10.1111/1755-0998.12929

Kolody, D., and Bravington, M. (2019). *Is close-kin mark recapture feasible for IOTC yellowfin tuna stock assessment. IOTC working party on methods document IOTC–2019–WPM10–25-rev.1* (Hobart, Australia: CSIRO).

Lugg, W. H., Griffiths, J., van Rooyen, A. R., Weeks, A. R., and Tingley, R. (2018). Optimal survey designs for environmental DNA sampling. *Methods Ecol. Evol.* 9, 1049–1059 doi: 10.1111/2041-210X.12951

Martinsohn, J. T., Jardim, E., Coscia, I., and Neat, F. (2015). *Deep Sea? close kin: A genetic approach for improved fisheries management* (Luxembourg (Luxembourg: Publications Office of the European Union.

Martinsohn, J. T., Raymond, P., Knott, T., Glover, K. A., Nielsen, E. E., Eriksen, L. B., et al. (2019). DNA-Analysis to monitor fisheries and aquaculture: Too costly? *Fish. Fish.* 20, 391–401. doi: 10.1111/faf.12343

Mauvisseau, Q., Parrondo, M., Fernández, M. P., García, L., Martínez, J. L., García-Vázquez, E., et al. (2017). On the way for detecting and quantifying elusive species in the sea: The octopus vulgaris case study. *Fish. Res.* 191, 41–48 doi: 10.1016/j.fishres.2017.02.023

Nkrumah-Elie, Y., Elie, M., and Reisdorph, N. (2018). "Chapter 14 - systems biology approaches to asthma management," in *Personalizing Asthma Management for the Clinician*. Eds. S. J. Szefler, F. Holguin, Wechsler, and M.E.B.T.-P.A.M. for the C, (Amsterdam: Elsevier), 151–160. doi: 10.1016/B978-0-323-48552-4.00014-7

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., and 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. Fish.* 16, 125–159. doi: 10.1111/faf.12052

Papa, Y., Oosting, T., Valenza-Troubat, N., Wellenreuther, M., and Ritchie, P. A. (2021). Genetic stock structure of new Zealand fish and the use of genomics in fisheries management: an overview and outlook. *New Zeal. J. Zool.* 48, 1–31. doi: 10.1080/03014223.2020.1788612

Pennino, M. G., Conesa, D., López-Quílez, A., Muñoz, F., Fernández, A., and Bellido, J. M. (2016). Fishery-dependent and -independent data lead to consistent estimations of essential habitats. *ICES J. Mar. Sci.* 73, 2302–2310. doi: 10.1093/icesjms/fsw062

Pita, A., Casey, J., Hawkins, S. J., Villarreal, M. R., Gutiérrez, M.-J., Cabral, H., et al. (2016). Conceptual and practical advances in fish stock delineation. *Fish. Res.* 173, 185–193. doi: 10.1016/j.fishres.2015.10.029

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., et al. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Mol. Ecol.* 25, 527–541 doi: 10.1111/mec.13481

Punt, A. E., Walker, T. I., and Prince, J. D. (2002). Assessing the management-related benefits of fixed-station fishery-independent surveys in australia's southern shark fishery. *Fish. Res.* 55, 281–295. doi: 10.1016/S0165-7836(01)00276-4

Qu, C., and Stewart, K. A. (2019). Evaluating monitoring options for conservation: comparing traditional and environmental DNA tools for a critically endangered mammal. *Sci. Nat.* 106, 1–9 doi: 10.1007/s00114-019-1605-1

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., and Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* 51, 1450–1459 doi: 10.1111/1365-2664.12306

Rodríguez-Rodríguez, G., Ballesteros, H. M., Martínez-Cabrera, H., Vilela, R., Pennino, M. G., and Bellido, J. M. (2021). On the role of perception: Understanding stakeholders' collaboration in natural resources management through the evolutionary theory of innovation. *Sustain.* 13, 1–11 doi: 10.3390/su13063564

Rotherham, D., Underwood, A. J., Chapman, M. G., and Gray, C. A. (2007). A strategy for developing scientific sampling tools for fishery-independent surveys of estuarine fish in new south Wales, Australia. *ICES J. Mar. Sci.* 64, 1512–1516. doi: 10.1093/icesjms/fsm096

Sartori, D., Catalano, G., Genco, M., Pancotti, C., Sirtori, E., Bo, S. V., et al. (2014). *Guide to cost-benefit analysis of investment projects: Economic appraisal tool for cohesion policy 2014-2020*, (Luxembourg: Publications Office of the European Union). doi: 10.2776/97516

Simpson, D. J., and Chandra, T. (2021). Epigenetic age prediction. *Aging Cell* 20, e13452. doi: 10.1111/acel.13452

Smart, A. S., Weeks, A. R., van Rooyen, A. R., Moore, A., McCarthy, M. A., and Tingley, R. (2016). Assessing the cost-efficiency of environmental DNA sampling. *Methods Ecol. Evol.* 7, 1291–1298 doi: 10.1111/2041-210X.12598

Stamatopoulos, C. (2002). "Sample-based fishery surveys," in *A tech. handbook*, vol. 425. (Rome FAO: FAO Fish. Tech. Pap), 1–132.

Stat, M., John, J., DiBattista, J. D., Newman, S. J., Bunce, M., and Harvey, E. S. (2019). Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. *Conserv. Biol.* 33, 196–205 doi: 10.1111/cobi.13183

Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E., and Zakharov, E. V. (2014). Is DNA barcoding actually cheaper and faster than traditional morphological methods: Results from a survey of freshwater bioassessment efforts in the united states? *PloS One*. 9, 1–10 doi: 10.1371/journal.pone.0095525

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., and Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PloS One*. 7, 1–9 doi: 10.1371/journal.pone.0041732

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., and Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PloS One*. 11, 1–22 doi: 10.1371/journal.pone.0165252

Thomsen, P. F., and Willerslev, E. (2015). Environmental DNA - an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183, 4–18 doi: 10.1016/j.biocon.2014.11.019

Underwood, A. J. (1996). *Experiments in ecology: Their logical design and interpretation using analysis of variance* (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511806407

Waples, R. S., Grewe, P. M., Bravington, M. W., Hillary, R., and Feutry, P. (2018). Robust estimates of a high Ne/N ratio in a top marine predator, southern bluefin tuna. *Sci. Adv.* 4, eaar7759. doi: 10.1126/sciadv.aar7759

Ward, R. D. (2000). Genetics in fisheries management. *Hydrobiologia* 420, 191–201. doi: 10.1023/A:1003928327503

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., et al. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Sci. Rep.* 7, 1–12 doi: 10.1038/srep40368

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., et al. (2016). Environmental DNA as a "snapshot" of fish distribution: A case study of Japanese jack mackerel in maizuru bay, Sea of Japan. *PloS One*. 11, 1–18 doi: 10.1371/journal.pone.0149786

Zimmermann, F., and Enberg, K. (2017). Can less be more? effects of reduced frequency of surveys and stock assessments. *ICES J. Mar. Sci.* 74, 56–68. doi: 10.1093/icesjms/fsw134

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 07 November 2022
DOI 10.3389/fmars.2022.1028863

OPEN ACCESS

EDITED BY
Fran Saborido-Rey,
Spanish National Research Council (CSIC), Spain

REVIEWED BY
Antonia Picornell,
University of the Balearic Islands, Spain
Juha Karjalainen,
University of Jyväskylä, Finland

*CORRESPONDENCE
Anti Vasemägi
anti.vasemagi@slu.se

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 26 August 2022
ACCEPTED 06 October 2022
PUBLISHED 07 November 2022

CITATION
López M-E, Bergenius Nord M, Kaljuste O, Wennerström L, Hekim Z, Tiainen J and Vasemägi A (2022) Lack of panmixia of Bothnian Bay vendace - Implications for fisheries management. *Front. Mar. Sci.* 9:1028863. doi: 10.3389/fmars.2022.1028863

Lack of panmixia of Bothnian Bay vendace - Implications for fisheries management

María-Eugenia López[1†], Mikaela Bergenius Nord[2†], Olavi Kaljuste[3], Lovisa Wennerström[3], Zeynep Hekim[3,4], Joni Tiainen[5] and Anti Vasemägi[1,6*]

[1]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Drottningholm, Sweden, [2]Department of Aquatic Resources, Institute of Marine Research, Swedish University of Agricultural Sciences, Lysekil, Sweden, [3]Department of Aquatic Resources, Institute of Coastal Research, Swedish University of Agricultural Sciences, Öregrund, Sweden, [4]Joint Research Centre (JRC), European Commission, Ispra, Italy, [5]Natural Resources Institute Finland, Natural Resources Unit, Helsinki, Finland, [6]Chair of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, Tartu, Estonia

Overexploitation of fisheries is recognized as a major environmental and socioeconomic problem that threats biodiversity and ecosystem functioning. Inappropriate management policies of fish stocks have been applied as a consequence of inadequate characterization of subtle genetic structure in many fish species. In this study, we aim to assess the extent of genetic differentiation and structure of vendace (*Coregonus albula*) collected from eight locations in the Bothnian Bay, the northernmost part of the Baltic Sea. Specifically, we test if this species forms a single panmictic population or is divided into several genetically distinct units. We used restriction site-associated DNA sequencing (RAD-seq) to identify 21,792 SNPs based on 266 individuals. We identified a clear pattern of genetic differentiation between River Kalix and the other sampling locations, and a weak structuring between samples from Swedish and Finnish coast. Outlier analysis detected 41 SNPs putatively under divergent selection, mainly reflecting the divergence between River Kalix and the other samples. The outlier SNPs were located within or near 25 genes, including voltage-dependent calcium channel subunit alpha-2 (CACNA2D2), cadherin 26 (CDH26) and carbonic anhydrase 4-like (CA4) that have earlier been associated with salt-tolerance and salinity stress. Our study provides the first genome-wide perspective on genetic structuring of Baltic Sea vendace and rejects the hypothesis of panmixia in the Bothnian Bay. As such, our work demonstrates the power of RAD-sequencing to detect low but significant genetic structuring relevant for fisheries management.

KEYWORDS

coregonids, effective size, population structure, genetic differentiation, fisheries management, *Coregonus albula*

Introduction

Overexploitation of fisheries has long been recognized as a major environmental and socioeconomic problem, threatening biodiversity and ecosystem functioning (Lotze et al., 2006; Worm et al., 2006; Worm et al., 2009). As a result, a third of commercial fish stocks are being harvested at biologically unsustainable levels and 90% are fully exploited (FAO, 2020), which strongly indicates that current approaches to fisheries management are not sufficiently successful in preventing the depletion of populations (Reiss et al., 2009). Successful management measures require that reproductively isolated populations are governed as independent stocks (Reiss et al., 2009). However, it is frequently assumed, without well-grounded scientific evidence, that the exploited species form single panmictic populations (Euclide et al., 2021). Management measures have consequently been inappropriate and in many cases led to unintentional overexploitation of local populations (Hutchinson, 2008; Li et al., 2015).

Testing for allele frequency differences between groups of individuals to characterize population genetic structuring should be the first step in fisheries management (Waples et al., 2008). However, the scarcity of obvious dispersal barriers, high levels of dispersal and gene flow, large population sizes generating low levels of random genetic drift in many marine species, make it challenging to identify low levels of genetic divergence (Waples, 1998; Allendorf et al., 2010; Waples et al., 2022). In the past, this has led to failed attempts to detect subtle genetic substructure using a limited number of neutral markers despite the ecological evidence for reproductive isolation (Waples, 1998; Allendorf et al., 2010). However, more recent studies have shown that by screening thousands or even millions of polymorphisms spread along the genome and identify specific variants affected by divergent natural selection, it is possible to identify genetic substructure, even when the overall level of genetic divergence across the majority of the genome is extremely low or non-existing (Lamichhaney et al., 2012; Guo et al., 2015; Barrio et al., 2016; Guo et al., 2016; Momigliano et al., 2017). For example, in Baltic herring, genome-wide screening of single nucleotide polymorphisms (SNPs) allowed identification of previously unknown highly divergent regions of the genome. These regions are most likely shaped by local adaptation linked to environmental factors, such as salinity and temperature, despite lack of genetic divergence at other parts of the genome (Guo et al., 2016; Han et al., 2020). Thus, the most promising feature of surveying genome-wide variation is that it provides insights into putative adaptive divergence and enables more precise and effective population delineation often unattainable with other methods (Bradbury et al., 2013).

Vendace (*Coregonus albula*) is a small salmonid fish that belongs to Eurasian cisco complex (Coregoninae) (Mehner et al., 2021; Sendek, 2021). This species inhabits deep and oligotrophic lakes in Western and Northern Europe (Vuorinen and Lankinen, 1978; Sendek, 2021; Karjalainen et al., 2022) and also occurs in the Baltic Sea, in brackish waters of the Bothnian Bay (Bergenius et al., 2013) and in the eastern part of the Gulf of Finland (Sendek, 2012). The southward distribution of vendace in the Baltic is limited by the increased salinity but it also occurs in estuaries. The upper salinity range of vendace is about 2 to 3 PSU (Lehtonen, 1981) and the Bothnian Bay salinity levels can increase to over 3 PSU in the southern part. Vendace is a short-lived species and can reach maturity already in their first year of life (Bergenius et al., 2013). Vendace spawns from October to December in river estuaries and shallow coastal areas (Lehtonen, 1981) and is known for its strong and unpredictable stock fluctuations caused by large fluctuations in recruitment (Karjalainen et al., 2000). Besides spawning stock biomass, hydro-climatic factors, in particular winter temperature and salinity, have shown to be important variables influencing the temporal variability of recruitment of vendace in the Bothnian Bay (Bergenius et al., 2013). Furthermore, Bergenius et al. (2013) demonstrated that recruitment variability in Bothnian Bay vendace is determined by density-independent factors (Karjalainen et al., 2000; Marjomäki, 2003). The role of trophic factors influencing vendace abundance in the Bothnian Bay have significantly changed with the increasing number of seals in the Bothnian Bay since the beginning of the 90s. The yearly consumption of vendace by seals is larger than the Finnish and Swedish landings combined (Lundström et al., 2014). The vendace in the Swedish part of the Bothnian Bay is mainly fished with pair bottom trawling for its roe, but also a smaller part of the fillets is sold for consumption (Bergenius et al., 2013). The main fishery takes place within the Luleå, Råneå and Kalix archipelagos during five weeks (as maximum) in September and October before spawning. The Swedish vendace roe fishery is one of the economically most important coastal fisheries in Sweden (Bergenius et al., 2018). In Finland, the vendace fishery is a small scale trawl and trap net fishery with increasing catches from early 2000's to recent years (373 tn in year 2021) (Luke, 2022). Both the roe and fillets are sold for consumption, but as in Sweden, the most of the vendace catch are taken in September and October.

Vendace in the Bothnian Bay (ICES Subdivision 31) is currently assessed and managed as two separate entities, one off the coast of Sweden and one off Finland. Vendace in Sweden has since 2017 been managed according to the concept of Fmsy (maximum sustainable yield) and the fishery regulated by a total allowable quota. The spawning stock biomass of vendace in the Swedish part of the Bothnian Bay peaked in 2003-2004 and 2013- 2014, but has in recent times decreased to about half of what it was during those peaks. The likely explanations for the decrease are weaker recruitment of juvenile fish and increased seal predation (Sundelöf et al., 2022). There are no specific management regulations in place for vendace in Finland and no stock assessment has been conducted. It is only since 2019 that Finland is obliged to collect biological samples and information

about discards as part of the EU's data collection framework also for the vendace in the Bothnian Bay.

The management of vendace in the Bothnian Bay is solely based on the national boundaries between Sweden and Finland, and has in this way been managed as two separate stocks. However, the movement patterns of vendace in the Bothnian Bay are not well understood. A tagging study conducted in the Luleå and Kalix archipelagos in the 1960s and 1970s show that vendace undertake natal homing, i.e. the adults return to their birthplace archipelago to reproduce (Enderlein, 1977; Enderlein, 1986). The studies also show that vendace migrates eastwards in summer to feed in more nutritious waters, during which individuals from different estuaries and bays mix (Enderlein, 1986). Furthermore, we currently do not know whether the spawning aggregations of vendace distributed along the coast of the Bothnian Bay are genetically structured into different sub-populations and thus potentially require a more fine-scale management than what we have today.

Restriction site-associated DNA sequencing (RAD-seq) is a cost-effective and flexible genome complexity reduction technique that enables to effectively screen tens to hundreds of thousands of regions in the genome (Davey et al., 2011; Bruneaux et al., 2013). RAD-seq therefore enables simultaneous characterization of both neutral and adaptive patterns of genetic variation (Andrews et al., 2016) and is particularly suitable for delineating low levels of population genetic differences relevant for exploited fish species with large population sizes (Pujolar et al., 2014; Guo et al., 2015; Euclide et al., 2021). In this study, we screened 21,792 SNPs in 266 vendace from Gulf of Bothnia, Baltic Sea collected at eight spawning locations to test if this economically and ecologically important species forms a single panmictic population or is divided into several genetically distinct units in the Bothnian Bay. We also tested if some loci deviate from neutral genetic patterns and show elevated genetic differentiation indicative of divergent natural selection. To identify potential targets of selection and understand their function, we carried out functional annotation of putative outlier loci. We discuss our results in the light of current fisheries management practices of Bothnian Bay vendace and identify key knowledge gaps for future research.

Methods

Sampling

Vendace were sampled from spawning grounds of the Bothnian Bay during spawning time in autumn 2019 and 2020. In 2019, samples from four sites were collected from passive gears (gillnets and trapnets) in mid-October to guarantee the catch of local spawning fish; two sites located in mouths of Piteå and Kalix rivers (River Pite and River Kalix, Figure 1) and two sites in the coastal area off Piteå and Kalix (Piteå coastal 1 and Kalix coastal, Figure 1). In 2020, two additional samples were collected from survey trawl catches in mid-October, comprising two sites in the coastal area off Piteå and Luleå (Piteå coastal 2 and Luleå coastal, Figure 1). Vendace were also sampled from two locations off Oulu in Finland in 2020 (Oulu costal 1 and Oulu costal 2, Figure 1). These samples were taken from the trawled catch in the harbour during spawning season (October and November).

From each site, 50 ripe individuals, either in their final stage of maturation or with running roe or sperm, were randomly collected and frozen immediately. Only from one site (Oulu coastal 1) did the sample contain of mostly small immature

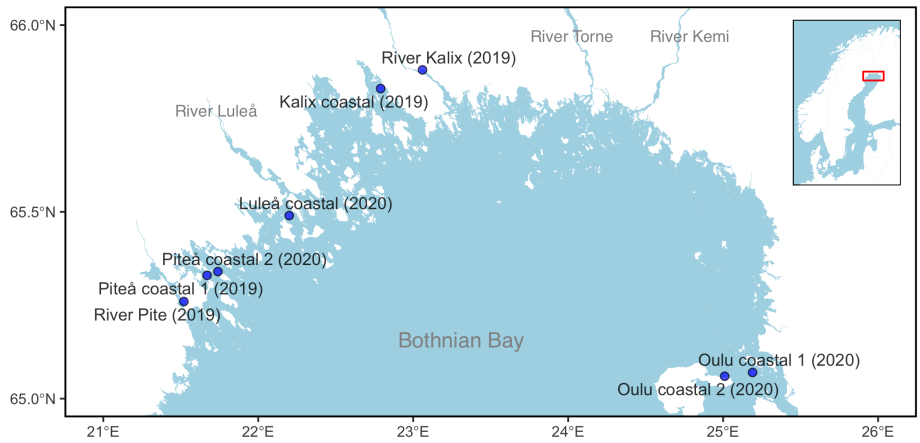


FIGURE 1
Geographical position of sampling locations for vendace. The main map shows the eight sampling sites in Gulf of Bothnia. The inset map gives an overview of the Scandinavian Peninsula and the location of the sampled area.

individuals. In the laboratory individuals were thawed and a small piece of the muscle tissue (ca 0.5 cm^3) was taken as soon as possible to ensure minimum degradation of the DNA. The small piece of muscle tissue was taken using a clean scalpel and stored in 95% ethanol.

Genome complexity reduction, sequencing and genotyping

A double-digestion restriction-site associated DNA (ddRAD) approach was employed for SNP discovery and *de novo* genotyping (Peterson et al., 2012). PstI and ApeKI were used for restriction digestion and a paired-end sequencing (2 × 150 bp) was carried out on an Illumina NextSeq 500/550 v2 platform by LGC Genomics GmbH (Berlin, Germany). SNP calling was carried out using Stacks v2.59 program (Catchen et al., 2013). Detected loci were filtered with Stacks v2.59 *populations* program setting option –r to 0.8 (minimum percentage of individuals per population required to process a locus), option -p 8 (minimum number of populations where a locus must be present) and option –min-maf to 0.05 (Minor allele frequency cutoff). For further analyses, additional filtering steps were applied with PLINK (Purcell et al., 2007), including samples call rate (–mind 0.2), global SNP call rate (–geno 0.1), and deviation from Hardy-Weinberg equilibrium (–hwe 2e-06) for all individuals.

Genetic diversity, structure and N_e estimation

We measured genetic diversity as the observed (H_O) and expected heterozygosity (H_E) calculated using *summary* function in the package R/adegenet (Jombart, 2008; Jombart et al., 2010). We also calculated the inbreeding coefficient (F_{IS}) and the number of private allele using the functions *basic.stats* in R/hierfstat (Goudet, 2005) and *private_alleles* in R/poppr (Kamvar et al., 2014), respectively. The genetic differentiation between samples was quantified using pairwise unbiased F_{ST} estimator (Weir and Cockerham, 1984) calculated using StAMPP R package (Pembleton et al., 2013). Significance of F_{ST} values and 95% confidence intervals were computed using bootstrap methods as implemented in StAMPP R. The number of population clusters was visualized using a discriminant analysis of principal components (DAPC) using the r package adegenet (Jombart, 2008; Jombart et al., 2010). We used the "optim.a.score" function to identify the best number of principal components (PCs) to retain. Too many or too few PCs can lead to low repeatability of results and over- or underfitting the data, respectively (Jombart et al., 2010). We also used the Bayesian clustering method implemented in STRUCTURE to infer genetic structure (Pritchard et al., 2000). All runs were made using PARALLELSTRUCTURE (Besnier and Glover, 2013). We performed three independent runs for each value of K, with 10,000 MCMC permutations and a burn-in of 1,000 permutations. We evaluated K = 1–8, with the largest K reflecting all eight sampling locations. STRUCTURE HARVESTER (Earl and VonHoldt, 2012) was used to infer the most likely value of K using the Evanno's method (Evanno et al., 2005). We calculated the effective population size (N_e) using the linkage disequilibrium (LD) method implemented in program NeEstimator 2.01 (Do et al., 2014). This method is based on a random mating model and assumes that all loci are physically unlinked (Waples, 2006). We obtained the N_e estimates and parametric 95% confidence intervals for each population a minimum allele frequency cutoff of 0.05 using 1650 SNPs filtered by call rate of 99%; outlier loci detected by BayeScan v.2.1 were also discarded from this analysis. We assessed the relationship between F_{ST} and migration rate (m) based on Wright (1943) island model of migration. We calculated F_{ST} as $1/[(4^*N_e^*m) +1)]$ by using the N_e estimates in each sample and several values of migration rates (m) from 0.02 to 0.4 in order to test possible demographic independence.

Signatures of selection

We identified candidate loci under selection using the Bayesian likelihood method implemented in BayeScan v.2.1, which uses differences in allele frequencies between populations to estimate the posterior probability of loci experiencing selection (Foll and Gaggiotti, 2008). The algorithm uses a reversible-jump Markov Chain Monte Carlo to explore models with or without selection, and applies a Bayes factor for two models: one assuming selection and another assuming neutrality given the data (Foll and Gaggiotti, 2008). Bayescan was carried out in two separated runs: (1) among the eight sampling locations and (2) among seven sampling sites with Kalix River excluded. We used BayeScan v.2.1 default parameters (Prior odds for the neutral model = 10). We considered candidate loci under selection as those that presented a Bayes factor of at least 32 (-log10 = 1.5) and a positive value of α (directional selection), corresponding to a posterior probability of 0.97 and expected as being "very strong" evidence of selection (Foll, 2012).

Gene annotation

Because of the lack of a reference genome of vendace, RAD loci harboring SNP putatively under selection were blasted against the reference genome of Swiss Alpine whitefish (*Coregonus* sp. *"Balchen"*) (GenBank: GCA_902810594.1) (De-Kayne et al., 2020) and lake whitefish (*Coregonus clupeaformis*) (GenBank: GCA_020615455.1) (Pasquier et al., 2016) using

BLAST® Command Line Applications 2.12.0 (NCBI and Camacho, 2008). The aligned sequences were then analyzed with BEDTools/2.29.2 adding 5000 bp upstream/downstream (Quinlan and Hall, 2010) to search genes within these regions using the reference genome annotation files (GFF).

Results

Sequencing and genotyping

Eighteen individuals with the lowest read count were excluded from Stacks analyses. Therefore, we used 270 individuals for SNP calling process. Mean locus coverage across all samples was 11.9x, ranging from 5.3x to 19.4x, with mean length of reads: 206.64 bp. Altogether, 21,792 variants and 266 individuals passed all filters and quality control steps and were used for subsequent population genetic analyses.

Genetic diversity, structure and N_e estimates

Heterozygosity estimates showed similar level of diversity among studied samples (Table 1). Slightly higher expected heterozygosity compared to observed heterozygosity suggested potentially some degree of inbreeding, which was shown by F_{IS} values that ranged 0.019 to 0.026. No private alleles were detected for any of the samples. The eight sampling locations yielded 28 possible pairwise F_{ST} comparisons of which 23 were significant (p-value < 0.05) ranging from 0.003 to 0.0096 (Figure 2 and Table 2). Additionally, the analysis of statistical power using POWSIM (Ryman and Palm, 2006) showed that the sample sizes and SNP markers used in this study were adequate for detecting very low level of genetic differentiation (F_{ST} = 0.0011) with a high probability (P ~ 1)). The highest genetic differentiation was observed between River Kalix and Oulu costal 2 locations (F_{ST}= 0.0096) and River Kalix sample was significantly differentiated from all other studied locations. This pattern was also evident from the DAPC plot (Figure 3), where most of individuals from River Kalix formed a differentiated cluster. In addition, individuals from Oulu costal 1 and 2 locations showed weak separation from the rest of the samples along the second axis of DAPC. Using the STRUCTURE analysis, the Evanno's method suggested the most likely number of genetic clusters was K=3, although K=2 showed a visually more distinct result (Figure 4 and Supplementary Table S1). Consistent with the DAPC and F_{ST}, STRUCTURE revealed that vendace from River Kalix were all clearly differentiated from the rest of the samples, showing a more homogenous pattern in terms of their ancestral proportions, being dominated by one cluster. Estimates of N_e with the RAD-derived SNPs varied across populations, ranging from close to 1288 (95%CI 1004-1794) in Oulu coastal 2 to 3439 (95%CI 1923-16019) for the Kalix coastal sample (Table 1). These estimates were calculated using SNPs that were filtered by call rate of 99%, providing over 1,000,000 pairwise comparisons between loci. Based on Wright (1943) island model of migration, the observed N_e estimates indicated that if the effective population size exceed 3000, low migration rates (m = 0.02-0.1) generate very low levels of differentiation (F_{ST} = 0.004-0.0008) (Figure 5). In contrast, if N_e <1000, low migration rates (m = 0.02-0.1) are expected to result in higher levels of differentiation (F_{ST} = 0.0025-0.012). Thus, the interpretation of low levels of divergence depends critically on the effective population sizes of studied populations.

Divergent selection and identification of outlier SNPs

BayeScan analysis detected 41 SNPs putatively under divergent selection when all samples were included to the analysis (Figure 6). However, when the River Kalix sample was removed, no SNP showed evidence of selection (Figure S1). Thus, all identified putative outliers reflected differences between River Kalix and other samples. Outlier loci were further investigated to identify genes by blasting the RAD loci against the Swiss Alpine whitefish and lake whitefish genome reference (S2, S3). The outlier SNPs genes were located within or near 25 genes, including voltage-dependent calcium channel subunit alpha-2 (CACNA2D2), cadherin 26 (CDH26) that have been associated with salt-tolerance in treefrog (Albecker et al., 2021),

TABLE 1 Observed (H_O), expected (H_E) heterozygosities, F_{IS} and effective population size (N_e) estimates for the studied vendace samples.

	Coordinate N	Coordinate E	N	H_O	H_E	F_{IS}	N_e(95%CI)
River Pite	65.26°	21.52°	36	0.272	0.274	0.020	2145 (1475-3914)
Piteå coastal 1	65.33°	21.67°	35	0.272	0.273	0.019	1789 (1273-3000)
Piteå coastal 2	65.34°	21.74°	35	0.271	0.273	0.021	1823 (1297-3056)
Luleå coastal	65.49°	22.20°	36	0.270	0.273	0.023	1432 (1094-2065)
Kalix coastal	65.83°	22.79°	34	0.270	0.272	0.020	3439 (1923-16019)
River Kalix	65.88°	23.06°	35	0.266	0.271	0.026	1449 (1084-2180)
Oulu coastal 1	65.07°	25.19°	19	0.266	0.267	0.026	2221 (1068-∞)
Oulu coastal 2	65.06°	25. 01°	36	0.268	0.272	0.025	1288 (1004-1794)

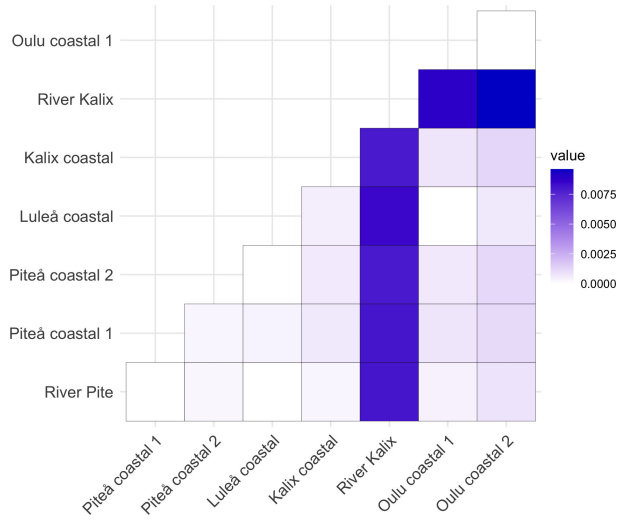


FIGURE 2
Heatmap for the pairwise F_{ST} between populations.

and the carbonic anhydrase 4-like (CA4) gene that showed to be upregulated in salinity stress in hybrid tilapia (Su et al., 2020). Genes within or adjacent (5000 bp) to the identified outlier loci for each reference genome are shown on Supplementary Files (S4, S5) and Figure 6.

Discussion

The successful management of fisheries depends on an adequate identification of biological populations, combined with the inclusion of spatial distribution information of the populations into management practices (Heath et al., 2014; Bernatchez et al., 2017). Pelagic species, however, that display certain life-history traits, such as high fecundity, large population sizes and high dispersal potential are expected to produce weak patterns of genetic differentiation (Ward et al., 1994; Waples, 1998; Palumbi, 2003; Hedgecock et al., 2007). This can lead to uncertainties about the level of structuring and gene flow between the populations, thus making sustainable fisheries management challenging (Cano et al., 2008). Here, we first evaluated the occurrence of a single panmictic population of vendace in the Bothnian Bay using tens of thousands of SNPs. After the rejection of panmixia, we further tested if some loci deviate from neutral genetic patterns indicative of putative

TABLE 2 Genetic differentiation measured as pairwise F_{ST} (Weir and Cockerham, 1984) and its associated *p*-values based on 21,792 SNPs.

	River Pite	Piteå coastal 1	Piteå coastal 2	Luleå coastal	Kalix coastal	River Kalix	Oulu coastal 1
Piteå coastal 1	0.0000	–	–	–	–	–	–
p-value	0.4357	–	–	–	–	–	–
Piteå coastal 2	0.0003	0.0003	–	–	–	–	–
p-value	**0.0403**	**0.0320**	–	–	–	–	–
Luleå coastal	0.0002	0.0003	-0.0001	–	–	–	–
p-value	0.1546	**0.0157**	0.7464	–	–	–	–
Kalix coastal	0.0003	0.0007	0.0006	0.0005	–	–	–
p-value	**0.0324**	**0.0000**	**0.0003**	**0.0009**	–	–	–
River Kalix	0.0081	0.0081	0.0080	0.0084	0.0080	–	–
p-value	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	–	–
Oulu coastal 1	0.0004	0.0008	0.0006	0.0003	0.0008	0.0089	–
p-value	**0.0301**	**0.0005**	**0.0031**	0.0852	**0.0014**	**0.0000**	–
Oulu coastal 2	0.0008	0.0011	0.0012	0.0007	0.0012	0.0096	0.0002
p-value	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	0.1957

Bold values denote statistical significance at the $p < 0.05$ level.

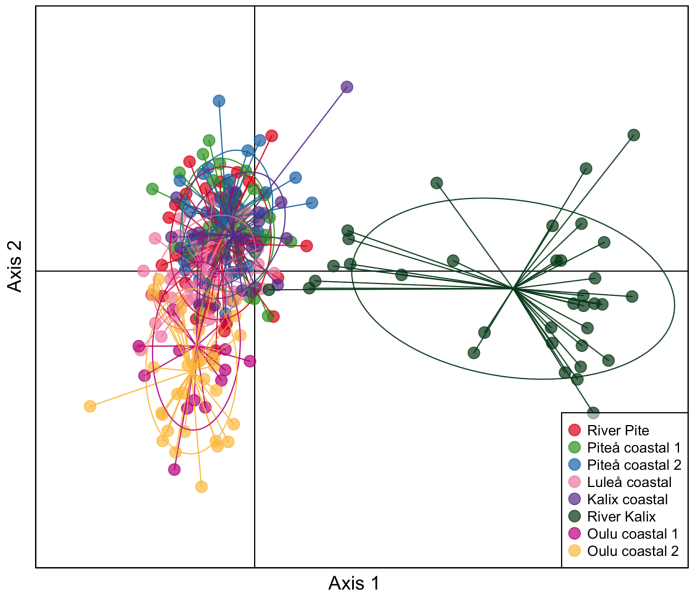


FIGURE 3
Discriminant Analysis of Principal Components (DAPC) of genetic differentiation of vendace. Individuals from different locations are represented by different colours.

adaptive divergence between populations driven by natural selection. In order to shed light on how much migration (m) is needed to generate the observed levels of low divergence between populations, we estimated the effective population sizes using linkage disequilibrium and inferred m based on Wright (1943) island model of migration. We discuss our findings in the light of demographic dependence, which has been suggested to occur if m>0.1 (Hastings, 1993) and the implications of low level of divergence to conservation and management of vendace in the Bothnian Bay.

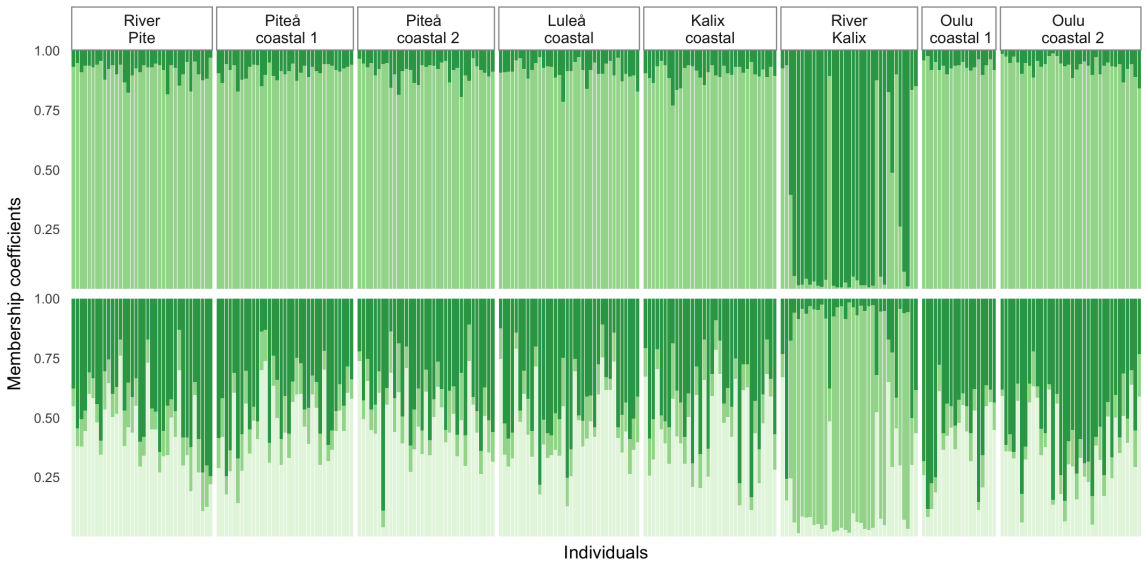


FIGURE 4
STRUCTURE estimated individual membership coefficients (q) of vendace for K=2 and K=3. Each color represents a cluster, and the ratio of vertical lines represent the membership coefficients for an individual.

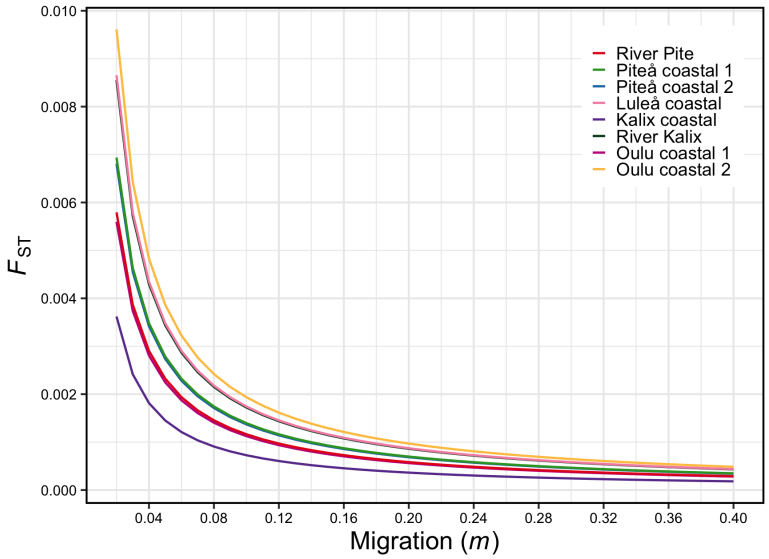


FIGURE 5
Relationship between F_{ST} and migration (m) calculated from N_e estimates of each sample of vendace used in this study. These calculations were based on Wright (1943) island model of migration.

Lack of panmixia and varying levels of divergence

Our study revealed weak but significant genetic structuring among studied samples of vendace in Bothnian Bay, ruling out the presence of panmixia. The most notable divergence was observed between River Kalix and the rest of the samples range (F_{ST} = 0.0080-0.0096). The Kalix river is one of the few large (drainage area 18 130 km^2, runoff 280 m^3) and unexploited (in terms of hydroelectric power) rivers in Sweden. The drop height is relatively even with many small rapids and falls. The genetic difference between the River Kalix, the Kalix coastal sample and

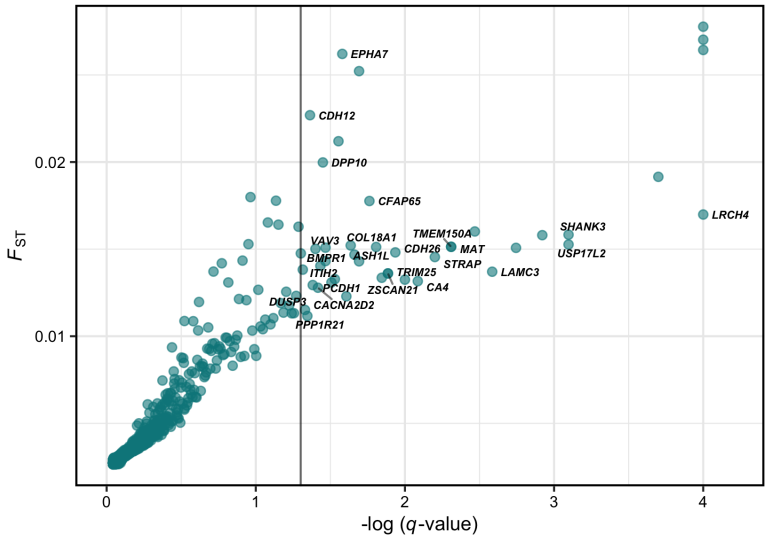


FIGURE 6
Analysis of divergent selection based on BayeScan. Points represent individual SNPs and their respective -log10(q-values) and F_{ST} values together with its closest gene. The vertical line represents the negative logarithm of a false discovery rate (FDR) threshold on 0.05 (1.30103). Dots greater than -log10(0.05) represent outlier loci.

the rest of the samples point to a complexity in the population structure of vendace not currently captured by the current management regime. The genetic differentiation found between the River Kalix and the other coastal samples likely reflects the relative isolation of the former, suggesting that lower reaches of rivers in the Bothnian Bay may support anadromous vendace populations which are genetically divergent from vendace spawning in coastal areas. In contrast to River Kalix, River Pite individuals were sampled very close to the river mouth, therefore, these samples may be closer to the coastal spawners than to vendace spawning in the river. Furthermore, outlier analysis revealed that the observed genetic divergence between River Kalix and other samples may be affected by divergent selection associated with the river environment. This was supported by identification of 41 outlier SNPs, from which, several of them were located near CACNA2D2 (voltage-dependent calcium channel subunit alpha-2), CDH26 (cadherin 26) and CA4 (carbonic anhydrase 4-like) genes that have previously been associated to salinity related traits in other species (Su et al., 2020; Albecker et al., 2021). For example, the osmoregulatory genes CACNA2D2 and CDH26 showed to be differentially expressed in costal populations of treefrog (i.e., populations with long-term saltwater exposure), while the CA4 gene showed to be upregulated in tilapia under salinity stress playing a role in osmoregulation-related signalling pathways. Thus, these genomic regions in vendace may be associated with divergent selection related to low salinity riverine environment. However, further analysis of other potential anadromous river populations in the Gulf of Bothnia is needed to confirm the occurrence of genetic divergence related to river-spawning and to further understand the impact of selective forces on the vendace genome. It may be that other large rivers, like the River Luleå, have their own genetically distinct population(s) of vendace. As the commercial fishery for vendace during spawning time is prohibited in Swedish rivers, it is possible that fisheries induced mortality is lower for those particular populations compared to coastal spawners.

From genetic differentiation to estimation of connectivity

The second most important finding of our study was the presence of low but statistically significant genetic differentiation (F_{ST} = 0.0004-0.0012) between vendace collected from different coastal areas in Sweden and Finland. Further, power analyses indicated our dataset has sufficient ability to uncover weak genetic differentiation. On one hand, the lack of genetic divergence or very low level of differentiation was expected for vendace given its pelagic lifestyle, large population size and dispersal ability, in waters with few physical barriers to gene flow. Such ecological attributes, shared among pelagic, as supposed to demersal, species are generally associated with low levels of genetic differentiation (Hauser and Ward, 1998). Unfortunately, any given F_{ST} value is consistent with a range of migration rates and therefore, knowledge on genetic divergence alone is insufficient for assessing demographic connectivity (Waples et al., 2022). Therefore, we inferred m based on Wright (1943) island model of migration [$F_{ST} \approx 1/(4N_e m + 1)$] by estimating the effective population sizes using linkage disequilibrium (Do et al., 2014). Our N_e estimates ranged from 1288 to 3439 which suggests that the migration rates do not necessarily need to exceed 0.1, the putative threshold suggested by Hastings (1993) for demographic independence (i.e., populations experiencing less migration are expected to act as demographically independent) to reach the current values of F_{ST}. Therefore, the observed low level of genetic divergence combined with large effective population size estimates may be associated with low or moderate levels of migration (Figure 5), indicative of some level of demographic independence. The occurrence of multiple genetically distinct populations is further supported by earlier mark-recapture studies which have shown strong homing of vendace to specific spawning grounds in the Bothnian Bay (Enderlein, 1989). However, since our N_e estimates were associated with considerable uncertainty, future studies based on larger sample sizes are needed to more precisely quantify the effective population size of the Bothnian Bay vendace. Furthermore, the Wright's Island model used here to translate F_{ST} and N_e to estimates of migration is based on many simplifying assumptions compared with real populations (Spies et al., 2018). Finally, in contrast to River Kalix, the outlier analyses did not provide support for adaptive divergence among coastal samples. This suggests that the observed low level of divergence may be driven by random genetic drift and not by selection. However, since the analysed 20 000 SNPs only cover a proportion of the variation in the whole genome of vendace, we cannot exclude the possibility that other regions of the genome still harbour variants influenced by divergent selection.

In contrast to the current genome-wide analysis, earlier population genetic studies on vendace have typically been carried out using small numbers of highly variable markers, such as microsatellites. For example, Delling et al. (2014) detected a clear genetic structure in samples collected from 23 localities in Swedish lakes, which grouped into two different clusters, with the mean F_{ST} among lakes exceeding 0.15 based on 9 microsatellite loci. Based on 13 microsatellites, Karjalainen et al. (2022) found low to moderate genetic divergence (F_{ST} = 0.005-0.059) between the vendace populations of Kymijoki and Vuoksi drainage systems in Finland. Within the highly spatially structured Lake Saimaa system, Karjalainen et al. (2022) also detected low but significant genetic divergence between some local populations (F_{ST} = 0.01-0.011). However, despite low levels of genetic structuring consistent with reduced connectivity, Karjalainen et al. (2022) advised that because of similar life history and low genetic diversity, management of local vendace populations of Saimaa to be carried out as one management unit

while allocating annually flexible fishing effort to the basins harbouring the strongest exploitation of local populations.

Management implications

The main rationale of this study was to investigate how genetic structure aligns with the current management units for vendace in the Bothnian Bay. Our results provide information that support low genetic divergence between River Kalix and all other samples, as well as weaker divergence between Finnish and Swedish costal locations. Genetic data are usually used to estimate genetic connectivity i.e., the degree to which gene flow affects evolutionary processes within populations. However, genetic methods alone provide limited information on demographic connectivity, i.e., the degree to which population growth and vital rates are affected by dispersal (Lowe and Allendorf, 2010). Furthermore, because of the nonlinear relationship between F_{ST} and N_e, estimations of demographic independence become straightforward only at moderate or large F_{ST} (Waples et al., 2022). Conversely, when genetic divergence is low, one cannot dismiss that there is significant isolation of stocks on timescales relevant to fisheries managements (Hauser and Carvalho, 2008). Hence, even though the divergence in the studied populations was low, it may still support separate management of vendace in the eastern and western Bothnian Bay, since we cannot rule out that these stocks are demographically independent. Furthermore, in the context of genetic conservation, the consequences of "oversplitting" are thought to be negligible, that is, there are no obvious costs if genetically homogeneous group of individuals are managed as if consisting of multiple divergent populations (Laikre et al., 2005). Tagging studies from the 1960s and 70s do in fact, show that vendace from the Swedish coast migrates eastwards, to feed in more nutritious waters, and on occasions all the way to the Finish coast (Enderlein, 1986). With the climate related changes of the biota in the Bothnian Bay during the last decades, including changes in potential food resources for vendace (Pekcan-Hekim et al., 2016) it is, however, possible that also the migration behaviour of vendace has changed since then. In addition, the location of the border between the two management areas is currently pragmatically following the jurisdictional boundary between the two countries. It may well be that the border for a potential eastern and western vendace stock in the Bothnian Bay should be placed somewhere else, for which the management of vendace would require an international agreement. The main ecological risk of not accounting for groups of individuals or stocks that are spatially distinct and have different productivities is the over harvesting of local spawning components, potentially resulting in reduced productivity and biodiversity combined with deteriorated local and regional stock dynamics (Ricker, 1958; Smedbol and Stephenson, 2001; Kerr et al., 2017). Moreover, not accounting for the degree of connectivity between local spawning components in stock assessments can result in the suboptimal use of the resource (Paulik et al., 1967).

Study design and methodological considerations

The observed level of genetic structuring observed among small number of samples clearly suggest the need for more extensive sampling and population genomic analysis of the Bothnian Bay vendace, as well as increasing the geographic scope including other regions of the Baltic Sea. Furthermore, our results indicate the importance of including rivers and estuaries, as potential habitats supporting genetically diverged Baltic vendace populations. As a result, it may well be that other rivers sustain local spawning populations that were not analysed in this study. Therefore, future work incorporating multidisciplinary analyses such as genomics, biometrics, geostatistics, oceanography as suggested in other studies (Abaunza et al., 2008; Cadrin, 2010; Zemeckis et al., 2014; McKeown et al., 2015; Mapp et al., 2017; McKeown et al., 2017; Cadrin, 2020) could yield additional insights into dispersal and connectivity of this species.

In addition, as no reference genome is available for vendace, the genomic positions of the SNPs of this study and gene annotation are still unknown. Therefore, a future reference genome assembly of this species will enable SNP discovery with greater power, and evade problems commonly met with the *de novo* RADseq SNP discovery (Díaz-Arce and Rodríguez-Ezpeleta, 2019). Moreover, future work incorporating whole genome resequencing data are expected to reduce ascertainment bias (Lachance and Tishkoff, 2013) allowing more comprehensive characterization of intra-specific genetic variation in vendace. The use of multiple statistical approaches including incorporation of environmental data, *i.e.* genetic-environment association analyses (GEAs) will likely provide more detailed understanding of key environmental factors driving local adaptation in vendace.

Conclusion

By utilizing 21,792 SNP loci, the present study provides the first genome-wide perspective on genetic differentiation in vendace populations within the northern most part of the Baltic Sea, the Bothnian Bay. Although the overall genome-wide genetic differentiation through the system studied here was low, we detected statistically significant genetic differentiation between a river and coastal samples, suggesting for the first time that at least some large river systems in Bothnian Bay may support anadromous vendace populations that are genetically divergent from vendace spawning in coastal areas. Accordingly, the results provide evidence for divergent selection in loci

potentially associated to local adaptation in the river environment. We have also shown here that a weak level of genetic differentiation combined with large effective population sizes may reflect the occurrence of two demographically separated populations of vendace on the Finnish and Swedish coast. However, more comprehensive sampling is needed to identify more accurately the population boundaries. Overall, our findings present new insights into evolutionary processes that shape vendace populations in Bothnian Bay and are expected to assist in the refinement of management policies applied to this species.

Data availability statement

The data presented in the study are deposited in the Dryad Digital Repository, accession number https://doi.org/10.5061/dryad.m37pvmd5q.

Ethics statement

The animal study was reviewed and approved by Swedish Board of Agriculture (Jorbruksverket).

Author contributions

MBN, OK, ZH, and AV conceived and designed the study. M-EL performed bioinformatics and genetic analyses. M-EL, MBN, and AV drafted the manuscript. MBN, OK and ZH acquired the financial support for the project leading to this publication. MBN, ZH, LW, JT, and OK revised the manuscript critically and provided important intellectual content. AV supervised the work. All authors significantly contributed to the improvement and approved the final manuscript.

Funding

This study was partially funded by the Swedish Agency for Marine and Water Management (projects 1706-2020, 2912-2020, and 1788-21).

Acknowledgments

The computations and data handling were enabled by resources in project SNIC 2022/5-147 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. We also thank Konrad Karlsson for his help with drawing the geographic map.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2022.1028863/full#supplementary-material

References

Abaunza, P., Murta, A., Campbell, N., Cimmaruta, R., Comesana, A., Dahle, G., et al. (2008). Considerations on sampling strategies for an holistic approach to stock identification: The example of the HOMSIR project. *Fish. Res.* 89 (2), 104–113. doi: 10.1016/j.fishres.2007.09.020

Albecker, M. A., Stuckert, A. M., Balakrishnan, C. N., and McCoy, M. W. (2021). Molecular mechanisms of local adaptation for salt-tolerance in a treefrog. *Mol. Ecol.* 30 (9), 2065–2086. doi: 10.1111/mec.15867

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11 (10), 697–709. doi: 10.1038/nrg2844

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17 (2), 81–92. doi: 10.1038/nrg.2015.28

Barrio, A. M., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., et al. (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *elife* 5, e12081. doi: 10.7554/eLife.12081

Bergenius, M. A., Gårdmark, A., Ustups, D., Kaljuste, O., and Aho, T. (2013). Fishing or the environment–what regulates recruitment of an exploited marginal vendace (Coregonus albula (L.)) population? *Adv. Limnol.* 64, 57–70. doi: 10.1127/1612-166X/2013/0064-0029

Bergenius, M., Ringdahl, K., Sundelöf, A., Carlshamre, S., Wennhage, H., and Valentinsson, D. (2018). "Atlas över svenskt kust-och havsfiske 2003-2015," in *Aqua reports*, vol. 3. (Drottningholm Lysekil Öregrund: Sveriges lantbruksuniversitet, Institutionen för akvatiska resurser).

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32 (9), 665–680. doi: 10.1016/j.tree.2017.06.010

Besnier, F., and Glover, K. A. (2013). ParallelStructure: AR package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PloS One* 8 (7), e70651. doi: 10.1371/journal.pone.0070651

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., et al. (2013). Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Appl.* 6 (3), 450–461. doi: 10.1111/eva.12026

Bruneaux, M., Johnston, S. E., Herczeg, G., Merilä, J., Primmer, C. R., and Vasemägi, A. (2013). Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Mol. Ecol.* 22 (3), 565–582. doi: 10.1111/j.1365-294X.2012.05749.x

Cadrin, S. X. (2010). Interdisciplinary analysis of yellowtail flounder stock structure off new England. *Rev. Fish. Sci.* 18 (3), 281–299. doi: 10.1080/10641262.2010.506251

Cadrin, S. X. (2020). Defining spatial structure for fishery stock assessment. *Fish. Res.* 221, 105397. doi: 10.1016/j.fishres.2019.105397

Camacho, C, Madden, T, Ma, N, Tao, T, Agarwala, R, Morgulis, A, et al (2008). *BLAST (r) command line applications user manual* (National Center for Biotechnology Information (US).

Cano, J. M., Shikano, T., Kuparinen, A., and Merilä, J. (2008). Genetic differentiation, effective population size and gene flow in marine fishes: Implications for stock management. *J. Integr. Field Biol.* 5, 1–10.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22 (11), 3124–3140. doi: 10.1111/mec.12354

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12 (7), 499–510. doi: 10.1038/nrg3012

De-Kayne, R., Zoller, S., and Feulner, P. G. (2020). A *de novo* chromosome-level genome assembly of coregonus sp."Balchen": One representative of the Swiss alpine whitefish radiation. *Mol. Ecol. Resour.* 20 (4), 1093–1109. doi: 10.1111/1755-0998.13187

Delling, B., Palm, S., Palkopoulou, E., and Prestegaard, T. (2014). Genetic signs of multiple colonization events in Baltic ciscoes with radiation into sympatric spring-and autumn-spawners confined to early postglacial arrival. *Ecol. Evol.* 4 (22), 4346–4360. doi: 10.1002/ece3.1299

Díaz-Arce, N., and Rodríguez-Ezpeleta, N. (2019). Selecting RAD-seq data analysis parameters for population genetics: The more the better? *Front. Genet.* 10, 533. doi: 10.3389/fgene.2019.00533

Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., and Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol. Ecol. Resour.* 14 (1), 209–214. doi: 10.1111/1755-0998.12157

Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4 (2), 359–361. doi: 10.1007/s12686-011-9548-7

Enderlein, O. (1977). Tre siklöjemärkningar. *Inf. Inst. Freshw. Res.* (Drottningholm) 1, 16.

Enderlein, O. (1986). Vendace (Coregonus albula (L.)) in the bothnian bay, the Baltic. *Inf. Inst. Freshw. Res.* (Drottningholm) 1, 130.

Enderlein, O. (1989). Migratory behaviour of adult cisco, coregonus albula l., in the bothnian bay. *J. fish Biol.* 34 (1), 11–18. doi: 10.1111/j.1095-8649.1989.tb02952.x

Euclide, P. T., MacDougall, T., Robinson, J. M., Faust, M. D., Wilson, C. C., Chen, K. Y., et al. (2021). Mixed-stock analysis using rapture genotyping to evaluate stock-specific exploitation of a walleye population despite weak genetic structure. *Evolutionary Appl.* 14 (5), 1403–1420. doi: 10.1111/eva.13209

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14 (8), 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

FAO (2020). *The State of World Fisheries and Aquaculture 2020*. Sustainability in action. Rome. https://doi.org/10.4060/ca9229en

Foll, M. (2012). BayeScan v2. 1 user manual. *Ecology* 20 (10), 1450–1462.

Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180 (2), 977–993. doi: 10.1534/genetics.108.092221

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical f-statistics. *Mol. Ecol. Notes* 5 (1), 184–186. doi: 10.1111/j.1471-8286.2004.00828.x

Guo, B., DeFaveri, J., Sotelo, G., Nair, A., and Merilä, J. (2015). Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol.* 13 (1), 1–18. doi: 10.1186/s12915-015-0130-8

Guo, B., Li, Z., and Merilä, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Mol. Ecol.* 25 (12), 2833–2852. doi: 10.1111/mec.13657

Han, F., Jamsandekar, M., Pettersson, M. E., Su, L., Fuentes-Pardo, A. P., Davis, B. W., et al. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *Elife* 9, e61076. doi: 10.7554/eLife.61076.sa2

Hastings, A. (1993). Complex interactions between dispersal and dynamics: Lessons from coupled logistic equations. *Ecology* 74 (5), 1362–1372. doi: 10.2307/1940066

Hauser, L., and Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish. Fish.* 9 (4), 333–362. doi: 10.1111/j.1467-2979.2008.00299.x

Hauser, L., and Ward, R. (1998). "Population identification in pelagic fish: the limits of molecular markers," in *Advances in molecular ecology*. Ed. G. R. Carvalho (Amsterdam: IOS Press).

Heath, M. R., Culling, M. A., Crozier, W. W., Fox, C. J., Gurney, W. S., Hutchinson, W. F., et al. (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 J. Mar. Sci.* 71 (4), 794–807. doi: 10.1093/icesjms/fst185

Hedgecock, D., Barber, P. H., and Edmands, S. (2007). Genetic approaches to measuring connectivity. *Oceanography* 20 (3), 70–79. doi: 10.5670/oceanog.2007.30

Hutchinson, W. F. (2008). The dangers of ignoring stock complexity in fishery management: The case of the north Sea cod. *Biol. Lett.* 4 (6), 693–695. doi: 10.1098/rsbl.2008.0443

Jombart, T. (2008). Adegenet: A r package for the multivariate analysis of genetic markers. *Bioinformatics* 24 (11), 1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* 11 (1), 94. doi: 10.1186/1471-2156-11-94

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an r package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281. doi: 10.7717/peerj.281

Karjalainen, J., Auvinen, H., Helminen, H., Marjomäki, T., Niva, T., Sarvala, J., et al. (2000). Unpredictability of fish recruitment: interannual variation in young–of–the–year abundance. *J. Fish. Biol.* 56 (4), 837–857. doi: 10.1111/j.1095-8649.2000.tb00876.x

Karjalainen, J., Sjövik, R., Väänänen, T., Sävilammi, T., Sundberg, L.-R., Uusi-Heikkilä, S., et al. (2022). Genetic-based evaluation of management units for sustainable vendace (Coregonus albula) fisheries in a large lake system. *Fish. Res.* 246, 106173. doi: 10.1016/j.fishres.2021.106173

Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L. W., Dickey-Collas, M., Goethel, D. R., et al. (2017). Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. *ICES J. Mar. Sci.* 74 (6), 1708–1722. doi: 10.1093/icesjms/fsw188

Lachance, J., and Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *Bioessays* 35 (9), 780–786. doi: 10.1002/bies.201300014

Laikre, L., Palm, S., and Ryman, N. (2005). Genetic population structure of fishes: Implications for coastal zone management. *AMBIO: A J. Hum. Environ.* 34 (2), 111–119. doi: 10.1579/0044-7447-34.2.111

Lamichhaney, S., Barrio, A. M., Rafati, N., Sundström, G., Rubin, C.-J., Gilbert, E. R., et al. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proc. Natl. Acad. Sci.* 109 (47), 19345–19350. doi: 10.1073/pnas.1216128109

Lehtonen, H. (1981). Biology and stock assessments of coregonids by the Baltic coast of Finland [Coregonus lavaretus, c. widegreni, c. albula]. *Finnish Fish. Res.* 3, 31–83

Li, Y., Bence, J. R., and Brenden, T. O. (2015). An evaluation of alternative assessment approaches for intermixing fish populations: a case study with great lakes lake whitefish. *ICES J. Mar. Sci.* 72 (1), 70–81. doi: 10.1093/icesjms/fsu057

Lotze, H. K., Lenihan, H. S., Bourque, B. J., Bradbury, R. H., Cooke, R. G., Kay, M. C., et al. (2006). Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* 312 (5781), 1806–1809. doi: 10.1126/science.1128035

Lowe, W. H., and Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Mol. Ecol.* 19 (15), 3038–3051. doi: 10.1111/j.1365-294X.2010.04688.x

Luke (2022). Available at: https://www.luke.fi/sv/statistik/statistikinfo/statistikdatabas.

Lundström, K., Bergenius, M., Aho, T., and Lunneryd, S. (2014). "Födoval hos vikaresäl i bottenviken: Rapport från den svenska forskningsjakten 2007–2009," in

Aqua reports 1. Institutionen för akvatiska resurser, Sveriges lantbruksuniversitet, Lysekil, 23

Mapp, J., Hunter, E., van der Kooij, J., Songer, S., and Fisher, M. (2017). Otolith shape and size: The importance of age when determining indices for fish-stock separation. *Fish. Res.* 190, 43–52. doi: 10.1016/j.fishres.2017.01.017

Marjomäki, T. J. (2003). "*Recruitment variability in vendace, coregonus albula (L.), and its consequences for vendace harvesting*," Jyväskylä Studies in Biological and Environmental Science, 127, 1–66

McKeown, N. J., Arkhipkin, A. I., and Shaw, P. W. (2015). Integrating genetic and otolith microchemistry data to understand population structure in the Patagonian hoki (Macruronus magellanicus). *Fish. Res.* 164, 1–7. doi: 10.1016/j.fishres.2014.10.004

McKeown, N. J., Arkhipkin, A. I., and Shaw, P. W. (2017). Regional genetic population structure and fine scale genetic cohesion in the southern blue whiting micromesistius australis. *Fish. Res.* 185, 176–184. doi: 10.1016/j.fishres.2016.09.006

Mehner, T., Palm, S., Delling, B., Karjalainen, J., Kiełpińska, J., Vogt, A., et al. (2021). Genetic relationships between sympatric and allopatric coregonus ciscoes in north and central Europe. *BMC Ecol. Evol.* 21 (1), 1–17. doi: 10.1186/s12862-021-01920-8

Momigliano, P., Jokinen, H., Fraimout, A., Florin, A.-B., Norkko, A., and Merilä, J. (2017). Extraordinarily rapid speciation in a marine fish. *Proc. Natl. Acad. Sci.* 114 (23), 6074–6079. doi: 10.1073/pnas.1615109114

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecol. Appl.* 13 (sp1), 146–158. doi: 10.1890/1051-0761(2003)013

Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., et al. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics* 17 (1), 1–10. doi: 10.1186/s12864-016-2709-z

Paulik, G., Hourston, A., and Larkin, P. (1967). Exploitation of multiple stocks by a common fishery. *J. Fish. Board Canada* 24 (12), 2527–2537. doi: 10.1139/f67-202

Pekcan-Hekim, Z., Gårdmark, A., Karlson, A. M., Kauppila, P., Bergenius, M., and Bergström, L. (2016). The role of climate and fisheries on the temporal changes in the bothnian bay foodweb. *ICES J. Mar. Sci.* 73 (7), 1739–1749. doi: 10.1093/icesjms/fsw032

Pembleton, L. W., Cogan, N. O. I., and Forster, J. W. (2013). StAMPP: An r package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 13 (5), 946–952. doi: 10.1111/1755-0998.12129

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One* 7(5), e37135. doi: 10.1371/journal.pone.0037135

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959. doi: 10.1093/genetics/155.2.945

Pujolar, J., Jacobsen, M., Als, T. D., Frydenberg, J., Munch, K., Jónsson, B., et al. (2014). Genome-wide single-generation signatures of local selection in the panmictic e uropean eel. *Mol. Ecol.* 23 (10), 2514–2528. doi: 10.1111/mec.12753

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi: 10.1086/519795

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi: 10.1093/bioinformatics/btq033

Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W. J. (2009). Genetic population structure of marine fish: Mismatch between biological and fisheries management units. *Fish. Fish.* 10 (4), 361–395. doi: 10.1111/j.1467-2979.2008.00324.x

Ricker, W. (1958). Maximum sustained yields from fluctuating environments and mixed stocks. *J. Fish. Board Canada* 15 (5), 991–1006. doi: 10.1139/f58-054

Ryman, N., and Palm, S. (2006). POWSIM: A computer program for assessing statistical power when testing for genetic differentiation. *Mol. Ecol. Notes* 6 (3), 600–602. doi: 10.1111/j.1471-8286.2006.01378.x

Sendek, D. (2012). Intra-species alliance of the European whitefish coregonus lavaretus l. and vendace coregonus albula l. from the Russian part of the gulf of Finland and the largest lakes of the eastern Baltic basin. *Oceanology* 52 (6), 790–796. doi: 10.1134/S0001437012060124

Sendek, D. S. (2021). Phylogenetic relationships in vendace and least cisco, and their distribution areas in Western Eurasia. *Annales Zool. Fennici* 58 (4-6), 289–306. doi: 10.5735/086.058.0413

Smedbol, R., and Stephenson, R. (2001). The importance of managing within-species diversity in cod and herring fisheries of the north-western Atlantic. *J. Fish. Biol.* 59, 109–128. doi: 10.1111/j.1095-8649.2001.tb01382.x

Spies, I., Hauser, L., Jorde, P. E., Knutsen, H., Punt, A. E., Rogers, L. A., et al. (2018). Inferring genetic connectivity in real populations, exemplified by coastal and oceanic Atlantic cod. *Proc. Natl. Acad. Sci.* 115 (19), 4945–4950. doi: 10.1073/pnas.1800096115

Su, H., Ma, D., Zhu, H., Liu, Z., and Gao, F. (2020). Transcriptomic response to three osmotic stresses in gills of hybrid tilapia (Oreochromis mossambicus female× o. urolepis hornorum male). *BMC Genomics* 21 (1), 1–20. doi: 10.1186/s12864-020-6512-5

Sundelöf, A. Florin A.-B. rogellB., bolund, E., and Vitale, F. Sundblad G. (2022). Fisk- och skaldjursbestånd i hav och sötvatten 2021, *Swedish Agency for Marine and Water Management* 342.

Vuorinen, J., and Lankinen, P. (1978). Genetic differentiation between vendace [Coregonus albula (L.)] populations in eastern Finland. *SIL Proceedings 1922-2010* 20 (3), 2111–2116. doi: 10.1080/03680770.1977.11896826

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *J. Hered.* 89, 439–450. doi: 10.1093/jhered/89.5.438

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* 7 (2), 167–184. doi: 10.1007/s10592-005-9100-y

Waples, R. S., Ford, M. J., Nichols, K., Kardos, M., Myers, J., Thompson, T. Q., et al. (2022). Implications of Large-effect loci for conservation: A review and case study with pacific salmon. *J. Hered.* 113 (2), 121–144. doi: 10.1093/jhered/esab069

Waples, R. S., Punt, A. E., and Cope, J. M. (2008). Integrating genetic data into management of marine resources: How can we do it better? *Fish. Fish.* 9 (4), 423–449. doi: 10.1111/j.1467-2979.2008.00303.x

Ward, R. D., Woodwark, M., and Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *J. Fish. Biol.* 44(2), 213–232. doi: 10.1111/j.1095-8649.1994.tb01200.x

Weir, B., and Cockerham, C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution* 38 (6), 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., et al. (2006). Impacts of biodiversity loss on ocean ecosystem services. *science* 314 (5800), 787–790. doi: 10.1126/science.1132294

Worm, B., Hilborn, R., Baum, J. K., Branch, T. A., Collie, J. S., Costello, C., et al. (2009). Rebuilding global fisheries. *science* 325 (5940), 578–585. doi: 10.1126/science.1173146

Wright, S. (1943). Isolation by distance. *Genetics* 28 (2), 114. doi: 10.1093/genetics/28.2.114

Zemeckis, D. R., Martins, D., Kerr, L. A., and Cadrin, S. X. (2014). Stock identification of Atlantic cod (Gadus morhua) in US waters: An interdisciplinary approach. *ICES J. Mar. Sci.* 71 (6), 1490–1506. doi: 10.1093/icesjms/fsu032

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 05 December 2022
DOI 10.3389/fmars.2022.1004674

OPEN ACCESS

EDITED BY
Fran Saborido-Rey,
Spanish National Research Council (CSIC), Spain

REVIEWED BY
Jianlong Li,
Hainan University, China
Simon Goodman,
University of Leeds, United Kingdom

*CORRESPONDENCE
Sergio Ramírez-Amaro
srgramirez08@gmail.com

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 27 July 2022
ACCEPTED 31 October 2022
PUBLISHED 05 December 2022

CITATION
Ramírez-Amaro S, Bassitta M, Picornell A, Ramon C and Terrasa B (2022) Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments. *Front. Mar. Sci.* 9:1004674. doi: 10.3389/fmars.2022.1004674

Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments

Sergio Ramírez-Amaro[1,2*], Marta Bassitta[1], Antònia Picornell[1], Cori Ramon[1] and Bàrbara Terrasa[1]

[1]Laboratori de Genètica, Universitat de les Illes Balears, Palma, Spain, [2]Instituto Español de Oceanografía (IEO-CSIC), Centre Oceanogràfic de les Balears, Palma, Spain

Fisheries management involves a broad and complex set of tasks that are necessary to prevent overfishing and to help the recovery of overfished stock. Monitoring fishing activities based on two main sources, landings data and scientific surveys, is a challenging task. Fisheries collection data is often limited, which compromises the accuracy of the results obtained. Therefore, several emerging applications of molecular methods have the potential to provide unique understanding of ecological processes in marine environments and to build stronger empirical underpinnings for the Ecosystem-Based Fisheries Management. Environmental DNA (eDNA) is a complex mixture of genetic material shed by those organisms that inhabit a given environment, whereby DNA is extracted from an environmental sample without accessing the target organism. eDNA studies can be categorized into two main approaches, i) eDNA metabarcoding or semi-targeted (community) approaches and ii) species-specific or targeted approaches (single). Although both categories are often discussed, they differ drastically in their methodology, interpretations and accuracy. Both approaches involve a series of steps that include eDNA capture, preservation, extraction and amplification. This detection will depend on the affinity to the targeted taxa sequences and completeness and accuracy of DNA reference collection databases. The eDNA method applied in marine environments are probably the most challenging aquatic environments for applying this technique. This is because of the extreme relationship between water-volume to biomass, dynamics and the physical and chemical properties of seawater that affect dispersion, dilution and preservation. Here, we review the present application of this novel method in fishery assessment in marine environments. To date, many studies suggest that this method offers the potential to revolutionize fisheries monitoring, which will contribute to improving the range of tasks involved in fisheries management. The compelling conclusion is that the methodological steps including in eDNA surveys should be standardized and that research efforts should focus on developing appropriately validated tests to address environmental and sampling factors

that may affect eDNA detection in marine environments in order to draw reliable conclusions. This bioassessment tool can assist fisheries professionals in achieve their research, management, and conservation objectives, but not as a replacement for time-proven assessment methods.

KEYWORDS

eDNA, fisheries, experimental design, marine environment, metabarcoding, monitoring, standardization, biodiversity

Fisheries

Marine fisheries

Fisheries management involves a broad and complex set of tasks (e.g., gathering, analyses, decision-making, resources allocation, implementation), which have the common goal of ensuring the continued productivity of the resources and the accomplishment of other fisheries objectives (Cochrane and Garcia, 2009). Knowledge and management of fisheries depends on accurate and precise data on the distribution and population status of exploited species. According to Cochrane (2000), there are key principles of fisheries management include i) fish stocks and communities, ii) biological production of a stock, iii) demand for human consumption, iv) multispecies fisheries, v) uncertainty of data collection, vi) the short-term dependence on society, vii) ownership of the resource and viii) appropriate communication.

These principles cannot be considered in isolation, in line with the integrated nature of fisheries ecosystems. Because of this, the concept of Ecosystem-Based Fisheries Management (EBFM) was incorporated. It is a holistic concept for managing fisheries and marine resources by considering the entire ecosystem of the species under management. The main objective of EBFM is to maintain the ecosystem in a healthy, productive and resilient condition so that it can provide the services that humans need (Pikitch et al., 2004). Since this concept was proposed, regular monitoring of the marine environment and its living resources has become increasingly necessary. Monitoring of fishing activities is based on two main sources of information: landings data (fishery-dependent) and scientific surveys (fishery-independent) (Dennis et al., 2015). The former type of data is often limited, compromising the accuracy of the results obtained, and the latter depends on the huge operational costs of the vessels (Dennis et al., 2015). Tools to monitor the interaction between fisheries and the environment are diverse, complex and difficult to characterize accurately. Thus, several emerging applications of molecular methods have the potential to provide unique insights into ecological processes in marine environments and to build stronger empirical underpinnings for EBFM (Ovenden et al., 2015).

What is eDNA?

Environmental DNA (eDNA) is defined as DNA captured from an environmental sample without first isolating any target organisms (Taberlet et al., 2012a; Deiner et al., 2017). This refers to any DNA that is collected from the environment rather than directly from an organism, originating from body cells or waste products of organisms. This DNA is released from organisms into a variety of environmental samples such as soil, seawater, snow, or even air (Ficetola et al., 2008; Epp et al., 2012; Turner et al., 2015). Environmental monitoring through DNA is a rapidly growing field largely driven by novel technological developments such as microarray analysis, high-throughput sequencing (HTS), quantitative Polymerase Chain Reaction (qPCR)/droplet digital PCR (ddPCR) and improved bioinformatics capabilities. For eDNA two approaches can be considered, i) based on DNA metabarcoding employing HTS to detect multi-species of mixed and large complex communities, and ii) generally based on qPCR and ddPCR, where the aim is to determine the presence or absence of a single species (e.g., invasive, rare, elusive or endangered species).

The eDNA surveys have recently been proposed as a tool to improve fisheries assessments, and thus conservation success by reducing systematic errors in species richness inference resulting from low detection probabilities and species misidentifications (Evans and Lamberti, 2018). The eDNA concept began in the 90s, with the development of a new method for the isolation of DNA from a variety of sediments (Ogram et al., 1987) and the use of barcoding to phylogenetically analyse clone libraries of eubacterial 16S ribosomal RNA genes amplified from natural populations of the Sargasso Sea (Giovannoni et al., 1990; Figure 1). At the beginning of the 21st century, there was an increase in eDNA-based studies for the detection of microorganisms (Handelsman et al., 1998) and macroorganism species from water and soil samples (e.g., Ficetola et al., 2008). In the

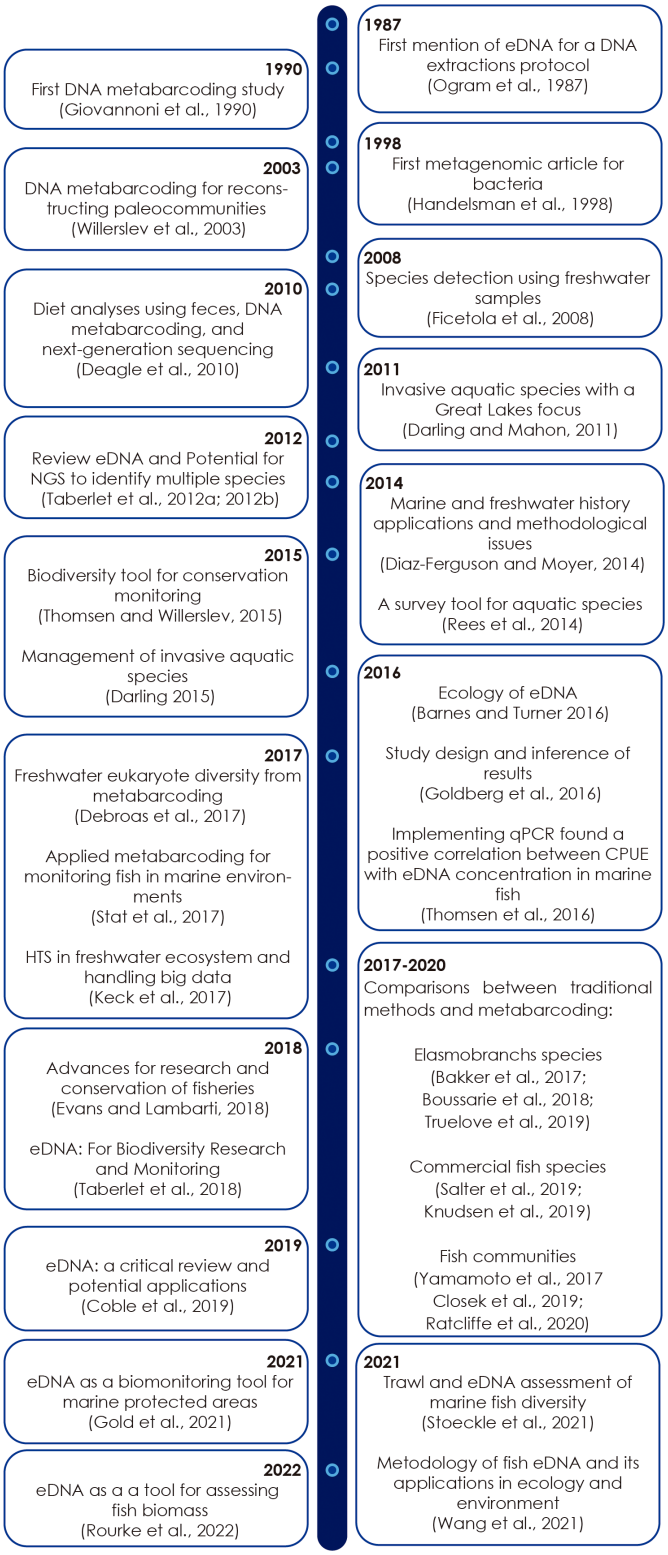


FIGURE 1
Timeline of representative eDNA literature between 1987 and 2022. The last five years only consider studies of biological monitoring in marine environment.

early 2000s, next-generation platforms emerged, leading to new techniques such as metabarcoding and metagenomics.

The first reviews related to eDNA were published in the early 2010s (e.g., Lodge et al., 2012; Taberlet et al., 2012a; Bohmann et al., 2014; Figure 1). The foremost uses of eDNA detection in the marine environments were conducted by Foote et al. (2012) for small-scale genetic monitoring of marine mammals, and by Thomsen et al. (2012a) for analyses of marine macrofauna and meiofauna using seawater samples. In addition, Kelly et al. (2014) published an aquatic eDNA analysis method and also developed a process to monitor the census of marine fishes in a large mesocosm. Furthermore, studies focusing on the comparison of traditional monitoring methods and eDNA (e.g., Boussarie et al., 2018; Knudsen et al., 2019; Stoeckle et al., 2021; Truelove et al., 2019; Figure 1) have helped to demonstrate the high utility of this novel method.

The number of scientific articles published on marine eDNA is lower than the rest of articles focused in the detection of freshwater organisms. However, a search in Google Scholar with both terms "eDNA and ocean", yielded over 6500 existing articles until the year 2017. And in the last five years (2017-2021) almost 9500 articles have been published, of which 3000 articles are related with fisheries topics, which demonstrate that eDNA is currently widely used in this field.

Applications in fisheries management

eDNA offers a potentially powerful method to improve different task related to fisheries assessment in marine environments, mainly in regards to marine biomonitoring by significantly increasing spatial and temporal biological monitoring in aquatic ecosystems due to the ease of water samples collection (Thomsen and Willerslev, 2015; Sassoubre et al., 2016), and by reducing labor-intensive routine taxonomic identification.

eDNA is an efficient tool for the accurate species identification, which is a key aspect of fisheries management. Determining the species compositions of local assemblages is a prerequisite for understanding how anthropogenic disturbances affect biodiversity. eDNA analysis has the potential to detect changes in the biological composition of communities in different ocean regions, which are much more challenging to measure because most biological monitoring methods focus on a limited taxonomic or size range, and is capable of identifying a huge phylogenetic range of organisms down to the species level. Several studies have demonstrated the utility of eDNA metabarcoding for assessing fish diversity (Supplementary Table S1). Biodiversity baselines can therefore be compiled through eDNA, which can help to understand short or long-term changes through comparison with future collections (Jarman et al., 2018; Djurhuus et al., 2020).

The eDNA method also has the potential to estimate abundance or biomass (see references in Supplementary Table S2), a key aspect for its application in stock assessments. The ability to estimate abundance based on concentrations of eDNA relies in part on the assumption that the release of eDNA from feces, secretions, or tissues correlates with the abundance or standing biomass of the respective individuals. Previous eDNA experiments using discrete static mesocosms (aquaria or ponds) showed an association between density and amplification rate (Ficetola et al., 2008), density and DNA concentration (Thomsen et al., 2012b), and biomass and DNA concentration (Takahara et al., 2012). Although several studies have found a positive relationship between eDNA concentration and abundance or biomass in lentic and lotic systems (Doi et al., 2017; Lacoursière-Roussel et al., 2016a; Takahara et al., 2012), there remains substantial variability around this relationship (Goldberg et al., 2015; Jerde and Mahon, 2015). In particular, the relationship between biomass and eDNA in flowing waters remains unclear due to contrasting results (Laramie et al., 2015; Spear et al., 2015; Doi et al., 2017; Hinlo et al., 2017). The variation in the relationship between eDNA and density could be due to differences in movement and retention of eDNA in the systems, for ecample., while eDNA in lentic systems is contained, transport of eDNA in lotic environments is complicated by flow and stream morphologies (Goldberg et al., 2011; Jerde et al., 2011; Olson et al., 2012).

The patchy distribution of fish DNA and the large variation in copy numbers in samples from the same location currently preclude making inferences about fish abundance (Hinlo et al., 2017). Some difficulties with the use of copy numbers are also related to variation in DNA source (e.g., multiple copies of mtDNA). Copy numbers may indicate some relationship with biomass, but this is also confounded with diversity in size distribution and life history stage of individuals. The correlation between eDNA copy number and catch per unit effort in flowing systems requires further study. Some publications suggest that temporal factors such as breeding and migration seasons, should also be considered in eDNA studies to increase the probability of detection (Thomsen and Willerslev, 2015; Barnes and Turner, 2016). Future studies on the temporal and spatial aspects of eDNA in various species and habitats should be carried out to evaluate the eDNA method in terms of conservation issues (Thomsen and Willerslev, 2015; Furlan et al., 2016). Some studies argue that results obtained with this method should be taken with caution, mainly regarding the inference of abundance through eDNA concentration, especially in natural flowing water bodies. The nature of each marine species, such as habitat preference, could affect the dispersal of eDNA in the water column and thus compromise the success of eDNA (Hinlo et al., 2017). The use of eDNA as a detection tool holds great promise, but as it entails additional cost and effort, studies comparing the performance of eDNA with conventional

tools during routine monitoring are needed for its adoption by management.

Using eDNA as a tool for generating population genetic data can be a preferred alternative to sampling biological tissues, which is often expensive and invasive. Knowledge of genetic structure helps conservation management to delineate Management Units and/or stocks and organize actions to preserve genetic diversity (Palsbøll et al., 2007; Abdul-Muneer, 2014). Population characteristics of large whale shark aggregations inferred from seawater eDNA showed that HTS of seawater eDNA can provide useful estimates of genetic diversity, thus extending the applications of eDNA to encompass population genetics of marine organisms (Sigsgaard et al., 2016).

Thomsen et al. (2016) reported eDNA metabarcoding of seawater samples and compare eDNA sequence reads with parallel trawl catch data. The two sampling methods showed generally good overlap. They presented results that suggest a correspondence between fish density (abundance and biomass) and marine eDNA sequence reads produced from Illumina HTS and show a possibility to assess marine fish stock using water samples. Calibration of this new technology against traditional methods is complex due to the fact that all marine monitoring techniques have "catchability" biases (Arreguín-Sánchez, 1996; Fraser et al., 2007). In this sense, Stoeckle et al. (2021) describe an example of bottom trawl fishery. These authors highlighted that bottom trawl catches are influenced by aspects of equipment such as net type, mesh size, and towing speed, and by biological factors such as patchy distribution and habitat preference. They suggested that eDNA surveys should be calibrated and compared with established methods in diverse habitats under a variety of hydrographic conditions.

eDNA analysis in fisheries science has also focused on the application of the method to the detection and monitoring of invasive fish and at-risk species. The sensitivity of eDNA-based methods makes them ideal for detecting the presence of species when efforts to detect low-density species would be unmanageable, such as the presence and distribution of low-density invasive, elusive, or threatened species (Dejean et al., 2012; Takahara et al., 2013; Rees et al., 2014). The non-invasive nature of eDNA analysis may provide advantages over traditional capture-based sampling by allowing the presence or absence of species to be determined without disturbing the fish or their environment. This approach could be particularly beneficial in endangered species situations, where there is a significant risk of injury to fish or damage to critical habitat (Evans and Lamberti, 2018). For endangered species in particular, a non-invasive genetic sampling technique that could provide information on absence/presence data and even estimates of population size, would therefore be of great use for the conservation and management of these species. In addition, eDNA approaches may offer a cost-effective way of obtaining basic distribution and abundance data, and allow limited conservation resources and taxonomic knowledge to be efficiently deployed to maximise returns (Rees et al., 2014). These points have highlighted the potential of eDNA analysis to improve assessments of rare species, as well as to assist in routine fisheries sampling. Several studies have illustrated that eDNA analysis can reliably detect the presence of fish in marine ecosystems where they are known to occur through catch-based sampling (see references in Supplementary Table S1).

The effectiveness of eDNA has been evaluated in comparison with traditional monitoring techniques (e.g., Knudsen and McDonald, 2019; Russo et al., 2021; Stoeckle et al., 2021). The results of most of these studies indicate that eDNA has higher detection rates and has a higher cost-effectiveness or higher catch per unit effort than traditional sampling methods. Scaling up long-term biomonitoring programs with eDNA could improve taxon detection and resolve long-term patterns or changes in species of interest (Berry et al., 2019). eDNA assessment can be used for biomonitoring of pelagic and benthic ecosystems targeting fish (Russo et al., 2021; Stoeckle et al., 2021; Valsecchi et al., 2021), mammals, seabirds (Ushio et al., 2018), and sea turtles (Kelly et al., 2014). Despite the advantages of eDNA detection, there is key ecological information, such as fish recruitment and size/age classes, that can only be obtained through traditional monitoring.

A virtual workshop was held on 28 May 2020, bringing together a broad cross-section of experts from the genetics and fisheries assessment communities to discuss the state-of-the-art and identify barriers and advantages for the application of HTS techniques on stocks developed under the FishGenome contract "Improving Cost-Efficiency of Fisheries Research Surveys and Fish Stocks Assessments using Next-Generation Genetic Sequencing Methods" (Figure 2). In particular, there was a section discussing the use of eDNA approaches in fisheries assessment. One of the main concerns with the use of eDNA that was pointed out is that, to date, the spatial distribution of eDNA in seawater column is unknown. The origin of this genetic material cannot be easily determined and its impact can usually be corrected and buffered by repetition and proper sampling design. The poor quality of existing public databases was also identified as a problem, as they contain significant errors that can mislead in the interpretation of analyses. For this reason, the construction of a specific database, containing the species that inhabit the studied environment, can be considered a recommendable alternative.

Regarding the possibility of using eDNA to estimate biomass abundance by qPCR, it seems that further research is needed to calibrate the tool. Once the amount of DNA representing a given fluorescence signal is determined, the biomass of that resource

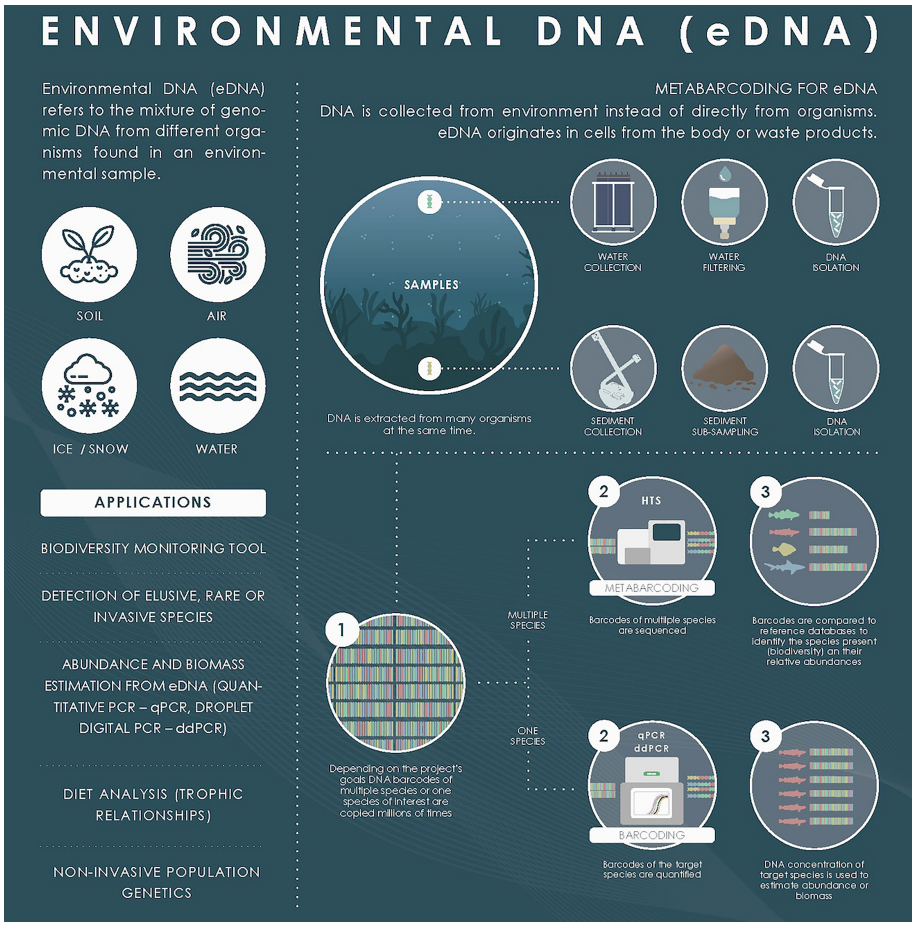


FIGURE 2
Scheme of eDNA used in the FishGenome contract virtual workshop "Improving Cost-Efficiency of Fisheries Research Surveys and Fish Stocks Assessments using Next-Generation Genetic Sequencing Methods". The definition, applicability and methodology of eDNA from samples collected from sediment and seawater are indicated.

must be calibrated with the amount of DNA, which may depend on many factors. However, some relationships have been found in metabarcoding, indicating that qPCR may work even better for certain species. If the objective is to quantify a species and continuous monitoring is performed, there is a possibility that qPCR can provide some relevant information (there is experience in this regard for detecting presence peaks). It is not yet known how close or how far we are from the goal of real biomass quantification that can be integrated into assessment models by metabarcoding.

As final notes during the workshop, experts suggested that beyond the measurement of species presence and distribution, the implementation of eDNA could provide information on species co-occurrence and relationships across ecological networks, with environmental parameters. This would improve quantifiable ecological information of great interest for biodiversity management and assessment.

eDNA behavior

Marine environments are probably the most difficult and challenging aquatic samples to apply the eDNA method. This is due to the extreme ratio of water-volume to biomass, the effects of sea currents and wave action on eDNA dispersion and dilution, the impact of salinity on eDNA preservation and extraction (Thomsen et al., 2012b). To date, studies of eDNA have mainly focused on proof-of-concept, and further research is needed on the 'ecology' of eDNA —release and concentration, degradation, and transport— and its influence on detection, quantification, analysis, and application of eDNA to assessment and conservation (Barnes and Turner, 2016; Thomsen et al., 2016; Hansen et al., 2018; Figure 3). Understanding the origin of eDNA and the physical, chemical, and biological factors that affect eDNA concentration and influence its production, persistence, and transport in marine ecosystems can contribute to our knowledge of the taxa and

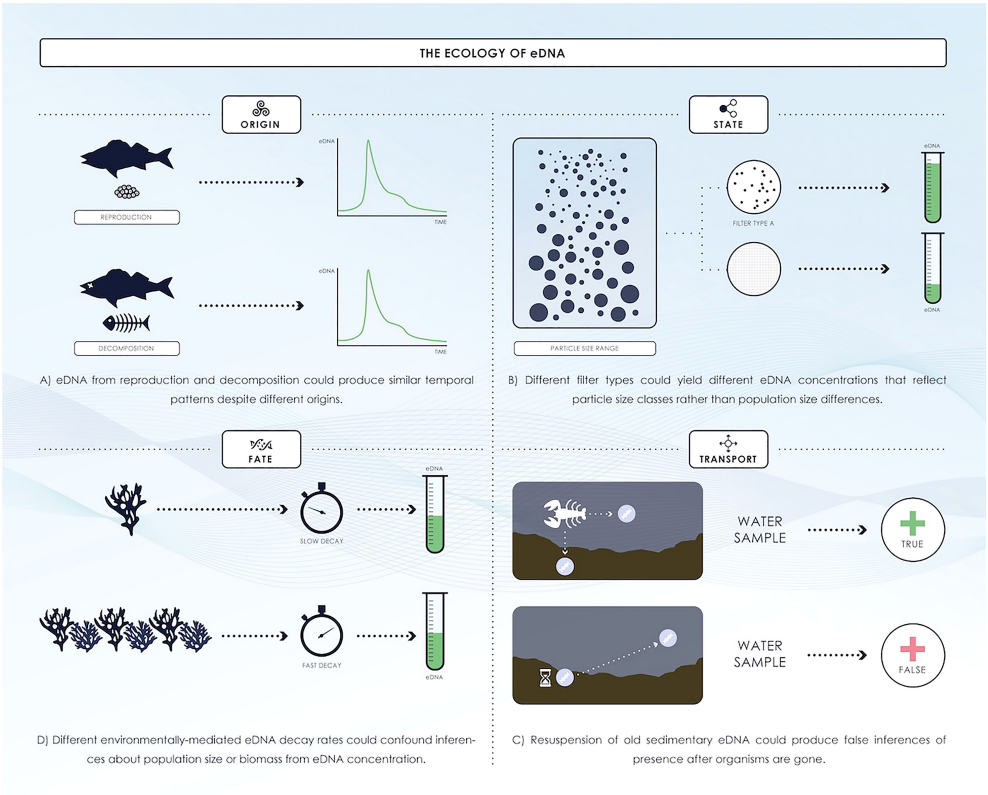


FIGURE 3
eDNA ecology affects population inferences: Origin, State, Fate and Transport (redrawing from Barnes and Turner, 2016).

environments for which eDNA represents an effective tool (Barnes and Turner, 2016).

eDNA release and concentration

The concentration of eDNA varies by several orders of magnitude between different environmental samples, reflecting the environmental conditions in which DNA is preserved. This concentration in the environment limits the scope of eDNA studies, as often only small segments of genetic material remain. The rate of release and degradation of eDNA, which can be strongly altered by environmental conditions, determine the concentration of eDNA in seawater samples (Lacoursière-Roussel et al., 2016b).

The production of eDNA depends on the biomass, age, and feeding activity of organisms, as well as physiology, life history, and space use (Barnes and Turner, 2016; Goldberg et al., 2016). Different experiments indicate that eDNA concentration is positively correlated with individual biomass (e.g., Pilliod et al., 2014; Evans et al., 2016; Piggott, 2016). However, a point of caution is suggested because differences in size between species (biological differences) and within a species (e.g., age structure or morphometrics) are likely to influence eDNA production and biomass interpretations (Stewart, 2019). Differences due to life history stage have been recurrently observed. For example, incongruities in eDNA production between juveniles and adults have been demonstrated in fish (Maruyama et al., 2014), amphibians (Goldberg et al., 2011; Thomsen et al., 2012b), and invertebrates (Tréguier et al., 2014). In aquaria-based experiments with bluegill sunfish (*Lepomis macrochirus*), juveniles were observed to have slightly higher excretion rate than adults, possibly due to ontogenetic reduction of metabolic activity in adults (Maruyama et al., 2014). The risk of using measures of eDNA abundance in wild populations without accounting for age-structure would lead to incorrect estimates of population abundance, especially if populations are dominated by one or another age class (Maruyama et al., 2014), such as in dwindling populations with low birth rates, or in newly founded populations monopolized by juveniles.

Species-specific differences may also affect the quantity and quality of eDNA production, which is strongly influenced not only by size but also by the ecology of the target taxa (Stewart, 2019). Indeed, studies have reported disparities in eDNA sources between and within taxonomic groups (e.g., Goldberg et al., 2011; Thomsen et al., 2012b; Sassoubre et al., 2016). For example, eDNA detection was shown to vary between both cownose ray (*Rhinoptera bonasus*) and Brazilian cownose ray

(*R. brasiliensis*) suggesting that disparities between species are related to their migration seasons (Stoeckle et al., 2020). Thomsen et al. (2012b) also found that eDNA production rates differed between two juvenile amphibian species and two fish species.

On the other hand, some studies found increases in the abundance of eDNA signals during breeding seasons (e.g., Bista et al., 2017; Stoeckle et al., 2017). For example, eDNA signals have been shown to be triggered during fish spawning, suggesting strong seasonal influences on eDNA detection (e.g., Laramie et al., 2015; Erickson et al., 2016). Similar results have also been observed in amphibians and reptiles, presumably reflecting species-specific behaviour (de Souza et al., 2016) such as male–male combat and mass release of gametes (e.g., Eastern Hellbender, *Cryptobranchus alleganiensis*; Spear et al., 2015). Genetic material such as gametes, blood, and other reproductive tissues (e.g., placenta, lactation, etc.) combine to make breeding events optimal for eDNA detection in wild populations. However, this overproduction of DNA at breeding times may also overestimate the relative presence of a species both temporally and spatially (Stewart, 2019).

The response of organisms to the presence of others is another factor to consider. For instance, it is well-known that predators exert physiological impacts on their prey, including reduced food intake, increased metabolic rate, and elevated stress (Boonstra, 2013; Van Dievel et al., 2016). Although the exact impact of this interaction on eDNA production rates is unknown to date, it is likely to affect these processes. It has been suggested that acute perturbations of physiological homeostasis, such as stress, may have effects on sources of genomic material (Pilliod et al., 2014; Klymus et al., 2015). Several studies have observed an increase in eDNA production due to osmotic changes or following the handling of animals (Maruyama et al., 2014; Pilliod et al., 2014). Increased density (McKenzie et al., 2012), social stress (Sloman et al., 2000), and shelter availability (Milidine et al., 1995) are other examples of stresses affecting the metabolic activity of aquatic organisms, with potential impacts on eDNA sources, that warrant further investigation.

eDNA degradation

eDNA degradation must be considered in eDNA studies, as it can reduce the detectability of species over time. The degradation rate can vary from hours to weeks, depending on the environment and target species (Thomsen et al., 2012b; Maruyama et al., 2014; Balasingham et al., 2017; Barnes et al., 2014; Sassoubre et al., 2016). In sediments and terrestrial soils, a very low proportion of DNA can persist for long periods, adsorbed to organic or inorganic particles that protect it from several potential degradation agents. Dell'Anno and Corinaldesi (2004) demonstrated that in marine sediments the turnover of extracellular DNA is ca. 200 times slower than in sea water (up to 93 days in sediments versus 10 h in seawater). DNA persistence is strictly dependent on the nature of the sediment, e.g., in loamy sediment, persistence time may be similar to persistence in the water column (Deere et al., 1996). However, under certain conditions, DNA can be preserved for hundreds of thousands of years, as observed by Coolen and Overmann (2007) who were able to analyse DNA in 217,000-year-old anoxic sediments. In marine environments, eDNA degrades faster than in freshwater environments, and can be used to obtain a "snapshot" of the species present in that particular environment at the time of sampling, or a few days or weeks earlier.

On the other hand, in marine and freshwater environments, different studies have evidenced that eDNA persistence can vary throughout the water column. In this sense, Matsui et al. (2001) reported a greater degradation of eDNA in the epilimnion (the upper, warmer layer of a thermally stratified lake and more exposed to UV radiation) than in the hypolimnion (the lower, colder layer of a thermally stratified lake). However, this difference in detection time appears to be due to current flow and related dilution, rather than DNA degradation, which is an important variable when dealing with a dynamic system such as the ocean.

eDNA transport

Once released into the environment, eDNA is transported away from organisms and begins to degrade. To better understand the distribution of eDNA in relation to species distribution, studies have begun to examine how this complex DNA signal is transported horizontally and vertically in aquatic environments, as the process of eDNA transport or diffusion is fundamental to sampling design and spatial inference. Understanding this transport is essential to relate the detected eDNA to the presence of species both in space (i.e. how close a species was to the location of eDNA detection) and in time (i.e. how recent the presence of the detected species was). For important conservation applications, being able to connect a positive eDNA detection within specific spatial and temporal boundaries is essential for drawing robust conclusions (Barnes and Turner, 2016).

In marine environments, where long-distance transport is possible, eDNA transport remains a problem (Thomsen et al., 2012a). Vertical transport (i.e. settling) of fish eDNA accumulation in sediments has also been described (Turner et al., 2015). Resuspension of sedimented eDNA within water, considering the high concentrations of sedimented fish eDNA that have been observed, could represent an important element of eDNA ecology (Turner et al., 2015). As eDNA is currently a tool used for contemporary biodiversity monitoring and conservation, it is crucial that the results reflect the current state of an ecosystem. In this regard, it is important to consider the fact that eDNA in soil appears to be able to persist for decades and centuries (Andersen et al., 2012; Yoccoz, 2012) and the potential release of "ancient" eDNA from bottom sediments

into water column could complicate the use of aquatic eDNA as strict contemporary biodiversity surveys.

Factors influencing the detectability of eDNA

The factors that influence the persistence of eDNA in aquatic environments have been the subject of different studies, concluding that they fall into two broad categories: (i the abiotic environment (i.e. temperature; UV radiation, salinity); and (ii the biotic environment (i.e. composition and activity of the microbial community and extracellular enzymes) (Barnes and Turner, 2016). These factors can affect both production and degradation of eDNA.

Abiotic environment

This factor plays a major role in the rate of DNA release and degradation in marine environments. Water temperature can affect the release of DNA from organisms and thus the availability of eDNA for detection. The effect of temperature on DNA release can be due to different reasons. For example, fish metabolism, growth, physiology, and immune function are influenced by water temperature (Engelsma et al., 2003; Person-Le Ruyet et al., 2004; Takahara et al., 2011). As a by-product of metabolic influences, evidence suggests that temperature additionally affects the production of feces and urine in fish (Selong et al., 2001; Gale et al., 2013), presumably the main component of eDNA sources. Fish mobility increases with water temperature (Petty et al., 2012), so genetic signals may also be more homogenised and/or spatially dispersed. The immune response can also facilitate mucus excretion and epithelial cells shedding. Studies to date have found conflicting results; in mesocosm experiments, no effect was found in two studies (common carp, *Cyprinus carpio*, Takahara et al., 2012; bighead carp, *Hypophthalmichthys* spp., Klymus et al., 2015), while there was a significant increase in production rates in Mozambique tilapia (*Oreochromis mossambicus*; Robson et al., 2016). In a field study, high water temperature significantly increased the amount of brook charr (*Salvelinus fontinalis*) eDNA within the water column and, moreover, biomass and thus predictability of population abundance increased at higher temperatures (Lacoursière-Roussel et al., 2016a, Lacoursière-Roussel et al., 2016b). In addition, temperature itself may affect the excretion of genetic material into the environment when phenologies simultaneously affect other physiological (e.g., metabolic regulation) or behavioural responses (e.g., temporal avoidance) of the organism.

Another such factor is the effect of UV radiation on detectability of eDNA. It was shown that eDNA was no longer detectable in samples exposed to full-sun after 8 days, but it was possible to detect eDNA in samples that were stored in the dark after 11 and 18 days, demonstrating the direct effect of UV radiation on eDNA (Pilliod et al., 2014).

Adaptation to saline environments also requires physiological compensation and acclimatization. In most marine fish species, egg fertilization and incubation, early embryogenesis, swim bladder inflation, and larval growth are salinity-dependent (Boeuf and Payan, 2001). In fact, studies have shown that up to 50% of the total energy of fish can be devoted to osmoregulation (Bushnell and Brill, 1992), and that food intake, feed conversion, and hormones associated with growth regulation depend on environmental salinity (Boeuf and Payan, 2001). Smoltification of salmon, for instance, demonstrated a drastic physiological adaptation to seawater, resulting in a significantly different metabolism than their freshwater counterparts (e.g., McCormick et al., 1989).

Widespread links between salinity and fish growth have been demonstrated for both marine and freshwater species, with general patterns suggesting that growth rates of marine species increase in slightly more saline environments, while development of freshwater species shows the opposite relationship (Boeuf and Payan, 2001). While assessment of marine species richness and approximate abundance is a relatively new foray for eDNA (e.g., Günther et al., 2018; Knudsen and McDonald, 2019), it has proven successful for accurate detection, but read abundance has failed to find to be easily correlated with DNA ratios (Günther et al., 2018) or traditional visual measures of biomass, such as trawling (Knudsen and McDonald, 2019).

Biotic environment

eDNA studies showed that DNA persistence in aquaria experiments, where water temperature and sunlight conditions were the same, ranged from one week to one month (Dejean et al., 2011; Thomsen et al., 2012b; Piaggio et al., 2014). Differences in DNA persistence time observed in these studies could be explained by differences in animal density. Endogenous nucleases are another factor influencing the amount of eDNA in the environment (Hebsgaard et al., 2005). Furthermore, disruption of the cell structure releases DNA and cellular fluids into the environment. This, in turn, stimulates the growth of microorganisms and leads to further degradation of DNA by their exogenous DNases (Hebsgaard et al., 2005; Willerslev and Cooper, 2005). Temperature also has an important influence on the action of endonucleases and microorganisms, as at low temperatures these activities can be slowed down or even inactivated (Hofreiter et al., 2001; Zhu, 2006).

Hydrolysis and interstrand crosslinks are another source of DNA damage (Herder et al., 2014). They influence the accessibility to DNA-polymerases and prevent DNA strand cleavage, which blocks DNA replication (Noll et al., 2006). Consequently, amplification of DNA extracted from an environmental sample (e.g., water, soil, sediment) is prevented

(Hansen et al., 2006), and the species will not be detected. Decreased DNA detection in the water column could also be due to uptake of DNA by sediments and organic matter present in the water (Deere et al., 1996). Corinaldesi et al. (2008) investigated which environmental factors (temperature, salinity, organic matter loads, and redox potentials) could affect extracellular DNA damage and degradation rates in various marine sediments, showing that extracellular DNA damage rates do not depend on a single factor (e.g., temperature) but on a complex interaction of different factors. In addition, fish physiology such as stress (Pilliod et al., 2014), breeding readiness (Spear et al., 2015), feeding behaviour (Klymus et al., 2015), and metabolic rate (Maruyama et al., 2014) may also play a role in eDNA degradation.

eDNA methodology notes

The widespread implementation of eDNA methods in species monitoring within fisheries assessments is currently prevented by the lack of rigorous standards for both sample collection and bioinformatics analysis.

Metabarcoding

This approach commonly uses universal primers that amplify DNA from a group of target species (e.g., fishes, crustaceans, echinoderms). The amplified fragments are then sequenced using an HTS platform. These new technologies allow to sequence DNA molecules present in the mixture and the simultaneous sequencing of millions or billions of molecules. Furthermore, several independent samples can be multiplexed in a single run. The workflow of the DNA metabarcoding approach is generally based on a double indexing and two-step PCR pipeline. In order to allow the clustering of several samples in the same sequencing run, double indexing is used, where sample-identification barcodes are incorporated in the Illumina adapter sequences, both forward and reverse. Barcodes should not be included in the region-specific primer in the first PCR. These barcodes are short sequences, 3 to 14 nucleotides in length, that should be distinct from each other and can have error-correcting properties to protect against the sequence alterations introduced during synthesis, amplification or sequencing (Krishnan et al., 2011). The first PCR is carried out with primers targeting the region of interest, but these primers already include the overhang adapters for the barcodes and the sequencing adapters. In the second PCR, the sequencing index adapters and barcodes are incorporated into the amplicons. For example, when Illumina indexes are used, a small multiplex identifier is added to the overhang to allow additional pooling of samples for sequencing. The final library should be quantified with qPCR or TapeStation or Bioanalyzer, otherwise the sequencing run will fail due to the inaccurate quantitative data.

The analysis of millions of sequences produced by the eDNA metabarcoding method requires efficient, automated and yet flexible analysis pipelines to translate the raw sequences into a statistically exploitable contingent matrix containing (Dufresne et al., 2019). Sequence clustering can be reference-based if sequences are assigned to a cluster because they are sufficiently similar to a sequence in a specific reference database. Therefore, taxa that are not included in the database will not be clustered, with the consequent loss of biological variation. The most commonly used clustering algorithms are based on the generation of clusters of sequences that differ by less than a fixed sequence dissimilarity threshold. The representative sequences of these clusters are referred to as molecular Operational Taxonomic Units (OTUs). However, the similarity thresholds used to delimit OTUs are arbitrary and depend on the variability of the genomic region and the targeted taxonomic groups and, more importantly, they have been shown to strongly affect molecular biodiversity inventories (Brown et al., 2015; Tapolczai et al., 2019). Recently, new methods have been developed to generate Amplicon Sequence Variants (ASVs), also known as Exact Sequence Variants (ESVs), Zero-radius OTUs (ZOTUs) or an OTU defined by 100% sequence similarity. The imposition of arbitrary dissimilarity thresholds is avoided by using a *de novo* process that control the errors contained in the dataset sufficiently such that the ASVs sequences only differ by as little as a single nucleotide (Eren et al., 2013; Callahan et al., 2016).

Different algorithms and software have been developed to perform single or multiple processing steps. New applications have even been developed on website platforms that help users to create and execute their own metabarcoding pipelines, such as OBITools (Boyer et al., 2016), DADA2 (Callahan et al., 2016), MiFish (Sato et al., 2018), Anacapa (Curd et al., 2019), Barque (Mathon et al., 2021), metaBEAT (metaBarcoding and eDNA Analysis Tool) v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT), Charybdis (https://github.com/cbirdlab/charybdis)). These pipelines can be modified according to the user's needs and in order to achieve more accurate and reliable results (Antich et al., 2021; Kim et al., 2021).

Once the HTS output data have been bioinformatically analysed, they can be compared with a reference database, although this can be a source of bias. When using public databases (e.g., GenBank, Ensembl, BOLD systems) as a reference database, the high number of sequencing errors (Harris, 2003) and mislabeled species (Santos and Branco, 2012) must be considered. One solution, to overcome this problem, is the construction of a private databases in which sequences, species labeling, and geographic origin are carefully verified, as in Meta-Fish-Lib reference library hosted at https://github.com/genner-lab/meta-fish-lib, or MetaZooGene Barcode

Atlas and Database (https://metazoogene.org/MZGdb), or use methods to identify mislabel species (e.g., SATIVA, https://github.com/amkozlov/sativa).

Quantitative PCR

This approach is commonly used to assess biomass and abundance of marine communities, including fish (Salter et al., 2019; Knudsen et al., 2019). Two types of chemistries are commonly used to detect PCR products using real-time PCR instruments: SYBR® Green and TaqMan®. Initially, intercalator dyes were used to measure real-time PCR products, with the main drawback that they detected the accumulation of both specific and non-specific PCR products. Currently, SYBR® Green method has two requirements for a DNA binding dye for real-time PCR detection: increased fluorescence when bound to double-stranded DNA, and no inhibition of PCR. On the other hand, the TaqMan® method uses a fluorogenic probe that allows the detection of a specific PCR product as it accumulates during PCR.

In addition, droplet digital PCR (ddPCR) is also recently being used for eDNA. This to nucleic acid detection and quantification method offers an alternative method to conventional real-time qPCR for absolute quantification and detection of rare alleles. Digital PCR works by partitioning DNA or cDNA sample into many individuals and performing parallel PCR reactions; some of these reactions contain the target molecule (positive) while others do not (negative). A single molecule can be amplified a million times or more. Nowadays, the most widely used method for quantification of abundance or biomass in aquatic environments is TaqMan chemistry (e.g., Sassoubre et al., 2016; Doi et al., 2017; Atkinson et al., 2018; Levi et al., 2019).

Workflow

In general, both eDNA approaches involve a series of steps that include eDNA capture, preservation, extraction, amplification, and sequencing to ensure detection of target species (Figure 4). Efficiency at each step is expected to affect DNA recovery and, consequently, detection. Researchers often choose methods based

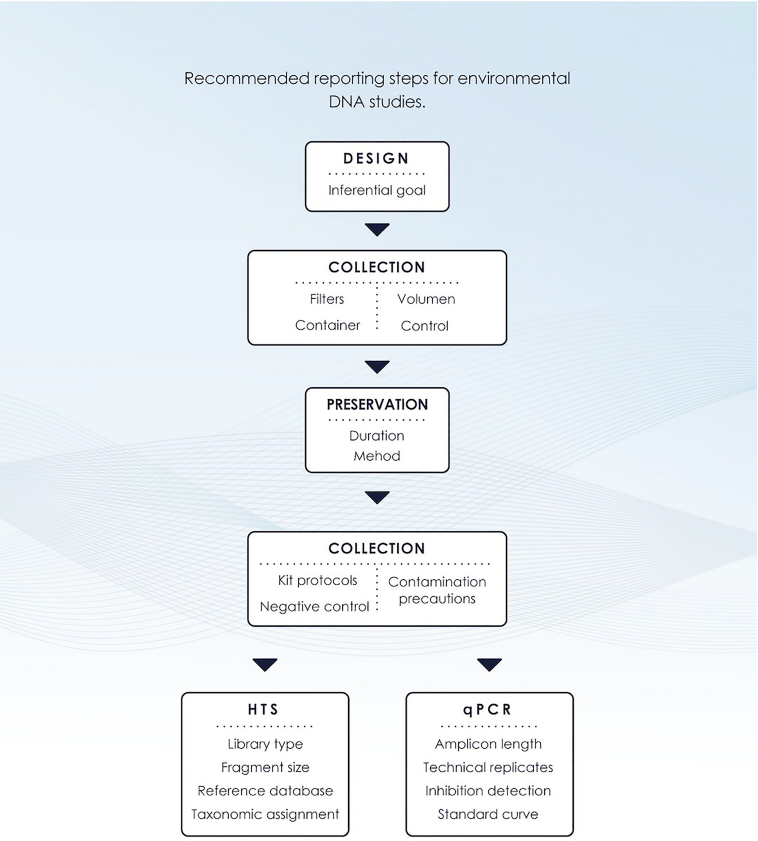


FIGURE 4
Schematic workflow designed for environmental DNA studies. In general, eDNA studies include the following steps: experimental design, collection, preservation, extraction, and analyses that include High-Throughput Sequencing (HTS) and/or quantitative PCR (qPCR).

on cost, ease of sampling, and availability of materials and equipment. Since eDNA typically exhibits high levels of degradation and is often of poor quality, DNA recovery could vary depending on the quality of the protocol used, so methods that maximise eDNA recovery in a cost-efficient manner are ideal (Hinlo et al., 2017). Several research groups have developed a great variety of protocols for eDNA detection from different marine species and environments. Currently, there is a wide variety of protocols for sampling and analysis of eDNA data, which must be optimized according to the objective of each study. This is a characteristic of emerging scientific fields, and we believe it is essential now, at this juncture, to develop minimum standards of quality assurance.

Samples

Aquatic samples: In general, two methods are used to capture eDNA from aquatic environments, filtration and precipitation. Filtration requires passing water samples through a filter to trap DNA, and allows larger volumes of water (commonly 250 ml - 5 L; Supplementary Tables S1, S2) to be processed. Water samples are filtered on-site or stored on ice for travel and then filtered in the laboratory. Meanwhile, the precipitation method uses ethanol to precipitate nucleic acids in water (Jerde et al., 2011; Hinlo et al., 2017). Although both methods have shown variable success rates in comparative studies, it is recognized that the filtration method recovers more eDNA from water samples than the precipitation method (e.g., Deiner et al., 2015; Eichmiller et al., 2016; Hinlo et al., 2017), and immediate preservation is generally recommended (Supplementary Tables S1, S2).

In addition to this, the filter material, such as pore size, filter material, and DNA extraction method, could affect the quality and quantity of eDNA, resulting in different final detection rates (Deiner et al., 2015; Renshaw et al., 2015). eDNA studies on marine environments use frequently pore sizes ranging from 0.2µm – 0.45 µm, while the filtration volume commonly ranges from 250 ml to 2 L (Supplementary Table S1, S2). Otherwise, eDNA samples have been effectively collected with different filter material such as cellulose nitrate, glass fiber, polycarbonate, nylon, polyethersulfone and cellulose acetate (Supplementary Tables S1, S2). Hinlo et al. (2017) conducted experiments to compare the recovery of eDNA at different stages of the analysis to determine which methods are most cost-effective, concluding that the most recommended filtration process was the use of cellulose nitrate filters. Finally, filters (and collected materials) can be preserved by freezing (Jerde et al., 2011), immersion in ethanol (Goldberg et al., 2011), drying, or immersion in cell lysis buffer (Renshaw et al., 2015). It is recommended to filter seawater samples within 24 hours, but if this cannot be done, short-term refrigeration (72 hours at 4°C) could be performed (Hinlo et al., 2017).

Sediment or soil samples: Recent studies have shown that, in general, extra membranous eDNA is found in higher concentration in sediments than in the overlying water column, as DNA from water columns can progressively accumulate in sediments (Corinaldesi et al., 2008; Turner et al., 2015). Marine sediments supposedly harbor one of the richest species reservoirs on Earth, but logistics and the scarcity of taxonomic specialist make it difficult to understand their biodiversity, and more so in the case of deep-sea sediments (Grassle and Maciolek, 1992; Snelgrove, 1999). The persistence of eDNA is strictly dependent on the nature of the sediment (Deere et al., 1996). In marine environments, DNA molecules degrade faster than in freshwater environments. Because of this, it can be used to give a "snapshot" of the species present in this particular environment at the time of sampling, or a few days or weeks earlier (Collins et al., 2018). eDNA preserved in sediment or soil samples can be used to obtain an integrative picture of present or past biodiversity (Herder et al., 2014). The high concentration and long persistence of fish eDNA in sediments can help to know the temporal and spatial scales from aquatic eDNA (Bloesch, 1995; Douville et al., 2007). According to Turner et al. (2015) fish eDNA in aquatic sediments may be a promising source of historical genetic materials.

In several studies, samples are collected from the sediment surface, which contains suspended material within the water column, including whole cells and extracellular DNA (Levy-Booth et al., 2007; Guardiola et al., 2015; Holman et al., 2019). Marine sediments are known to have key ecological functions and ecosystem services, and are sensitive to anthropogenic disturbances. However, they have been poorly studied by traditional means, and are not well understood because taxonomic work to describe species found in marine sediments is difficult and rarely undertaken (Guardiola et al., 2015). The potentially extensive persistence of DNA bound to sediments is very valuable, but can be difficult to identify when the target species was present at the sampling site. In fact, detection of eDNA also varies according to sediment texture, so the amount to process is also variable. Typically, the volume of sediment samples processed is between 10-50 g of sediment surface (Guardiola et al., 2015; Holman et al., 2019).

To obtain accurate results from seawater and sediment samples, strict and clean collection protocols must be followed. Decontamination of collecting equipment is essential to maintain sample independence and samples should be preserved in sterile containers. In fact, single use supplies for eDNA collection can significantly reduce the risk of contamination (Goldberg et al., 2016). Bottles for eDNA sample collection should be cleaned with 10% bleach and washed with DNA-free distilled water.

Equipment requirements

Environmental DNA samples present the same contamination challenges as other low-quantity DNA samples, such as ancient, forensic, and non-invasive genetic samples (Herder et al., 2014). eDNA samples should be handled and stored in a dedicated room that is physically separate from rooms where high quantity DNA extraction and PCR products are handled (Taberlet et al., 1999). Laboratories should be organized in order to avoid contamination as much as possible. All eDNA extractions must be carried out in an isolated room ideally equipped with positive air pressure, overnight UV treatment, and air renewal. Pre-amplification and post-amplification work should be performed in separate rooms, distant from each other, ideally in different rooms. DNA extraction and PCR mix preparation should be carried out in the pre-amplification rooms, while PCRs will be performed and the PCR results analysed in the post-amplification room. Control samples without DNA should be extracted at the same time and used as negative controls. Positive PCR controls and qPCR standards should ideally be added in a third room assigned as an intermediate DNA level room, between the pre-amplification and the post-amplification room (Herder et al., 2014). For all laboratory procedures, filter pipette tips and clean gloves should be used.

eDNA extraction

Several DNA extraction protocols have been optimized and applied in eDNA studies. The choice of protocols may affect species detectability and sample diversity, and different approaches may be required, depending on the objective of the study. Different aspects should be considered when selecting the best strategy and protocol for DNA extraction: i) the proportion of sample used for extraction; ii) the sampling or subsampling strategy for DNA extraction; and iii) the detection of the presence of inhibitors. The most frequently used protocols for eDNA studies are: 1) Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germany); 2) MO BIO's Qiagen PowerWater DNA Isolation Kit (Qiagen, Germany), 3) CTAB (Cetyltrimethylammonium bromide)-chloroform and phenol-chloroform, 4) Qiagen DNeasy Powermax Soil (Qiagen, Germany), and 5) Qiagen DNeasy Powersoil (Qiagen, Germany). Phase separation and precipitation methods for DNA extraction (e.g., CTAB protocol) typically yield more DNA than silica column methods (e.g., Qiagen kits). An initial step is recommended for sediment samples, which consists of mixing the sediment sample with an equivalent volume of phosphate buffer (0.12 M Na_2HPO_4; pH=8) and then homogenizing the mixture in a shaker for 15 minutes (Holman et al., 2019).

Final product of eDNA extraction is commonly quantified using the NanoDrop (Thermo Scientific) or Qubit (ThermoFisher) system in order to determine DNA concentration. With a Qubit fluorometer, eDNA studies are enhanced by more accurate measurements, as it detects fluorescent dyes that are specific to the target of interest (e.g., DNA, RNA, or protein) in the sample, even at low concentration.

Marker selection: DNA mitochondrial vs nuclear

Mitochondrial DNA (mtDNA) is often targeted because of its high number of copies compared to nuclear DNA, its efficiency in identifying organism to species level by DNA barcoding and its accessibility through universal sequence databases on public servers (e.g., GenBank and BOLD systems). Amplified mitochondrial eDNA can come from extracellular DNA fragments, mitochondria, cells, excretion, or eggs, and the amount of quantified eDNA is likely to vary depending on the target genetic material collected (Herder et al., 2014; Goldberg et al., 2016). Although there is a clear preference for the use of markers based on mtDNA in metabarcoding studies, other types of eDNA were found to be available for this purpose, as multi-copy nuclear eDNA (e.g., ribosomal RNA genes, microsatellites; Günther et al., 2018; Andres et al., 2021; Jo et al., 2022) even though nuclear genes evolve slowly and may diminish diversity (Hillis and Dixon, 1991; Castro et al., 1998). According to Jo et al. (2021) copies of nu-eDNA may provide a more recent estimation of species abundance if its production and degradation rates are higher than those of mtDNA. In addition, using nu-eDNA may enable the evaluation of genetic diversity in a population with a higher resolution than mt-eDNA (Sigsgaard et al., 2020).

The goal of marker selection is to use a portion of the mitochondrial genome of the target species that is species-specific, with the appropriate fragment size, and amenable to accurate primer binding (Bohmann et al., 2014; Rees et al., 2014). Therefore, one of the most important considerations in eDNA studies is the design of PCR primers. Different primers and regions differ in coverage, resolution, and bias between taxa. Short DNA fragments (around 150 bp) degrade slowly and are easier to recover from environmental samples (Herder et al., 2014), so primers for eDNA studies need to amplify a short fragment. This fragment should be variable to amplify a variety of species without sacrificing the specificity of the target group (Epp et al., 2012).

For both metabarcoding and qPCR, species-specific and general primers can be used, depending on the goal of the study and the available budget. Species-specific primers may be needed to ensure detection of specific species (endangered, invasive, elusive, or rare taxa). In contrast, general primers allow detection of a wide range of species, but may not detect less abundant taxa (Thomsen et al., 2012a). The detection power of general primers will depend on i) the affinity to the sequences of target taxa, and ii) the availability of databases of DNA reference collection necessary for species identification. In addition, the choice of primers may bias the results by preferentially amplifying some target sequences more than others, as well as amplifying non-target groups (Cristescu, 2014). One potential solution to this issue is the use of multiple primer sets, in particular evolutionarily independent primer sets that match standardised barcodes for the target

taxonomic groups (Drummond et al., 2015). In this sense, many universal primers have been designed for a short fragment containing sufficient sequence variation to correctly assign fish communities (Supplementary Table S3).

Another important factor in PCR and primer design is the use of appropriate "replicates" since they increase species detection and decrease the likelihood of false negatives. The number of replicates used often differs between studies depending on detection probabilities, research objectives, sequencing depth, primer choice, cost constraints, and sequencing platform (Ficetola et al., 2015; Alberdi et al., 2018). In PCR-based amplifications for HTS, primers can be labeled with short nucleotide sequences to uniquely identify their origin in a process commonly referred to as multiplexing. These tags, while useful, also have the potential to bias results, particularly when located at the 5' end, and therefore, also require rigorous testing prior to implementation (Binladen et al., 2007; Berry et al., 2012). Because of this, the number of cycles in the indexing PCR is typically kept low to minimize PCR errors (Bohmann et al., 2022).

On the other hand, the qPCR approach is usually performed for species detection and involves the use of species-specific primer sets. Detection of a wide range of species by qPCR would involve the use of a high number of primers sets and increased costs (Lodge et al., 2012; Thomsen et al., 2012a; Lacoursière-Roussel et al., 2016a). For qPCR, short primers that recognize and flank the sequence of interest should be used. The resulting amplicons for qPCR assays are usually short (typically 50-150 bp; Supplementary Table S4), even shorter than those used in HTS platforms. DNA sequence databases for species-specific assay design are also far from complete (Kwong et al., 2012); but, qPCR assays require the development of additional sequence databases.

Quality control

In any DNA metabarcoding and qPCR experiments, it is crucial to include appropriate multiple controls in both sample collection and laboratory procedures in order to monitor for potential contamination and interpret the results correctly (Goldberg et al., 2016). The number of negative controls required at each stage should be determined according to the number of samples and the confidence required in the inference (Sepulveda et al., 2020). In this sense, different types of control can be performed, such as those suggested below:

a) Negative collection control: it is recommended to incorporate an "equipment blank" as a negative control for each filtering and sampling step. In the case of the water sampler, DNA-free distilled water, as a negative control, should be treated identically to the sampled water bottles. The DNA from this negative control shall be analysed with the sample filters, which will allow the identification of any field/transport, filter equipment, or background contamination.

b) Negative extraction controls: a negative control corresponds to a mock DNA sample, a mixture of nucleic acid molecules created *in vitro* to simulate the composition of a nucleic acid isolated therefrom, which are carried out at the same time and using the same consumables as a normal extraction, except that the sample is omitted.

c) Negative PCR controls: a PCR reaction in which the addition of template DNA is replaced by the addition of water (the same DNA-free water used to dilute the PCR reagents).

d) Positive PCR controls: the ideal positive control should be comparable to the samples analysed (similar concentration and complexity). They can be used to detect contaminants, and artifacts.

e) Tagging system controls: due to the considerable number of samples involved in DNA metabarcoding studies, it is generally necessary to implement a tagging system in which each sample shows a unique combination of forward and reverse tags.

f) Internal controls: an internal control is a template DNA that is added to the PCR mix at a low concentration to produce a small percentage of the final PCR product to act as a internal positive control (IPC). Internal controls are especially useful for estimating the relative amount of target DNA in different samples. They are also suitable for identifying PCR inhibition, even if the amount of PCR inhibitors varies between samples. In the case of eDNA samples, the use of a low amount of IPC that matches the expected concentrations of eDNA (e.g., 100 copies) may better reflect the degree of inhibition affecting the samples, as these will result in non-amplification with approximately the same level of PCR inhibition.

Advantages and limitations

As stated above, the eDNA approaches offer distinct advantages and limitations over traditional monitoring methods.

Advantages

In general, eDNA is a non-invasive method, so it does not damage or alter the species or habitats under study (Thomsen and Willerslev, 2015; Senapati et al., 2019) and is a sensitive method that can detect the target species in its environment. These advantages make this method particularly useful for

detecting cryptic and sibling species that are often difficult to identify (Senapati et al., 2019). eDNA-based methods are probably, on a medium to large scale, more cost-effective than traditional methods (Dejean et al., 2012; Herder et al., 2014). However, it will depend on the target species or the community assemblage being studied.

Several studies report shorter handling time and lower cost using eDNA compared to traditional monitoring techniques (Jerde et al., 2011; Biggs et al., 2015; Sigsgaard et al., 2015). In fact, sampling can be performed by one or two persons, thus reducing the cost of sampling. Environmental sample collection is relatively simple compared to traditional monitoring methods, which could simplify a standardised sampling scheme, which can be used in different areas (Thomsen and Willerslev, 2015). Standardisation is a requirement for all monitoring methods in order to compare different sampling areas.

In addition, the high-throughput nature of recent sequencing platforms allows multiplexing of hundreds of samples, thereby providing the means to increase new species records in a given environment (Piper et al., 2019).

Limitations

One of the main limitations of eDNA methods is the high rate of DNA degradation that clearly limits the success of these methods. In some cases, the resulting short fragments may not have sufficient genetic information to allow discrimination between species (Herder et al., 2014). Another important limitation, mainly for metabarcoding technique, is that public genetic databases are incomplete. Therefore, species-level assignment of some fish groups is of low reliability. In the case of qPCR, although the relationship between the density of a species and the amount of eDNA it releases into its environment has been proven in several experiments (e.g., Takahara et al., 2012; Thomsen et al., 2012a), little is known about how external factors (e.g., temperature, depth) influence the persistence and dilution of eDNA, making difficult to have a realistic quantification. Another factor to consider is contamination, which is a serious pitfall of eDNA methods, since it raises the possibility of false positive results. Contamination of samples can occur from sample collection in the field to each step of analysis in the laboratory. The use and sensitivity of HTS has further complicated the contamination issue, as it produces a very high yields of DNA sequences that are likely to reveal tiny amounts of lab-source PCR products (Thomsen and Willerslev, 2015; Ficetola et al., 2016).

Similar to contaminations, erroneous DNA sequences can also lead to biased results. These errors can occur prior to sampling in long-term conserved DNA, during PCR (mutations and formation of chimeric molecules), or during sequencing process (Acinas et al., 2005; Hansen et al., 2006; Thomsen and Willerslev, 2015). Because of this, raw sequence data must be carefully filtered to limit false positives and generate a reliable taxon list. There are multiple sources of error for all detection technologies and it is possible to distinguish between errors attributable specifically to the DNA-based method employed (method errors), and errors that arise during the monitoring process despite the effectiveness of that method (process errors) (Darling and Mahon, 2011). In addition, environmental samples may also contain PCR inhibitors, which can be co-extracted with eDNA. In marine environments, suspended particles (e.g., organic matter and sediment) can clog filters and increase the concentrations of PCR inhibitors (Tsai and Olson, 1992). These external substances could interact with PCR by binding to DNA or prohibiting DNA-polymerase binding, Opel et al., 2010). High concentrations of non-target DNA in marine environmental samples could strongly inhibit enzymes such as Taq Polymerase used in PCR reactions (Matheson et al., 2010; McKee et al., 2015), resulting in failed or delayed amplification of DNA from target species. Both false positive and false negative results can have consequences for the subsequent conservation effort, leading to overestimation or underestimation of the presence of a species, respectively (Goldberg et al., 2016). Furthermore, since eDNA is often distributed in its environment in patches, the sampling strategy can also strongly influence the amount of DNA found in the samples. Traditional monitoring methods provide valuable biological data such as population structure, fecundity and fish condition, this type of data, for now, cannot be provided by eDNA (Herder et al., 2014; Evans and Lamberti, 2018).

Future challenges

At present, there is no consensus on eDNA preservation and isolation protocols, nor on the choice of DNA barcodes and PCR primers, not to mention the debate concerning the parameters for clustering molecular operational taxonomic units and their taxonomic assignment. Standardization of molecular protocols is an urgent needed given the constant evolution and parallel development of new biotechnological tools for DNA data acquisition and analysis. Research efforts should focus on the development of appropriately validated tests to address environmental and sampling factors that may affect eDNA detection, develop competency and proficiency testing for laboratory accreditation, promote inter-laboratory comparisons, and improve and enhance reference databases for DNA analysis. In addition, and given the dynamic nature of marine ecosystems

novel studies are required that focus on better understanding the temporal and spatial distribution of eDNA in different marine habitats, to know exactly the relationship between eDNA concentration and species abundance, as well as to understand how biotic and abiotic factors influence the persistence and dispersal of eDNA in different environments. Furthermore, it is important to note that it cannot be stressed enough that the interpretation of eDNA results must go through well-trained taxonomists and ecologists to meaningfully interpret the results and recommend subsequent actions.

It should also be noted that the reference database of bioindicator taxa is far from complete, despite the continuing efforts of numerous national barcoding initiatives. Most existing metabarcoding data are only available locally and are geographically dispersed, which is hampering the development of useful tools at the global level. Considerable effort is still needed to ensure coverage of a range of stressor values at least as wide as that of the development of the traditional methods. In this respect, we believe that eDNA approaches will complement, rather than replace, traditional monitoring. This bioassessment tool can assist fisheries professionals achieve their research, management, and conservation objectives, but not replace time-tested assessment methods. If the costs of eDNA analysis continues to fall, and experience becomes commonplace, eDNA will become an increasingly viable option to complement fisheries monitoring and conservation programs.

It is appealing to imagine the possibilities that eDNA could open up in fisheries ecology and monitoring if advances in molecular ecology, bioinformatics, and sequencing technologies continue to accelerate. Several of the studies mentioned in this work suggest that it will one day be possible to accurately quantify the relative abundance of fish using eDNA analysis (qPCR). In addition to these promising results from eDNA analysis of target fish species, advances in eDNA metabarcoding are now making possible the simultaneous detection of multiple species and the estimation of total species richness from seawater samples. The eDNA methodology, which has already been successfully applied, is likely to be further developed in the near future, leading to increased detection probabilities and reliability (Ruppert et al., 2019).

Third-generation sequencing is likely to provide us with new opportunities to improve the performance of eDNA methods. New generations of powerful technologies, such as single molecular real-time sequencing (SMRT) (from Pacific Bioscience), carbon nanotube chips, and real-time laser transmission spectroscopy, are waiting to be fully tested for their promising potential in eDNA approaches (Lee et al., 2022). The new technology is also being applied in the eDNA collection process. Researchers at NOAA's Atlantic Oceanographic and Meteorological Laboratory have designed a new instrument that will provide valuable data on biodiversity in marine environments (Formel et al., 2021). These authors have designed a low-cost automated subsurface sampler for eDNA (SASe), which is submersible to 55 m and can filter a programmable volume of seawater and store the eDNA at a collection site.

In the future, if the technology to transmit live data is combined with currently developed technology, the sequence information of interest could be transmitted remotely, with additional overlaid that could help identify how long the eDNA has been in the environment and where it probably originated. In the short term, there is the possibility of using eDNA in population genetics, with, for example, applications for conservation genetics and phylogeography.

Author contributions

SR-A, BT, AP, and CR conceived ideas and obtained specific funding. SR-A and MB searched and collected information related to metabarcoding and qPCR, respectively. All authors contributed to the article and approved the submitted version.

Funding

The information and views set out in this publication are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein." SRA (Vicenç Mut Estabilidad) and MB (ForInDoc) were supported by postdoctoral contracts co-funded by the Regional Government of the Balearic Islands and the European Social Fund.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2022.1004674/full#supplementary-material

References

Abdul-Muneer, P. M. (2014). Application of microsatellite markers in conservation genetics and fisheries management: Recent advances in population structure analysis and conservation strategies. *Genet. Res. Int.*, 691759. doi: 10.1155/2014/691759

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M. F. (2005) PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* 71 (12), 8966–8969. doi: 10.1128/AEM.71.12.8966-8969.2005

Alberdi, A., Aizupura, O., Bohmann, K., Gopalakrishnan., S., Lynggard, C., Nielsen, M., et al. (2018). promises and pitfalls of using high-throughput sequencing for diet analysis. *Mol. Ecol. Resour.* 19, 327–348. doi: 10.1111/1755-0998.12960

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K. H., et al. (2012). Metabarcoding of 'dirt'DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* 21 (8), 1966–1979. doi: 10.1111/j.1365-294X.2011.05261.x

Andres, K. J., Sethi, S. A., Lodge, D. M., and Andrés, J. (2021). Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. *Mol. Ecol.* 30, 685–697. doi: 10.1111/mec.15765

Antich, A., Palacin, C., Wangensteen, O. S., and Turon, X. (2021). To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. *BMC Bioinf.* 22 (1), 1–24. doi: 10.1186/s12859-021-04115-6

Arreguín-Sánchez, F. (1996). Catchability: a key parameter for fish stock assessment. *Rev. Fish. Biol. Fish.* 6, 221–242. doi: 10.1007/BF00182344

Atkinson, S., Carlsson, J. E. L., Ball, B., Egan., D., Kelly-Quinn, M., Whelan, K., et al. (2018). . a quantitative PCR-based environmental DNA assay for detecting Atlantic salmon (*Salmo salar* L.). *Aquatic. Conserv.: Mar. Freshw. Ecosyst.* 28, 1238–1243. doi: 10.1002/aqc.2931

Bakker, J., Wangensteen, O. S., Chapman, D. D., Boussarie, G., Buddo, D., Guttridge, T.L., et al. (2017). Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact. *Sci. Rep.* 7, 16886. doi: 10.1038/s41598-017-17150-2

Balasingham, K. D., Walter, R. P., and Heath, D. D. (2017). Residual eDNA detection sensitivity assessed by quantitative real-time PCR in a river ecosystem. *Mol. Ecol. Resour.* 17 (3), 523–532. doi: 10.1111/1755-0998.12598

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

Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., Boussarie, G., et al. (2014). . environmental conditions influence eDNA persistence in aquatic systems. *Environ. Sci. Technol.* 48, 1819–1827. doi: 10.1021/es404734p

Berry, D., Mahfoudh, B., Wagner, M., and Loy, A. (2012). Barcode primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* 77, 7846–7849. doi: 10.1128/AEM.05220-11

Berry, T. E., Saunders, B. J., Coghlan, M. L., Stat, M., Jarman, S., Richardson, A. J., et al. (2019). Marine environmental DNA biomonitoring reveals seasonal patterns in biodiversity and identifies ecosystem responses to anomalous climatic events. *PloS Genet.* 15 (2), e1007943. doi: 10.1371/journal.pgen.1007943

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., et al. (2015). Using eDNA to develop a national citizen science-based monitoring program for the great crested newt (*Triturus cristatus*). *Biol. Conserv.* 183, 19–28. doi: 10.1016/j.biocon.2014.11.029

Binladen, J., Gilbert, M. T. P., Blolback, J. P., Panitz, F., Bendixen, C., Nielsen, R., et al. (2007). The use of code PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS One* 2, e197. doi: 10.1371/journal.pone.0000197

Bista, I., Carvalho, G., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., et al. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nat. Commun.* 8, 14087. doi: 10.1038/ncomms14087

Bloesch, J. (1995). Mechanisms, measurement and importance of sediment resuspension in lakes. *Mar. Freshw. Res.* 46 (1), 295–304. doi: 10.1071/MF9950295

Boeuf, G., and Payan, P. (2001). How should salinity influence fish growth? comp. biochem. *Physiol. Part C.* 130 (4), 411–423. doi: 10.1016/s1532-0456(01)00268-x

Bohmann, K., Elbrecht, V., Carøe, C., Bista, I., Leese, F., Bunce, M., et al. (2022). Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Mol. Ecol. Resour.* 22 (4), 1231–1246. doi: 10.1111/1755-0998.13512

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., et al. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29 (6), 358–367. doi: 10.1016/j.tree.2014.04.003

Boonstra, R. (2013). Reality as the leading cause of stress: rethinking the impact of chronic stress in nature. *Funct. Ecol.* 27 (1), 11–23. doi: 10.1111/1365-2435.12008

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J. N., et al. (2018). Environmental DNA illuminates the dark diversity of sharks. *Sci. Adv.* 4 (5), eaap9661. doi: 10.1126/sciadv.aap9661

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., and Coissac, E. (2016). obitools: a unix-inspired software package for DNA metabarcoding mol. *Ecol. Resour.* 16 (20), 176–182. doi: 10.1111/1755-0998.12428

Brown, E. A., Chain, F. J. J., Crease, T. J., MacIsaac, H. J., and Cristescu, M. E. (2015). Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecol. Evol.* 5 (11), 2234–2251. doi: 10.1002/ece3.1485

Bushnell, P. G., and Brill, R. W. (1992). Oxygen transport and cardiovascular responses in skipjack tuna (Katsuwonus pelamis) and yellowfin tuna (Thunnus albacares) exposed to acute hypoxia. *J. Comp. Physiol. B.* 162 (2), 131–143. doi: 10.1007/BF00398338

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High resolution sample inference from amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Castro, J. A., Picornell, A., and Ramon, M. (1998). Mitochondrial DNA: a tool for populational genetics studies. *Int. Microbiol.* 1 (4), 327–332.

Closek, C. J., Santora, J. A., Starks, H. A., Schroeder, I. D., Andruszkiewicz, E. A., Sakuma, K. M., et al. (2019). Marine vertebrate biodiversity and distribution within the central California current using environmental DNA (eDNA) metabarcoding and ecosystem surveys. *Front. Mar. Sci.* 6, 732. doi: 10.3389/fmars.2019.00732

Coble, A. A., Flinders, C. A., Homyack, J. A., Penaluna, B. E., Cronn, R. C., and Weitemier, K. (2019). eDNA as a tool for identifying freshwater species in sustainable forestry: a critical review and potential future applications. *Sci. Total Environ.* 649, 1157–1170. doi: 10.1016/j.scitotenv.2018.08.370

Cochrane, K. L. (2000). Reconciling sustainability, economic efficiency and equity in fisheries: the one that got away? *Fish. Fish.* 1, 3–21. doi: 10.1046/j.1467-2979.2000.00003.x

Cochrane, K. L., and Garcia, S. M. (2009). "Introduction – fisheries management," in *A fishery manager's guidebook*. Eds. K. L. Cochrane and S. M. Garcia (West Sussex, United Kingdom: FAO). doi: 10.1002/9781444316315.ch1

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., Genner, J., et al. (2018). Persistence of environmental DNA in marine systems. *Commun. Biol.* 1, 185. doi: 10.1038/s42003-018-0192-6

Coolen, M. J., and 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. *Environ. Microbiol.* 9 (1), 238–249. doi: 10.1111/j.1462-2920.2006.01134.x

Corinaldesi, C., Beolchini, F., and Dell'Anno, A. (2008). Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Mol. Ecol.* 17 (17), 3939–3951. doi: 10.1111/j.1365-294X.2008.03880.x

Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* 29, 556–571. doi: 10.1016/j.tree.2014.08.001

Curd, E. E., Gold, Z., Kandlikar, G. S., Gomer, J., Ogden, M., O'Conell, T., et al. (2019). Anacapa toolkit: An environmental DNA toolkit for processing multilocus metabarcode datasets. *Methods Ecol. Evol.* 10, 1469 – 1475. doi: 10.1111/2041-210X.13214

Darling, J. A. (2015). Genetic studies of aquatic biological invasions: closing the gap between research and management. *Biol. Invasions* 17, 951–971. doi: 10.1007/s10530-014-0726-x

Darling, J. A., and Mahon, A. R. (2011). From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environ. Res.* 111 (7), 978–988. doi: 10.1016/j.envres.2011.02.001

Deagle, B. E., Chiaradia, A., McInnes, J. , and Jarman, S. N. (2010). Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conserv. Genet.* 11, 2039–2048. doi: 10.1007/s10592-010-0096-6

Debroas, D., Domaizon, I., Humbert, J. F., Jardillier, L., Lepère, C., Oudart, A., et al. (2017). Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiol.* 93, (4). doi: 10.1093/femsec/fix023

Deere, D., Porter, J., Pickup, R. W., and Edwards, C. (1996). Survival of cells and DNA of aeromonas salmonicida released into aquatic microcosms. *J. Appl. Bacteriol.* 81 (3), 309–318. doi: 10.1111/j.1365-2672.1996.tb04333.x

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., et al. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* 26 (21), 5872–5895. doi: 10.1111/mec.14350

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

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., et al. (2011). Persistence of environmental DNA in freshwater ecosystems. *PloS One* 6 (8), e23398. doi: 10.1371/journal.pone.0023398

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., Miaud, C., et al. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. J. *Appl. Ecol.* 49 (4), 953–959. doi: 10.1111/j.1365-2664.2012.02171.x

Dell'Anno, A., and Corinaldesi, C. (2004). Degradation and turnover of extracellular DNA in marine sediments: ecological and methodological considerations. *Appl. Environ. Microbiol.* 70 (7), 4384–4386. doi: 10.1128/AEM.70.7.4384-4386.2004

Dennis, D., Plagányi, E., Van Putten, I., Hutton, T., and Pascoe, S. (2015). Cost benefit of fishery-independent surveys: Are they worth the money? *Mar. Policy* 58, 108–115. doi: 10.1016/j.marpol.2015.04.016

de Souza, L. S., Godwin, J. C., Renshaw, M. A., and Larson, E. (2016). Environmental DNA (eDNA) detection probability is influenced by seasonal activity of organisms. *PloS One* 11, e0165273. doi: 10.1371/journal.pone.0165273

Díaz-Ferguson, E. E., and Moyer, G. R. (2014). History, applications, methodological issues and perspectives for the use of environmental DNA (eDNA) in marine and freshwater environments. *Rev. Biol. Trop.* 62 (4), 1273–1284. doi: 10.15517/rbt.v62i4.13231

Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., et al. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nat. Commun.* 11, 254. doi: 10.1038/s41467-019-14105-1

Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., et al. (2017). Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshw. Biol.* 62, 30–39. doi: 10.1111/fwb.12846

Douville, M., Gagné, F., Blaise, C., and André, C. (2007). Occurrence and persistence of bacillus thuringiensis (Bt) and transgenic bt corn cry1Ab gene from an aquatic environment. ecotoxicol. *Environ. Saf.* 66, 195–203. doi: 10.1016/j.ecoenv.2006.01.002

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., et al. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* 4, 46. doi: 10.1186/s13742-015-0086-1

Dufresne, Y., Lejzerowicz, F., Perret-Gentil, L. A., Pawlowski, J., and Cordier, T. (2019). SLIM: a flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinf.* 20, 88. doi: 10.1186/s12859-019-2663-2

Eichmiller, J. J., Miller, L. M., and Sorensen, P. W. (2016). Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Mol. Ecol. Resour.* 16 (1), 56–68. doi: 10.1111/1755-0998.12421

Engelsma, M. Y., Hougee, S., Nap, D., Hofenk, M., Rombout, J. H., van Muiswinkel, W. B., et al. (2003). Multiple acute temperature stress affects leucocyte populations and antibody responses in common carp, *Cyprinus carpio* L. *Fish. Shellfish Immun.* 15 (5), 397–410. doi: 10.1016/S1050-4648(03)00006-8

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., et al. (2012). New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 21, 1821–1833. doi: 10.1111/j.1365-294X.2012.05537.x

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., et al. (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* 4, 1111–1119. doi: 10.1111/2041-210X.12114

Erickson, R. A., Rees, C. B., Coulter, A. A., Merkes, C. M., McCalla, S. G., Touzinsky, K. F., et al. (2016). Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Mol. Ecol. Resour.* 16 (4), 957–965. doi: 10.1111/1755-0998.12533

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

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., et al. (2016). Quantification of mesocosm fish and amphibian species diversity *via* environmental DNA metabarcoding. *Mol. Ecol. Resour.* 16 (1), 29–41. doi: 10.1111/1755-0998.12433

Ficetola, G. F., Miaud, C., and Pompanon F and Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biol. Lett.* 4 (4), 423–425. doi: 10.1098/rsbl.2008.0118

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., et al. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol. Ecol. Resour.* 15, 543–556. doi: 10.1111/1755-0998.12338

Ficetola, G. F., Taberlet, P., and Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? *Mol. Ecol. Resour.* 16 (3), 604–607. doi: 10.1111/1755-0998.12508

Foote, A. D., Thomsen, P. F., Sveegaard, S., Wahlberg, M., and Kielgast, J. (2012). Investigating the potential use ofof Environmental DNA (eDNA) for genetic monitoring of marine mammals. *PloS One* 7 (8), e41781. doi: 10.1371/journal.pone.0041781

Formel, N., Enochs, I. C., Sinigalliano, C., Anderson, S. R., and Thompson, L. R. (2021). Subsurface automated samplers for eDNA (SASe) for biological monitoring and research. *HardwareX* 10, e00239. doi: 10.1016/j.ohx.2021.e00239

Fraser, H. M., Greenstreet, S. P. R., and Piet, G. J. (2007). Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass. *J. Mar. Sci.* 64 (9), 1800–1819. doi: 10.1093/icesjms/fsm145

Furlan, E. M., Gleeson, D., Hardy, C. M., and Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. *Mol. Ecol. Resour.* 16, 641–654. doi: 10.1111/1755-0998.12483

Gale, N. K., Heath, G., Cameron, E., Rashid, S., and Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med. Res. Methodol.* 13 (1), 117. doi: 10.1186/1471-2288-13-117

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., and Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345(6270), 60–63. doi: 10.1038/345060a0

Goldberg, C. S., Pilliod, D. S., Arkle, R. S., and Waits, L. P. (2011). Molecular detection of vertebrates in stream water: a demonstration using rocky mountain tailed frogs and Idaho giant salamanders. *PloS One* 6 (7), e22746. doi: 10.1371/journal.pone.0022746

Goldberg, C. S., Strickler, K. M., and Pilliod, D. S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biol. Conserv.* 183, 1–3. doi: 10.1016/j.biocon.2014.11.040

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., et al. (2016). Critical considerations for the application of environmental DNA methods to detect 0aquatic species. *Methods Ecol. Evol.* 7 (11), 1299–1307. doi: 10.1111/2041-210X.12595

Gold, Z., Sprague, J., Kushner, D. J., Zerecero, M. E., and Barber, P. H. (2021). eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PloS One* 16 (2), e0238557. doi: 10.1371/journal.pone.0238557

Grassle, J. F., and Maciolek, N. J. (1992). Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *Am. Nat.* 139 (2), 313–341. Available at: https://www.jstor.org/stable/2462414

Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S., and Turon, X. (2015). Deep-Sea, deep-sequencing: Metabarcoding extracellular DNA from sediments of marine canyons. *PloS One* 10 (10), e0139633. doi: 10.1371/journal.pone.0139633

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., et al. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PloS Biol.* 16 (1), e2003703. doi: 10.1371/journal.pbio.2003703

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5 (10), R245–R249. doi: 10.1016/s1074-5521(98)90108-9

Hansen, B. K., Bekkevold, D., Clausen, L. W., and Nielsen, E. E. (2018). The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish. Fish.* 19 (5), 751–768. doi: 10.1111/faf.12286

Hansen, A. J., Mitchell, D. L., Wiuf, C., Paniker, L., and Brand, T. B. (2006). Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173 (2), 1175–1179. doi: 10.1534/genetics.106.057349

Harris, J. D. (2003). Can you bank on GenBank? *Trends Ecol. Evol.* 18, 317–319. doi: 10.1016/S0169-5347(03)00150-2

Hebsgaard, M. B., Phillips, M. J., and Willerslev, E. (2005). Geologically ancient DNA: fact or artefact? *Trends Microbiol.* 13 (5), 212–220. doi: 10.1016/j.tim.2005.03.010

Herder, J., Valentini, A., Bellemain, E., Dejean, T., Van Delft, J. J. C. W., Thomsen, P. F., et al. (2014). *Environmental DNA. a review of the possible applications for the detection of (invasive) species* (Stichting RAVON, Nijmegen: Netherlands Food and Consumer Product Safety Authority). Report 2013-104. doi: 10.13140/RG.2.1.4002.1208

Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev. Biol.* 66, 411–453. doi: 10.1086/417338

Hinlo, R., Gleeson, D., Lintermans, M., and Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PloS One* 12 (6), e0179251. doi: 10.1371/journal.pone.0179251

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. *Nat. Rev. Genet.* 2 (5), 353. doi: 10.1038/35072071

Holman, L. E., Bruyn, M., Creer, S., Carvalho, G., Robidart, J., and Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Sci. Rep.* 9, 911559. doi: 10.1038/s41598-019-47899-7

Jarman, S. N., Berry, O., and Bunce, M. (2018). The value of environmental DNA biobanking for long-term biomonitoring. *Nat. Ecol. Evol.* 2, 1192–1193. doi: 10.1038/s41559-018-0614-3

Jerde, C. L., and Mahon, A. R. (2015). Improving confidence in environmental DNA species detection. *Mol. Ecol. Resour.* 15, 461–463. doi: 10.1111/1755-0998.12377

Jerde, C. L., Mahon, A. R., Chadderton, W. L., and Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv. Lett.* 4 (2), 150–157. doi: 10.1111/j.1755-263X.2010.00158.x

Jo, T., Sakata, M. K., Murakami, H., Masuda, R., and Minamoto, T. (2021). Universal performance of benzalkonium chloride for the preservation of environmental DNA in seawater samples. *Limnol. Oceanogr-Methods* 19, 758–768. doi: 10.1002/lom3.10459

Jo, T., Takao, K., and Minamoto, T. (2022). Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. *Environ. DNA* 4, 271–283. doi: 10.1002/edn3.253

Keck, F., Vasselon, V., Tapolczai, K., Rimet, F., and Bouchez, A. (2017). Freshwater biomonitoring in the information age. *Front. Ecol. Environ.* 15 (5), 266–274. doi: 10.1002/fee.1490

Kelly, R. P., Port, J. A., Yamahara, K. M., and Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *PloS One* 9 (1), e86175. doi: 10.1371/journal.pone.0086175

Kim, A. R., Yoon, T. H., Lee, C. I., Kang, C. K., and Kim, H. W. (2021). Metabarcoding analysis of ichthyoplankton in the East/Japan Sea using the novel fish-specific universal primer set. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.614394

Klymus, K. E., Richter, C. A., Chapman, D. C., and Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biol. Conserv.* 183, 77–84. doi: 10.1016/j.biocon.2014.11.020

Knudsen, S. W., Ebert, R. B., Hesselsøe, M., Kuntke, F., Hassingboe, J., Bondgaard, P., et al. (2019). Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic sea. *J. Exp. Mar. Biol. Ecol.* 510, 31–45. doi: 10.1016/j.jembe.2018.09.004

Knudsen, E. E., and McDonald, D. (2019). *Sustainable fisheries management* (Boca Ratón, United States: CRC Press), ISBN: .

Krishnan, A., Sweeney, M., Vasic, J., Galbraith, D., and Vasic, B. (2011). Barcodes for DNA sequencing with guaranteed error-correction capability. *Electron. Lett.* 47 (4), 236–237. doi: 10.1049/el.2010.3546

Kwong, S., Srivathsan, A., and Meier, R. (2012). An update on DNA barcoding: low species coverage and numerous unidentified sequences. *Cladistics* 28, 639–644. doi: 10.1111/j.1096-0031.2012.00408.x

Lacoursière-Roussel, A., Côté, G., Leclerc, V. , and Bernatchez, L. (2016a). Quantifying relative fish abundance with eDNA: A promising tool for fisheries management. *J. Appl. Ecol.* 53, 1148–1157. doi: 10.1111/1365-2664.12598

Lacoursière-Roussel, A., Rosabal, M. , and Bernatchez, L. (2016b). Estimating fish abundance and biomass from eDNA concentrations: Variability among capture methods and environmental conditions. *Mol. Ecol. Resour.* 16, 1401–1414. doi: 10.1111/1755-0998.12522

Laramie, M. B., Pilliodm, D. S., and Goldberg, C. S. (2015). Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biol. Conserv.* 183, 29–37. doi: 10.1016/j.biocon.2014.11.025

Lee, J. H., La, H. S., Kim, J. H., Son, W., Park, H., Kim, Y. M., et al. (2022). Application of dual metabarcoding platforms for the meso- and macrozooplankton taxa in the Ross sea. *Genes* 13, 922. doi: 10.3390/genes13050922

Levi, T., Allen, J. M., Bell, D., Joyce, J., Russuell, J. R., Tallmon, D. A., et al. (2019). Environmental DNA for the enumeration and management of pacific salmon. *Mol. Ecol. Resour.* 19, 597–608. doi: 10.1111/1755-0998.12987

Levy-Booth, D. J., Campbell, R. G., Gulden, R. H., Hart, M. M., Powell,, Klironomos, J. N., et al. (2007). Cycling of extracellular DNA in the soil environment. *Soil Biol. Biochem.* 39 (12), 2977–2991. doi: 10.1016/j.soilbio.2007.06.020

Lodge, D. M., Turner, C. R., Jerde, C. L., Barnes, M. A., Chadderton, L., Egan, S. P., et al. (2012). Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Mol. Ecol.* 21 (11), 2555–2558. doi: 10.1111/j.1365-294X.2012.05600.x

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., and Minamoto, T. (2014). The release rate of environmental DNA from juvenile and adult fish. *PloS One* 9 (12), e114639. doi: 10.1371/journal.pone.0114639

Matheson, C. D., Gurney, C., Esau, N., and Lehto, R. (2010). Assessing PCR inhibition from humic substances. *Open Enzym. Inhid. J.* 3 (1), 38–45. doi: 10.2174/1874940201003010038

Mathon, L., Guérin, P. E., Normandeau, E., Valentini, A., Noel, C., Lionnet, C., et al. (2021). Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Mol. Ecol. Resour.* 21 (7), 2565–2579. doi: 10.1111/1755-0998.13430

Matsui, K., Honjo, M., and Kawabata, Z. (2001). Estimation of the fate of dissolved DNA in thermally stratified lake water from the stability of exogenous plasmid DNA. *Aquat. Microb. Ecol.* 26 (1), 95–102. doi: 10.3354/ame026095

McCormick, S. T., Pinedo, M. L., Shenker, S., and Wolf, B. (1989). Sequencing in an assembly line with blocking to minimize cycle time. *Oper. Res.* 37 (6), 925–935. doi: 10.1287/opre.37.6.925

McKee, A. M., Spear, S. F., and Pierson, T. W. (2015). The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol. Conserv.* 183, 70–76. doi: 10.1016/j.biocon.2014.11.031

McKenzie, D. J., Höglund, E., Dupont-Prinet, A., Larsen, B. K., Skov, P. V., Pedersen, P. B., et al. (2012). Effects of stocking density and sustained aerobic exercise on growth, energetics and welfare of rainbow trout. *Aquaculture* 338, 216–222. doi: 10.1016/j.aquaculture.2012.01.020

Milidine, K., Armstrong, J., and Metcalfe, N. (1995). Presence of shelter reduces maintenance metabolism of juvenile salmon. *Funct. Ecol.* 20 (5), 839–845. doi: 10.1111/j.1365-2435.2006.01166.x

Noll, D. M., Mason, T. M., and Miller, P. S. (2006). Formation and repair of interstrand cross-links in DNA. *Chem. Rev.* 106 (2), 277–301. doi: 10.1021/cr040478b

Ogram, A., Sayler, G. S., and Barkay, T. (1987). The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* 7 (2-3): 57-66. doi: 10.1016/0167-7012(87)90025-X

Olson, Z. H., Briggler, J. T., and Williams, R. N. (2012). An eDNA approach to detect eastern hellbenders (Cryptobranchus a. alleganiensis) using samples of water. *Wildl. Res.* 39 (7), 629–636. doi: 10.1071/WR12114

Opel, K. L., Chung, D., and McCord, B. R. (2010). A study of PCR inhibition mechanisms using real time PCR. *J. Forensic Sci.* 55 (1), 25–33. doi: 10.1111/j.1556-4029.2009.01245.x

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R., and Dichmont, M. (2015). Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish. Fish.* 16, 125–159. doi: 10.1111/faf.12052

Palsbøll, P., Berube, M., and Allendorf, F. (2007). . identification of management units using population genetic data. *Trends Ecol. Evol.* 22, 11–16. doi: 10.1016/j.tree.2006.09.003

Person-Le Ruyet, J., Mahé, K., Le Bayon, N., and Le, Delliou, H. (2004). Effects of temperature on growth and metabolism in a Mediterranean population of European sea bass, *Dicentrarchus labrax*. *Aquaculture* 237 (1-4), 269–280. doi: 10.1016/j.aquaculture.2004.04.021

Petty, J. T., Hansbarger, J. L., Huntsman, B. M., and Mazik, P. M. (2012). Brook trout movement in response to temperature, flow, and thermal refugia within a complex Appalachian riverscape. *Trans. Am. Fish. Soc* 141 (4), 1060–1073. doi: 10.1080/00028487.2012.681102

Piaggio, A. J., Engeman, R. M., Hopken, M. W., Humphrey, J. S., Keacher, K. L., Bruce, W. E., et al. (2014). Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Mol. Ecol. Resour.* 14 (2), 374–380. doi: 10.1111/1755-0998.12180

Piggott, M. P. (2016). Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecol. Evol.* 6 (9), 2739–2750. doi: 10.1002/ece3.2083

Pikitch, E. K., Santora, C., Babcock, E. A., Bakun, A., Bonfil, R., Dayton, P., et al. (2004). Ecology. ecosystem-based fishery management. *Science* 305 (5682), 346–347. doi: 10.1126/science.1098222

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., and Waits, L. P. (2014). Factors influencing detection of eDNA from a stream-dwelling amphibian. *Mol. Ecol. Resour.* 14 (1), 109–116. doi: 10.1111/1755-0998.12159

Piper, A. M., Batovska, J., Cogan, N. O. I., Weiss, J., Cunningham, J. P., Rodoni, B. C., et al. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience* 8 (8), giz092. doi: 10.1093/gigascience/giz092

Ratcliffe, F. C., Uren Webster, T. M., Garcia de Leaniz, C., and Consuegra, S. (2020). A drop in the ocean: Monitoring fish communities in spawning areas using environmental DNA. *Environ. DNA* 3, 43–54. doi: 10.1002/edn3.87

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., and Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* 51 (5), 1450–1459. doi: 10.1111/1365-2664.12306

Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., and Lodge., D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Mol. Ecol. Resour.* 15 (1), 168–176. doi: 10.1111/1755-0998.12281

Robson, H. L., Noble, T. H., Saunders, R. J., Robson, S. K., Burrows, D. W., Jerry, D. R., et al. (2016). Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Mol. Ecol. Resour.* 16 (4), 922–932. doi: 10.1111/1755-0998.12505

Rourke, M. L., Fowler, A. M., Hughes, J. M., Broadhurst, M., DiBattista, J., Fielder, S., et al. (2022). Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. *Environ. DNA*. 4, 9–33. doi: 10.1002/edn3.185

Ruppert, K. M., Kline, R. J., and Rahman, M. S. (2019). . past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* 17, e00547. doi: 10.1016/j.gecco.2019.e00547

Russo, T., Maiello, G., Talarico, L., Baillie, C., Colosimo, G., D'Andrea, L., et al. (2021). All is fish that comes to the net: metabarcoding for rapid fisheries catch assessment. *Ecol. Appl.* 31 (2), e02273. doi: 10.1002/eap.2273

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., and Vestergaard, P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Commun. Biol.* 2, 461. doi: 10.1038/s42003-019-0696-8

Santos, A. M., and Branco, M. (2012). The quality of name-based species records in databases. *Trends Ecol. Evol.* 27, 6–7. doi: 10.1016/j.tree.2011.10.004

Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., and Boehm, A. B. (2016). Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. *Environ. Sci. Technol.* 50 (19), 10456–10464. doi: 10.1021/acs.est.6b03114

Sato, Y., Miya, M., Fukunaga, T., Sado, T., and Iwasaki, W. (2018). MitoFish and MiFish pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Mol. Biol. Evol.* 35, 1553–1555. doi: 10.1093/molbev/msy074

Selong, J. H., McMahon, T. E., Zale, A. V., and Barrows, F. T. (2001). Effect of temperature on growth and survival of bull trout, with application of an improved method for determining thermal tolerance in fishes. *Trans. Am. Fish. Soc* 130 (6), 1026–1037. doi: 10.1577/1548-8659(2001)130<1026:EOTOGA>2.0.CO;2

Senapati, D., Bhattacharya, M., Kar, A., Chini, D. S., Das, B. K., and Patra, B. C. (2019). Environmental DNA (eDNA): A promising biological survey tool for aquatic species detection. *Proc. Zool. Soc* 72 (3), 211–228. doi: 10.1007/s12595-018-0268-9

Sepulveda, A. J., Hutchins, P. R., Forstchen, M., Mckeefry, M. N., and Swigris, A. M. (2020). The elephant in the Lab (and field): Contamination in aquatic environmental DNA studies. *Front. Ecol. Evol.* 8. doi: 10.3389/fevo.2020.609973

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., and Thomsen, P. F. (2020). Population-level inferences from environmental DNA–current status and future perspectives. *Evol. Appl.* 13, 245 – 262. doi: 10.1111/eva.12882

Sigsgaard, E. E., Carl, H., Møller, P. R., and Thomsen, P. F. (2015). Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biol. Conserv.* 183, 46–52. doi: 10.1016/j.biocon.2014.11.023

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., et al. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nat. Ecol. Evol.* 1 (1), 4. doi: 10.1038/s41559-016-0004

Sloman, K. A., Motherwell, G., O'connor, K., and Taylor, A. C. (2000). The effect of social stress on the standard metabolic rate (SMR) of brown trout, salmo trutta. fish physiol. *Biochem* 23 (1), 49–53. doi: 10.1023/A:1007855100185

Snelgrove, P. V. (1999). Getting to the bottom of marine biodiversity: sedimentary habitats: ocean bottoms are the most widespread habitat on earth and support high biodiversity and key ecosystem services. *BioScience* 49 (2), 129–138. doi: 10.2307/1313538

Spear, S. F., Groves, J. D., Williams, L. A., and Waits, L. P. (2015). Using environmental DNA methods to improve detectability in a hellbender (Cryptobranchus alleganiensis) monitoring program. *Biol. Conserv.* 183, 38–45. doi: 10.1016/j.biocon.2014.11.016

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., et al. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Sci. Rep.* 7, 12240. doi: 10.1038/s41598-017-12501-5

Stewart, K. A. (2019). Understanding the effects of biotic and abiotic factors on sources of aquatic environmental DNA. *Biodivers. Conserv.* 28 (5), 983–1001. doi: 10.1007/s10531-019-01709-8

Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z., Dunton, K. J., Hinks, G., VanMorter, S. M., et al. (2021). Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal new Jersey, USA. ICES. *J. Mar. Sci.* 78, (1) 293–304. doi: 10.1093/icesjms/fsaa225

Stoeckle, M. Y., Das, M. M., and Charlop-Powers, Z. (2020). Improved environmental DNA reference library detects overlooked marine fishes in new Jersey, united states. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00226

Stoeckle, M. Y., Soboleva, L., and Charlop-Powers, Z. (2017). Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PloS One* 12 (4), e0175186. doi: 10.1371/journal.pone.0175186

Taberlet, P., Bonin, A., Zinger, L., and Coissac, E. (2018). *Environmental DNA* (UK: Oxford University Press), ISBN: . doi: 10.1093/oso/9780198767220.001.0001

Taberlet, P., Coissac, E., Hajibabaei, M. , and Rieseberg, L. H. (2012a). Environmental DNA. *Mol. Ecol.* 21 (8), 1789–1793. doi: 10.1111/j.1365-294X.2012.05542.x

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J. , Miquel, C., Shehzad, W., et al. (2012b). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* 21 (8), 1816–1820. doi: 10.1111/j.1365-294X.2011.05317.x

Taberlet, P., Waits, L. P., and Luikart, G. (1999). Noninvasive genetic sampling: look before you leap. *Trends Ecol. Evol.* 14, 323–327. doi: 10.1016/s0169-5347(99)01637-7

Takahara, T., Minamoto, T., and Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PloS One* 8 (2), e56584. doi: 10.1371/journal.pone.0056584

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., and Kawabata, Z. I. (2012). Estimation of fish biomass using environmental DNA. *PloS One* 7 (4), e35868. doi: 10.1371/journal.pone.0035868

Takahara, T., Yamanaka, H., Suzuki, A. A., Honjo, M. N., Minamoto, T., Yonekura, R., et al. (2011). Stress response to daily temperature fluctuations in common carp, *Cyprinus carpio l. Hydrobiologia* 675 (1), 65. doi: 10.1007/s10750-011-0796-z

Tapolczai, K., Vasselon, V., Bouchez, A., Stenger-Kovács, C., Padisák, J., and Rimet, F. (2019). The impact of OTU sequence similarity threshold on diatom-based bioassessment: a case study of the rivers of Mayotte (France, Indian ocean). *Ecol. Evol.* 9, 166–179. doi: 10.1002/ece3.4701

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., and Willerslev, E. (2012a). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PloS One* 7 (8), e41732. doi: 10.1371/journal.pone.0041732

Thomsen, P. F., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., et al. (2012b). Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* 21 (11), 2565–2573. doi: 10.1111/j.1365-294X.2011.05418.x

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., and Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PloS One* 11 (11), e0165252. doi: 10.1371/journal.pone.0165252

Thomsen, P. F., and Willerslev, E. (2015). Environmental DNA–an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183, 4–18. doi: 10.1016/j.biocon.2014.11.019

Tréguier, A., Paillisson, J. M., Dejean, T., Valentini, A., Schlaepfer, M. A., and Roussel, J. M. (2014). Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *J. Appl. Ecol.* 51 (4), 871–879. doi: 10.1111/1365-2664.12262

Truelove, N. K., Andruszkiewicz, E. A., and Block, B. A. (2019). A rapid environmental DNA method for detecting white sharks in the open ocean. *Methods Ecol. Evol.* 10, 1128–1135. doi: 10.1111/2041-210X.13201

Tsai, Y. L., and Olson, B. H. (1992). Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl. Environ. Microbiol.* 58 (7), 2292–2295. doi: 10.1128/aem.58.7.2292-2295.1992

Turner, C. R., Uy, K. L., and Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol. Conserv.* 183, 93–102. doi: 10.1016/j.biocon.2014.11.017

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., et al. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding Metagenom.* 2, e23297. doi: 10.3897/mbmg.2.23297

Valsecchi, E., Arcangeli, A., Lombardi, R., Boyse, E., Carr, I. M., Galli, P., et al. (2021). Ferries and environmental DNA: underway sampling from commercial vessels provides new opportunities for systematic genetic surveys of marine biodiversity. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.704786

Van Dievel, M., Janssens, L., and Stoks, R. (2016). Short-and long-term behavioural, physiological and stoichiometric responses to predation risk indicate chronic stress and compensatory mechanisms. *Oecologia* 181 (2), 347–357. doi: 10.1007/s00442-015-3440-1

Wang, Y., Chen, K., Gao, J., Wang, M., Dong, J., Xie, Y., et al. (2021). Environmental DNA of preservative ethanol performed better than water samples in detecting macroinvertebrate diversity using metabarcoding. *Divers. Distrib.* 27, 1989–2002. doi: 10.1111/ddi.13284

Willerslev, E., and Cooper, A. (2005). Ancient DNA. *Proc. R. Soc B. Sci.* 272 (1558), 3–16. doi: 10.1098/rspb.2004.2813

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., et al. (2003). Diverse plant and animal genetic records from Holocene and pleistocene sediments. *Science* 300 (5620), 791–795. doi: 10.1126/science.1084114

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., et al. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Sci. Rep.* 7, 40368. doi: 10.1038/srep40368

Yoccoz, N. G. (2012). The future of environmental DNA in ecology. *Mol. Ecol.* 21 (8), 2031–2038. doi: 10.1111/j.1365-294X.2012.05505.x

Zhu, B. (2006). Degradation of plasmid and plant DNA in water microcosms monitored by natural transformation and real- time polymerase chain reaction (PCR). *Water Res.* 40 (17), 3231–3238. doi: 10.1016/j.watres.2006.06.040

frontiers | Frontiers in Marine Science

TYPE Perspective
PUBLISHED 06 January 2023
DOI 10.3389/fmars.2022.1014361

OPEN ACCESS

EDITED BY
Fran Saborido-Rey,
Spanish National Research Council (CSIC), Spain

REVIEWED BY
Jitendra Kumar Sundaray,
Central Institute of Freshwater Aquaculture (ICAR), India
Wenbin Zhou,
University of North Carolina at Chapel Hill, United States

*CORRESPONDENCE
Miguel Baltazar-Soares
miguelalexsoares@gmail.com

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 08 August 2022
ACCEPTED 06 December 2022
PUBLISHED 06 January 2023

CITATION
Baltazar-Soares M, Lima ARA, Silva G and Gaget E (2023) Towards a unified eco-evolutionary framework for fisheries management: Coupling advances in next-generation sequencing with species distribution modelling.
Front. Mar. Sci. 9:1014361.
doi: 10.3389/fmars.2022.1014361

Towards a unified eco-evolutionary framework for fisheries management: Coupling advances in next-generation sequencing with species distribution modelling

Miguel Baltazar-Soares[1]*, André R. A. Lima[2], Gonçalo Silva[2] and Elie Gaget[1]

[1]Department of Biology, University of Turku, Turku, Finland, [2]MARE–Marine and Environmental Sciences Centre, ARNET–Aquatic Research Network, ISPA – Instituto Universitário de Ciências Psicológicas, Lisboa, Portugal

The establishment of high-throughput sequencing technologies and subsequent large-scale genomic datasets has flourished across fields of fundamental biological sciences. The introduction of genomic resources in fisheries management has been proposed from multiple angles, ranging from an accurate re-definition of geographical limitations of stocks and connectivity, identification of fine-scale stock structure linked to locally adapted sub-populations, or even the integration with individual-based biophysical models to explore life history strategies. While those clearly enhance our perception of patterns at the light of a spatial scale, temporal depth and consequently forecasting ability might be compromised as an analytical trade-off. Here, we present a framework to reinforce our understanding of stock dynamics by adding also a temporal point of view. We propose to integrate genomic information on temporal projections of species distributions computed by Species Distribution Models (SDMs). SDMs have the potential to project the current and future distribution ranges of a given species from relevant environmental predictors. These projections serve as tools to inform about range expansions and contractions of fish stocks and suggest either suitable locations or local extirpations that may arise in the future. However, SDMs assume that the whole population respond homogenously to the range of environmental conditions. Here, we conceptualize a framework that leverages a conventional Bayesian joint-SDM approach with the incorporation of genomic data. We propose that introducing genomic information at the basis of a joint-SDM will explore the range of suitable habitats where stocks could thrive in the future as a function of their current evolutionary potential.

KEYWORDS

high-throughput sequencing, genomics, species distribution model (SDMs), fisheries applications, evolutionary ecology

Introduction

There is unequivocal evidence that marine biodiversity is declining, with severe impacts on marine ecosystems that reverberate at ecological, social and economic scales (Cardinale et al., 2012). As 40% of the human population lives within 100 km from the coast (Stead, 2018), marine fish constitute one of the most accessible food bases and the main source of protein. Overfishing, habitat destruction and pollution have been pointed as the main responsible causes for marine biodiversity loss both regionally and globally, but the role of climate change in shifting distribution ranges and promoting local extinctions is becoming more and more evident (Brander, 2010; Lam et al., 2020). Climate change is reflected in temperature increments inducing modifications both at biochemical and geological levels, including ocean acidification, hypoxia, sea level rises, and more frequent droughts, storms, or heat waves (Lohbeck et al., 2012; Frölicher and Laufkötter, 2018). Environmental oscillations and anthropogenic pressures such as fisheries have direct effects on the trophodynamic structure and function of marine systems, and consequently a massive impact on worldwide fisheries (Cheung et al., 2009). At the light of such events, fisheries research has become pivotal in understanding stocks response to climatic shifts and attempts to predict their future distribution and abundance. Currently, there are perhaps two major tools at scientists' disposal to do it so. The first is genomic resources, which became increasingly more reachable for non-model species after the advent of next-generation sequencing a decade ago (Allendorf et al., 2010; Benestan et al., 2016). Genetics has since decades delivered valuable outputs to fisheries management, though restrained to a handful of well-established systems (Bernatchez et al., 2017). Early implementations of genetic information came tackle some of the wider gaps on fisheries management, such as estimates of stock connectivity and delimitation, development of monitoring programs, and design of marine protected areas (Hauser and Carvalho, 2008; Verspoor et al., 2008; Casey et al., 2016; Blasco et al., 2020). Characterizing genetic diversity as respective distribution patterns became more enticing with NGS as it permitted to expand the search beyond the traditional stock structure and connectivity towards signatures of selection and inference of the putative adaptive potential of those same stocks (Therkildsen et al., 2013; Baltazar-Soares et al., 2021b). Inferring adaptive potential on top of stock connectivity is pivotal to estimate the repertoire of genetic-based adaptive responses and build expectations on the spatial reshuffling of adaptive alleles (Eizaguirre and Baltazar-Soares, 2014; Capblancq et al., 2020).

The second one is the development of algorithms to infer species distribution forecasts, which is being performed by using Species Distribution Models (SDM) based on assumptions of niche conservatism (Guisan and Thuiller, 2005). These have been created to provide an understanding of the factors and processes that may cause fluctuations in local populations, and to facilitate climate-ready management of living marine resources under social, economic, and ecological perspectives (Porfirio et al., 2014; Villero et al., 2017). Distribution models can describe essential habitats for early (egg, larval, and pelagic juvenile) and later (juvenile and adult) life history stages of marine fishes (Zurell et al., 2016; Laman et al., 2017). They can also be used to predict potential spawning habitats (Planque et al., 2007), provide a basis to define new fishing areas by evaluating gains and losses in species suitable areas over time, and propose better management options in areas where habitat contractions are predicted in the future (Lima et al., 2022). Different levels of complexity and data integration already exist in SDMs, but efforts are still needed to offer more than correlative outputs (Zurell et al., 2016). Indeed, the integration of complementary methodologies and multidisciplinary approaches, such as genomics and SDM, constitute promising advances on predicting species response to environmental changes by incorporating "genomic vulnerability" or "genomic offset" in the model (Fitzpatrick et al., 2007; Laman et al., 2018; Nielsen et al., 2021; Layton and Bradbury, 2022).

Here we proposed an entirely different approach on how SDMs could utilize genomic information. Briefly, the framework we propose consists of 1) utilizing high-throughput sequencing to detect candidate genomic variants whose frequency is associated with environmental conditions at spawning areas, 2) explore links between genotype and phenotype to infer functionality of the relationship and selective value and 3) utilize allelic frequencies as inputs to joint-SDMs, using allele selective values as prior on allele-environmental conditions relationships, taking into account for allele co-occurrence patterns and phylogeny to, 4) predict allele distribution, and *de facto* species distribution. Our focus on spawning areas is justified by their relevance to define the viability of fish populations as enhancers of reproductive success. It has been demonstrated that broadcast spawning (a reproduction strategy common in the marine environment) have evolved to optimize spawning timing and location to target optimal environmental conditions (Thorrold et al., 2001; Planque et al., 2007; Baltazar-Soares et al., 2018) . For the large majority of marine species, early life stages prior to first feeding check also occurs in the vicinity of spawning areas and thus we hold this framework in the premise that selective pressures at spawning are extremely high and thus require the evolution of adaptive responses. We will first cover how genetics is a tool utilized to understand evolutionary responses within natural populations and revise how high-throughput sequencing revolutionized the statistical power of genomic signatures. We will then approach concepts and methodologies behind SDMs and lastly, we will briefly illustrate ongoing efforts to conjugate both research areas and present our own suggestion to do it so.

High-throughput sequencing on the search of genetic signatures in natural populations

Screening genomes has become a routine task in evolutionary genetics in the recent decade, but that has not always been the case. The early 2000s experienced the first steps towards characterizing genome-wide diversity in no-model species with the discovery of short-tandem repeats (STRs) (Ellegren, 2004; Vieira et al., 2016). STRs loci are usually captured across random locations in the genome, each optimally exhibiting high levels of sequence length polymorphism (Jarne and Lagoda, 1996). Therefore, inferences from polymorphic-STRs analyses are based on multiple independent observations of the target genome´s evolution. In general terms, HTS relates to massive parallel production of DNA sequences. Its implementation expanded our capacity to collect multiple molecular markers across the genome (Mardis, 2008a). Nowadays, HTS techniques have evolved to sequence full genomes both in the form of short-reads DNA se, i.e., 250 base-pair sized reads commonly produced with Illumina, or long-reads DNA strings, where average read-lengths are as high as dozens of kilo base-pairs such as those commonly produced with PacBio or Oxford Nanopore technology (Mardis, 2008b; Hu et al., 2021). While the abovementioned strategies aim to sequence the whole of the nuclear DNA molecule, HTS have further expanded to sequence the full spectrum of DNA replication *via* the characterizing of transcriptomes and proteomes. Sequencing these DNA provides strong evidence of functionality, reinforcing the putative links between genotypic and phenotypic variation that can be observed at individual, population, or species level (Oomen and Hutchings, 2022).

The working-horse on any population genetic analysis are allelic frequencies. Thus, the baseline output after data processing (which involves cleaning of raw sequence data, curation, and variant calling) is a panel of genetic markers with allelic or haplotype frequencies distributed either by loci, individuals, or populations (R. Nielsen and Slatkin, 2013). Population genetics theory holds on principles of mendelian inheritance, evolution, and mathematics to devise how allelic frequencies vary across generations (R. Nielsen and Slatkin, 2013). It is against theoretical expectations of allelic frequency distributions that observed patterns of genetic variation are interpreted at population scales. One of the main goals of applying population genetics theory has been linking evolutionary and demographic processes, where the linear relationship between indices of genetic diversity such as heterozygosity or allelic richness and effective population sizes is well established (Reed and Frankham, 2003). Inferring population structure, effective population size and historical demography with genetic variation that has no impact on individual fitness became staple examples of applied population genetics. However, next generation sequencing effortlessly pushed the limits of population genetic inferences and greatly facilitated the search for molecular signatures of selection (Frankham, 2010).

Genome-wide screens usually target > 20 individuals from as many locations of the species' distribution as possible. The objective at this stage is to obtain representativity of the overall genetic diversity of the natural population. Here, the most common methodology to identify signatures of selection are environmental associations (EA). These frameworks identify signatures of selection by comparing allelic frequencies against gradients of environmental variables. At the light of evolutionary theory, significant correlations indicate selection for the presence of the candidate allele (in a population) in numbers higher than those expected by chance. Thus, candidate loci (under selection) will be those whose allelic frequencies vary consonant to the hypothesized environmental gradient (Forester et al., 2016). For instances, Benestan et al. (2016) candidate loci putatively involved in the response to thermal adaptation of lobsters by reporting an environmental association of allelic frequencies across a latitudinal gradient (Benestan et al., 2016). Currently, identification of loci under selection is an active and fertile research ground. There exist multiple methodologies, frameworks and even ideologies that are beyond the scope of this manuscript to discuss in detail (Günther and Coop, 2013; Whitlock and Lotterhos, 2015). Still, the major caveat of environmental correlations is arguably the absence of causality. The fact that a functional link cannot be established between genetic variation and a successful response to selective pressures renders environmental associations insufficient to argue with confidence about demographic impacts (Lotterhos and Whitlock, 2015).

Validating the adaptive potential estimated from molecular signatures can be achieved experimentally. Manipulating environmental settings to test fitness effects of candidate genomic variants in different environmental conditions should provide conclusive evidence of functionality (Lenz et al., 2013; Kaufmann et al., 2014). The challenge is seldom rearing is possible, which is an essential step to a) observe the functionality and b) validate true positives. If rearing is possible, then the method is partially quite established. It consists in exposing specimens to specific selective pressures measuring reproductive success and subsequent fitness of F1 and/or F2 (to mitigate the noise of natural genetic variation and family effects), and genotype individuals either at the end point of the experiment or through biologically established timestamps (Huang et al., 2016; Heckwolf et al., 2020). Genotyping is commonly performed *via* collecting and sequencing transcripts of individuals exposed to different conditions, where fold-differences or structural variants such as SNPs, copy number and/or splicing variants are indicators of selection and respective adaptive responses (Lenz et al., 2013; Heckwolf et al., 2020).

Understanding SDMs and respective predictive potential for fisheries

Species distribution models are valuable statistical tools providing management and conservation supports (Zurell et al., 2022). In addition to describe and explain relationships between species and environmental characteristics based on the niche-biotope duality, they are extensively used to map species' present-day distributions and to forecast changes over space and time (Zurell et al., 2020). SDMs traditionally follow a bottom-up approach using geo-referenced species records (presence-only, presence-absence [or presence-pseudo-absence], abundance) to estimate species niche in a correlative framework. This framework is based on the extraction of environmental characteristics from a stack of physical and biogeochemical climate model projections (past, present and (or) future) for each sampling point, including potentially the environmental characteristics where the species does not occur (absence or pseudo-absence) (Hollowed et al., 2013; Lima et al., 2022). The chosen algorithm then creates response functions exhibiting the environmental optima and explaining the relationship of the species' occurrence and the environment (Hollowed et al., 2013). Such relationship is returned as an index of habitat suitability which can be, under several assumptions (e.g., constant detection probability), considered as occurrence probability (Royle et al., 2012). Final SDM outputs are estimated relationships between species occurrence, environmental variables, and habitat suitability maps. SDMs can eventually be used to evaluate whether the spatial variability of a species' environmental optima will shift under different climate change scenarios over large scale spatial projections (Raybaud et al., 2017; Jghab et al., 2019; Schickele et al., 2020). From these maps it is also possible to define range expansion and contraction in the distribution of the species over time by calculating suitable areas (km^2) over the entire distribution range or in specific habitats (Lima et al., 2022). Recent developments have been done to increase SDM' ecological reliability and tackle many of its caveats (Zurell et al., 2016; Ovaskainen et al., 2017). For example, introducing Bayesian inference allows to cover the simplified vision of niche conservatism where individuals' occurrences remain fixed throughout space and time in a correlative approach. Bayesian inference has been used to improve the estimation of species-environmental relationships by integrating *a priori* knowledge on species niche dimensions. In Bayesian theory, the probability of an event to occur is mediated by information on past occurrences of the event. Technically, it translates into the use of priors with a respective distribution density (Gaussian, Poisson, etc. distribution) representing the event's past occurrence, to build a range of probable future events, or the so-called posterior density distribution (Bolstad and Curran, 2016). Posterior density distributions are constructed with Markov-Chain-Monte-Carlo (MCMC) samplers which, briefly, are a chained repetition of the model utilizing randomly picked priors from the proposed distribution (Van Ravenzwaaij et al., 2018). Bayesian inference coupled with MCMC simulations allow the efficient computation of thousands or millions of scenarios based on prior-posterior conjugations and are regarded as a major improvement in statistical computing frameworks in biology (Huelsenbeck et al., 2001; Yau and Campbell, 2019). Bayesian inferences have been shown to improve the explicative and predictive powers of SDMs (Vermeiren et al., 2020). Another aspect traditionally limiting the ecological reliability of SDMs it their reliance on abiotic variables, when it is factual that biotic variables, such as species interactions, also play a critical role to shape species niche (Zurell et al., 2016). Joint-SDMs have been built to tackle this issue, by considering multi-species co-occurrences inside the same model, using functional traits or phylogenetic relatedness to investigate the dependance pattern between species (Ovaskainen et al., 2017). Estimating multi-species niche, while combining different sources of data improve prediction accuracy, and that is why exploring possibilities to capitalize the amount of information produced by HTS is flourishing research area (Ovaskainen et al., 2017; Peel et al., 2019; Vermeiren et al., 2020; Andrello et al., 2022).

Riding the wave of evolutionary-based SDMs: Current integrations of genetics and SDMs

Adaptations accounting from genomic interaction with environmental conditions are usually ignored but might be important to explain and predict species distribution (DeMarche et al., 2019). SDMs usually assume a homogeneity of the genomic composition inside the distribution of a focal species. Relationships between genomic diversity and environmental factors have been used to map the vulnerability of the species at intra-specific level (Ruegg et al., 2018). This can help understanding where individuals of a same species will be more impacted by climate change or will need faster adaptive capacity/facilitation/resilience. To date, most studies have considered that evaluating different SDMs of genetically defined populations of a given species would be a sufficient proxy to integrate adaptive potential (Ikeda et al., 2017; Chardon et al., 2020). These studies used predictions of species' niche space by considering that genetically distinct populations would respond differently to present-day climate, and thus future climates, to then test for similarity in the climatic niche of the groups (DeMarche et al., 2019). Alternatively, researchers now have been focusing on evolutionary algorithms to combine multiple information in one single learning cycle (Gobeyn et al., 2019). Those frameworks consider that the

geographical range of a species is defined by its ability to track favourable environmental conditions depending on its physiology, evolutionary adaptation and the inter and intraspecific biotic interactions (Thuiller et al., 2013). SDMs also span the divide between correlative and mechanistic models using eco-evolutionary forecasting frameworks (Bush et al., 2016; Cotto et al., 2017). The framework combines niche-based projections and individual-based, genetically and spatially explicit stochastic simulations (Cotto et al., 2017). Modelling frameworks such as *AdaptR* are useful to predict the distribution of species through time steps known as generations (Bush et al., 2016). This hybrid approach allows the incorporation of adaptive capacity as phenotypic plasticity, evolutionary adaptation and adaptive capacity through physiological limits into the same framework (Bush et al., 2016). 2016). In eco-evolutionary dynamic models, local populations on a grid cell are assumed to adapt to local environmental conditions, whilst accounting for stochastic processes of individual life cycle, such as birth, death and migration (i.e. age-structured demographic model). Here, static niche models are used to predict the current distribution of a species based on environmental conditions. Then, the predicted distribution is used to initialize simulated changes in the distribution accounting to adaptation as driven by scenarios of climatic change (Cotto et al., 2017).

Integrating genomic at the onset of SDMs: Predicting stock distribution as a function of standing adaptive variation

Our proposal places adaptive potential – from genomics – at the basis of an SDM approach (Figure 1). Conceptually, we propose to adjust the Bayesian J-SDM framework presented in the previous section to utilize information at genotype level, or specifically, allelic frequencies. In essence, we are literally transfiguring the concept of "species" into "alleles". This means that instead of a pool of species (or individuals of a given species), we would have a pool of allelic frequencies from the candidate loci obtained either *via* environmental associations or transcriptomes of experimentally exposed individuals. The biological input variable would thus be allele presence-absence or frequency between populations of either specimens collected across the species distribution range or experimental groups. Naturally, the concept can be extended to a pool of candidate loci (and respective alleles), as the identification of several or dozens of candidate loci either through environmental associations or experimental work is nevertheless common (Hoban et al., 2016). The main goal is to consider adaptive genomic information as valid predictors of shifts in species

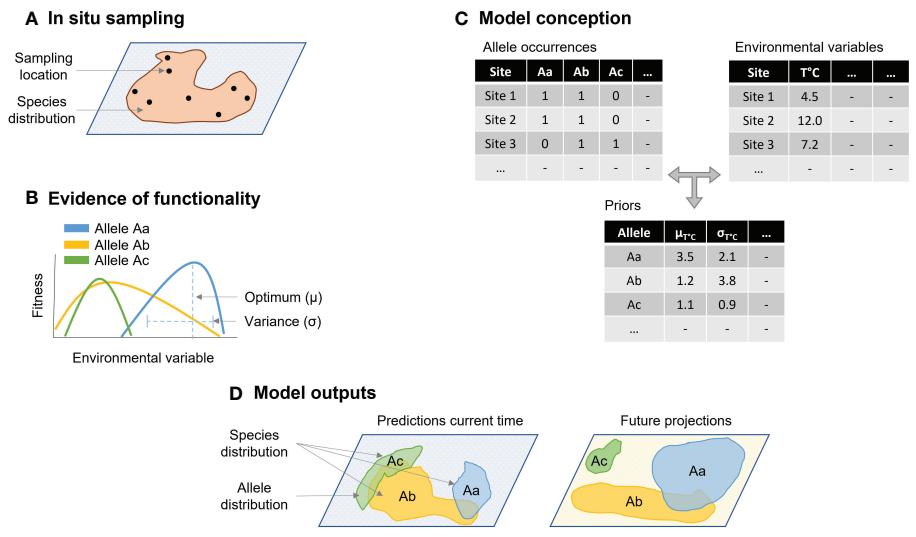


FIGURE 1
Conceptual eco-evolutionary framework adapted from a Bayesian joint-species distribution model (J-SDM) to estimate species distribution from allele occurrence patterns and fitness. The framework is divided in four steps, **(A)** fish sampling at spawning locations representative of the species distribution, **(B)** identification of the candidate genes and if possible, empirical assessment of allele reaction norm for a focal environmental variable, **(C)** model conception based on allele occurrence and environmental conditions at sampling sites, using optimum and variance of the reaction norm to inform prior distributions and phylogenetic trees to identify similarities between sampled locations genetic composition to characterize allele cooccurrence patterns, and **(D)** explain allele relationships with environmental variables, predict allele distributions and eventually forecast species spawning distribution as the stacked allele distributions changes in response to environmental variable changes.

distribution, implying that to be a canonical response to change of environmental conditions. Here we can make use of the more informative power of experimental approaches by utilizing reaction norms - or the range of phenotypic expression of a specific genotype - as priors of the Bayesian-*JSDM* framework we are proposing. Because reactions norms relate to the phenotypic response of the measured trait of which the genotype would be responsible for, these transcriptome-derived priors are key to model posterior distribution of adaptive genetic variation. When candidate loci are identified with environmental associations, then the priors for allele association with environmental variable can be the correlation statistics reported by currently utilized environmental association software. To name some, Bayenv2 (Günther and Coop, 2013) reports Bayes factors and Spearman's *rho* statistic for each allele-environment variable association, and the latent factor mixed models (lfmm) incorporated in R package *LEA* (Frichot and François, 2015) identifies candidate loci considering a z-score distribution of correlation values also for each allele-environment variable association. This candidate-detection software also employs Bayesian statistics and thus correlation statistics outputs of several runs consecutively identifying the same candidate loci can be utilized as prior distribution for allelic response. Certainly, these priors are far less indicative of an allelic response than reaction norms obtained from experimental set-ups. Nevertheless, having a prior on allele-abiotic predictor relationships can reduce the risk for spurious correlation in the SDM framework and become useful in case of sampling bias.

Lastly, the abiotic input variables would be those relevant to explain species niche, such as temperature, salinity, oxygen concentrations, etc. Priors for these variables are key - because environmental conditions are highly dynamic in the marine environment - and can be obtained from several databases such as BioOracle (Assis et al., 2018). At the end, the output variable is the occurrence probability, or frequency, of candidate loci's alleles in response to abiotic predictors. Species distribution is eventually obtained by stacking together genotype distributions, returning the potential full habitat suitability map at species level as a function of the adaptive potential if its populations (Figure 1). Adding phylogenetic trees to the J-SDM like framework can help identifying similarities between sampled locations genetic composition and characterize allele co-occurrence patterns.

Applicability to fisheries management

The combined framework we propose here explores the adaptive potential of fish stocks while projecting it into future species distributions. As such, the framework not only has the potential to inform about the suitability of areas beyond the currently known viable distribution but also that species might adapt to environmental shifts within current distribution boundaries. In this context, mobility (to follow environmental optima) and adaptation (to remain in the same environment) are not mutually exclusive. For fisheries management, the realization of a dynamic fish stock both in terms of mobility and adaptive potential would necessarily translate in the following considerations:

1 – Local or regional extinction of stocks – often assumed when only mobility is considered – might not necessarily occur as a function of environmental shifts. This means that management bodies should not "give up" on the protection of specimens that are already moving to outside their respective areas of governance. If anything, it becomes even more important to protect mobile stocks because selective pressure posed by environmental shifts acts synergistically with anthropogenic pressure of fishing (Baltazar-Soares et al., 2021a; Hočevar and Kuparinen, 2021). In practice, identifying adaptive potential of stocks potentiates the development of measures to maintain local stocks levels despite species' migratory capacity. Notably, mechanisms to implement measures stemming from the above premise already exist and are commonly applied upon stock crashes: fishing restrictions and imposition of quotas. The context however would differ, because in our particular example fish stocks are not necessarily in a vulnerable state in present conditions.

2 - New areas might be deemed as suitable on top of those that are currently being proposed by traditional SDMs approaches. The implications for management here are multiple and range from an increase of monitoring activities to detect the species presence at those sites, to the design of measures to preserve/protect of future spawning areas, and to a deeper investigation of biotic interactions that might result from those expansions. Perhaps in a first stage, biomonitoring activities appears to be immediately applicable measure in scientific fisheries surveys. Identifying this type of climate migrants and where/when they start to occur would be an indicator of change and evidence of shifting climatic conditions. Within a management framework, it would also likely lead to access the impact of these newcomers to nonnative to the sustainability of local fisheries.

We acknowledge that it might be challenging to accept the above considerations in the absence of the hindsight the future inevitably offers. Still, tackling climate-related issues requires enhancing of our predictive capacity and arguably no other framework would do it better besides one that incorporates ecological and evolutionary components. Still, it is important to remember that a framework like ours holds on assumptions that are in themselves speculations in the respective area of research. The most notable are those associated with the assignment of genetic variation to selective advantage and with habitat suitability disregarding biotic interactions in new habitats. It is obvious that

our framework will certainly benefit from knowledge increments on genotype-phenotype-environment interactions in providing conclusive evidence for genomic signatures of selection, and in SDMs considering biotic interactions (HilleRisLambers et al., 2013; Pigot and Tobias, 2013). For example, expanding the spatial scope to other areas critical to marine organism's life cycle such as nurseries or feeding grounds will certainly increment the resolution of analytical strategies focused on evolutionary responses. Likewise, considering genetic other architectures more complex than the simplistic genotype x phenotype view we illustrated here would also offer a more realistic picture of evolutionary responses. Until complexity arises, frameworks such as the one we propose here remain highly exploratory but nevertheless a stepping stone to the flourishing field of advanced fisheries research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

All authors contributed to the article and approved the submitted version.

Funding

The authors acknowledge the Portuguese Foundation for Science and Technology (FCT) through the strategic project MARE-UIDB/04292/2020 granted to MARE (Marine and Environmental Sciences Centre) and through the project LA/P/0069/2020 granted to the Associate Laboratory ARNET. AL acknowledges the European Union's (EU's) Horizon 2020 (H2020) research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101038057. EG acknowledges the Turku Collegium for Science, Medicine and Technology. MBS acknowledges the Academy of Finland funding decision 321417.

Acknowledgments

We also thank members of the ICES-PICES Working Group on Small Pelagic Fish (WGSPF) for helpful discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11 (10), 697–709. doi: 10.1038/nrg2844

Andrello, M., D'Aloia, C., Dalongeville, A., Escalante, M. A., Guerrero, J., Perrier, C., et al. (2022). Evolving spatial conservation prioritization with intraspecific genetic data. *Trends Ecol. Evol* 37 (6), 5537–564. doi: 10.1016/j.tree.2022.03.003

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., and De Clerck, O. (2018). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecol. biogeogr.* 27 (3), 277–284. doi: 10.1111/geb.12693

Baltazar-Soares, M., Hans-Harald, H., and Eizaguirre, C. (2018). Integrating population genomics and biophysical models towards evolutionary-based fisheries management, *ICES J. Mar. Sci.* (2018) (Wiley Online Library), 75.4 1245–1257.

Baltazar-Soares, M., Brans, K. I., and Eizaguirre, C. (2021a). *Human-induced evolution: Signatures, processes and mechanisms underneath anthropogenic footprints on natural systems*. (Wiley Online Library) 14, 2335–2341.

Baltazar-Soares, M., de Araújo Lima, A. R., and Silva, G. (2021b). Targeted sequencing of mitochondrial genes reveals signatures of molecular adaptation in a nearly panmictic small pelagic fish species. *Genes* 12 (1), 91. doi: 10.3390/genes12010091

Benestan, L. M., Ferchaud, A. L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J., Baums, I. B., et al. (2016). Conservation genomics of natural and managed populations: Building a conceptual and practical framework. *Mol. Ecol.* 25 (13), 2967–2977. doi: 10.1111/mec.13647

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32 (9), 665–680. doi: 10.1016/j.tree.2017.06.010

Blasco, G. D., Ferraro, D. M., Cottrell, R. S., Halpern, B. S., and Froehlich, H. E. (2020). Substantial gaps in the current fisheries data landscape. *Front. Mar. Sci.* 7, 612831. doi: 10.3389/fmars.2020.612831

Bolstad, W. M., and Curran, J. M. (2016). *Introduction to Bayesian statistics* (Hoboken, New Jersey: John Wiley & Sons, Inc.,).

Brander, K. (2010). Impacts of climate change on fisheries. *J. Mar. Syst.* 79 (3-4), 389–402. doi: 10.1016/j.jmarsys.2008.12.015

Bush, A., Mokany, K., Catullo, R., Hoffmann, A., Kellermann, V., Sgrò, C., et al. (2016). Incorporating evolutionary adaptation in species distribution modelling

reduces projected vulnerability to climate change. *Ecol. Lett.* 19 (12), 1468–1478. doi: 10.1111/ele.12696

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., and Keller, S. R. (2020). Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annu. Rev. Ecol. Evol System.* 51 (1), 245–269. doi: 10.1146/annurev-ecolsys-020720-042553

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., et al. (2012). Biodiversity loss and its impact on humanity. *Nature* 486 (7401), 59–67. doi: 10.1038/nature11148

Casey, J., Jardim, E., and Martinsohn, J. T. (2016). The role of genetics in fisheries management under the EU common fisheries policy. *J. Fish. Biol.* 89 (6), 2755–2767. doi: 10.1111/jfb.13151

Chardon, N. I., Pironon, S., Peterson, M. L., and Doak, D. F. (2020). Incorporating intraspecific variation into species distribution models improves distribution predictions, but cannot predict species traits for a wide-spread plant species. *Ecography* 43 (1), 60–74. doi: 10.1111/ecog.04630

Cheung, W. W. L., Lam, V. W. Y., Sarmiento, J. L., Kearney, K., Watson, R., and Pauly, D. (2009). Projecting global marine biodiversity impacts under climate change scenarios. *Fish Fish.* 10 (3), 235–251. doi: 10.1111/j.1467-2979.2008.00315.x

Cotto, O., Wessely, J., Georges, D., Klonner, G., Schmid, M., Dullinger, S., et al. (2017). A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nat. Commun.* 8 (1), 1–9. doi: 10.1038/ncomms15399

DeMarche, M. L., Doak, D. F., and Morris, W. F. (2019). Incorporating local adaptation into forecasts of species' distribution and abundance under climate change. *Global Change Biol.* 25 (3), 775–793. doi: 10.1111/gcb.14562

Eizaguirre, C., and Baltazar-Soares, M. (2014). Evolutionary conservation–evaluating the adaptive potential of species. *Evol. Appl.* 7 (9), 963–967. doi: 10.1111/eva.12227

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5 (6), 435–445. doi: 10.1038/nrg1348

Fitzpatrick, M. C., Weltzin, J. F., Sanders, N. J., and Dunn, R. R. (2007). The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecol. biogeogr.* 16 (1), 24–33. doi: 10.1111/j.1466-8238.2006.00258.x

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., and Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol. Ecol.* 25 (1), 104–120. doi: 10.1111/mec.13476

Frankham, R. (2010). Where are we in conservation genetics and where do we need to go? *Conserv. Genet.* 11 (2), 661–663. doi: 10.1007/s10592-009-0010-2

Frichot, E., and François, O. (2015). LEA: An r package for landscape and ecological association studies. *Methods Ecol. Evol.* 6 (8), 925–929. doi: 10.1111/2041-210X.12382

Frölicher, T. L., and Laufkötter, C. (2018). Emerging risks from marine heat waves. *Nat. Commun.* 9 (1), 1–4. doi: 10.1038/s41467-018-03163-6

Gobeyn, S., Mouton, A. M., Cord, A. F., Kaim, A., Volk, M., and Goethals, P. L. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecol. Model.* 392, 179–195. doi: 10.1016/j.ecolmodel.2018.11.013

Guisan, A., and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8 (9), 993–1009. doi: 10.1111/j.1461-0248.2005.00792.x

Günther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics* 195 (1), 205–220. doi: 10.1534/genetics.113.152462

Hauser, L., and Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish. Fish.* 9 (4), 333–362. doi: 10.1111/j.1467-2979.2008.00299.x

Heckwolf, M. J., Meyer, B. S., Häsler, R., Höppner, M. P., Eizaguirre, C., and Reusch, T. B. (2020). Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. *Sci. Adv.* 6 (12), eaaz1138. doi: 10.1126/sciadv.aaz1138

HilleRisLambers, J., Harsch, M. A., Ettinger, A. K., Ford, K. R., and Theobald, E. J. (2013). How will biotic interactions influence climate change–induced range shifts? *Ann. New York Acad. Sci.* 1297 (1), 112–125. doi: 10.1111/nyas.12182

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., et al. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am. Nat.* 188 (4), 379–397. doi: 10.1086/688018

Hočevar, S., and Kuparinen, A. (2021). Marine food web perspective to fisheries-induced evolution. *Evol. Appl.* 14 (10), 2378–2391. doi: 10.1111/eva.13259

Hollowed, A. B., Barange, M., Beamish, R. J., Brander, K., Cochrane, K., Drinkwater, K., et al. (2013). Projected impacts of climate change on marine fish and fisheries. *ICES J. Mar. Sci.* 70 (5), 1023–1037. doi: 10.1093/icesjms/fst081

Huang, Y., Chain, F. J. J., Panchal, M., Eizaguirre, C., Kalbe, M., Lenz, T. L., et al. (2016). Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. *Mol. Ecol.* 25 (4), 943–958. doi: 10.1111/mec.13520

Hu, T., Chitnis, N., Monos, D., and Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Hum. Immunol.* 82 (11), 801–811. doi: 10.1016/j.humimm.2021.02.012

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001). Bayesian Inference of phylogeny and its impact on evolutionary biology. *Science* 294 (5550), 2310–2314 doi: 10.1126/science.1065889.

Ikeda, D. H., Max, T. L., Allan, G. J., Lau, M. K., Shuster, S. M., and Whitham, T. G. (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biol.* 23 (1), 164–176. doi: 10.1111/gcb.13470

Jarne, P., and Lagoda, P. J. (1996). Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* 11 (10), 424–429. doi: 10.1016/0169-5347(96)10049-5

Jghab, A., Vargas-Yañez, M., Reul, A., Garcia-Martínez, M., Hidalgo, M., Moya, F., et al. (2019). The influence of environmental factors and hydrodynamics on sardine (Sardina pilchardus, walbaum 1792) abundance in the southern alboran Sea. *J. Mar. Syst.* 191, 51–63. doi: 10.1016/j.jmarsys.2018.12.002

Kaufmann, J., Lenz, T. L., Milinski, M., and Eizaguirre, C. (2014). Experimental parasite infection reveals costs and benefits of paternal effects. *Ecol. Lett.* 17 (11), 1409–1417. doi: 10.1111/ele.12344

Lam, V. W., Allison, E. H., Bell, J. D., Blythe, J., Cheung, W. W., Frölicher, T. L., et al. (2020). Climate change, tropical fisheries and prospects for sustainable development. *Nat. Rev. Earth Environ.* 1 (9), 440–454. doi: 10.1038/s43017-020-0071-9

Laman, E. A., Rooper, C. N., Rooney, S. C., Turner, K. A., Cooper, D. W., and Zimmermann, M. (2017). *Model-based essential fish habitat definitions for Bering Sea groundfish species*.

Laman, E. A., Rooper, C. N., Turner, K., Rooney, S., Cooper, D. W., and Zimmermann, M. (2018). Using species distribution models to describe essential fish habitat in Alaska. *Can. J. Fish. Aquat. Sci.* 75 (8), 1230–1255. doi: 10.1139/cjfas-2017-0181

Layton, K. K., and Bradbury, I. R. (2022). Harnessing the power of multi-omics data for predicting climate change response. *J. Anim. Ecol.* 91 (6), 1064–1072. doi: 10.1111/1365-2656.13619

Lenz, T. L., Eizaguirre, C., Rotter, B., Kalbe, M., and Milinski, M. (2013). Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Mol. Ecol.* 22 (3), 774–786. doi: 10.1111/j.1365-294X.2012.05756.x

Lima, A. R., Baltazar-Soares, M., Garrido, S., Riveiro, I., Carrera, P., Piecho-Santos, A. M., et al. (2022). Forecasting shifts in habitat suitability across the distribution range of a temperate small pelagic fish under different scenarios of climate change. *Sci. Total Environ.* 804, 150167. doi: 10.1016/j.scitotenv.2021.150167

Lima, A. R., Garrido, S., Riveiro, I., Rodrigues, D., Borges, R., Angélico, M. M., et al. (2022). Seasonal approach to forecast the suitability of spawning habitats of a temperate small pelagic fish under a high-emission climate change scenario. *Front. Mar. Sci.*, 1523. doi: 10.3389/fmars.2022.956654

Lohbeck, K. T., Riebesell, U., and Reusch, T. B. H. (2012). Adaptive evolution of a key phytoplankton species to ocean acidification. *Nat. Geosci.* 5 (5), 346–351. doi: 10.1038/ngeo1441

Lotterhos, K. E., and Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24 (5), 1031–1046. doi: 10.1111/mec.13100

Mardis, E. R. (2008a). The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24 (3), 133–141. doi: 10.1016/j.tig.2007.12.007

Mardis, E. R. (2008b). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9 (1), 387–402. doi: 10.1146/annurev.genom.9.081307.164359

Nielsen, E. S., Henriques, R., Beger, M., and von der Heyden, S. (2021). Distinct interspecific and intraspecific vulnerability of coastal species to global change. *Global Change Biol.* 27 (15), 3415–3431. doi: 10.1111/gcb.15651

Nielsen, R., and Slatkin, M. (2013). *An introduction to population genetics* (Sunderland, MA: Sinauer Associates).

Oomen, R. A., and Hutchings, J. A. (2022). Genomic reaction norms inform predictions of plastic and adaptive responses to climate change. *J. Anim. Ecol.* 91 (6), 1073–1087. doi: 10.1111/1365-2656.13707

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., et al. (2017). How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecol. Lett.* 20 (5), 561–576. doi: 10.1111/ele.12757

Peel, S. L., Hill, N. A., Foster, S. D., Wotherspoon, S. J., Ghiglione, C., and Schiaparelli, S. (2019). Reliable species distributions are obtainable with sparse,

patchy and biased data by leveraging over species and data types. *Methods Ecol. Evol.* 10 (7), 1002–1014. doi: 10.1111/2041-210X.13196

Pigot, A. L., and Tobias, J. A. (2013). Species interactions constrain geographic range expansion over evolutionary time. *Ecol. Lett.* 16 (3), 330–338. doi: 10.1111/ele.12043

Planque, B., Bellier, E., and Lazure, P. (2007). Modelling potential spawning habitat of sardine (Sardina pilchardus) and anchovy (Engraulis encrasicolus) in the bay of Biscay. *Fish. Oceanogr.* 16 (1), 16–30. doi: 10.1111/j.1365-2419.2006.00411.x

Porfirio, L. L., Harris, R. M., Lefroy, E. C., Hugh, S., Gould, S. F., Lee, G., et al. (2014). Improving the use of species distribution models in conservation planning and management under climate change. *PloS One* 9 (11), e113749. doi: 10.1371/journal.pone.0113749

Raybaud, V., Bacha, M., Amara, R., and Beaugrand, G. (2017). Forecasting climate-driven changes in the geographical range of the European anchovy (Engraulis encrasicolus). *ICES J. Mar. Sci.* 74 (5), 1288–1299. doi: 10.1093/icesjms/fsx003

Reed, D. H., and Frankham, R. (2003). Correlation between fitness and genetic diversity correlación entre adaptabilidad y diversidad genética. *Conserv. Biol.* 17 (1), 230–237. doi: 10.1046/j.1523-1739.2003.01236.x

Royle, J. A., Chandler, R. B., Yackulic, C., and Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods Ecol. Evol.* 3 (3), 545–554. doi: 10.1111/j.2041-210X.2011.00182.x

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., et al. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecol. Lett.* 21 (7), 1085–1096. doi: 10.1111/ele.12977

Schickele, A., Goberville, E., Leroy, B., Beaugrand, G., Hattab, T., Francour, P., et al (2020). European Small pelagic fish distribution under global change scenarios. *Fish. Fish.* 22 (1), 212–225. doi: 10.1111/faf.12515

Stead, S. M. (2018). Rethinking marine resource governance for the united nations sustainable development goals. *Curr. Opin. Environ. Sustainability* 34, 54–61. doi: 10.1016/j.cosust.2018.12.001

Therkildsen, N. O., Hemmer-Hansen, J., Hedeholm, R. B., Wisz, M. S., Pampoulie, C., Meldrup, D., et al. (2013). Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod gadus morhua. *Evol. Appl.* 6 (4), 690–705. doi: 10.1111/eva.12055

Thorrold, Simon R., et al. (2001). Natal homing in a marine fish metapopulation. *Science* 291.5502, 297–299.

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K., et al. (2013). A road map for integrating eco-evolutionary processes into biodiversity models. *Ecol. Lett.* 16, 94–105. doi: 10.1111/ele.12104

Van Ravenzwaaij, D., Cassey, P., and Brown, S. D. (2018). A simple introduction to Markov chain Monte–Carlo sampling. *Psychonomic. Bull. Rev.* 25 (1), 143–154. doi: 10.3758/s13423-016-1015-8

Vermeiren, P., Reichert, P., and Schuwirth, N. (2020). Integrating uncertain prior knowledge regarding ecological preferences into multi-species distribution models: Effects of model complexity on predictive performance. *Ecol. Model.* 420, 108956. doi: 10.1016/j.ecolmodel.2020.108956

Verspoor, E., Stradmeyer, L., and Nielsen, J. L. (2008). *The Atlantic salmon: genetics, conservation and management* (Oxford: : Blackwell Pub).

Vieira, M. L. C., Santini, L., Diniz, A. L., and Munhoz, C. (2016). Microsatellite markers: what they mean and why they are so useful. *Genet. Mol. Biol.* 39, 312–328. doi: 10.1590/1678-4685-GMB-2016-0027

Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., and Brotons, L. (2017). Integrating species distribution modelling into decision-making to inform conservation actions. *Biodiver. Conserv.* 26 (2), 251–271. doi: 10.1007/s10531-016-1243-2

Whitlock, M. C., and Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of f ST. *Am. Nat.* 186 (S1), S24–S36. doi: 10.1086/682949

Yau, C., and Campbell, K. (2019). Bayesian Statistical learning for big data biology. *Biophys. Rev.* 11 (1), 95–102. doi: 10.1007/s12551-019-00499-1

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., et al. (2020). A standard protocol for reporting species distribution models. *Ecography* 43 (9), 1261–1277. doi: 10.1111/ecog.04960

Zurell, D., König, C., Malchow, A. K., Kapitza, S., Bocedi, G., Travis, J., et al. (2022). Spatially explicit models for decision-making in animal conservation and restoration. *Ecography* 2022 (4). doi: 10.1111/ecog.05787

Zurell, D., Thuiller, W., Pagel, J., Cabral, J. S., Münkemüller, T., Gravel, D., et al. (2016). Benchmarking novel approaches for modelling species range dynamics. *Global Change Biol.* 22 (8), 2651–2664. doi: 10.1111/gcb.13251

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 30 January 2023
DOI 10.3389/fmars.2023.1062151

OPEN ACCESS

EDITED BY
Rodrigo Vidal,
University of Santiago, Chile

REVIEWED BY
Holly Shiels,
The University of Manchester,
United Kingdom
Frédéric Silvestre,
University of Namur, Belgium

*CORRESPONDENCE
Francesc Piferrer
piferrer@icm.csic.es

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 05 October 2022
ACCEPTED 13 January 2023
PUBLISHED 30 January 2023

CITATION
Piferrer F and Anastasiadi D (2023) Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology. *Front. Mar. Sci.* 10:1062151. doi: 10.3389/fmars.2023.1062151

Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology

Francesc Piferrer[1]* and Dafni Anastasiadi[2]

[1]Institut de Ciències del Mar, Spanish National Research Council (CSIC), Barcelona, Spain, [2]Seafood Technologies, The New Zealand Institute for Plant and Food Research, Nelson, New Zealand

The distribution of age classes is a key demographic parameter of populations and thus proper age estimation is crucial for fisheries management and for conservation biology. Age estimation in fishes has traditionally relied on the analysis of growth marks in hard structures such as otoliths. However, besides being lethal this method is time-consuming, can have low accuracy in some species and cannot be applied in others. Thus, there is a need for the development of new methods. DNA methylation is an epigenetic modification consisting in the addition of a methyl group in cytosine-guanine loci. Aging is associated with changes in DNA methylation. Among a background of global and weak genome hypomethylation, there are some loci in which age-associated DNA methylation changes are of a "clock-like" nature and thus predictable. Chronological age estimators built from DNA methylation are termed 'epigenetic clocks'. Epigenetic clocks have been developed in the last ten years for many species, notably vertebrates, including already several fish species. Here, we review the piscine epigenetic clocks built so far and outline the major considerations to be taken into account for the development of new epigenetic clocks for additional species, which include the number of samples to be collected and tissues to be targeted. The steps on how to develop such a clock and the techniques available to do so are also discussed. Next, we focus on the features of epigenetic clocks as measuring devices, considering aspects such as accuracy, precision and reproducibility. Finally, we discuss the possibility of developing a multi-species piscine epigenetic clock and how processing automation can greatly reduce the cost per sample. One important knowledge gap is to determine how environmental changes, especially temperature and food availability, may affect the tick rate of piscine epigenetic clocks. Improved age prediction through the use of piscine epigenetic clocks can contribute to better fisheries management practices in a context of overexploited fish stocks worldwide, and in the estimation of age classes in endangered species.

KEYWORDS

age estimation, age predictor, epigenetic clocks, fisheries management, conservation biology, DNA methylation, machine learning, penalized regressions

1 Introduction

1.1 Rationale for age estimation in fish

Age is one of the most influential of biological variables. Age-class distribution is considered a key population parameter, with important influences on biomass distribution, intra-specific interactions and reproductive potential, among others (Campana, 2001). Effective and accurate age estimation in fishes is crucial to know the demographic structure of populations, which needs to be assessed in a representative, informative and robust way. This assessment is of paramount importance in species of commercial interest for fisheries management (Pardo et al., 2013). Also, it is valuable for the estimation of age classes in populations of endangered species in order to design proper conservation strategies (Heydenrych et al., 2021). More specifically, and according to New Zealand's National Institute of Water and Atmospheric Research (NIWA), age estimation for the monitoring and management of fishery resources enables the following (Sutton et al., 2023):

1) Calculation of growth rates. By ageing many fish over the whole size-range in a population, it is possible to generate a growth curve (i.e., average length at each age) for that population. This method allows to assess how productive species are and to estimate the variability in their growth rates over time.
2) Calculation of mortality rates. By ageing a random sample of individuals from a population, and evaluating how the frequency of fish in each year class (all fish born in the same year) diminishes with age, it is possible to estimate: a) the rate of natural mortality given the sample size is very large, or, b) in a heavily fished population, the combined rate of mortality due to natural causes and fishing.
3) Assessment of population age-class structures. The aged random sample described above also estimates the distribution of age classes in a population if the sample is from a particular fishery, allowing to infer what group of age classes are particularly targeted by fisheries.
4) Estimation of annual spawning success. After accounting for mortality over time, an aged random sample can show the relative strengths of individual year classes and illustrate how successful spawning was in each year. The more estimates that can be derived for the strength of a particular year class (e.g., from consistent sampling in consecutive years), the greater the confidence in the estimate for that year class. These estimates of year class strength can explain variations in total stock abundance and are used to predict future recruitment to the fishery.
5) Investigations of stock structure. In the same geographic region, there may be more than one stock of a given species, for example, if distinct populations of the same species with little or no genetic or physical mixing are present. Differences in growth rates or population age structures between areas can be suggestive of multiple stocks (Sutton et al., 2023).

To illustrate the importance of age estimation in fish, it was estimated that about two million fish were aged worldwide in 1999 (Campana and Thorrold, 2001). Despite some efforts we have not been able to obtain a more current figure. However, Campana (2001) considered that fish aging efforts dwarf those routinely applied to non-fish species, and highlighted the importance attributed to age-structured information in fisheries science.

1.2 Shortcomings of current methods for age estimation and limitations

Age estimation in fishes has traditionally relied on the analysis of growth marks in hard structures such as vertebra, scales and otoliths (fish ear stones; Figure 1), in an analogous manner that rings are used to estimate the age of trees (Campana, 2001). Traditional methods require well-trained personnel, are time-consuming, often lethal depending upon choice of material (vertebra and otoliths), and can have low accuracy or are impossible to apply in some species. Otolith-based methods can nowadays be aided by deep neural network procedures of machine learning. However, despite recent technical improvements, traditional methods still have many shortcomings, including the fact that knowledge about the internal structure of the otolith is needed to improve the results for the youngest age groups, and that the contour shape and size attributes are sometimes not good enough for the younger age groups (Ordoñez et al., 2020).

Error in fish age estimation has two major sources: 1) Error associated with what is being measured, i.e., the structure being examined. Thus, some of the bony structures may not exhibit a complete growth sequence throughout lifetime, and this type of error can result in under- or over-age estimation. 2) Error due to sample preparation and interpretation of growth marks in calcified structures, which can vary markedly among readers and laboratories (Campana, 2001). This type of error can be biased or random. Together, both types of errors can result in estimates that can differ up to a factor of three among investigators (Campana, 2001).

There are some fish species of worldwide economic importance for which age estimation based on these traditional methods is very difficult or impossible. For example, the monkfish (*Lophius americanus*), supports an important commercial fishery in the Northeastern United States. Despite healthy stock status, annual catch limits are typically low. This is in part due to uncertainty in the stock assessment since age estimation in monkfish traditionally suffers from lack of effective validation due to irregular growth of the otolith. In consequence, this hampers its use as a reliable age structure assessment method (Bank, 2016). As a second option, annual growth rings are counted on the vertebrae and are assumed to follow a seasonal pattern but, again, they failed to accurately determine the age of monkfish (Bank, 2016). A similar situation occurs in hake, *Merluccius merluccius* (Morales-Nin et al., 1998). Thus, there is a need for the development of new, widely applicable methods.

Taking into account all the above and the rapid development of molecular methods to interrogate different aspects of the living organisms, exploring the possibility of applying some of these methods for age estimation in fish is warranted.

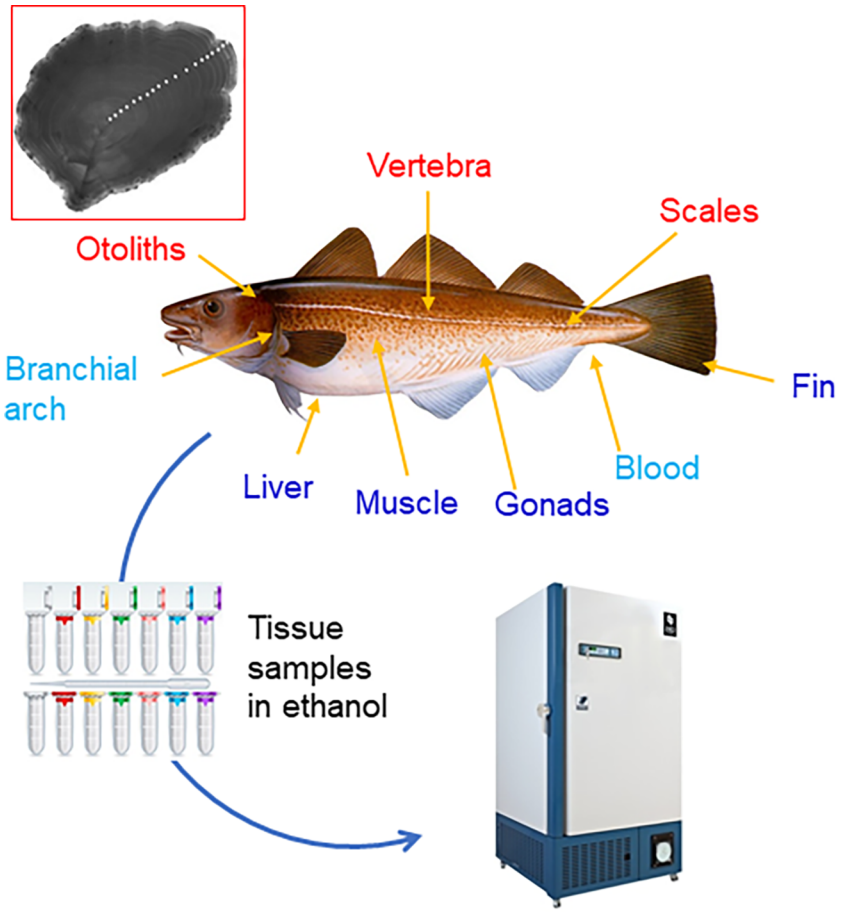


FIGURE 1
Hard structures traditionally used to estimate age in fish (names above the fish figure and in red), particularly otoliths (insert); and tissues either already used (names below the fish figure in dark blue) or with potential use (light blue) to develop epigenetic clocks. Fin clips are the preferred choice because they are easy to access and collect, and represent a non-invasive and non-lethal technique. Insert: the otolith image corresponds to yellowfin flounder, aged 19 years. The annual increments are marked by white dots. Source: Wikimedia Commons.

2 Epigenetic clocks

2.1 The definition and main mechanisms of epigenetics

Epigenetics can be defined as "the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to DNA sequence." (Deans and Maggert, 2015). The term "heritable" has two different implications because these changes can be passed from one cell to their daughter cells during mitotic cell division, but also during the formation of gametes through meiotic cell division, and thus can be transmitted from parents to offspring. The main epigenetic mechanisms for the regulation of gene expression are: DNA methylation, modification of histones and histone variants, and the presence of non-coding RNAs (ncRNAs) (Brock and Fisher, 2005). The epigenetic clocks built to date are based on DNA methylation analysis. Therefore, from now on this review will discuss only DNA methylation with more detail, while the other epigenetic mechanisms will not be further considered.

2.2 DNA methylation

DNA methylation is a chemical modification of the DNA chain itself, whereby the 5' carbon of cytosine is modified by the addition of a methyl (CH_3) group ("methylated"), becoming 5'-methylcytosine (5mC). In vertebrates, DNA methylation essentially occurs in a CpG context, i.e., when a cytosine is followed by a guanine.

The enzymes responsible for DNA methylation are called DNA methyltransferases (DNMTs) (Goll and Bestor, 2005). DNMTs are classified according to their target. DNMT3 methylates previously unmethylated CpGs, and is responsible for *de novo* DNA methylation. DNMT1 methylates the unmethylated opposing pair of a hemimethylated site (Goll and Bestor, 2005). DNMT1 is called maintenance DNMT, and is responsible for copying the existing methylation profile during cell division. Thus, DNMT1 participates in the transmission of epigenetic marks and contributes to the epigenetic inheritance mechanism (Hermann et al., 2004). Methylated DNA can be demethylated in either an active or passive manner. Active demethylation occurs mainly through the action of ten-eleven translocation (TET) deoxygenases, which can remove methyl

groups independently of replication (Kohli and Zhang, 2013). Passive demethylation occurs when newly synthesized DNA strands are not methylated by DNMT1 either because DNMT1 activity is lost or because DNMT1 concentration is diluted during several replication rounds (Rasmussen and Helin, 2016). The same effect, i.e., inhibition of DNMT1 activity, can be achieved by the use of demethylating agents such as 5-azacytidine (Ribas et al., 2017). In fishes, because of several genome duplications, there can be several isoforms of each type of DNMT and TET enzymes (Chen et al., 2017) although with no new unique functions known so far as a result of these genome duplications. In vertebrate genomes, CpGs are usually methylated and evenly distributed, except in regions where there is an elevated content of CpGs (Illingworth and Bird, 2009; Moore et al., 2013). These regions are called CpG islands (CGIs), and they are normally associated with promoter or regulatory regions (Illingworth and Bird, 2009; Moore et al., 2013). Changes in methylation levels in these CGIs are associated with gene expression regulation (Illingworth and Bird, 2009; Moore et al., 2013). When comparing groups, a difference in the methylation level of a given CpG loci is referred to as differentially methylated cytosine (DMC), and many of them as DMCs. Likewise, genomic regions with different DNA methylation are called differentially methylated regions (DMRs).

2.3 Epigenetic changes with age

Aging is associated with changes in DNA methylation. Variation in DNA methylation is influenced by both intrinsic and extrinsic factors, plus stochastic events (Jung and Pfeifer, 2015). In general, there is an age-dependent change in DNA methylation that may be summarized in global genomic hypomethylation (Heyn et al., 2012), accompanied by hypermethylation of specific CpGs. However, aside from this "epigenetic drift", there are changes that are of a "clock-like" nature and thus are predictable (De Paoli-Iseppi et al., 2019) (Figure 2A). Chronological age (referred to as cAge in some publications) predictors built with information on DNA methylation at several CpG loci are termed "epigenetic clocks" (Guevara and Lawler, 2018; Zhang et al., 2019). Epigenetic clocks integrate age-associated (hyper- and hypomethylation) changes from a group of carefully selected loci across the genome that provide what is termed estimated or predicted age and, in some cases, epigenetic age (eAge). High correlation between eAge against cAge underlies the use of epigenetic clocks and their ability of predicting cAge with high accuracy.

There are still many unknown aspects regarding what epigenetic clocks actually measure (Bell et al., 2019). It has been argued that they do actually measure aspects of age-related epigenetic drift but the fact that eAge can be measured in both proliferating and non-proliferating tissues suggest that passive demethylation unlikely underpins epigenetic clocks (Simpson and Chandra, 2021). Thus, the nature of eAge is still unknown but does not preclude the use of epigenetic clocks to estimate age in different taxa, mostly vertebrates. Therefore, in this review we will use the term "age estimation" rather than eAge. It has also been argued that epigenetic clocks can provide both chronological and biological age information (i.e., as measured now by variables such as creatinine, fasting glucose, telomere length, etc.) in different relative proportions, which depend on the CpG loci used to build the clock (Bell et al., 2019).

Although DNA methylation changes with age provides a strong foundation for the development of epigenetic clocks, consideration could perhaps be paid to other epigenetic mechanisms or modifications such as histone modifications and, specially, non-coding RNAs (ncRNAs), including micro-RNAs (miRNAs) (van Gelderen et al., 2022). In particular, quantifying levels of miRNAs is relatively easy (Lee and Shin, 2012). However, although fish miRNAs can change along developmental stages, e.g., during flatfish metamorphosis (Campos et al., 2014), to the best of our knowledge, variations of a clock type have not been described yet.

2.4 Vertebrate epigenetic clocks

A tissue-independent epigenetic clock was first developed in humans in a landmark study (Horvath, 2013) and was made of 353 CpGs. These CpGs can be divided into two sets according to their correlation with age. Thus, the 193 positively and 160 negatively correlated CpGs become hypermethylated and hypomethylated with age, respectively. Subsequently, additional epigenetic clocks have been

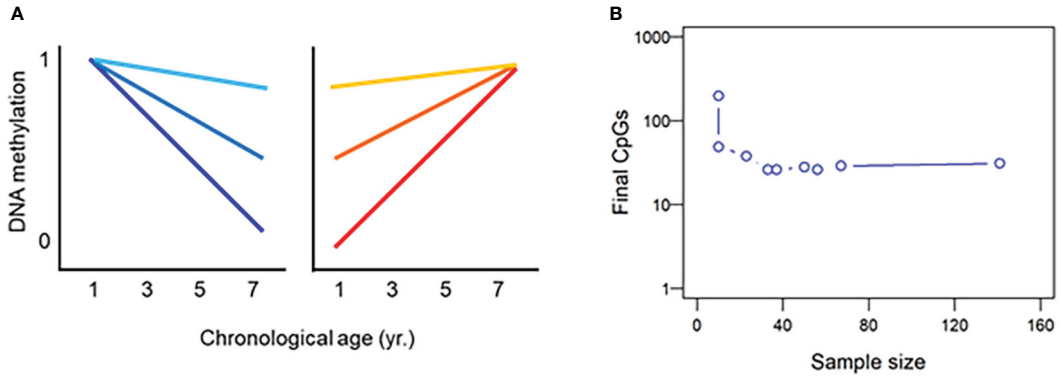


FIGURE 2
Principles behind the development of epigenetic clocks. **(A)** There are some CpG loci the methylation of which is of a clock nature becoming progressively hypo- (blue) or hypermethylated (red) in different degrees. **(B)** Relationship between sample size and number of informative loci in the piscine epigenetic clocks developed so far.

developed for several vertebrates (De Paoli-Iseppi et al., 2019), mostly mammals, but also birds, a marine turtle (Mayne et al., 2022) and fishes (Table 1). More recently, epigenetic clocks have also been developed in an invertebrate, the European lobster (Fairfield et al., 2021), and in a tree, the loblolly pine (Gardner et al., 2022), suggesting that age-dependent DNA methylation changes may be a universal phenomenon in eukaryotes. Of note, recently, a multi-species epigenetic clock was built to age eight species of toothed whales and dolphins (odontocetes) (Robeck et al., 2021). Further still, a single universal clock was built based on 185 mammalian species that is capable of predicting accurately (average correlation > 0.91) the age across species and tissues, indicating that aging is an evolutionarily conserved developmental process (Lu et al., 2021).

Epigenetic clocks are characterized by very high accuracy when compared to more traditional methods of age estimation in mammals such as assessment of telomere length. Since epigenetics bridges genomic and environmental information many epigenetic changes are sensitive to environmental influences (O'Dea et al., 2016). Thus, building an epigenetic clock is based on finding CpGs the methylation of which is dependent only on age, not external influences.

2.5 DNA methylation changes with age in fishes

In accordance with the general pattern that seems to be conserved in vertebrates, and shortly after the publication of Horvath's human clock in 2013, it was found that zebrafish (*Danio rerio*) exhibited a general hypomethylation of CpGs with age (Shimoda et al., 2014). Subsequently, changes in DNA methylation were identified and were found to be gene-, tissue- and age-dependent in Chinook salmon (*Oncorhynchus tshawytscha*) (Venney et al., 2016). These results indicated that epigenetic clocks could be developed in fishes.

2.6 Piscine epigenetic clocks

The first piscine epigenetic clock (and the first one in poikilothermic vertebrates) was developed in the European sea bass (*Dicentrarchus labrax*), a modern teleost distributed in the NE Atlantic, Mediterranean and Black Sea and important for both fisheries and aquaculture (Anastasiadi and Piferrer, 2020). Although it is possible to estimate age based on otolith readings in this species (Aguilera et al., 2009), we took advantage of fish of different ages in different research facilities. In this way, chronological age was known with absolute precision and thus, in principle, there was one less possible interfering source of error. We faced two difficulties compared to epigenetic clocks developed until that time. First, in contrast to birds and mammals, fish are poikilothermic vertebrates and thus we did not know how environmental factors such as temperature could affect the results. Second, in contrast to mammals, which exhibit determinate growth, many fish exhibit indeterminate growth, particularly long-lived species of temperate and cool regions (Dutta, 1994). We used epaxial muscle samples and a candidate gene approach, i.e., we started with a set of CpGs distributed in four genes analyzed in another experiment (Anastasiadi et al., 2018). A total of 48 CpGs were tested of which finally 28 CpGs were retained for clock construction. In a parallel test of clock building, 299 CpGs (including the 48 initial) in 12 samples of muscle were used for age prediction and across tissues for testing accuracy. Additional proof that the retained CpGs are reliable age predictors comes from the fact that overall methylation of the 299 CpGs showed no significant differences between the extreme age classes (1.28 vs. 10.5 years) in the muscle, testis or ovary, indicating that differences in the selected CpGs were specific and not due to an overall age-related trend. The performance of the fish epigenetic clock in terms of accuracy and precision (Table 1) compared favorably with clocks developed for other vertebrates (De Paoli-Iseppi et al., 2019). In juvenile European sea bass, accelerated growth due to elevated temperatures had no effect in age prediction, indicating that the clock is able to predict cAge independently of at least some major environmentally-driven perturbations. This study is important not only because it was the first epigenetic clock in a poikilothermic animal and with indeterminate growth, but also because different tissues were assayed and the effects of temperature tested. Nevertheless, environmental influences, especially the early environment, should be assessed since in mammals environmental stress is associated with epigenetic age acceleration (Parrott and Bertucci, 2019). Specifically, glucocorticoid-induced epigenetic changes have been shown to be the mechanisms linking chronic stress with accelerated aging and higher disease risk in mammals (Zannas et al., 2015). In the last years, other piscine epigenetic clocks have been built (Table 1), and are discussed below. Most likely many more are currently being built.

Using the zebrafish (*Danio rerio*) model and Reduced Representation Bisulfite Sequencing (RRBS), a genome-wide technique to assess DNA methylation (see section 3.4.2. below), and starting with more than half a million CpG loci (Mayne et al., 2020) built a epigenetic clock from fin clips, of high accuracy and consisting of 29 final informative CpG loci. Further, using 26 of these CpGs they successfully implemented a PCR-based method followed by sequencing, retaining clock performance (r = 0.97). However, methylation-sensitive PCR did not work as well (r = 0.62) to predict age with sufficient accuracy. This study demonstrated that age can be epigenetically determined in fish with high accuracy at low cost (Mayne et al., 2020). The same authors then used a PCR-based approach and the CpGs of the zebrafish clock to determine age in three endangered species, the Australian lungfish (*Neoceratodus forsteri*), the Mary river cod (*Maccullochella mariensis*) and the Murray cod (*Maccullochella peelii*) using also fin clips. Remarkably, accuracy was high (r > 0.9 in the three species). This study is important because demonstrates that an epigenetic clock developed in one species can be still used with good performance to determine age in phylogenetically distant species and in a non-lethal manner (Mayne et al., 2021b).

Using medaka (*Oryzias latipes*) between 2 and 12 months of age and RRBS (Bertucci et al., 2021) built an epigenetic clock using liver samples. Furthermore, they showed that exposure to ionizing radiation during early development caused epigenetic age acceleration. This study is also important because it sets the basis to further explore aging-by-environment interactions (Bertucci et al., 2021). More recently, Weber et al. (2022) developed epigenetic clocks for two species present in the Gulf of Mexico, the Northern red snapper (*Lutjanus campechanus*) and the Red grouper (*Epinephelus*

TABLE 1 Characteristics and performance of piscine epigenetic clocks.

Species scientific name	Age range included	Number of fish used	Tissue used	DNA methylation method	Initial no. of CpGs	Final no. of CpGs	Method of Validation	Accuracy	Precision (MAE)	Equivalent % lifespan	Reference
Dicentrarchus labrax	0.5-10.5 yr	50	Muscle	MBS	48	28	LOOCV	0.82	2.14 yr	8.5	Anastasiadi and Piferrer (2020)
Danio rerio	11.9-60.1 wk	67	Fin	RRBS	524038	29	10FCV	0.92	3.7 wk	3	Mayne et al. (2020)
Neoceratodus forsteri	0.1-77 yr	141	Fin	Multiplex PCR	31	31	10FCV	0.98	0.86 yr	< 1	Mayne et al. (2021b)
Maccullochella mariensis	0.5-2.88 yr	37	Fin	Multiplex PCR	26	26	10FCV	0.92	0.34 yr	N/A	Mayne et al. (2021b)
Maccullochella peelii	1.1-12.1 yr	33	Fin	Multiplex PCR	26	26	10FCV	0.92	0.34 yr	0.7	Mayne et al. (2021b)
Oryzias latipes	2-12 mo	47	Liver	RRBS	45273	38	10FCV	0.94	22 d	4	Bertucci et al. (2021)
Lutjanus campechanus	1-26 yr	10	Muscle	bis-RADseq	49189	199	10FCV	0.98	N/A	N/A	Weber et al. (2022)
Epinephelus morio	2-14 yr	10	Fin	bis-RADseq	9834	49	10FCV	0.98	N/A	N/A	Weber et al. (2022)
Mean		49			78558	53		0.93	0.87	3.40	
Median		42			4941	30		0.93	0.69	3.0	

d, days; wk, weeks; mo, months; yr, years; LOOCV, leave-one-out cross validation; 10FCV, 10-fold cross validation; MBS, multiplex bisulfite sequencing; RRBS, reduced representation bisulfite sequencing; MAE, mean absolute error; N/A, it does not apply. In all cases the penalyzed regression used was Elastic net.

morio). The low sample size used (10 fish in each case) prevented the application of sample splitting and rigorous testing to prevent model overfitting, probably explaining the high accuracy obtained (r = 0.98) but lack of proper precision estimation. Nevertheless, the study is important for at least two aspects. First, in contrast to previous studies, this was the first time that an epigenetic clock was developed using wild-caught fish where age had to be also determined by otolith readings. Second, because also for the first time bisulfite RAD-sequencing (bis-RAD-seq) is used. This technique has the advantage that it starts with a still considerable number of CpGs but fewer than with WGBS or RRBS (see section 3.4.2. below) and hence the cost per sample is cheaper.

Regardless of the target species, there are a series of important aspects common to all epigenetic clocks: the age range available to build the clock, the number of samples used, the tissue biopsied to obtain DNA and the method used to analyze its methylation level, the number of CpG loci available for clock construction, the type of penalized (or regularized) regression used, how samples are split between training and test samples, the method of result validation, and then the performance of the clock itself, mainly in terms of accuracy and precision. These characteristics roughly correspond to columns in Table 1 and because of their importance they are further discussed below. But first we want to highlight four aspects worth mentioning.

First, penalized regressions are used for epigenetic clock building because they consist of a type of regression that is "penalized" for having a number of variables (methylation at each CpG loci in epigenetic clocks) much, much higher than the number of samples typically available. Penalization is achieved by adding a constraint in the equation that reduces the coefficients of the less contributing variables towards zero (Fu, 1998). The three most common penalized regression types are: Ridge, LASSO and elastic net, discussed in detail in the review on the bioinformatics aspects of epigenetic clocks (Anastasiadi and Piferrer, 2023). Second, when looking at the informative CpGs used in piscine epigenetic clocks, it seems that the number of CpGs that become hypomethylated with age is higher than the number that become hypermethylated (Weber et al., 2022), but whether this is actually related to the global hypomethylation of vertebrate genomes with age remains to be established. Third, the importance that environmental factors may have on the performance of piscine epigenetic clocks. In this regard, the information available so far seems contradictory. On one hand, as explained above, temperature changes during the thermosensitive period, i.e., the period in which temperature changes are able to elicit important physiological responses such as to affect the sex ratio (Navarro-Martín et al., 2009), did not seem to affect the European sea bass epigenetic clock (Anastasiadi and Piferrer, 2020). However, in the medaka exposure to ionizing radiation affected epigenetic aging by changing the rate of DNA methylation in age-associated CpGs (Bertucci et al., 2021). Environmental effects on ageing in poikilotherms certainly deserve more research to obtain robust conclusions. Fourth, it is worth to mention that age-associated CpG loci in zebrafish were used to build two epigenetic clocks for three threatened freshwater fish species: one for the Australian lungfish and another for the Murray cod and Mary River cod, showing that even phylogenetically divergent species share the same CpG sites that change methylation with age (Mayne et al., 2021b). This significant observation suggests that it may be possible to develop a pan-fish epigenetic clock like the pan-mammalian one (Lu et al., 2021).

3 Outlook and important considerations

Epigenetic clocks constitute a new age estimation tool that is less than 10 years old taking into account that the human epigenetic clock, the first one, was published in December of 2013 (Horvath, 2013). Further, in contrast to other molecular approaches that can be applied to fisheries management such as environmental DNA (eDNA) or Close-kin mark-recapture (CKMR), reviewed elsewhere in this volume (Ramírez-Amaro et al., 2022), no literature reviews exist on the subject of epigenetic clocks to estimate age in fishes, although genomic predictors of lifespan do exist for vertebrates (Mayne et al., 2020). Also, to the best of our knowledge, there are only epigenetic clocks for eight fish species so far (Table 1). Therefore, some considerations relevant in the development of new epigenetic clocks for additional species, with the aim to aid in age estimation in fisheries management and in conservation biology will be discussed.

3.1 General outline for the development of a new piscine epigenetic clock

The steps towards the development of piscine epigenetic clocks for new species can be summarized in the following points below (Figure 3), which do not significantly differ from the ones that should be used for any other vertebrate where age estimation is sought for management or conservation purposes:

1) Species selection
2) Sampling scheme
3) Molecular procedures
4) Bioinformatics and clock-building machine learning methods
5) Validation and implementation: ready-to-use kit

Points above are dealt in detail in the following sections below, with the exception of point 4, which are dealt in an *ad hoc* review purely focused on the bioinformatics aspects of epigenetic clock building (Anastasiadi and Piferrer, 2023).

3.2 Species selection

Since changes in DNA methylation with age is a conserved feature across vertebrates (Jung and Pfeifer, 2015), and since several epigenetic clocks have been already built in fishes (Table 1) an epigenetic clock should be possible to build, in principle, for any fish species, not to mention the possibility to build an all-fish epigenetic clock, as advanced above. However, the actual species to be targeted should have commercial relevance for fisheries or for conservation programs.

One interesting aspect is that fisheries target species of both short (ca. 1 yr.) and long lifespan, e.g., > 10 years. Thus, it should be proved whether the clock performs equally well in both types of species, although *a priori* there is no reason to suspect otherwise due to the conserved nature of DNA methylation changes. In a comparison involving six mammalian taxa, epigenetic aging seemed faster in species with shorter lifespans, suggesting that the underlying mechanisms of epigenetic aging might contribute to our understanding of interspecific lifespan diversity (Parrott and Bertucci, 2019). Similarly, to study DNA methylation changes associated with age in fishes, species selected should include those with different morphologies (i.e., large and small species in terms of attained maximum length) and also those with differences in life-history traits (i.e., semelparous vs. iteroparous species), to determine whether epigenetic clocks perform well in all of these cases. Lastly, the species considered should come from different geographic locations. Epigenetic modifications have a strong underlying genetic basis and, in addition, can be influenced by changes in the environment such as by temperature changes (Anastasiadi et al., 2018), diet composition (Dhanasiri et al., 2020) and stress (Krick et al., 2021). Therefore, it is possible to find differences among different populations of the same species. If that were the case, clock construction should take into account these differences by selecting loci that are not influenced by genetic variation such as single nucleotide polymorphism (SNPs).

3.3 Sampling scheme

3.3.1 Sample size

In the epigenetic clocks developed so far in fishes, the number of individuals used has been usually lower than 100 (mean: 49; range 10-141). The important aspect of the required sample size to build a reliable epigenetic clock has been addressed by Mayne and colleagues (Mayne et al., 2021a). These authors determined the performance of epigenetic clocks by mean Pearson correlation (r) and maximum absolute error (MAE) in simulations using different sample sizes ranging from 15 to 394 for humans and 15 to 96 for zebrafish with a step size of one. The minimum sample size was determined when the correlation approached the true correlation, defined as the correlation achieved that does not increase more with increasing sample size. These Monte-Carlo simulations included determination of the point of stability (POS), defined when the correlation is within a pre-defined range or width (w) around the true correlation with a certain degree of confidence. Mayne et al. (2021a) showed that the minimum sample size required to develop an epigenetic clock with appropriate accuracy and precision was 70 (w = 0.20 and 80% level of confidence interval) but that for better performance it should approach 134 (w = 0.10 and 95% confidence interval). Of note, the approach used by (Mayne et al., 2021a) provides a robust framework to determine whether future epigenetic clocks have the required statistical power.

There is not a particular number of age-classes that should be sampled per se, as far as the number of samples are evenly distributed and adequately cover the age range for which the clock is meant to be used to estimate age. However, DNA methylation in a given CpG will have a certain level of variation among individuals. If the samples available for clock construction are divided into discrete age classes, then their number per age class might become too low if the total number is reduced. Therefore, the number of age classes should be calculated based on the number of replicates per age class and considering the ideal total estimated by simulations. For example, if the minimum sample size is, say, 72, then not less than four age classes with 18 replicates each and ideally six to eight age classes with 12 and 9 replicates each, respectively, should be used. The availability of enough fish per age class becomes less relevant as far as specimens

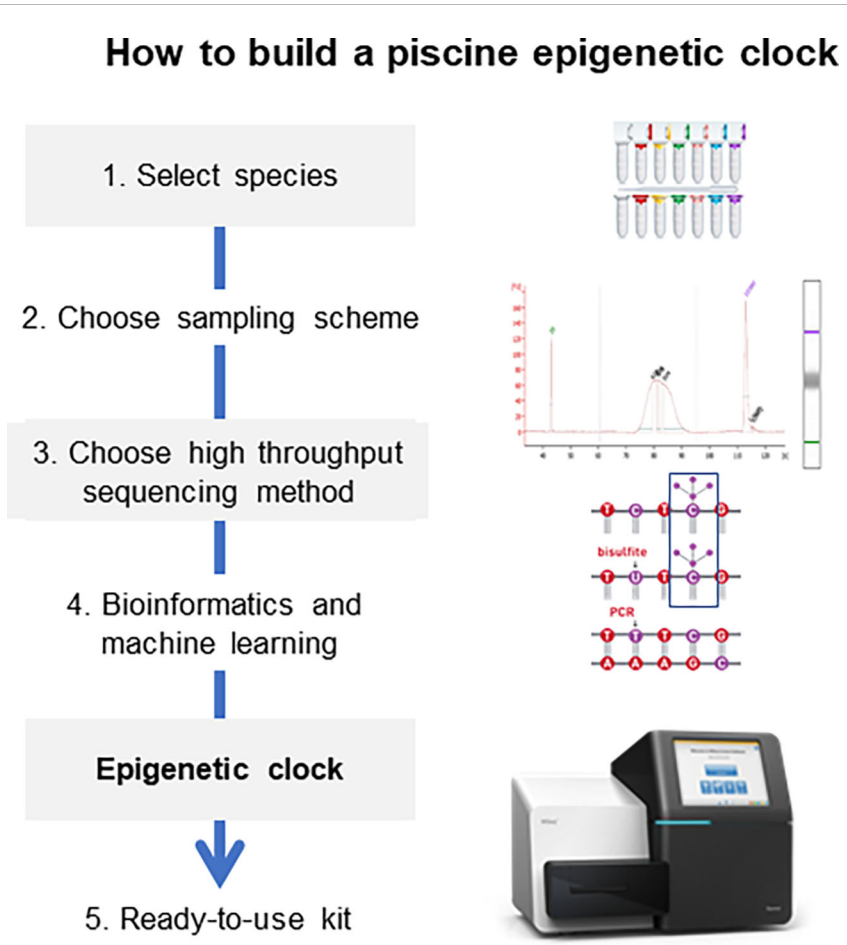


FIGURE 3
General procedure for the development of epigenetic clocks. Notice that in the bioinformatics pipeline not all samples are used in the same manner; some are used for the train set while others are used for the test set (Anastasiadi and Piferrer, 2023).

covering a wide spectrum of sizes are sampled. In any case, it is important to include individuals representative of the top and bottom 10% of known ages (Mayne et al., 2021b). Attention should also be paid to the relationship between time points and age classes as determined by other age-estimation methods such as otolith analysis.

Studies carried out using human blood and saliva samples as tissue sources to develop epigenetic clocks showed that almost 100% of the variation in chronological age could be effectively captured by the DNA methylation probes on the arrays used. Development of epigenetic clocks implies the use of machine learning procedures (see Anastasiadi and Beemelmanns, 2023, for review). This means that the accuracy of a chronological age predictor should increase with a very large training set (Zhang et al., 2019). In practice, actual sample size will determine aspects of machine learning model building, i.e., how samples will be split into training and test and how the models will be evaluated.

3.3.2 Data to be collected from the same individuals

In order to develop piscine epigenetic clocks, it is important to gather information on other variables. Usually, fish age is determined by a combination of size-age relationship, i.e., year classes or the range of length for individuals of a given species at a given age class, and/or the assessment of the hard structures mentioned above (Morales-Nin et al., 1998). Weight usually has a higher coefficient of variation than length because at a given length fish can exhibit quite different weights depending upon food abundance and other environmental factors. In order to relate length and weight the condition factor (K, typically defined as K = 100* W/L3, where W is the weight and L is the length) is used (Karuppiah et al., 2022). There are three ways to determine length: a) standard length (SL), the length measured from the tip of the snout to the posterior end of the last vertebra, i.e., the length without the caudal tail, b) fork length (FL), the length measured from the tip of the snout to the end of the middle caudal fin rays, used in fishes in which it is difficult to tell where the last vertebra ends, and c) total length (TL), the length measured from the tip of the snout to the tip of the longer lobe of the caudal fin (Froese, 2006). It does not matter which one is used for epigenetic age estimation as far it is clearly stated. However, the measure that allows comparison with historical series of data should be favored.

Sex and maturity status are routinely determined during fish sampling and these are important because there are many species in which after a certain age/length there is sexual growth dimorphism (SGD), i.e., one sex grows more than the other (Parker, 1992). Thus, in order to account for the influence of SGD, sex and, if possible, maturity status (e.g., immature, maturing, mature, or spent gonads, after visual

inspection) should be included. In the epigenetic clocks developed so far in fish sex has not been properly taken into account, although in the European sea bass muscle epigenetic clock was tested in ovaries and testes (Anastasiadi and Piferrer, 2020). Thus, it is interesting to know about the possible existence of sex-related differences in the tick rate of piscine epigenetic clocks. In fact, for many species of commercial importance such as the Atlantic cod, *Gadus morhua* (Armstrong et al., 2004), the European hake, *Merluccius merluccius* (de Pontual et al., 2006) and the Pacific halibut, *Hippoglossus stenolepis* (Drinan et al., 2018), just to name a few, there are large differences in growth rate and maximum size depending on the sex. Development of epigenetic clocks in these species will therefore need to take into account these differences and to seek whether a sex-independent clock can be built that is also highly accurate.

Geographic location (latitude and longitude) of wild caught samples should also be recorded. This may be helpful in species with large geographical variation, where polymorphisms at certain loci may exist due to the genetic structure of populations. Taking into account the possible effect of genetics may help in excluding polymorphic loci. Also, location can help to know whether there are differences in the magnitude of certain environmental variables that would need to be taken into account as having possible influences on DNA methylation at certain environmentally responsive loci (ERL), loci the methylation of which depends on the magnitude of an environmental factor (Garg et al., 2018).

Fishes are poikilothermic animals and as such their body temperature depends on that of the water. There are some exceptions, such as tunas, which do generate a certain amount of body heat and therefore their body temperature, although still dependent on the water temperature, can be a few degrees above (Barrett and Hester, 1964). Temperature impacts biological processes at different levels of organization by changing the rates of chemical reactions and physiological processes, or by changing the three-dimensional shapes of biomolecules (Schulte et al., 2011). Growth, therefore, depends on temperature and seasonal cycles of growth in fish are particularly apparent in, for example, species living in temperate latitudes or cold latitudes with a short growing season (Boltaña et al., 2017).

Thus, in order to properly construct an epigenetic clock in fish, when sampling the following information should be collected from each fish: 1) Length 2) Weight 3) Condition factor (K), based on the combination of #1 and #2; 4) Sex; 5) Maturity status; 6) Location; and 7) Environmental information, mainly water temperature. Having this data collected will allow to correlate variation in these variables with age or vice versa and, importantly, to have them as potential covariates for optimal clock development.

3.3.3 Selection of tissues from which DNA samples should be obtained

The two major criteria for selection of tissues to be sampled would be: 1) easy access to the tissue and, if possible, that its collection is not lethal to the animal, and 2), that their inclusion should not represent a significant increase in labor in the context of routine samplings during periodic surveys.

In the piscine epigenetic clocks built so far (Table 1), basically three tissues have been used: fin, liver and muscle (Figure 1). Fin clips are very easy to obtain, only a pair of scissors and tweezers are required, and are commonly preferred for genotyping, so they are already routinely taken. Fin clips, in addition, do not damage the aspect of the fish and therefore their commercial value should be retained for sales, and sampling them is not lethal. The liver is a large organ very easy to distinguish once the internal organs have been exposed. During routine sampling cruises, the body cavity of fish is typically opened up since their sex and maturity status is usually assessed. Taking advantage of this to also collect a fragment of the liver would require minimal extra work, while the aspect of fish is not further damaged either, especially if destined for filleting. Other tissues potentially useful would be a muscle fragment or the branchial arch. If the muscle is chosen, care should be taken not to include skin and scales. Blood can be also considered a good candidate since, in fact, has been the tissue of choice in the development of epigenetic clocks in mammals (Bell et al., 2019; De Paoli-Iseppi et al., 2019). Mayne et al. (2020) suggested that, despite its diverse cellular heterogeneity, blood would provide a more homogeneous cell population than fin clips and therefore require less CpGs to build an epigenetic clock but fin clips have already provided very good results (Table 1). The size of samples does not have to be large since DNA is the only thing is needed, so a cube of a few millimeters in volume is sufficient. Dissected tissues should be placed in 95% ethanol (1:3 volume ratio at least) with no methanol content, avoiding prolonged exposure to sunlight, flash frozen with liquid nitrogen or dry ice, if available on board, and in any case stored at 4°C, -20°C or -80°C depending on the possibilities until further analysis.

The tissue should be also selected based on having low propensity to epigenetic age acceleration, i.e., the difference between eAge vs. cAge. This is referred to as the tick rate of the epigenetic clock, usually higher in younger individuals and diminishing in older ones (Horvath, 2013). This means that the relationship between eAge and cAge can be non-linear over the lifespan as shown, for example, in teenager humans (Bell et al., 2019) although the underlying cause and possible external influences are not well understood. With the data available so far it is still soon to know whether age acceleration is also present in fish. Until this aspect becomes clearer, some precautions seem appropriate. Thus, the pronephros or anterior kidney of fish is an analogue of the bone marrow of higher vertebrates and functions as the primary hematopoietic tissue and lymphoid organ in bony fishes (Kondera, 2019). Therefore, the pronephros should probably be avoided when developing an epigenetic clock in fish in order to avoid possible age-related changes in hematopoiesis. Likewise, the brain should also probably be avoided, the reason being that 5-hydoxymethyl-cytosine (5hmC) is an epigenetic modification particularly prevalent in the brain and indistinguishable from 5mC after bisulfite conversion. It is possible that this characteristic of the brain could cause errors in age prediction (Simpson and Chandra, 2021). In humans, in order to aim for higher accuracy for a particular cell-type or specific application, several epigenetic clocks have been developed for single tissues, since cell heterogeneity in tissue composition can compromise the accuracy of epigenetic clocks (Simpson and Chandra, 2021). However, a multi-tissue epigenetic clock like the one developed for humans (Horvath, 2013) has not been developed in fish.

Finally, an aspect worth considering is that the use of fin clips and thus non-lethal sampling theoretically allows repeated measures in

the same individuals (e.g., if kept in tanks or through mark-recapture) and hence to expand the age of the models as well as to study DNA methylation changes as fish age (Mayne et al., 2021b). This offers an opportunity that is not possible or very difficult for most tissues, except blood, and hence allows the study of age-related methylation changes on an individual basis, something that perhaps could help to refine models.

3.4 Molecular procedures

3.4.1 DNA samples

DNA quality parameters should be in line with those typically used in molecular biology applications. Therefore, ideally at least 1 µg of clean DNA of high molecular weight (>10 kb) per sample at 20 ng/µl in at least 50 µl volume should be available for clock building. Samples should be checked for degradation. Bisulfite conversion of DNA samples is an essential step of the procedure, nowadays routinely done with commercial kits allowing for > 99.5% bisulfite conversion rates. Targeting specific genomic parts may be performed *via* enzyme digestion or PCRs or other methods and would typically be followed by sequencing to obtain information at single nucleotide resolution. Downstream procedures are explained in the Bioinformatics review published elsewhere (Anastasiadi and Piferrer, 2023).

3.4.2 High throughput sequencing methods

A whole-genome or genome-wide method should be employed to simultaneously interrogate the methylation of hundreds or thousands of candidate CpGs with the aim to find the most informative CpGs. In the common ancestor of teleosts there was an additional genome duplication (referred to as 3R or teleost-specific genome duplication, TGD) when compared to the rest of vertebrates Thus, teleost genomes tend to contain 10–15% more duplicated genes and pseudogenes than those of other vertebrates (Volff, 2005). Whole Genome Bisulfite Sequencing (WGBS), Reduced Representation Bisulfite Sequencing (RRBS) (Gu et al., 2011) or Bisulfite RAD-seq (bis-RAD-seq;Trucchi et al., 2016) are appropriate techniques for these steps since they provide single nucleotide resolution and wide genome coverage. Here it would suffice to mention that two of these three methods have been used to develop epigenetic clocks in fish and below we made recommendations as which one can be used to develop new clocks. It should be pointed out that by using WGBS one deals with all CpG loci present in the genome, whereas RRBS and bis-RAD-seq are genome-wide but not whole genome methods since they provide only small (< 5%) representation of the genome. Too many CpGs imply also a heavy workload for subsequent bioinformatics processing of the data. For this reason, it is best to use bis-RAD-seq or epiGBS or a similar approach as the first option and RRBS as the second option. Bis-RAD-seq can provide 10.000 to 50.000 CpGs per sample. Notice that this is in the range of initial CpGs used for the development of the human epigenetic clock, but much higher than targeted approaches such as multiplex PCR-based protocols and much lower than RRBS- or WGBS- based protocols (Table 1). Finally, long-read sequencing platforms such as the Oxford nanopore should be mentioned because a major advantage is that it skips the treatment of DNA with bisulfite. This, along the fact that there are portable devices to implement this technique makes it an approach to consider for future development and implementation to estimate age in fish. This review is not intended to discuss details of each method but the reader interested in further details can consult any of the excellent reviews available, e.g., Kurdyukov and Bullock (2016).

DMRs would theoretically be possible to use, i.e., a group of CpGs closely located in a given genomic region the methylation of which changes with age in the same manner. However, in epigenetic clocks DMRs may add noise instead of power since three good CpGs were sufficient in the mice to predict age (Han et al., 2019). During filtering to find informative CpGs, at step 1 low stringency is used and CpGs can be loosely ranked by P-values.

The use of genes or gene parts with as little as possible sequence variation across species, i.e., genes with high nucleotide sequence conservation, can be considered, so informative CpGs in one species could also be informative in other species. Evidence along these lines has recently been provided (Mayne et al. (2021b), where informative CpGs in zebrafish were used to build epigenetic clocks for three fish species phylogenetically quite apart. In the European lobster, the evolutionary conserved ribosomal DNA was targeted for constructing an epigenetic clock (Fairfield et al., 2021). Finally, if many loci are available for selection, one possibility is to target loci with function related to developmental genes, since the ageing process was shown to be evolutionary conserved and linked to developmental processes across mammalian species (Lu et al., 2021).

3.5 Bioinformatics and clock-building machine learning methods

A typical epigenetic biomarker discovery procedure should be followed in general for epigenetic clock construction (Anastasiadi and Beemelmanns, 2023), although some steps can be skipped or adapted depending on resources and phylogenetic information. The aim is to build a model that addresses the question of which epigenetic features (CpGs) show a higher association (correlation) with the outcome quantitative variable (age). Detailed explanations and examples on bioinformatics and clock-building machine learning methods can be found elsewhere in this volume (Anastasiadi and Piferrer, 2023), while expanded details on epigenetic biomarker bioinformatics and statistics can be found elsewhere (Anastasiadi and Beemelmanns, 2023). Briefly, the outcome variable will be predicted based on machine learning methods. Penalized regressions have been used for almost all epigenetic clocks developed so far because they fit well the structure of data and the final objective. These result in the best model or epigenetic clock based on a set of few CpGs with highest accuracy and precision. The number of final informative CpGs in epigenetic clocks developed so far ranges between 3 (mice) to 353 (human) with an average around 50. Among the final, informative, age-correlated CpGs, there tends to be a higher frequency of them that are hypomethylated than hypermethylated (Weber et al., 2022).

3.6 Validation and implementation: Ready-to-use kit

The CpGs that best correlate with age are selected to develop a targeted assay. Methods that can interrogate a few number of loci in a high number of samples are adequate, e.g. multiplex bisulfite sequencing, MBS (Anastasiadi et al., 2018) and MeDIP-qPCR (Jacinto et al., 2008). Other promising methods for application of epigenetic piscine clock for age prediction in thousands of samples include high-throughput multiplex PCR similar to Genotyping-in-Thousands by sequencing (GT-seq) (Kelsey et al., 2017), Tagmentation-based Indexing for Methylation Sequencing (TIMEseq) (Griffin et al., 2021) and cross-species methylation arrays based on conserved CpGs (Arneson et al., 2022). Finally, the practical utility of the approach should be assessed using independent samples.

With the data available from the piscine epigenetic clocks built so far, there seems to be no relationship between the number of samples used to build the clock and the final number of informative CpGs that predict age. Likewise, there seems to be no relationship between the initial number of available CpG sites and the number of informative sites (Figure 2B). However, it is interesting to note that many of the piscine epigenetic clocks built so far end up consisting of around 30 informative CpGs (Table 1). Thus, 30 seems to be an appropriate number of informative CpG able to estimate age in fish with high accuracy. Undoubtedly, more data is needed to determine whether the above affirmations hold as epigenetic clocks are developed for additional species. In this regard, it has been suggested that increasing sample size may capture more biological variation and thus require more CpG sites to accurately predict age (Mayne et al., 2020).

Any epigenetic clock should be, at the end, downscaled to a simple test so cost per sample could be drastically reduced. In this regard, microarrays could provide a good solution. Arrays are available that enable quantitative interrogation of selected methylation sites across the genome, offering high-throughput capabilities that minimize the cost per sample. Thus, Illumina's Infinium Methylation EPIC BeadChip Kit covers ~850,000 CpGs of the human genome and comes in different formats to analyze 16 (8x2), 32 (8x4) or 96 (8x12) samples arranged in groups of eight. The cost per sample is around 245 € because of the large amount of CpGs covered. However, this cost can be drastically reduced to ~10 € or even less in custom arrays with a significantly lower coverage of CpG sites (Smith, 2022). For example, a 9K Illumina BeadChip for polar bears (*Ursus maritimus*) was designed from RAD and transcriptome sequencing (Malenfant et al., 2015). Also, a 21K Illumina custom-microarray containing ~80% of the protein coding genes was developed in the European sea bass (Ribas et al., 2019). Therefore, it is possible to envisage a multi-species array for age estimation in fish that could provide reliable results at low cost and able to process a large amount of samples coming from fisheries survey cruises. For example, a five-species array with an average of 50 CpG sites per species done in triplicate plus hybridization controls would require less than one thousand probes. Another alternative would be the construction of high-throughput multiplex PCR assays to determine the methylation level of several age-correlated CpG sites. In this regard, multiplex PCR has been used to estimate the methylation levels of conserved CpGs exhibiting age-related changes of a clock type to build an epigenetic clock for three endangered species (Mayne et al., 2021b).

4 Features of piscine epigenetic clocks as a measuring device

An epigenetic clock to estimate age is, after all, a measuring device based on molecular methods. As such, in this section some characteristics common to all measuring devices are discussed and adapted to the specifics of epigenetic age estimation and, in particular, in fish. These characteristics are: accuracy, precision, repeatability, reproducibility, resolution, sensitivity, and range. The definitions used below are taken from, and in accordance with, accepted international standards (ISO, 1994).

4.1 Accuracy

Accuracy refers to how close is the measured age to the actual "true" age. In epigenetic clocks, accuracy is measured by the degree of correlation between cAge and estimated age. Thus, estimates of model performance are used to report accuracy as Pearson correlation (r), the most popular metric to evaluate regressions. A useful piscine epigenetic clock should be able to correctly classify different age classes. According to Horvath and Raj (2018), an age estimator can be called an "epigenetic clock" only if the correlation between chronological age and estimated age is > 0.80 in a large independent validation data set with a broad range of chronological ages. The piscine epigenetic clocks developed so far have mean accuracy of 0.93 but values of 0.98 are not unusual (Table 1 and Figure 4). Thus, they are able to tell chronological age with high accuracy and nowadays represent the best way to estimate age in fish, much better than what the traditional methods can achieve (Luo et al., 2016). For comparison, in the grey mullet, *Mugil cephalus*, otolith image analysis coupled with deep learning produced accuracy values of 0.64 at best (Politikos et al., 2021).

An inherent aspect of accuracy is bias. Bias is constituted by non-random effects (systematic error) on measured values, and its magnitude should be known and be as low as possible (ISO, 1994). Differences between estimated and true values can be corrected by calibration, e.g., by using a different method of measure. In species where age estimation by otoliths is not possible, this would not be possible either, unless another independent method of age estimation is used (e.g., rearing artificially the species, if feasible). Thus, the selection of the calibration method could include: 1) Length/weight data for independent length-age estimations across a range of environmental conditions, 2) Comparison with other age estimation methods such as otolith or scale readings, and 3) The use of different aquaculture species to fine calibrate representative species, although this possibility can be risky unless age-growth curves among the two species to be compared are well known and are as similar as possible.

The accuracy of epigenetic clocks can be reduced under certain circumstances. For example, in humans cancer is known to change the tick rate of the epigenetic clock (Horvath, 2013). This results in age acceleration, i.e., estimated age is higher than actual chronological

age and leads to wrong age estimations. At present, it is not clear whether age acceleration can occur in fishes and, if it does, be of such magnitude as to impose the need for correction measures. For example, one possibility would be that overall changes in temperature or high levels of contamination could affect the tick rate of epigenetic clocks. Addressing such a possibility should be part of the priorities in the development of few piscine epigenetic clocks since we know that coastal or offshore waters, from different regions, will be differentially impacted by contaminants, and clearly, it is vital to be able to exclude such possible impacts from wild-caught fish.

Other potential sources of error can be bias in the sampling, but this is not related to the epigenetic approach per se and, if it is detected, e.g., by observing unusual age-size distributions or that some age classes are under- or over-represented, this is something that should be corrected by improving sampling procedures or by using bootstrapping techniques (Hoyle and Cameron, 2003).

4.2 Precision

Precision is defined as how close to each other are repeated measures of the same variable (ISO, 1994). Thus, given a set of measurements, they are said to be precise if the values are close to each other (e.g., different fish of the same age class in our case). Notice that precision has nothing to do with accuracy, since a series of measurements can be both accurate and precise, accurate but not precise, precise but not accurate or neither.

In the same way that bias was associated with accuracy, there are also aspects associated with precision. In this regard, error is defined as random effects on measured value, and it should be also estimated. Measurement of errors should be reported in combination with accuracy to properly evaluate any regression model. Precision of an epigenetic clock is usually reported as mean absolute error (MAE), which indicates how well the model (the epigenetic clock) fits the actual data, and can usually expressed in time units without conversion, most commonly years. Other typical error measurements for regression models, such as Root Squared Mean Error (RMSE) would require conversion of their units and are thus less intuitive (Hodson, 2022). Error as proportion of the total species lifespan is another intuitive way of reporting, but estimation of MAE would still precede it.

Error can be introduced in many different points of the procedure including during sampling, sample preparation, sequencing and bisulfite conversion. Error can be associated with the specific technique used to assess DNA methylation. Error due to bisulfite treatment should be low, since conversion rates >99% are usually required nowadays (Leontiou et al., 2015). In any case, technologies for measuring DNA methylation are constantly improving and deeper sequencing or other more accurate methodologies are expected to further improve the precision of the epigenetic clocks (Polanowski et al., 2014). Aside from the European sea bass epigenetic clock, built as a proof-of-concept piscine epigenetic clock (Anastasiadi and Piferrer, 2020), current epigenetic clocks have MAE typically of less than 1 year and sometimes even of less than 6 months (Table 1). Minimizing error is very important in age estimation for fisheries management because otherwise poor decisions may be made. The best hard structure-based methods currently available, i.e., Fourier transform near-infrared spectroscopy (FT-NIRS) that can derive age estimates from the vertebral centre provide error values of about 1.5 years (Helser et al., 2019).

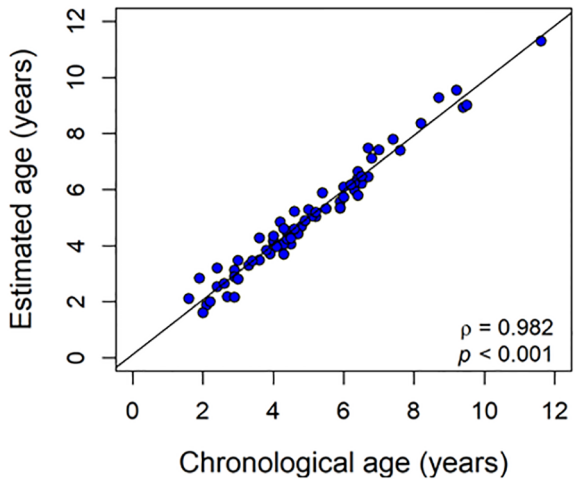


FIGURE 4
Typical epigenetic clock in fish. Estimated age is regressed against chronological age. In this hypothetical example, the sample size is 70, a number capable of producing good results while accuracy, represented by the correlation coefficient, is 0.95 or higher, common in the latest developed piscine epigenetic clocks.

4.3 Repeatability and reproducibility

Repeatability is the precision obtained by making the same measure on the same set of samples with the same method. In addition, these measures should be carried out in the same laboratory by the same operator using the same equipment within short intervals of time (Pryseley et al., 2010). This leads to the intra-assay error, expressed as the coefficient of variation (the ratio of the standard deviation to the mean). On the other hand, reproducibility is the precision obtained by making the same measure on the same set of samples in different laboratories with different operators using different equipment and at different times (Pryseley et al., 2010). In addition, measures can include testing complimentary but different methods of measure. This leads to the estimation of inter-assay error. These are inherent parts of precision and should also be considered. In particular, clocks can provide good repeatability but loose reproducibility (e.g., if the tick rate changes), and this is something known to occur in biological clocks as well (Horvath, 2013). Thus, these are aspects worth considering because it is conceivable that samples obtained in a sampling cruise could be processed in different labs. However, repeatability and reproducibility have not been properly reported in the piscine epigenetic clocks developed so far.

4.4 Resolution

The resolution is the minimum difference of age (in months/years) that can be detected and is related to the precision of the clock. A desirable property of a piscine epigenetic clock would be the

possibility to identify differences of less than a year, for example, to distinguish 1 from 1+, 1+ from 2, 2 from 2+, etc. age classes. This would allow to gather fine-scale information that would represent an advantage for fisheries management. Assessing differences of less than one year is possible in some species using otoliths (Luo et al., 2016) but is not the norm. However, this requires an exhaustive analysis of the daily growth rings, if feasible, using scanning electron microscopy and it is done only for research purposes, but not for routine age estimation. It should be noted that the tick rate of vertebrate epigenetic clocks is higher with younger individuals, as assessed both in different human cell lines (Horvath, 2013) and so it seems to be in sea bass (Anastasiadi and Piferrer, 2020). This may imply higher variability in DNA methylation at a given series of CpGs and thus it may require including more samples of these early age classes when developing epigenetic clocks for new species.

4.5 Sensitivity

The sensitivity is the minimum value (age) that can be detected. This is a challenge because in the epigenetic clocks developed in vertebrates age estimation around one year seems difficult. From a fisheries/conservation point of view, ability to distinguish 0, 0+ and 1 age classes would be considered an important aspect because these are crucial stages for the recruitment towards future exploited populations, especially when it comes to highly dynamic populations or very short-lived species. Clocks developed specifically for younger individuals might provide a solution to this. Having precise information of these age classes is an important priority for fisheries management. Not counting the model species like zebrafish and medaka, which are short-lived, and based in what has been reported for the Australian lungfish and Mary river cod (Mayne et al., 2021b), discrimination among the 0, 0+ and 1 ages classes seems feasible.

4.6 Range

The range, or the difference between maximum and minimum age predicted, should also be defined for a piscine epigenetic clock. This is a very important aspect because otolith measurement accuracy ceases to work once fish reach a certain age/length due to the peculiarities of the growth curve (Campana and Thorrold, 2001).

This limitation of current methods may be circumvented with epigenetic clocks since DNA methylation changes still take place even when growth slows down with age. In this regard, the vertebrate epigenetic clocks developed so far have been used with wide age ranges: 0.1-77 years in fish (Table 1); 0-21 years in a sea bird (De Paoli-Iseppi et al., 2019); 0-30 years in whales (Polanowski et al., 2014); 0-14 years in bats (Wright et al., 2018), 2-39 years in chimpanzees (IIto et al., 2018), 0-23 years in naked mole-rats (Lowe et al., 2020). Therefore, epigenetic clocks in fishes should allow age estimation during the whole lifespan and thus including the age classes commonly targeted by fisheries. An interesting aspect of the DNA methylation-based clocks is that the differences in known and estimated age show little heteroscedasticity, meaning that variance in the differences is similar throughout the range of ages assayed (Polanowski et al., 2014).

5 How to address the integration of age estimation using DNA methylation-based biomarkers in fisheries management?

In the sections above, some of the drawbacks and limitations of the methods commonly used to estimate age in fish for fisheries management have been explained. It has been shown that DNA methylation changes occur as animals age and that some of these changes are of a clock-like type. Consequently, identifying loci that exhibit DNA methylation changes highly correlated with time is the basis for the development of epigenetic clocks. Their development in humans (Horvath, 2013) and other vertebrates (De Paoli-Iseppi et al., 2019) suggested that epigenetic clocks should be possible to develop in fish. The major doubt about the success of such enterprise was the fact that all previous clocks in vertebrates were generated in several mammalian and a bird species, i.e., warm-blooded species with determinate growth, while fish are poikilothermic and have indeterminate growth. The first epigenetic clock in fish was developed in the European sea bass, taking advantage of the fact that age could be known with accuracy because fish were of hatchery origin (Anastasiadi and Piferrer, 2020). This helped to pave the way for the development of similar clocks in other fish species of commercial importance (Table 1). The current state of the art constitutes, in our opinion, a very promising framework upon which additional applications can be developed, especially now that we know that epigenetic clocks across many species can be developed (Lu et al., 2021; Robeck et al., 2021).

To summarize, in our opinion the major advantages of the new DNA methylation-based epigenetic clocks for age estimation in fish over traditional methods such as otolith analysis are:

1. High accuracy
2. High precision
3. Non-lethal
4. Does not damage the commercial value of the specimens
5. Potential transferability to other species since informative loci developed in one species can be used in a different species (with adjustments) reducing cost of development for the new species.
6. Higher consistency between labs/investigators, since standardized measurements rely on computed values with clear protocols, and thus do not rely on interpretation like otolith readings.
7. Age can be estimated independently of growth patterns by the identification of CpGs with clock-like methylation changes with age.

Points of caution and strategies for mitigation include:

1. Personnel needs specific knowledge but training in general molecular biology/sequencing techniques and general bioinformatics should be enough.
2. The process of sampling tissue, extracting DNA, assessing DNA methylation, analyzing, predicting age can be time-consuming at the beginning but the whole process is prone to

be automated and eventually will allow the processing of hundreds or thousands of samples together.

3. A ready-to-use kit for commercial application has not been developed yet. Cost per sample can be expensive until such type of kit is ready.
4. The influence of environmental and other cues needs to be better studied. Information is still too fragmentary to make strong conclusions about the independence of such external influences.

Some aspects that are important to take into account towards the development of piscine epigenetic clocks for their application in fisheries management have also been presented. These included the number of samples to be collected, tissues to be targeted, handling DNA sample preservation and the type of epigenetic modifications to be measured, clearly favoring DNA methylation. The steps on how to develop such a clock and the techniques available to do so have also been explained. Finally, a section has been devoted on the different characteristics such as accuracy, precision, etc., that as any measuring device the piscine epigenetic clock should have. Here it should be emphasized that most likely a set of < 100 carefully selected loci would be enough to predict age with both accuracy and precision. Expertise in epigenetics, machine learning methods and fisheries management are essential to bring this endeavor to success.

No major technical difficulties are envisaged, perhaps the most important aspect is, once developed, the actual technique can then be used in large sample sizes at a very low cost. Nowadays age estimation involves otolith extraction, processing and analysis and therefore constitutes a significant part of the cost of stock assessment (Helser et al., 2019). The cost benefit of implementing epigenetic clocks is dealt with in great detail in another review of this special volume that also includes other molecular-based techniques for fisheries management (Rodríguez-Rodríguez et al., 2022).

Importantly, improved age prediction would contribute to better fisheries management in a context of overexploited fish stocks worldwide. It can be envisaged that one day regional, national and supranational organizations involved in fisheries management will use epigenetic clocks for age estimation in commercial fisheries. Further, advances in techniques aimed at measuring DNA methylation will make it possible to estimate age in large amounts of fish at a very low cost.

Challenges to be resolved include to experimentally verify that, as expected, there will be specific CpG loci with age-related methylation changes that are conserved across fish species. If so, this would facilitate the development of multi-species epigenetic clocks. A particularly important aspect is to determine how environmental changes, especially temperature and food availability may affect the tick rate of piscine epigenetic clocks. Recall that the evidence gathered so far does not allow firm conclusions since temperature did not affect the performance of the European sea bass clock (Anastasiadi and Piferrer, 2020) but ionizing radiation affected the medaka epigenetic clock (Bertucci et al., 2021). Further, it may be difficult to obtain individual samples from the whole age range, especially towards the maximum lifespan, for some fish species, or if age estimation by otoliths or other current method is problematic, calibration of the epigenetic clock may be challenging.

Fish are the most diverse group of vertebrates in many aspects, including life-history traits, morphology, habitat occupancy, reproduction, etc. This diversity may be regarded as a challenge, but in fact piscine epigenetic clocks have been built from both short-lived species such as zebrafish or medaka and very long-lived species such as the Australian lungfish, although the error of the clock of the Australian lungfish was lower in the oldest fish (Mayne et al., 2021b). Thus, despite their high degree of diversity in many aspects of their biology, including longevity, this does not seem an obstacle for the development of epigenetic clocks.

Last but not least, estimation of fish age would also be a very useful molecular resource for ecological studies and, in particular, for those aimed at conservation purposes. Thus, epigenetic clocks could be of practical use in the estimation of age classes in endangered species, as well as for fish biology in general. Epigenetic age estimation as well as estimation of other essential parameters such as species population structure will likely improve stock assessment and fisheries management in a significant manner in the years to come.

Author contributions

FP conceived and wrote the manuscript. DA conceived and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was funded through EU Contract – EMFF/2018/015. Research at the F.P. laboratory was also supported by Spanish Ministry of Science Grant No. PID2019-108888RB-I00. Also with funding from the Spanish government through the 'Severo Ochoa Centre of Excellence' Accreditation (CEX2019-000928-S).

Acknowledgments

We would like to thank Drs. Gary Carvalho, Sissel Jentoft, Laszlo Orban and Allan Tucker and the two reviewers for constructive criticisms and suggestions for improvement.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

The information and views set out in this publication are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

References

Aguilera, B., Catalán, I. A., Palomera, I., and Olivar, M. P. (2009). Crecimiento del otolito en larvas de lubina europea (*Dicentrarchus labrax, l.*) bajo régimen de alimentación constante o variable. *Sci. Mar.* 73, 183–197. doi: 10.3989/scimar.2009.73n1173

Anastasiadi, D., and Beemelmanns, A. (2023). "Development of epigenetic biomarkers in Aquatic organisms," in F. Piferrer and H. P. Wang *Epigenetics in Aquaculture* (Hoboken, NJ: John Wiley & Sons Ltd).

Anastasiadi, D., and Piferrer, F. (2020). A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. *Mol. Ecol. Resour.* 20, 387–397. doi: 10.1111/1755-0998.13111

Anastasiadi, D., and Piferrer, F. (2023). Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. *Front. Mar. Sci.* 10, 1096909. doi: 10.3389/fmars.2023.1096909

Anastasiadi, D., Vandeputte, M., Sánchez-Baizán, N., Allal, F., and Piferrer, F. (2018). Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics* 13, 988–1011. doi: 10.1080/15592294.2018.1529504

Armstrong, M. J., Gerritsen, H. D., Allen, M., McCurdy, W. J., and Peel, J. A. D. (2004). Variability in maturity and growth in a heavily exploited stock: cod (*Gadus morhua* l.) in the Irish Sea. *ICES J. Mar. Sci.* 61, 98–112. doi: 10.1016/j.icesjms.2003.10.005

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H., et al. (2022). A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat. Commun.* 13, 783. doi: 10.1038/s41467-022-28355-z

Bank, C. (2016). *Validation of age determination methods for monkfish (Lophius americanus)* M.Sc. thesis. (Dartmouth: University of Massachusetts) 104. doi: 10.13140/RG.2.2.36759.57761

Barrett, I., and Hester, F. J. (1964). Body temperature of yellowfin and skipjack tunas in relation to sea surface temperature. *Nature* 203, 96–97. doi: 10.1038/203096b0

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., et al. (2019). DNA Methylation aging clocks: challenges and recommendations. *Genome Biol.* 20, 249. doi: 10.1186/s13059-019-1824-y

Bertucci, E. M., Mason, M. W., Rhodes, O. E., and Parrott, B. B. (2021). Exposure to ionizing radiation disrupts normal epigenetic aging in Japanese medaka. *Aging* 13, 22752–22771. doi: 10.18632/aging.203624

Boltaña, S., Sanhueza, N., Aguilar, A., Gallardo-Escarate, C., Arriagada, G., Valdes, J. A., et al. (2017). Influences of thermal environment on fish growth. *Ecol. Evol.* 7, 6814–6825. doi: 10.1002/ece3.3239

Brock, H. W., and Fisher, C. L. (2005). Maintenance of gene expression patterns. *Dev. Dyn.* 232, 633–655. doi: 10.1002/dvdy.20298

Campana, S. E. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *J. Fish Biol.* 59, 197–242. doi: 10.1111/j.1095-8649.2001.tb00127.x

Campana, S. E., and Thorrold, S. R. (2001). Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? *Can. J. Fish. Aquat. Sci.* 58, 30–38. doi: 10.1139/f00-177

Campos, C., Sundaram, A. Y., Valente, L. M., Conceição, L. E., Engrola, S., and Fernandes, J. M. (2014). Thermal plasticity of the miRNA transcriptome during Senegalese sole development. *BMC Genomics* 15, 525. doi: 10.1186/1471-2164-15-525

Chen, X., Wang, Z., Tang, S., Zhao, Y., and Zhao, J. (2017). Genome-wide mapping of DNA methylation in Nile tilapia. *Hydrobiologia* 791, 247–257. doi: 10.1007/s10750-016-2823-6

Deans, C., and Maggert, K. A. (2015). What do you mean, "Epigenetic"? *Genetics* 199, 887–896. doi: 10.1534/genetics.114.173492

De Paoli-Iseppi, R., Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A., et al. (2019). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Mol. Ecol. Resour* 19, 411–425. doi: 10.1111/1755-0998.12981

de Pontual, H., Groison, A. L., Piñeiro, C., and Bertignac, M. (2006). Evidence of underestimation of European hake growth in the bay of Biscay, and its relationship with bias in the agreed method of age estimation. *ICES J. Mar. Sci.* 63, 1674–1681. doi: 10.1016/j.icesjms.2006.07.007

Dhanasiri, A. K. S., Chen, X., Dahle, D., Siriyappagouder, P., Fæste, C. K., and Fernandes, J. M. O. (2020). Dietary inclusion of plant ingredients induces epigenetic changes in the intestine of zebrafish. *Epigenetics* 15, 1035-1051. doi: 10.1080/15592294.2020.1747777

Drinan, D. P., Loher, T., and Hauser, L. (2018). Identification of genomic regions associated with sex in pacific halibut. *J. Heredity* 109, 326–332. doi: 10.1093/jhered/esx102

Dutta, H. (1994). Growth in fishes. *Gerontology* 40, 97–112. doi: 10.1159/000213581

Fairfield, E. A., Richardson, D. S., Daniels, C. L., Butler, C. L., Bell, E., and Taylor, M. I. (2021). Ageing European lobsters (*Homarus gammarus*) using DNA methylation of evolutionarily conserved ribosomal DNA. *Evol. Appl.* 14, 2305–2318. doi: 10.1111/eva.13296

Froese, R. (2006). Cube law, condition factor and weight-length relationships: History, meta-analysis and recommendations. *J. Appl. Ichthyol* 22, 241–253. doi: 10.1111/j.1439-0426.2006.00805.x

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *J. Comput. Graphical Stat* 7, 397. doi: 10.2307/1390712

Gardner, S. T., Bertucci, E. M., Sutton, R., Horcher, A., Aubrey, D., and Parrott, B. B. (2022). Development of DNA methylation-based epigenetic age predictors in loblolly pine (*Pinus taeda*). *bioRxiv*. doi: 10.1101/2022.01.27.477887

Garg, P., Joshi, R. S., Watson, C., and Sharp, A. J. (2018). A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. *PloS Genet.* 14, e1007707. doi: 10.1371/journal.pgen.1007707

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic Cytosine Methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Griffin, P. T., Kane, A. E., Trapp, A., Li, J., McNamara, M. S., Meer, M. V., et al. (2021). TIME-seq enables scalable and inexpensive epigenetic age predictions (preprint). *bioRxiv*. doi: 10.1101/2021.10.25.465725

Guevara, E. E., and Lawler, R. R. (2018). Epigenetic clocks. *Evol. Anthropol* 27, 256–260. doi: 10.1002/evan.21745

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* 6, 468–481. doi: 10.1038/nprot.2010.190

Han, Y., Franzen, J., Stiehl, T., Gobs, M., Kuo, C.-C., Nikolić, M., et al. (2019). New targeted approaches for epigenetic age predictions. *BMC Biol* 18, 71. doi: 10.1186/s12915-020-00807-2

Helser, T. E., Benson, I. M., and Barnett, B. K. (2019). Proceedings of the research workshop on the rapid estimation of fish age using Fourier transform near infrared spectroscopy (FT-NIRS). *AFSC Processed Rep.* (Seattle WA: Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv.,) 2019-06, 195. Available at: https://www.afsc.noaa.gov/Publications/ProcRpt/PR2019-06.pdf.

Hermann, A., Gowher, H., and Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. CMLS, cell. *Mol. Life Sci.* 61, 2571–2587. doi: 10.1007/s00018-004-4201-1

Heydenrych, M. J., Saunders, B. J., Bunce, M., and Jarman, S. N. (2021). Epigenetic measurement of key vertebrate population biology parameters. *Front. Ecol. Evol.* 9. doi: 10.3389/fevo.2021.617376

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., et al. (2012). Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10522–10527. doi: 10.1073/pnas.1120658109

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geosci. Model. Dev.* 15, 5481–5487. doi: 10.5194/gmd-15-5481-2022

Horvath, S. (2013). DNA Methylation age of human tissues and cell types. *Genome Biol.* 14, R115. doi: 10.1186/gb-2013-14-10-r115

Horvath, S., and Raj, K. (2018). DNA Methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371–384. doi: 10.1038/s41576-018-0004-3

Hoyle, S. D., and Cameron, D. S. (2003). Confidence intervals on catch estimates from a recreational fishing survey: a comparison of bootstrap methods. *Fisheries Manage. Ecol.* 10, 97–108. doi: 10.1046/j.1365-2400.2003.00321.x

Illingworth, R. S., and Bird, A. P. (2009). CpG islands - 'A rough guide.' FEBS letters 583, 1713–1720. doi: 10.1016/j.febslet.2009.04.012

ISO (1994). Accuracy (trueness and precision) of measurement methods and results — part 1: General principles and definitions (Vernier, Switzerland: International Standards

Organization) Document no. 5725-1. Available at: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en.

Ito, H., Udono, T., Hirata, S., and Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Sci. Rep.* 8, 9998. doi: 10.1038/s41598-018-28318-9

Jacinto, F. V., Ballestar, E., and Esteller, M. (2008). Methyl-DNA immunoprecipitation (MeDIP): Hunting down the DNA methylome. *BioTechniques* 44, 35–43. doi: 10.2144/000112708

Jung, M., and Pfeifer, G. P. (2015). Aging and DNA methylation. *BMC Biol.* 13, 7. doi: 10.1186/s12915-015-0118-4

Karuppiah, K., Ethiraj, K., Sekar, S., Rajendran, K., Krishnamoorthy, M., and Dharmaraj, D. (2022). Weight–length relationships and fulton's condition factors of ten commercially important scombridae fish species in southeast coast of India, bay of Bengal. *Thalassas* 38, 709–713. doi: 10.1007/s41208-021-00390-5

Kelsey, G., Stegle, O., and Reik, W. (2017). Single-cell epigenomics: Recording the past and predicting the future. *Science* 358, 69–75. doi: 10.1126/science.aan6826

Kohli, R. M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479. doi: 10.1038/nature12750

Kondera, E. (2019). Haematopoiesis and haematopoietic organs in fish. *Roczniki Naukowe Polskiego Towarzystwa Zootechnicznego* 15, 9–16. doi: 10.5604/01.3001.0013.4535

Krick, M. V., Desmarais, E., Samaras, A., Guéret, E., Dimitroglou, A., Pavlidis, M., et al. (2021). Family-effects in the epigenomic response of red blood cells to a challenge test in the European sea bass (*Dicentrarchus labrax*, l.). *BMC Genomics* 22, 111. doi: 10.1186/s12864-021-07420-9

Kurdyukov, S., and Bullock, M. (2016). DNA Methylation analysis: Choosing the right method. *Biology* 5, 3. doi: 10.3390/biology5010003

Lee, D., and Shin, C. (2012). MicroRNA–target interactions: New insights from genome-wide approaches. *Ann. New York Acad. Sci.* 1271, 118–128. doi: 10.1111/j.1749-6632.2012.06745.x

Leontiou, C. A., Hadjidaniel, M. D., Mina, P., Antoniou, P., Ioannides, M., and Patsalis, P. C. (2015). Bisulfite conversion of DNA: Performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing. *PloS One* 10, e0135058. doi: 10.1371/journal.pone.0135058

Lowe, R., Danson, A. F., Rakyan, V. K., Yildizoglu, S., Saldmann, F., Viltard, M., et al. (2020). DNA Methylation clocks as a predictor for ageing and age estimation in naked mole-rats, *Heterocephalus glaber*. *Aging* 12, 4394–4406. doi: 10.18632/aging.102892

Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. A., Li, C. Z., et al. (2021). Universal DNA methylation age across mammalian tissues (preprint). *bioRxiv*. doi: 10.1101/2021.01.18.426733

Luo, W.-W., Liu, C.-S., Cao, X.-J., Huang, L.-F., and Huang, S.-Q. (2016). Precision of age estimations from scales, otoliths, vertebrae, opercular bones and cleithra of two loaches, *Misgurnus anguillicaudatus* and *Paramisgurnus dabryanus*. *Folia Zoologica* 65, 183–188 . doi: 10.25225/fozo.v65.i3.a2.2016

Malenfant, R. M., Coltman, D. W., and Davis, C. S. (2015). Design of a 9K illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Mol. Ecol. Resour* 15, 587–600. doi: 10.1111/1755-0998.12327

Mayne, B., Berry, O., and Jarman, S. (2021a). Optimal sample size for calibrating DNA methylation age estimators. *Mol. Ecol. Resour.* 21, 2316–2323. doi: 10.1111/1755-0998.13437

Mayne, B., Espinoza, T., Roberts, D., Butler, G. L., Brooks, S., Korbie, D., et al. (2021b). Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary river cod. *Mol. Ecol. Resour.* 21, 2324–2332. doi: 10.1111/1755-0998.13440

Mayne, B., Korbie, D., Kenchington, L., Ezzy, B., Berry, O., and Jarman, S. (2020). A DNA methylation age predictor for zebrafish. *Aging* 12, 24817–24835. doi: 10.18632/aging.202400

Mayne, B., Mustin, W., Baboolal, V., Casella, F., Ballorain, K., Barret, M., et al. (2022). Age prediction of green turtles with an epigenetic clock. *Mol. Ecol. Resour.* 22, 2275–2284. doi: 10.1111/1755-0998.13621

Moore, L. D., Le, T., and Fan, G. (2013). DNA Methylation and its basic function. *Neuropsychopharmacol* 38, 23–38. doi: 10.1038/npp.2012.112

Morales-Nin, B., Torres, G. J., Lombarte, A., and Recasens, L. (1998). Otolith growth and age estimation in European hake. *J. Fish Biol.* 53, 1155–1168. doi: 10.1111/j.1095-8649.1998.tb00239.x

Navarro-Martín, L., Blázquez, M., Viñas, J., Joly, S., and Piferrer, F. (2009). Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture* 296, 347–358. doi: 10.1016/j.aquaculture.2009.07.022

O'Dea, R. E., Noble, D. W. A., Johnson, S. L., Hesselson, D., and Nakagawa, S. (2016). The role of non-genetic inheritance in evolutionary rescue: Epigenetic buffering, heritable bet hedging and epigenetic traps. *Environ. Epigenet* 2, dvv014. doi: 10.1093/eep/dvv014

Ordoñez, A., Eikvil, L., Salberg, A.-B., Harbitz, A., Murray, S. M., and Kampffmeyer, M. C. (2020). Explaining decisions of deep neural networks used for fish age prediction. *PloS One* 15, e0235013. doi: 10.1371/journal.pone.0235013

Pardo, S. A., Cooper, A. B., and Dulvy, N. K. (2013). Avoiding fishy growth curves. *Methods Ecol. Evol.* 4, 353–360. doi: 10.1111/2041-210x.12020

Parker, G. A. (1992). The evolution of sexual size dimorphism in fish*. *J. Fish Biol.* 41, 1–20. doi: 10.1111/j.1095-8649.1992.tb03864.x

Parrott, B. B., and Bertucci, E. M. (2019). Epigenetic aging clocks in ecology and evolution. *Trends Ecol. Evol.* 34, 767–770. doi: 10.1016/j.tree.2019.06.008

Polanowski, A. M., Robbins, J., Chandler, D., and Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Mol. Ecol. Resour* 14, 976–987. doi: 10.1111/1755-0998.12247

Politikos, D. V., Petasis, G., Chatzispyrou, A., Mytilineou, C., and Anastasopoulou, A. (2021). Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fisheries Res.* 242, 106033. doi: 10.1016/j.fishres.2021.106033

Pryseley, A., Mintiens, K., Knapen, K., van der Stede, Y., and Molenberghs, G. (2010). Estimating precision, repeatability, and reproducibility from Gaussian and non- Gaussian data: a mixed models approach. *J. Appl. Stat* 37, 1729–1747. doi: 10.1080/02664760903150706

Ramírez-Amaro, S., Bassitta, M., Picornell, A., Ramon, C., and Terrasa, B. (2022). Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1004674

Rasmussen, K. D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* 30, 733–750. doi: 10.1101/gad.276568.115

Ribas, L., Crespo, B., Sánchez-Baizán, N., Xavier, D., Kuhl, H., Rodríguez, J. M., et al. (2019). Characterization of the European sea bass (*Dicentrarchus labrax*) gonadal transcriptome during sexual development. *Mar. Biotechnol.* 21, 359–373. doi: 10.1007/s10126-019-09886-x

Ribas, L., Vanezis, K., Imués, M. A., and Piferrer, F. (2017). Treatment with a DNA methyltransferase inhibitor feminizes zebrafish and induces long-term expression changes in the gonads. *Epigenet. Chromatin* 10, 59. doi: 10.1186/s13072-017-0168-7

Robeck, T. R., Fei, Z., Lu, A. T., Haghani, A., Jourdain, E., Zoller, J. A., et al. (2021). Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun. Biol.* 4, 642. doi: 10.1038/s42003-021-02179-x

Rodríguez-Rodríguez, G., Ballesteros, H. M., Sánchez-Llamas, E., Bande, R., and Otero, R. F. (2022). The state of the art in cost-benefit of HTS methods for stock assessment: An overview. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1005534

Schulte, P. M., Healy, T. M., and Fangue, N. A. (2011). Thermal performance curves, phenotypic plasticity, and the time scales of temperature exposure. *Integr. Comp. Biol.* 51, 691–702. doi: 10.1093/icb/icr097

Shimoda, N, Izawa, T, Yoshizawa, A, Yokoi, H, Kikuchi, Y, Hashimoto, N., et al (2014). Decrease in cytosine methylation at CpG island shores and increase in DNA fragmentation during zebrafish aging. *Age (Dodr)* 36 (1), 103–15. doi: 10.1007/s11357-013-9548-5

Simpson, D. J., and Chandra, T. (2021). Epigenetic age prediction. *Aging Cell* 20. doi: 10.1111/acel.13452

Smith, C. (2022). DNA Microarrays: A trusted tool keeps evolving. *Biocompare. The buyer's guide for Life scientists*. Available at: https://www.biocompare.com/Editorial-Articles/172195-DNA-Microarrays-A-Trusted-Tool-Keeps-Evolving/.

Sutton, C., Maolagáin, C.Ó., and McKenzie, J. (2023). *Determining the age of fish. National Institute of Water and Atmospheric Research (NIWA)*. Available at: https://niwa.co.nz/fisheries/research-projects/determining-the-age-of-fish.

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., and Paun, O. (2016). Bs RAD seq: screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* 25, 1697–1713. doi: 10.1111/mec.13550

van Gelderen, T. A., Montfort, J., Alvarez-Dios, J. A., Thermes, V., Piferrer, F., Bobe, J., et al. (2022). Deciphering sex-specific miRNAs as heat-recorders in zebrafish. *Sci. Rep.* 12, 18722. doi: 10.1038/s41598-022-21864-3

Venney, C. J., Johansson, M. L., and Heath, D. D. (2016). Inbreeding effects on gene-specific DNA methylation among tissues of Chinook salmon. *Mol Ecol.* 25(18), 4521–33. doi: 10.1111/mec.13777

Volff, J.-N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity* 94, 280–294. doi: 10.1038/sj.hdy.6800635

Weber, D. N., Fields, A. T., Patterson, W. F., Barnett, B. K., Hollenbeck, C. M., and Portnoy, D. S. (2022). Novel epigenetic age estimation in wild-caught gulf of Mexico reef fishes. *Can. J. Fish. Aquat. Sci.* 79, 1–5. doi: 10.1139/cjfas-2021-0240

Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., et al. (2018). Application of a novel molecular method to age free-living wild bechstein's bats. *Mol. Ecol. Resour* 18, 1374–1380. doi: 10.1111/1755-0998.12925

Zannas, A. S., Arloth, J., Carrillo-Roa, T., Iurato, S., Röh, S., Ressler, K. J., et al. (2015). Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol.* 16, 266. doi: 10.1186/s13059-015-0828-5

Zhang, Q., Vallerga, C. L., Walker, R. M., Lin, T., Henders, A. K., Montgomery, G. W., et al. (2019). Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* 11, 54. doi: 10.1186/s13073-019-0667-1

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 13 February 2023
DOI 10.3389/fmars.2023.1096909

OPEN ACCESS

EDITED BY
Montse Pérez,
Spanish Institute of Oceanography (IEO), Spain

REVIEWED BY
Waldir M. Berbel-Filho,
Swansea University, United Kingdom
Deiene Rodriguez Barreto,
University of La Laguna, Spain

*CORRESPONDENCE
Dafni Anastasiadi
dafanast@gmail.com

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 13 November 2022
ACCEPTED 23 January 2023
PUBLISHED 13 February 2023

CITATION
Anastasiadi D and Piferrer F (2023) Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology.
Front. Mar. Sci. 10:1096909.
doi: 10.3389/fmars.2023.1096909

Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology

Dafni Anastasiadi[1]* and Francesc Piferrer[2]

[1]Seafood Technologies, The New Zealand Institute for Plant and Food Research, Nelson, New Zealand, [2]Institut de Ciències del Mar, Spanish National Research Council (CSIC), Barcelona, Spain

Epigenetic clocks are accurate tools for age prediction and are of great interest for fisheries management and conservation biology. Here, we review the necessary computational steps and tools in order to build an epigenetic clock in any species focusing on fish. Currently, a bisulfite conversion method which allows the distinction of methylated and unmethylated cytosines is the recommended method to be performed at single nucleotide resolution. Typically, reduced representation bisulfite sequencing methods provide enough coverage of CpGs to select from for age prediction while the exact implemented method depends on the specific objectives and cost of the study. Sequenced reads are controlled for their quality, aligned to either a reference or a deduced genome and methylation levels of CpGs are extracted. Methylation values are obtained in biological samples of fish that cover the widest age range possible. Using these datasets, machine learning statistical procedures and, in particular, penalized regressions, are applied in order to identify a set of CpGs the methylation of which in combination is enough to accurately predict age. Training and test datasets are used to build the optimal model or "epigenetic clock", which can then be used to predict age in independent samples. Once a set of CpGs is robustly identified to predict age in a given species, DNA methylation in only a small number of CpGs is necessary, thus, sequencing efforts including data and money resources can be adjusted to interrogate a small number of CpGs in a high number of samples. Implementation of this molecular resource in routine evaluations of fish population structure is expected to increase in the years to come due to high accuracy, robustness and decreasing costs of sequencing. In the context of overexploited fish stocks, as well as endangered fish species, accurate age prediction with easy-to-use tools is much needed for improved fish populations management and conservation.

KEYWORDS

age estimation, epigenetic clock, fisheries management, conservation biology, DNA methylation, machine learning, penalized regressions, bioinformatics

1 Introduction

Epigenetics can be defined as "the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to DNA sequence" (Deans and Maggert, 2015). Epigenetics has emerged as a powerful discipline in the study of the integration of genomic and environmental information, both intrinsic and extrinsic factors, to bring about a specific phenotype (Turner, 2009; Vogt, 2017). There are three major epigenetic molecular mechanisms widely accepted as such: 1) DNA methylation, 2) the modifications of histones and histone variants, and 3) the abundance and distribution of regulatory non-coding RNA (for review, see Carlberg and Molnár (2014)). One of the best studied epigenetic mechanisms is DNA methylation. Methylation can occur in two of the four nucleotides of DNA, cytosine and adenine. The former is the process by which a methyl-group (CH3) is transferred from a methyl donor, S-adenosyl-L-methionine (SAM), to the fifth position of a cytosine, converting it to 5-methylcytosine (5mC) or to the sixth position of an adenine converting it to N6-methyladenine (Ratel et al., 2006; Grosjean, 2013; Pfeifer, 2016). 5mCs are the most abundant modifications, are present in most species and therefore the most studied.

According to the Biomarkers Definitions Working Group, a biomarker is defined as "a characteristic that is objectively measured and evaluated as indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention" (Biomarkers Definitions Working Group, 2001). Biomarkers have been developed for a variety of purposes, including medicine and environmental assessment (Liu et al., 2019). Epigenetic modifications have been suggested recently as good candidates for biomarkers because they can be stable, frequent, abundant and accessible (Costa-Pinheiro et al., 2015). Details on the development of epigenetic biomarkers in aquatic organisms can be found elsewhere (Anastasiadi and Beemelmanns, 2023). An epigenetic clock is a set of biomarkers used to predict age, or in other words a "highly accurate age estimator based on CpG DNA methylation levels". In the last years they have been developed for about half a dozen fish species and it is expected that in the years to come epigenetic clocks will be of common use for both fisheries management and conservation biology. To the best of our knowledge, epigenetic clocks have been developed for: European sea bass (Anastasiadi and Piferrer, 2020), zebrafish (Mayne et al., 2020), Australian lungfish (Mayne et al., 2021b), Mary river cod (Mayne et al., 2021b), Murray cod (Mayne et al., 2021b), medaka (Bertucci et al., 2021), northern red snapper (Weber et al., 2021) and red grouper (Weber et al., 2021). For details on piscine epigenetic clocks including accuracy, techniques, CpGs covered and biological aspects to consider for new clocks please see Piferrer and Anastasiadi (2023). However, a crucial aspect for epigenetic clock development is how DNA methylation data is actually used to build the age predictor. This is of importance because a proper model building is essential to take out the most of the capabilities that epigenetic clocks may offer. There are several reviews that cover the factors causing, modulating and accelerating epigenetic clocks, mainly focusing on humans and mammals (Field et al., 2018; Guevara and Lawler, 2018; Bell et al., 2019; Simpson and Chandra, 2021). However, to the best of our knowledge, there are no reviews on the necessary computational steps and tools in order to build an epigenetic clock in any species, while these steps will be essentially the same. The issues dealt with below will thus be very helpful not only to fisheries managers and conservation biologists but to scientists that want to develop epigenetic clocks for new species.

1.1 Methods to analyze DNA methylation

The methods used to analyze DNA methylation can be categorized at three broad levels [Table 1 (Anastasiadi, 2016; Barros-Silva et al., 2018; Ortega-Recalde and Hore, 2023]. These three levels are based on how methylated loci are identified (level 1), at what resolution they are identified (level 2) and what portion of the genome is interrogated (level 3). For epigenetic clocks

TABLE 1 Overview of methodologies for the analysis of DNA methylation (updated from (Anastasiadi, 2016).

		Resolution				
		Low	Medium	High		
^{5m}Cs distinction	**Restriction enzymes**			*Hpa*II-PCR	**Locus-specific**	**Genome coverage**
		MSAP, RLGS	CHARM	RRBS, MRE-seq	**Genome-wide**	
				EM-seq	**Whole genome**	
	Affinity enrichment		MeDIP/MBD-PCR		**Locus-specific**	
			MeDIP/MBD-chip		**Genome-wide**	
		HPLC	MeDIP/MBD-seq		**Whole genome**	
	Bisulfite treatment		Infinium BeadChip	MBS, MSP, Bis-PCR, Sanger BS	**Locus-specific**	
				RRBS, bis-RAD-seq, epi-GBS	**Genome-wide**	
				WGBS	**Whole genome**	

MSAP, Methylation Sensitive Amplified Polymorphism; RLGS, Restriction Landmark Genomic Scanning; HPLC, High Performance Liquid Chromatography; CHARM, Comprehensive High-throughput Arrays for Relative Methylation; MeDIP, Methylated DNA ImmunoPrecipitation; MBD, Methyl-CpG-Binding Domain; RRBS, Reduced Representation Bisulfite Sequencing; MRE, Methyl-sensitive Restriction Enzyme; EM-seq, Enzymatic Methyl-seq; MBS, Multiplex Bisulfite Sequencing; MSP, Methylation Specific PCR; BS, Bisulfite Sequencing; bis-RAD, Bisulfite Restriction site Associated DNA; WGBS, Whole Genome Bisulfite Sequencing.

construction, methods that make use of bisulfite (level 1) at single nucleotide resolution (level 2) are used. However, the portion of the genome to be interrogated depends on the resources and available knowledge on the species of target or closely related species. Importantly, advances in sequencing using Oxford Nanopore Technologies MinION render this a powerful alternative to other methods. Thus, direct detection at single nucleotide resolution of 5mCs using portable devices is possible without the need of bisulfite conversion. This technology has been used recently to construct an epigenetic clock in cattle (Hayes et al., 2021).

1.1.1 Level 1. How are methylated loci identified?

5mCs must be identified and separated from the unmethylated ones (Cs). The processes of distinction between the two types of cytosines can be further divided into three general sub-levels, detailed below (Table 1 for an overview), that are not mutually exclusive and that in some cases are used in combination (Rauluseviciute et al., 2019):

1) Restriction enzymes. There are restriction enzymes which function differently when they encounter 5mCs and Cs. This property can be used to distinguish between the two types of Cs and ultimately identify their methylation status. Common isoschizomers, like *Msp*I and *Hpa*II, are used. For instance, these enzymes recognize the same sequence pattern (5'-CCGG-3'), however, *Msp*I cuts at those sites where the internal C is methylated in the two complementary DNA strands, while *Hpa*II is functional in those with methylation of the external C in one or both of the complementary DNA strands. The Methylation Sensitive Amplified Polymorphism (MSAP) (Reyna-Lopez et al., 1997; Xu et al., 2000) and the Restriction Landmark Genomic Scanning (RLGS (Hatada et al., 1991); are examples of approaches using methylation-sensitive restriction enzyme (Table 1).
2) Antibodies. This approach is based on the use of antibodies that show specificity against 5mC or of recombinant proteins which have been developed to contain a methyl-CpG binding domain (MBD; e.g (Aberg et al., 2012). These processes end up enriching the fraction of chromatin that is methylated. Methylated DNA ImmunoPrecipitation (MeDIP) (Jacinto et al., 2008) and Methyl-CpG-Binding Domain (MBD) (Jacinto et al., 2008; Nair et al., 2011) are examples of affinity-based approaches (Table 1) with MeDIP using a monoclonal antibody specific for 5mCs and MBD-based strategies using methyl-CpG binding domain-based proteins (MBDCap) (Nair et al., 2011).
3) Bisulfite. The treatment of DNA with bisulfite involves a chemical reaction that converts unmethylated Cs into uracils in 3 steps. Methylated 5mCs also react with bisulfite but this reaction is extremely slow and 5mCs are favoured by the equilibrium. Thus, 5mCs essentially escape conversion and remain intact (Clark et al., 1994). This reaction functions, therefore, as a recorder of the original methylation status and downstream steps allow to register and recall it. Several techniques, ranging from locus-specific to whole-genome, take advantage of the bisulfite properties in order to analyze the DNA methylation levels, like the Methylation-specific PCR (MSP) or the Whole Genome Bisulfite Sequencing (WGBS; Table 1). Bisulfite conversion of DNA is considered the "gold standard" in DNA methylation analysis because it allows the identification of the methylation status of each interrogated cytosine. However, limitations exist for bisulfite conversion methods as well. Methylation of a cytosine is a binary state (methylated or not methylated) in a given cell at a given time. Bisulfite sequencing reflects the relative proportion of Us/Cs at a given position, when sequencing tissues due to cell heterogeneity, and not the binary state of a specific cytosine unless single cell sequencing is performed. Methods based on bisulfite treatment of DNA are used for epigenetic clock construction.

1.1.2 Level 2. What is the resolution used?

The methylation profiling methods can have variable resolution, where higher resolution means information retrieval at the level of nucleotide and lower resolution means information retrieval at a larger genomic scale. In Table 1, an overview of the different methodologies for the analysis of DNA methylation with their corresponding resolution is provided. The resolution can broadly be grouped into the following three categories:

1) Low resolution. These techniques typically allow to obtain information on the global 5mC content. This is useful in order to conclude whether there are overall differences in the global methylation content or not, e.g., between control and treatment or disease group. Nevertheless, where exactly in the genome these differences occur remains unknown.
2) Medium resolution. Here, apart from global differences, an approximate location of the 5mCs is obtained. This is the case, for instance, of MeDIP-seq, where the methylated fraction of the immunoprecipitated DNA is sequenced and the differences can be located within a region that corresponds to the length of the sequenced fragment.
3) Single nucleotide resolution. In this case, the precise location of both 5mCs and Cs is obtained. This means that the exact position of 5mCs and Cs can be mapped to genomic coordinates that include 3 numbers: chromosome, start position, end position. For example, one obtains the information that in chromosome 1, start position=253, end position=254, there is a 5mC. Single nucleotide resolution is needed to construct an epigenetic clock.

1.1.3 Level 3. Which part of the genome is targeted?

The part of the genome that is investigated following the separation of Cs is also variable. In Table 1 an overview of the different methodologies for the analysis of DNA methylation with their relative CpG/genome coverage is provided. They can be broadly grouped into three categories according to this criterion as well:

1) Locus-specific. The amount of 5mCs and Cs is measured within target regions of interest typically spanning 10^{-1}–10^{2} CpGs. The target region of interest can be a specific gene, regulatory region of a gene, genomic regions within a gene such as exons, introns or 5'UTRs, or any other genomic region that is *a priori* interesting and therefore can be a target for the analysis of its DNA methylation.
2) Genome-wide. The amount of 5mCs and Cs is measured within a part of the genome that is considered representative of the overall genome. The part of the genome is in the order of 10^{5}–10^{6} CpGs and is representative because usually it is enriched for sites that can be methylated. For example, after digestion with enzymes that specifically recognize sites that include CpGs.
3) Whole-genome. The amount of 5mCs and Cs is measured across the whole genome covering more than 10^{6} CpGs. The entire genome is interrogated for its methylation levels, there is no reduction for specific regions or representative parts, but rather information on every single basis is obtained.

2 DNA methylation analysis using bisulfite sequencing

In the last years, high throughput sequencing (HTS) approaches have been used extensively to analyze the DNA methylation patterns in many different situations. The technique that combines the best possible way to distinguish 5mCs, single nucleotide resolution and whole genome coverage is Whole Genome Bisulfite Sequencing (WGBS), which is a HTS-based approach that uses bisulfite conversion to allow the distinction between 5mCs and Cs and interrogates the whole genome at single nucleotide resolution (Bock, 2012). Other HTS-based approaches that use bisulfite conversion, but analyze only a part of the genome are: Reduced Representation Bisulfite Sequencing (RRBS) (Gu et al., 2011; Klughammer et al., 2015), Bisulfite-converted Restriction site Associated DNA sequencing (bis-RAD-seq) (Trucchi et al., 2016), epiRADseq (Schield et al., 2016) and epi-GBS (van Gurp et al., 2016; Gawehns et al., 2022). Furthermore, targeted approaches such as BisPCR^{2} (Bernstein et al., 2015), Multiplex Bisulfite Sequencing (MBS) (Anastasiadi et al., 2018) or others (Masser et al., 2013; Korbie et al., 2015; Roeh et al., 2018) are also HTS-based techniques that make use of bisulfite conversion but for a targeted part of the genome. Oxford Nanopore Technology sequencing is expected to vary for the basic bioinformatics steps, however, the statistical analysis including machine learning model building will be essentially the same (section 3). The HTS methods used for epigenetic clocks in fish species until now include RRBS, bis-RAD-seq, MBS and BisPCR^{2} (see Table 1, Piferrer and Anastasiadi, 2023).

Different epigenetic clocks can be developed using different CpGs across the genome in different combinations, depending on the original dataset and the machine learning model. Around 20% of the Illumina 450K CpGs (90000 CpGs) can be used for epigenetic clocks (Porter et al., 2021). Taking this information into account, WGBS (>10^{6} CpGs) or RRBS (10^{3}-10^{6} CpGs) may produce a large amount of unnecessary data and workload, while bis-RAD-seq or epiRADseq (10^{3}-10^{5} CpGs) are expected to produce a good balance of informative CpGs without excess. Targeted approaches (e.g., MBS 10^{-1}-10^{2} CpGs) will be more relevant when prior information is available.

2.1 Bioinformatics workflow for bisulfite sequencing

All HTS-derived data produced from methods that use bisulfite conversion share some common characteristics. A summary workflow allows to distinguish the steps of quality controls, filtering/trimming, alignment/mapping, methylation extraction and analysis (Figure 1).

Processing of HTS-derived data always initiates with the appropriate quality controls of the raw sequencing data obtained followed by filtering of the data that fall below the specified thresholds. Adapters or indices have usually been added to the DNA fragments during the preparation of the libraries and are used to demultiplex the samples if needed. Their sequences also usually need to be removed from the data (trimming) otherwise they might influence the downstream steps of the workflow. Once adapters have been trimmed and low quality reads filtered out, the reads are aligned against the genome which, importantly, must have been previously bisulfite converted *in silico*. This is because bisulfite treatment converts the unmethylated cytosines of the genome into uracils, which are in turn converted into thymines (Ts) after amplification by PCR. This process results in many genomic sites in the sequenced reads that fail to map to the genome because the original sites have been lost and thus, they cannot match. Moreover, after PCR amplification, the complementary DNA strand contains adenines (As) instead of guanines (Gs) in the positions where the C was unmethylated and has been converted into T. Thus, the procedure through which the sequenced

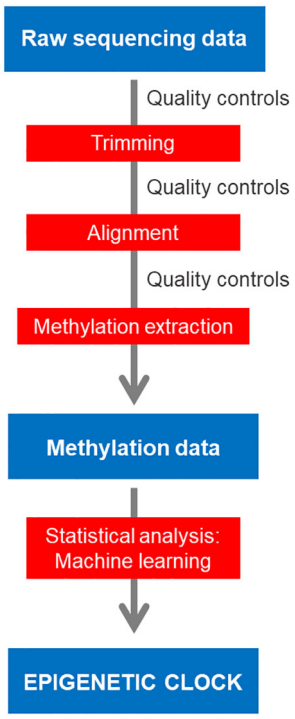


FIGURE 1
Workflow for bioinformatics analysis of bisulfite sequencing data for epigenetic clock construction.

reads are aligned to the genome needs to take into account these mismatches and considerations. Different tools have been developed to convert and map bisulfite sequencing data (See section 2.4.). A reference genome is not a pre-requisite for applying bisulfite sequencing since alternative bioinformatics procedures have been developed to assist the analysis, e.g., *ad hoc* genomes can be deduced from RRBS reads (Klughammer et al., 2015) or for bs-RAD-seq, a standard RAD-seq reduced representation genome can be used (Trucchi et al., 2016).

Once reads have been successfully mapped, 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. Usually, the final methylation of a given C is calculated according to the proportion of 5mCs and Cs found in that position: their sum equals to the coverage of a position and is the denominator in the equation where the numerator is the number of 5mCs. This value may be expressed as (5mCs/5mCs+Cs) and thus the methylation values will range from 0 to 1, or can be expressed as percentage, (5mCs/5mCs+Cs) multiplied by 100, and thus the methylation values will range from 0 to 100 (more details in section 2.5).

2.2 Quality controls

Modern sequencing platforms (e.g., Illumina) usually include the corresponding software which automatically performs the demultiplexing steps required prior to sample analysis and thus the corresponding set of files for each sample are obtained. In the case of single-end sequencing one file per sample is produced, while in the case of paired-end sequencing two files are produced per sample, each one of them refers to the forward and reverse read.

The standard format for these files is the FASTQ. FASTQ is a text-based format to store the sequences which includes more information than the older FASTA format which included only the sequence. In FASTQ, each read is unique and contains a sequence identifier and there is further information on the specific quality of the read. The quality of the reads is mainly measured by the Phred score which is a property logarithmically related to the base-calling error probabilities. A Phred score of 10, means that there is a 1 in 10 probability of incorrect base call and a 90% accuracy in base calls. A Phred score of 30 means that there is a 1 in 100 probability of incorrect base call and a 99.9% accuracy in base calls. Typically, Phred scores below 20, which equals to 99% accuracy of base call, are excluded from downstream analysis.

Quality controls are usually performed by a range of open source software packages, the most common of which is FASTQC (Andrews, 2010). In case several samples are to be evaluated at once, the MultiQC (Ewels et al., 2016) can be useful for simultaneous assessment of quality (see indicators below). These tools allow to assess the data quality *via* a variety of plots and statistics (Figures 2A–D), namely:

1) Sequence counts. The number of sequences counted for each sample.
2) Sequence quality histograms. The mean Phred score across each base position in the read.
3) Per sequence quality scores. The total number of reads plotted against the average Phred quality scores over the full read.
4) Per base sequence content. The percent of bases called for each of the four nucleotides (e.g., 30% A, 40% T, 20% C, 10% G) at each position (e.g, position 1-150 for a 150 bp read sequencing) across all reads.
5) Per sequence GC content. The number of reads plotted against the GC% per read.
6) Per base N content. The percent of bases at each position of the read for which no base could be called and are therefore coded as "N".
7) Sequence length distribution. The distribution of lengths across all reads.
8) Sequence duplication levels. The percent of reads of a specific sequence that are present repeatedly inside the file and can be an indicator of PCR duplication.
9) Overrepresented sequences. Sequences that appear more times than expected.
10) Adapter content. Cumulative plot of the fraction of reads where the adapters used for library construction are identified.

Sequence quality scores below 30 are nowadays considered unacceptable. However, interpretation of the rest of the metrics will be specific to the technique and sequencing platform used, since high duplication levels are inherent to enrichment (e.g., RRBS) or targeted techniques, but may indicate a problem with WGBS data.

On the other hand, the simultaneous visualization of multiple quality controls (QC) can be obtained by the MultiQC software (Figure 2E). A drop of quality below the available threshold at the end of the read is expected in general and for long reads (300 bp) in particular.

Example code of running FASTQC in all available fq files:

```
fastqc –nogroup -q -t 2 -o output_fastq_raw *.fq.gz
```

Example code of running MultiQC in the output of FASTQC:

```
multiqc output_fastqc_raw -i Fastqc-Raw
```

2.3 Trimming

Several open source packages are available for trimming (Table 2). Trimming of low quality reads can significantly improve the quality of the data to process and all downstream workflow, as a minimum in terms of Phred scores Low quality Phred scores (<20) would be associated with too high probabilities of erroneously called bases, one nucleotide, e.g,. A, instead of another, e.g., C. Thus, they cannot be accepted in a HTS experiment. Quality controls are performed again after the trimming procedure as well and are useful to visualize the improvements.

Example code of running Trimmomatic on WGBS data:

```
java -jar -Xms8G -Xmx8G
/software/Trimmomatic-0.36/trimmomatic.jar PE -threads 3
/raw_data/sample1_1.fq.gz
/raw_data/sample1_2.fq.gz
/trimmed_data/sample1_trimmed_R1.fastq
/trimmed_data/sample1_trimmed_R2.fastq
ILLUMINACLIP:/trimmed_data/adapters.fasta:2:30:10
SLIDINGWINDOW:5:20 MINLEN:50 HEADCROP:10
LEADING:5 TRAILING:5
```

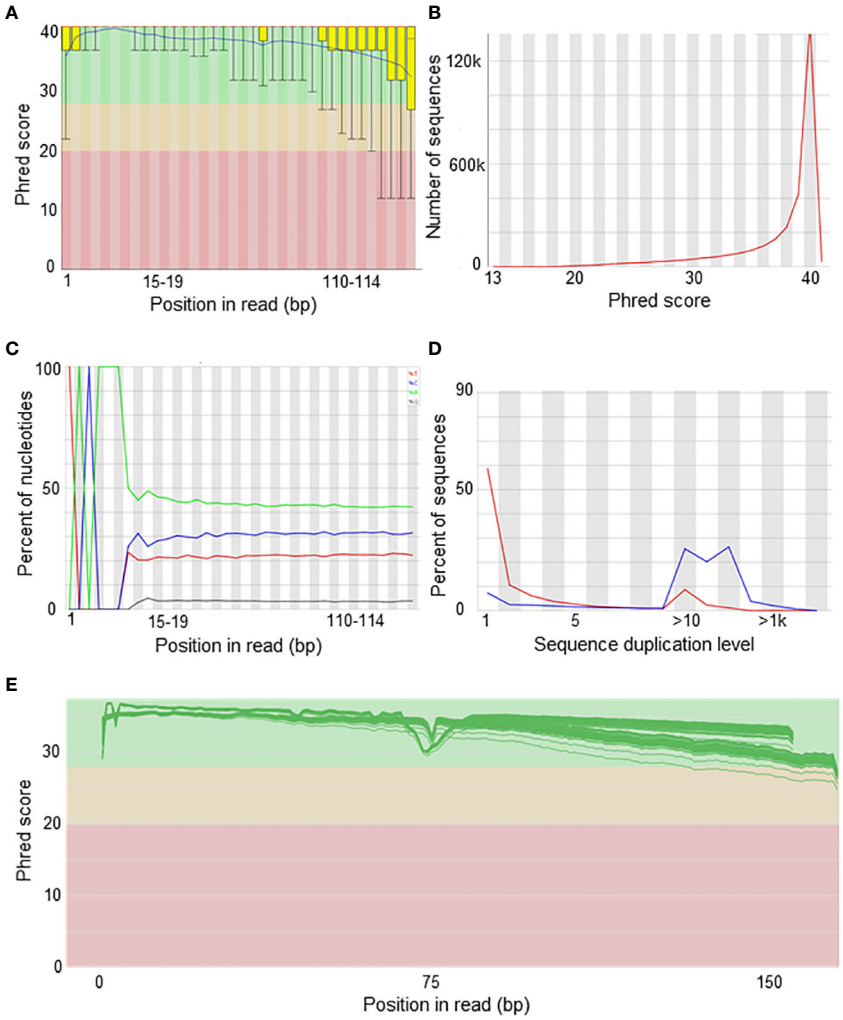


FIGURE 2
Quality control of sequencing data. Examples of plots from bis-RAD-seq experiment (own data). **(A)** Per base sequence quality shows the distribution of Phred quality of the bases (y-axis) along the length of the reads from base 0 to 150 (x-axis). **(B)** Per sequence quality scores shows the mean sequence quality as assessed by the Phred score (x-axis) in the number of overall sequences (y-axis). **(C)** Per base sequence content show the percentage of the four bases (T in red, C in blue, A in green and G in brown) along the length of the read from position 0 to 150 (x-axis). **(D)** The sequence duplication levels show the percent of sequences and their corresponding duplication levels (x-axis). **(E)** Simultaneous visualization of per base sequence quality from multiple samples by MultiQC software. The distribution of Phred quality of the bases (y-axis) along the length of the reads from base 0 to 150 (x-axis) is shown and green lines represent multiple samples.

TABLE 2 Trimming software.

Software package	Website link	Reference
Trim Galore!	https://github.com/FelixKrueger/TrimGalore	
Trimmomatic	https://github.com/timflutre/trimmomatic	(Bolger et al., 2014)
Cutadapt	https://github.com/marcelm/cutadapt/	(Martin, 2011)
NGS QC Toolkit	http://www.nipgr.res.in/ngsqctoolkit.html	(Patel and Jain, 2012)
ngsShoRT	http://research.bioinformatics.udel.edu/genomics/ngsShoRT/	(Chen et al., 2014)
UrQt	https://lbbe.univ-lyon1.fr/-UrQt-.html	(Modolo and Lerat, 2015)
Flexbar	https://github.com/seqan/flexbar	(Dodt et al., 2012)

2.4 Alignment

Bisulfite conversion depletes the genome of unmethylated cytosines which represents a challenge for the normal alignment procedure of reads to a large reference genome. Softwares developed for standard alignment procedures are not adequate in this case due to the conversion effect (Laird, 2010). This challenge has been circumvented by two different algorithms:

1) Wild card aligners. In this case, the Cs of the genome are replaced by Y which is the wild-card letter that is able to match both Cs and Ts, equivalent to Cs and 5mCs in the original molecule. Otherwise, these aligners modify the alignment score matrix in a manner that allows mismatch between Cs in the original molecule and Ts in the sequence of the read. Examples of wild card aligners include BSMAP and RRBSMAP (Xi and Li, 2009; Xi et al., 2012).
2) Three-letter aligners. In this case, all the Cs are converted into Ts in both the reads to be aligned and in the genomic sequence. The alignment is simplified and carried out using only three-letters of the nucleotide alphabet with C excluded. In this case, any standard aligner can be used at the lower level of the package, such as Bowtie or Bowtie2 (Langmead et al., 2009). Examples of three-letter aligners include Bismark, bwa-meth and BS-Seeker (Chen et al., 2010; Krueger and Andrews, 2011; Krueger et al., 2012, Pedersen et al., 2014).

Example code using bwa-meth:
Index reference genome
python/software/bwa-meth/bwameth.py index/genome/species-genome.fasta
Align reads to reference genome
python
/software/bwa-meth/bwameth.py–threads 16
–reference/genome/species-genome.fasta
/trimmed_data/sample1_trimmed_R1.fastq
/trimmed_data/sample1_trimmed_R2.fastq
| samtools view -Sb -q 10 - >/alignments/sample1.bam
Example code using Bismark:
bismark
/reference/genome/
-1 sample1_trimmed_R1.fastq
-2 sample1_trimmed_R2.fastq
–non_directional –un -o alignments
bismark
/software/bismark/Genome/
-1 sample1_trimmed_R1. fastq
-2 sample1_trimmed_R2.fastq
–non_directional –un -o alignments

Wild card aligners typically result in higher genomic coverage, but also in the introduction of bias towards higher DNA methylation as compared to three-letter aligners. This is relevant mainly in parts of the genome such as repetitive sequences. When selecting an aligner, considerations such as speed, computer memory and program use are more important (Bock, 2012). A recent comprehensive comparison of the most commonly used aligners should be consulted before executing this step (Nunn et al., 2021). In any case, mapping of the reads to the genome needs to be precise because otherwise it would result in biased DNA methylation levels calculated on the basis of methylated and unmethylated reads (Bock et al., 2010).

2.5 Methylation extraction

The methylation state of each C is extracted according to the alignments. Cytosines from the aligned sequences are transcribed into a table format where each row corresponds to a cytosine and its genomic position according the chromosome and position, methylation state and strand. Coding of this information within the table depends on the software used. For example, the Bismark (Krueger and Andrews, 2011) primary alignment output codes cytosines depending on the context, as z in CpG context, x in CHG context and h in CHH context. Methylation status is coded as uppercase (Z, X, H) for methylated and lowercase (z, x, h) for unmethylated. This information is transcribed into + for methylated and - for unmethylated cytosines in the methylation extraction file.

Example code using MethylDackel for use with methylKit:
MethylDackel extract –OT 0,0,0,145 –OB 3,0,6,0 –methylKit
-o/methylation_extraction/sample1.methylKit
/genome/species-genome.fasta
sample1-aligned.bam
Example code using Bismark:
bismark_methylation_extractor sample1_aligned.bam -p –merge_non_CpG -o extraction –bedGraph –cutoff 1

2.6 Bisulfite conversion rate

Evaluation of bisulfite conversion efficiency is an important step of the whole procedure because if this fails, then conclusions on the methylation status of the cytosines are erroneous. Spike-in sequences of known methylation status may have been introduced during library preparation to assist with bisulfite conversion rate estimation. If not, bisulfite conversion rate can be estimated *in silico*. Tools like the 'bsrate' script of the MethPipe pipeline (Song et al., 2013) allow for an automatic calculation of the bisulfite conversion ratio. Otherwise, one can make use of the percent of Cs methylated in a CHH context where C is cytosine and H can be any nucleotide except of Gs. In this case, the percent of these Cs is subtracted from 100 and the result is the bisulfite conversion rate. In current DNA methylation analysis procedures, bisulfite conversion rates should be as high as possible. Typical good values are >99%.

3 Statistical analysis

3.1 Objective

The objective of this step is to identify CpGs the methylation levels of which allow to predict the age of an individual. The methylation of these CpGs may be decreasing or increasing with

age with different slopes. The methylation of each CpG will be given a specific weight (coefficient) and their combination will be sufficient to predict age. These coefficients were shown to differ in the same CpGs between broad age groups in mammals, including humans (younger, middle-aged and older). Therefore, the extreme age groups should be considered with caution (Field et al., 2018) but nevertheless included for the development of the clock. The statistical analysis includes a typical machine learning model building (Figure 3). Building of machine learning models for age prediction follows the same principles as for any biological feature predicted from epigenetic biomarkers (Anastasiadi and Beemelmanns, 2023). The outcome variable is quantitative (age) and thus we deal with a regression problem aiming to predict the outcome variable on the basis of the independent variable(s) by means of a fitting curve explaining the input. When running a regression trying to predict a quantitative value (i.e. age) with many predictors (CpGs) results tend to overfit, reducing the predictive value. Penalized regression circumvent this problem by shrinking values of the predictors, being the recommended for age prediction based on CpG methylation (See Section 3.3).

3.2 Data structure

The dataset consists of:

1) Biological samples that cover a defined age range. The total number of samples should ensure covering the full age range of the species considered, and may vary between species in the extremes of lifespan. In the published literature of fish epigenetic clocks, the total number of samples range between 10 in Northern red snapper and red grouper (Weber et al., 2021) and 141 in Australian lungfish (Mayne et al., 2021b) with mean 46 samples (Piferrer and Anastasiadi, 2023). These numbers maybe suboptimal, since the minimum sample size according to simulations using human and zebrafish (*Danio rerio*) data is 70 (Mayne et al., 2021a). If feasible, 134 samples should be ideally included according to the same simulations, a recommendation for all new piscine epigenetic clocks (Mayne et al., 2021a). In order to build a prediction model, these will have to be divided into training and test sets. The training set is used for the machine to learn, to fit the parameters of the model. The test set is an independent set of data which the model built predicts and thus serves as an evaluation dataset of the model fit. Usually, the original dataset is split in 70-80% of the observations into training and 20-30% of the observations into the test set, using random procedures.

2) The methylation levels of target CpGs. Depending on the technique used, the number of CpGs will be in the order of hundreds (e.g. MBS), thousands (e.g. bis-RAD-seq), hundreds of thousands (e.g. RRBS) or millions (e.g. WGBS). Since many epigenetic clocks across a genome are possible (Porter et al., 2021) and extremely accurate epigenetic clocks with only 3 carefully selected CpGs have been constructed in mice (Han et al., 2018), the number of CpGs analyzed are not expected to affect the overall accuracy. However, each epigenetic clock or model will be unique as will be the coefficients attributed to each CpG of the clock. This type of data is not independent, since the methylation of one CpG may depend on the methylation of its neighboring CpG and are characterized by strong multicollinearity, where a large number of CpGs may be closely related to each other. Genome-wide patterns of DNA methylation in vertebrates are bimodal with a specific CpGs showing 0 or 100% methylation.

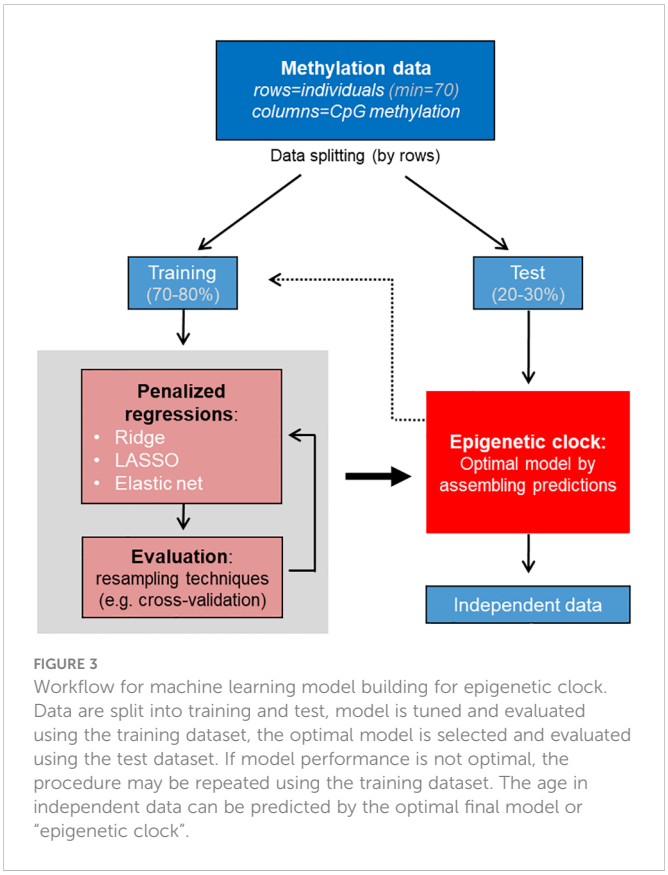


FIGURE 3
Workflow for machine learning model building for epigenetic clock. Data are split into training and test, model is tuned and evaluated using the training dataset, the optimal model is selected and evaluated using the test dataset. If model performance is not optimal, the procedure may be repeated using the training dataset. The age in independent data can be predicted by the optimal final model or "epigenetic clock".

3.3 Penalized regressions

In the development of epigenetic clocks we are dealing with a large multivariate dataset, where the number of variables (the different CpGs, at least the ones initially analyzed) is much, much higher than the number of samples (biological samples). Thus, the standard linear model is not suitable to use. A way to circumvent the structure of the dataset is to use penalized regressions. This approach was already implemented by Horvath (2013) when constructing the first epigenetic clock in humans. Penalized regressions allow to construct linear regression models that are penalized when they have too many variables (Kassambara, 2018). The penalization occurs *via* the addition of a constraint in the equation (Bruce and Bruce, 2017). This increases bias but, importantly, reduces variance. The methodology to achieve this is shrinkage or regularization, which results in the shrinkage of some coefficients values to zero. This allows

for exclusion of the variables (i.e., individual CpGs) that contribute less by shrinking their coefficient or in other words, to retain the minimum number of CpGs that are valuable for age prediction.

There are three most commonly used methods of penalized regression and typically they are all tested when constructing an epigenetic age prediction clock for a new species:

1) Ridge regression. The least contributing variables will have their coefficient very close to zero.
2) LASSO regression (Least Absolute Shrinkage and Selection Operator). The variables with the least contribution will be forced to be zero. This will produce models with reduced complexity as compared to ridge regression, where all variables are kept.
3) Elastic net regression. This type of penalized regression stands in between the previous two types, where some coefficients will be shrank, as in ridge regression, and some coefficients will be set to zero, as in LASSO.

There are advantages and disadvantages of each penalized regression type over the other that depend on the specific dataset. LASSO will perform better when there are few predictors with large coefficients and a lot of predictors with small coefficients, while ridge will perform better where there are a lot of predictors with similar coefficients. Ridge regression keeps all variables, therefore, is not recommended when genome-wide techniques have been used. In any case, parameters of the model have to be tuned and the model has to be selected by evaluating its performance, as explained below.

3.4 Machine learning model building

Penalized regressions are machine learning models and thus, to build them, a standard machine learning model building procedure should be followed (Figure 3). In aquatic organisms, machine learning methods for developing epigenetic biomarkers have been applied in limited cases, while the procedure has been recently reviewed in details (Anastasiadi and Beemelmanns, 2023).

Below we explain the typical workflow of the procedure that can be implemented in R using the specialized caret (*Cl*assification *A*nd *RE*gression *T*raining) package (Kuhn, 2008). Nevertheless, other packages or programming language (e.g., Python) can also be used to navigate the same workflow.

1) Data splitting. Data are split into at least 2 datasets that allow to later evaluate model performance. The training dataset contains 70-80% of the samples and is used to for algorithm training and parameter tuning. The test datasets contains the remaining 20-30% of samples and is used once the right model has been trained and selected to test whether the model can be generalized in unseen data. Ideally, training dataset is sufficiently large to be split further into training and validation dataset during model performance assessment. However, this is rarely the case and instead resampling techniques are used. With resampling, iterative splitting into training and validation datasets occurs and prediction errors of all splitting cycles are averaged at the end. K-fold cross-validation (CV) has been extensively used in fish epigenetic clock building. Data splitting can be performed using specific functions that randomly splits the dataset, while keeping track of the randomness by setting the seed to a specific number in R.

R code example:

```
library(tidymodels)
library(readr)
set.seed(123)
splits      <- initial_split(meth.age.df, strata = age)
age_other <- training(splits)
age_test  <- testing(splits)
Training set proportions by age class
age_other %>%
count(age) %>%
mutate(prop = n/sum(n))
Test set proportions by age class
age_test %>%
count(age) %>%
mutate(prop = n/sum(n))
```

2) Data preparation and pre-processing. This step may include a) exclusion of CpGs the methylation of which has zero or near-zero variance across ages in the training dataset; b) dealing with multicollinearity by identifying CpGs with correlated methylation –a common feature in this type of data–; c) data transformation of centering and scaling variables to mean 0 and standard deviation 1; d) imputation of missing values if necessary. Imputations can be performed by the mice (Multivariate Imputation by Chained Equations) package in R (van Buuren and Groothuis-Oudshoorn, 2011).

Correlation of CpG methylation with other biological parameters that we want to account for, such as diet, sex or other environmental factors, can be dealt with by exclusion of the correlated CpGs when lots of CpGs. This type of correlation is likely to be confounding factor in the model if these biological parameters are parallel to age (i.e., we have many samples of younger males and older females).

R code example:

a) Excluding features with zero or near-zero variance among groups

```
library(caret)
library(dplyr)
## Detect features and visualize them
nzv.cpg <- nearZeroVar(age_other, saveMetrics= TRUE, names=TRUE, freqCut = 85/15, uniqueCut = 50)
boxplot(nzv.cpg$percentUnique)
boxplot(nzv.cpg$freqRatio)
## Detect features, exclude them and save the object
nzv.cpg.list <- nearZeroVar(age_other, freqCut = 85/15, uniqueCut = 50) filteredDescr <- age_other[, -nzv.cpg.list] dim(filteredDescr)
```

b) Exclude highly correlated variables

```
highlyCorDescr <- findCorrelation(filteredDescr, cutoff = 0.8)
filteredDescr.cor <- filteredDescr[,-highlyCorDescr]
```

c) Transformation *via* preProcess data

```
preProcValues <- preProcess(filteredDescr.cor, method = c ("center", "scale"))
trainTransformed <- predict(preProcValues, filteredDescr.cor)
```

d) Imputation of missing values

Method 1 using package "mice" (Multiple Imputation by Chained Equation)

```
library(mice)
```

```
init = mice(meth.age.df, maxit=0)
meth = init$method
predM = init$predictorMatrix
colnames(meth.age.df)
predM[, c("age")]=0
meth[c("age")]=""
set.seed(100)
imputed = mice(meth.age.df, method=meth, predictorMatrix=predM, m=5)
```

Method 2 using package "zoo" (Missing values replaced by the mean or other function of its group)

```
library(zoo)
meth.age.df.na <- na.aggregate(meth.age.df)
```

3) Model tuning. The best tuning parameters for alpha and lambda of the penalized regression algorithm are selected. Alpha defines the type of regression with α=0 ridge, α=1 LASSO and $0<\alpha<1$ elastic net, while lambda defines the amount of shrinkage. Lambda will be automatically selected to minimize prediction error. Simultaneously feature selection, i.e., selection of the most informative CpGs, is performed.

4) Model evaluation is performed using resampling techniques, k-fold CV, repeated CV or leave-one-out CV (LOOCV). The error will be minimized after several repeated rounds of dataset splitting and finally, the optimal model is selected.

R code example using caret (steps 3-4):

Define resampling technique to be used. Here we choose repeated cross-validation

```
fitControl <- trainControl(method = 'repeatedcv', number=10, repeats=10)
```

Define range of lambda to be tested

```
lambda <- 10^seq(-3, 3, length = 100)
```

Run penalized regressions. Examples of ridge, LASSO and elastic net regressions are shown here.

Ridge regression. This regression may not be relevant in cases of RRBS or WGBS data since it keeps all CpGs available, but may be worth in cases of targeted methods (e.g., MBS).

```
set.seed(123)
ridge_model <- train(age ~., data = trainTransformed, method = "glmnet", trControl = fitControl, tuneGrid = data.frame(alpha = 0, lambda = 10^seq(-3, 3, length = 100)), tuneLength = 10)
```

LASSO

```
set.seed(123)
lasso_model <- train(age ~., data = trainTransformed, method = "glmnet", trControl = fitControl, tuneGrid = data.frame(alpha = 1, lambda = 10^seq(-3, 3, length = 100)), tuneLength = 10)
```

In Elastic net best tuning of both lambda and alpha will be automatically selected

```
set.seed(123)
elastic_model <- train(age ~., data = trainTransformed, method = "glmnet", trControl = fitControl, tuneLength = 10)
```

Elastic net with alpha set to 0.5 and best tuning of lambda will be automatically selected

```
set.seed(123)
elastic_model.05 <- train(age ~., data = trainTransformed, method = "glmnet", trControl = fitControl, tuneGrid = data.frame (alpha = 0.5, lambda = 10^seq(-3, 3, length = 100)), tuneLength = 10)
```

Compare metrics of the models

```
models_compare <- resamples(list(R=ridge_model, LM=lasso_model, EM=elastic_model, EM05=elastic_model.05))
summary(models_compare)
```

Count features (CpGs) kept by each model. An ideal piscine epigenetic clock with wide application would contain as few CpGs as possible without compromising accuracy and precision. Example using elastic net.

```
sum(coef(elastic_model$finalModel, elastic_model$bestTune $lambda)!=0)
```

Compare metrics in the training datasets

Ridge

```
predicted.age <- predict.train(ridge_model)
postResample(pred = predicted.age, trainTransformed$age)
cor.test(predicted.age, trainTransformed$age)
```

LASSO

```
predicted.age <- predict.train(lasso_model)
postResample(pred = predicted.age, trainTransformed$age)
cor.test(predicted.age, trainTransformed$age)
```

Elastic net

```
predicted.age <- predict.train(elastic_model)
postResample(pred = predicted.age, trainTransformed$age)
cor.test(predicted.age, trainTransformed$age)
```

5) Assembling predictions. The optimal model needs to be further evaluated using the test dataset in order to assess how well it can generalize. The final model will be then built using the optimal model run on the whole training dataset.

R code example: Compare metrics in the test dataset

Ridge

```
predict.ridge.test <- predict(ridge_model, testTransformed)
postResample(pred = predict.ridge.test, testTransformed$age)
cor.test(predict.ridge.test, testTransformed$age)
```

LASSO

```
predict.lasso.test <- predict(lasso_model, testTransformed)
postResample(pred = predict.lasso.test, testTransformed$age)
cor.test(predict.lasso.test testTransformed$age
```

Elastic net

```
predict.enet.test <- predict(elastic_model, testTransformed)
postResample(pred = predict.enet.test, testTransformed$age)
cor.test(predict.enet.test, testTransformed$age)
```

Build and evaluate the final model

```
finalmodelCtrl <- trainControl(method = "none")
set.seed(123)
final <- train(age ~., data = trainTransformed, method = "glmnet", trControl=finalmodelCtrl, tuneGrid = expand.grid(alpha = bestalpha, lambda = bestlambda))
predicted.final.train <- predict(final, trainTransformed)
cor.test(predicted.final.train, trainTransformed$age)
```

Evaluation of models during training as well as at the final model is done by assessing the predictive accuracy *via* loss functions

comparing predicted age vs actual age. The measures to take into account and report include:

a) Root Mean Squared Error (RMSE) = average deviation of the predictions from the observations.
b) Mean Absolute Error (MAE) = average of the absolute differences between the observed and predicted values.
c) R^2 = the squared correlation between the observed and predicted values. This value shows how well the selected variables (methylation of CpGs) explain the variability of the dependent variable (age).

The errors should be minimized while the R^2 should be maximized.

Epigenetic clocks are considered valid if the correlation (R) is higher than 0.80 in large independent data covering a broad age range (Horvath and Raj, 2018). Piscine epigenetic clocks show a mean correlation of 0.93 (Piferrer and Anastasiadi, 2023), while higher values are also possible. Precision reported as MAE is used with actual time units (days, months or years) and shows a mean of 0.87 years in piscine clocks, or an average of about 3.5% of the total lifespan (Piferrer and Anastasiadi, 2023).

4 Conclusions

Epigenetic clocks for age prediction are typically constructed using DNA methylation sequencing technologies that involve the use of bisulfite conversion and provide information at single nucleotide resolution. Bioinformatic analysis of the data follows mostly standard procedures of sequencing reads analysis, however, care should be taken to account for C to T conversion during the alignment step. Methylation values are extracted per base and this results in the dataset consisting of individual fish aged samples as rows and methylation values of interrogated CpGs as columns. This multivariate dataset is submitted to machine learning procedures aiming to select features, i.e., CpGs the methylation of which is enough to predict age. The machine learning procedures used are penalized (or regularized) regressions which fit well the structure of the multivariate dataset. At the end of the procedure, the optimal model or "epigenetic clock" is constructed. This constitutes a molecular resource to be implemented by scientists and managers for accurate age prediction of fish. The simultaneous interrogation of the methylation of a few target CpGs forming the epigenetic clock of a large amount of samples in a ready-to-use kit constitutes the ultimate goal for application of this HTS to fisheries and conservation.

Author contributions

DA: Conceived and wrote the manuscript. FP: Conceived and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Funded through EU Contract – EASME/EMFF/2017/1.3.2.10/SI2.790889 to FP.

Acknowledgments

We would like to thank Drs. Gary Carvalho, Sissel Jentoft, Laszlo Orban, Allan Tucker and two reviewers for constructive criticisms and suggestions for improvement to an earlier version of this manuscript.

Conflict of interest

DA is employed by The New Zealand Institute for Plant and Food Research Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The information and views set out in this publication are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

References

Aberg, K. A., McClay, J. L., Nerella, S., Xie, L. Y., Clark, S. L., Hudson, A. D., et al. (2012). MBD-seq as a cost-effective approach for methylome-wide association studies: Demonstration in 1500 case–control samples. *Epigenomics* 4, 605–621. doi: 10.2217/epi.12.59

Anastasiadi, D. (2016). Intrinsic and environmental influences on DNA methylation and gene expression in fish. In: *TDX (Tesis doctorals en xarxa)*. Available at: http://www.tdx.cat/handle/10803/511360 (Accessed April 8, 2020).

Anastasiadi, D., and Beemelmanns, A. (2023). Development of epigenetic biomarkeres in aquatic organisms in *Epigenetics in aquaculture*. Eds. F. Piferrer and H. Wang (Hoboken, NJ: John Wiley & Sons Ltd).

Anastasiadi, D., and Piferrer, F. (2020). A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. *Mol. Ecol. Resour* 20, 387–397. doi: 10.1111/1755-0998.13111

Anastasiadi, D., Vandeputte, M., Sánchez-Baizán, N., Allal, F., and Piferrer, F. (2018). Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics* 13, 988–1011. doi: 10.1080/15592294.2018.1529504

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Accessed January 29, 2023. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barros-Silva, D., Marques, C. J., Henrique, R., and Jerónimo, C. (2018). Profiling DNA methylation based on next-generation sequencing approaches: New insights and clinical applications. *Genes (Basel)* 9, 429. doi: 10.3390/genes9090429

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., et al. (2019). DNA Methylation aging clocks: Challenges and recommendations. *Genome Biol.* 20, 249. doi: 10.1186/s13059-019-1824-y

Bernstein, D. L., Kameswaran, V., Le Lay, J. E., Sheaffer, K. L., and Kaestner, K. H. (2015). The BisPCR2 method for targeted bisulfite sequencing. *Epigenet. Chromatin* 8, 27. doi: 10.1186/s13072-015-0020-x

Bertucci, E. M., Mason, M. W., Rhodes, O. E., and Parrott, B. B. (2021). Exposure to ionizing radiation disrupts normal epigenetic aging in Japanese medaka. *Aging* 13, 22752–22771. doi: 10.18632/aging.203624

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69, 89–95. doi: 10.1067/mcp.2001.113989

Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* 13, 705–719. doi: 10.1038/nrg3273

Bock, C., Tomazou, E. M., Brinkman, A. B., Müller, F., Simmer, F., Gu, H., et al. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* 28, 1106–1114. doi: 10.1038/nbt.1681

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bruce, P., and Bruce, A. (2017). *Practical statistics for data scientists: 50 essential concepts* (Sebastopol, CA: O'Reilly Media, Inc).

Carlberg, C., and Molnár, F. (2014). The epigenome in *Mechanisms of gene regulation* (Springer, Netherlands).

Chen, P. Y., Cokus, S. J., and Pellegrini, M. (2010). BS seeker: Precise mapping for bisulfite sequencing. *BMC Bioinf.* 11, 203. doi: 10.1186/1471-2105-11-203

Chen, C., Khaleel, S. S., Huang, H., and Wu, C. H. (2014). Software for pre-processing illumina next-generation sequencing short read sequences. *Source Code Biol. Med.* 9, 8. doi: 10.1186/1751-0473-9-8

Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22, 2990–2997. doi: 10.1093/nar/22.15.2990

Costa-Pinheiro, P., Montezuma, D., Henrique, R., and Jerónimo, C. (2015). Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics* 7, 1003–1015. doi: 10.2217/epi.15.56

Deans, C., and Maggert, K. A. (2015). What do you mean, "epigenetic"? *Genetics* 199, 887–896. doi: 10.1534/genetics.114.173492

Dodt, M., Roehr, J. T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. *Biol. (Basel)* 1, 895–905. doi: 10.3390/biology1030895

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354

Field, A. E., Robertson, N. A., Wang, T., Havas, A., Ideker, T., and Adams, P. D. (2018). DNA Methylation clocks in aging: Categories, causes, and consequences. *Mol. Cell* 71, 882–895. doi: 10.1016/j.molcel.2018.08.008

Gawehns, F., Postuma, M., van Antro, M., Nunn, A., Sepers, B., Fatma, S., et al. (2022). epiGBS2: Improvements and evaluation of highly multiplexed, epiGBS-based reduced representation bisulfite sequencing. *Mol. Ecol. Resour.* 22, 2087–2104. doi: 10.1111/1755-0998.13597

Grosjean, H. (2013) *Nucleic acids are not boring long polymers of only four types of nucleotides: A guided tour* (Landes Bioscience). Available at: http://www.ncbi.nlm.nih.gov/books/NBK6489/ (Accessed July 26, 2016).

Guevara, E. E., and Lawler, R. R. (2018). Epigenetic clocks. *Evolutionary Anthropol: Issues News Rev.* 27, 256–260. doi: 10.1002/evan.21745

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* 6, 468–481. doi: 10.1038/nprot.2010.190

Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., et al. (2018). Epigenetic age-predictor for mice based on three CpG sites. *eLife* 7, e37462. doi: 10.7554/eLife.37462

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., and Mukai, T. (1991). A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl. Acad. Sci.* 88, 9523–9527. doi: 10.1073/pnas.88.21.9523

Hayes, B. J., Nguyen, L. T., Forutan, M., Engle, B. N., Lamb, H. J., Copley, J. P., et al. (2021). An epigenetic aging clock for cattle using portable sequencing technology. *Front. Genet.* 12. doi: 10.3389/fgene.2021.760450

Horvath, S. (2013). DNA Methylation age of human tissues and cell types. *Genome Biol.* 14, R115. doi: 10.1186/gb-2013-14-10-r115

Horvath, S., and Raj, K. (2018). DNA Methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371–384. doi: 10.1038/s41576-018-0004-3

Jacinto, F. V., Ballestar, E., and Esteller, M. (2008). Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques* 44, 35. doi: 10.2144/000112708

Kassambara, A. (2018). *Machine learning essentials: Practical guide in r* (France: STHDA).

Klughammer, J., Datlinger, P., Printz, D., Sheffield, N. C., Farlik, M., Hadler, J., et al. (2015). Differential DNA methylation analysis without a reference genome. *Cell Rep.* 13, 2621–2633. doi: 10.1016/j.celrep.2015.11.024

Korbie, D., Lin, E., Wall, D., Nair, S. S., Stirzaker, C., Clark, S. J., et al. (2015). Multiplex bisulfite PCR resequencing of clinical FFPE DNA. *Clin. Epigenet.* 7, 28. doi: 10.1186/s13148-015-0067-3

Krueger, F., and Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167

Krueger, F., Kreck, B., Franke, A., and Andrews, S. R. (2012). DNA Methylome analysis using short bisulfite sequencing data. *Nat. Methods* 9, 145–151. doi: 10.1038/nmeth.1828

Kuhn, M (2008). Building predictive models in R using the caret package. *J Stat Softw.* 28, 1–26. doi: 10.1038/nrg2732

Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.* 11, 191. doi: 10.1038/nrg2732

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25

Liu, W., Cui, Y., Ren, W., and Irudayaraj, J. (2019). Epigenetic biomarker screening by FLIM-FRET for combination therapy in ER+ breast cancer. *Clin. Epigenet.* 11, 16. doi: 10.1186/s13148-019-0620-6

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

Masser, D. R., Berg, A. S., and Freeman, W. M. (2013). Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenet. chromatin* 6, 33. doi: 10.1186/1756-8935-6-33

Mayne, B., Berry, O., and Jarman, S. (2021a). Optimal sample size for calibrating DNA methylation age estimators. *Mol. Ecol. Resour.* 21, 2316–2323. doi: 10.1111/1755-0998.13437

Mayne, B., Espinoza, T., Roberts, D., Butler, G. L., Brooks, S., Korbie, D., et al. (2021b). Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary river cod. *Mol. Ecol. Resour.* 21, 2324–2332. doi: 10.1111/1755-0998.13440

Mayne, B., Korbie, D., Kenchington, L., Ezzy, B., Berry, O., and Jarman, S. (2020). A DNA methylation age predictor for zebrafish. *Aging* 12, 24817–24835. doi: 10.18632/aging.202400

Modolo, L., and Lerat, E. (2015). UrQt: An efficient software for the unsupervised quality trimming of NGS data. *BMC Bioinf.* 16, 1–8. doi: 10.1186/s12859-015-0546-8

Nair, S. S., Coolen, M. W., Stirzaker, C., Song, J. Z., Statham, A. L., Strbenac, D., et al. (2011). Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* 6, 34–44. doi: 10.4161/epi.6.1.13313

Nunn, A., Otto, C., Stadler, P. F., and Langenberger, D. (2021). Comprehensive benchmarking of software for mapping whole genome bisulfite data: From read alignment to DNA methylation analysis. *Brief Bioinform.* 22, bbab021. doi: 10.1093/bib/bbab021

Ortega-Recalde, O., and Hore, T. A. (2023). Analythical methods to study the epigenome in *Epigenetics in aquaculture*. Eds. F. Piferrer and H. Wang (Hoboken, NJ: John Wiley & Sons Ltd).

Patel, R. K., and Jain, M. (2012). NGS QC toolkit: A toolkit for quality control of next generation sequencing data. *PloS One* 7, e30619. doi: 10.1371/journal.pone.0030619

Pedersen, B. S., Eyring, K., De, S., Yang, I. V., and Schwartz, D. A. (2014). Fast and accurate alignment of long bisulfite-seq reads. *arXiv*. doi: 10.48550/arXiv.1401.1129

Pfeifer, G. P. (2016). Epigenetics: An elusive DNA base in mammals. *Nature* 532, 319–320. doi: 10.1038/nature17315

Piferrer, F, and Anastasiadi, D (2023). Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology. *Front. Mar. Sci.* 10:1062151. doi: 10.3389/fmars.2023.1062151

Porter, H. L., Brown, C. A., Roopnarinesingh, X., Giles, C. B., Georgescu, C., Freeman, W. M., et al. (2021). Many chronological aging clocks can be found throughout the epigenome: Implications for quantifying biological aging. *Aging Cell* 20, e13492. doi: 10.1111/acel.13492

Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. (2006). N6-methyladenine: The other methylated base of DNA. *Bioessays* 28, 309–315. doi: 10.1002/bies.20342

Rauluseviciute, I., Drabløs, F., and Rye, M. B. (2019). DNA Methylation data by sequencing: Experimental approaches and recommendations for tools and pipelines for data analysis. *Clin. Epigenet.* 11, 193. doi: 10.1186/s13148-019-0795-x

Reyna-Lopez, G. E., Simpson, J., and Ruiz-Herrera, J. (1997). Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Mol. Gen. Genet. MGG* 253, 703–710. doi: 10.1007/s004380050374

Roeh, S., Wiechmann, T., Sauer, S., Ködel, M., Binder, E. B., and Provençal, N. (2018). HAM-TBS: high-accuracy methylation measurements *via* targeted bisulfite sequencing. *Epigenet. Chromatin* 11, 1–10. doi: 10.1186/s13072-018-0209-x

Schield, D. R., Walsh, M. R., Card, D. C., Andrew, A. L., Adams, R. H., and Castoe, T. A. (2016). EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods Ecol. Evol.* 7, 60–69. doi: 10.1111/2041-210X.12435

Simpson, D. J., and Chandra, T. (2021). Epigenetic age prediction. *Aging Cell* 20, e13452. doi: 10.1111/acel.13452

Song, Q., Decato, B., Hong, E. E., Zhou, M., Fang, F., Qu, J., et al. (2013). A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PloS One* 8, e81148. doi: 10.1371/journal.pone.0081148

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., and Paun, O. (2016). BsRADseq: Screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* 25, 1697–1713. doi: 10.1111/mec.13550

Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philos. Trans. R Soc. Lond B Biol. Sci.* 364, 3403–3418. doi: 10.1098/rstb.2009.0125

van Buuren, S., and Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in r. *J. Stat. Software* 45. doi: 10.18637/jss.v045.i03

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., and Verhoeven, K. J. F. (2016). epiGBS: Reference-free reduced representation bisulfite sequencing. *Nat. Methods* 13, 322–324. doi: 10.1038/nmeth.3763

Vogt, G. (2017). Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: Insights from clonal, invasive, polyploid, and domesticated animals. *Environ. Epigenet.* 3, dvx002. doi: 10.1093/eep/dvx002

Weber, D. N., Fields, A. T., Patterson, W. F., Barnett, B. K., Hollenbeck, C. M., and Portnoy, D. S. (2021). Novel epigenetic age estimation in wild-caught gulf of Mexico reef fishes. *Can. J. Fish. Aquat. Sci.* 79, 1–5. doi: 10.1139/cjfas-2021-0240

Xi, Y., Bock, C., Muller, F., Sun, D., Meissner, A., and Li, W. (2012). RRBSMAP: A fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. *Bioinformatics* 28, 430–432. doi: 10.1093/bioinformatics/btr668

Xi, Y., and Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinf.* 10, 232. doi: 10.1186/1471-2105-10-232

Xu, M., Li, X., and Korban, S. S. (2000). AFLP-based detection of DNA methylation. *Plant Mol. Biol. Rep.* 18, 361–368. doi: 10.1007/BF02825064

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 24 March 2023
DOI 10.3389/fmars.2023.1058354

OPEN ACCESS

EDITED BY
Marty Riche,
Florida Atlantic University, United States

REVIEWED BY
Luca Mirimin,
Galway-Mayo Institute of Technology, Ireland
Stephan Koblmüller,
University of Graz, Austria

*CORRESPONDENCE
Yassine Kasmi
yassine.kasmi@thuenen.de

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 30 September 2022
ACCEPTED 13 March 2023
PUBLISHED 24 March 2023

CITATION
Kasmi Y, Blancke T, Eschbach E, Möckel B, Casas L, Bernreuther M, Nogueira P, Delfs G, Kadhim S, Meißner T, Rödiger M, Eladdadi A, Stransky C and Hanel R (2023) Atlantic cod (*Gadus morhua*) assessment approaches in the North and Baltic Sea: A comparison of environmental DNA analysis versus bottom trawl sampling. *Front. Mar. Sci.* 10:1058354. doi: 10.3389/fmars.2023.1058354

Atlantic cod (*Gadus morhua*) assessment approaches in the North and Baltic Sea: A comparison of environmental DNA analysis versus bottom trawl sampling

Yassine Kasmi[1*†], Tina Blancke[1†], Erik Eschbach[1], Benita Möckel[1], Laura Casas[2], Matthias Bernreuther[3], Pedro Nogueira[1], Gertrud Delfs[3], Samira Kadhim[3], Timo Meißner[3], Marcellus Rödiger[3], Amina Eladdadi[4,5], Christoph Stransky[3] and Reinhold Hanel[1]

[1]Thünen Institute of Fisheries Ecology, Bremerhaven, Germany, [2]Institute of Marine Research, Instituto de Investigaciones Marinas (IIM-CSIC), Vigo, Spain, [3]Thünen Institute of Sea Fisheries, Bremerhaven, Germany, [4]Department of Mathematics, The College of Saint Rose, Albany, NY, United States, [5]Division of Mathematical Sciences, National Science Foundation, Alexandria, VA, United States

The assessment of fish stocks is often dependent on scientific trawl fisheries surveys, which are both invasive and costly. The analysis of environmental DNA (eDNA) from water samples is regarded as a non-invasive and cost-effective alternative, but meaningful performance evaluations are required for a wider application. The goal of this study was to comparatively analyze a newly developed, more sensitive real-time PCR based eDNA approach with bottom trawl fisheries catches to locally detect and quantify Atlantic cod (*Gadus morhua)* in the North and Baltic Seas. With a species-specificity of the qPCR assay of 100%, a minimal limit of 15 Cytochrome b eDNA copies was determined for the detection of cod. In addition, a Gaussian processing regression proved a significant correlation (95%) between eDNA (copies per L of water) and cod biomass (CPUE/Ha) found by bottom trawling. The results presented here prove the potential of eDNA analyses for quantitative assessments of commercial fish stocks in the open ocean, although additional comparative analyses are needed to demonstrate its performance under different oceanographic conditions.

KEYWORDS

Environmental DNA (eDNA), quantitative eDNA analysis, *bottom trawl sampling*, Gadus morhua, North Sea, Baltic Sea, eDNA modeling

1 Introduction

A sustainable fishery requires accurate and up-to-date information on the status of the fished stocks. Over the last decade, several studies have demonstrated the potential of exploiting eDNA for marine biodiversity assessment and monitoring (Tillotson et al., 2018; Andruszkiewicz et al., 2019; Jerde, 2019; Jo et al., 2021; Stoeckle et al., 2021). The increased interest in applying eDNA tools is mainly due to their non-invasive properties with a half-life of the eDNA in sea water of up to 48 h (Tsuji et al., 2017; Collins et al., 2018).

The monitoring and surveillance of fish diversity and the estimation of fish abundance are usually carried out with methods based on visual census, remotely operated vehicles (ROVs), as well as fishing techniques, such as bottom trawls (Thompson et al., 1982; Groeneveld, 2000; Sward et al., 2019; Trenkel et al., 2019). Despite all the standardization and optimization protocols, these traditional sampling methods provide information only on short sampling/monitoring time within the stations that span over a few minutes to an hour (Baudrier et al., 2018; Sward et al., 2019; Jourdain et al., 2020) and can be altered by behavioral responses, like observer or gear avoidance. In addition, fishing methods are invasive and cause disturbances and disequilibrium in the marine ecosystems, and can therefore threaten the conservation efforts for marine species.

Advances in molecular biology have allowed the extraction of eDNA from water samples avoiding any invasive effects (Lacoursière-Roussel et al., 2016; Dickie et al., 2018; Jerde, 2019). Water-extracted eDNA is therefore increasingly used for biodiversity assessments as well as for semi-quantitative biomass surveys based on eDNA copy numbers of specific species.

Knudsen et al. (2019) reported on the development of a new PCR assay for quantifying eDNA copy numbers with a Limit of Quantification (LoQ) of 665 copies per reaction, equivalent to a minimum detection limit of 200 kg/h of trawling for Atlantic cod (*Gadus morhua*). However, a significant correlation between eDNA concentration and bottom trawl catches could not be found. A similar study by Salter et al. (2019), based on a commercial qPCR kit (Techne) to target Atlantic cod mitochondrial control region, showed significantly positive correlations between regional integrals of cod biomass (kg) and eDNA copy number (R2 = 0.79, P = 0.003) as well as between catch per unit effort, normalized by sampling effort (kg/h), and eDNA concentrations (copies L^{-1}) (R2 = 0.71, P = 0.008). Nonetheless, despite first promising results in the interpretation of eDNA copy number analysis, the technology is still in its infancy and needs a much more robust calibration to assist or even replace invasive routine methods for quantitative fish stock assessments. Furthermore, eDNA based methods can also not yet provide important stock structure information such as length or age class distributions, even though a few studies tried to tackle this issue with methylome approaches (Sigsgaard et al., 2020; Minamoto, 2022; Zhao et al., 2023). The goal of this study was to refine the existing quantitative eDNA based assessment approaches for Atlantic cod, by focusing on the development of a more sensitive qPCR assay with an improved LoQ value compared to existing methods. In addition, the suitability of various computational models to establish a correlation between eDNA copy number and bottom trawl sampling are evaluated.

2 Methods

2.1 Primer development

2.1.1 Sequence selection

Reference sequences of whole mitochondrial genomes of northeastern Atlantic fishes were downloaded from the public databases **Aquagene** (Thünen Institute of Fisheries Ecology) (Hanel, 2019) and **NCBI-GenBank** (Benson et al., 2013) together with single mitochondrial sequences of cytochrome oxidase (COI) subunit 1, small unit ribosomal RNA (12S rRNA), NADH dehydrogenase (NADH) subunits, cytochrome b (CYTB), ATPase6 genes and control region (D-loop) of Atlantic cod (*Gadus morhua*) and untargeted gadid species (Table S1: Sequences Accession).

2.1.2 Primer development

The collected sequences were aligned using Bioedit (Hall et al., 2011) and MAFFT (Katoh et al., 2018). Plotcon plots were produced for every alignment to check for conserved regions between species and the regions that distinguish Atlantic cod from other gadid species. Subsequently, primers and probes were designed for the COI, 12S rRNA, NADH5, NADH3 and CYTB regions manually or by using PrimerMiner (Elbrecht and Leese, 2017) and the R script DECIPHER (Wright et al., 2012). All primers and probes targeted regions with low intraspecific divergence while maximizing mismatches between related species at the 3′ end as described by (Wilcox et al., 2013). Primers were designed to amplify fragments in the size range of 80 to 250 bp. Annealing temperatures and cross-amplifications of untargeted species were verified using Primer Blast (Ye et al., 2012) and were adjusted to be adequate to the PCR/qPCR kit used.

All designed primers and probes underwent a screening according to their ability to amplify the target species (*G. morhua*) versus cross-amplification of various untargeted species, abundant in the region (*Brosme brosme, Chelidonichthys lucerna, Clupea harengus, Enchelyopus cimbrius, Engraulis encrasiolus, Helicolenus dactylopterus, Lota lota, Melanogrammus aeglefinus, Merlangius merlangus, Merluccius merluccius, Micromesistius poutassou, Molva molva, Pollachius virens, Scophthalmus maximus, Trachurus trachurus, Trisopterus luscus, Trisopterus minutus*) (Table S2).

A preliminary primer screening was performed in a conventional PCR using Phusion® High-Fidelity PCR Master Mix with HF Buffer, 2X (New England Biolabs, Germany) on a BioRad T100™ PCR system in a final volume of 20 µl: 10 µl of Phusion Master Mix (2X), 0.5 µl of each primer (10 µM), 3 µl of DNA (5 -10 ng) and 6 µl of ultrapure distilled water, following these conditions: 98°C for 30 s, 34 cycles × [98°C for 10 s, 54-70*°C for 30 s], 72°C for 20s, in which (*) refers to a gradient PCR. Then, the PCR products were visualized on a 1% electrophoresis agarose gel.

Those primer pairs that showed strong amplification of the target and weak or even no amplification of the non-targeted species were selected and passed on for a second screening. This was performed on SYBR Green qPCR, using Luna® Universal qPCR Mastermix (New England Biolabs, Germany) in a qTOWER³ real-time PCR thermal cycler (Analytik Jena, Germany), in a final volume of 20 µl: 10 µl of Luna® Universal qPCR Master Mix (2X), 0.5 µl of each primer (10 µM), 2 µl of DNA (5 -10 ng) and 7 µl of ultrapure distilled water, following these conditions: 95°C for 60 s, 44 cycles × [95°C for 15 s, 54-70°C for 30 s]. Data collection was enabled at each combined annealing/extension step. The amplification cycle was followed by a melting curve protocol: 60-95°C each for 15 s with an increment of 1°C.

At this stage, the selection of the potential primers was done using two criteria: the specificity to the target species (i. e. Atlantic cod) and in which cycle the first amplification signal started. The two primers/probe pairs that fulfilled these criteria were then tested on different polymerases: TaqMan assay with TaqMan™ Environmental Master Mix 2.0 (ThermoFisher, Germany), KAPA PROBE FORCE qPCR kit (Roche, Germany) and Luna Universal probe qPCR mastermix (New England Biolabs, Germany), according to the manufacturer's protocols. All probes carried 5'FAM fluorescence modifications and BHQ1 as a 3' quencher. Some probes were additionally tested with alternative fluorescence technologies (Locked Nucleic Acids (LNA) and Minor Groove Binder (MGB)) to increase specificity.

All PCR preparations were performed in a designated DNA-free hood in a pre-PCR room.

2.2 Standard curve and assay sensitivity

A standard curve experiment was performed using purified and diluted target amplicons as templates. The PCR products were first purified by Monarch® PCR & DNA Cleanup Kit (New England Biolabs, Frankfurt, Germany) and after 24 h at 4°C quantified by Qubit using Qubit™ dsDNA BR-Assay-Kits (Thermofisher).

PCR-products of Atlantic cod amplified with universal primers for the complete CytB gene (Sevilla et al., 2007): FishCytbF (5' ACC ACC GTT GTT ATT CAA CTA CAA GAA C-3'), TruccytbR (5' CGA CTT CCG GAT TAC AAG ACC G-3') served as a target amplicon for CytB standard curves. For the NADH5 gene, the corresponding qPCR primers were used to create the template including the target amplicon sequence.

From the amplicon stock solutions, diluted to $1.00E^{+10}$ copies/µl, a series of tenfold dilutions (1,000,000, 1,00,000, 10,000, 1,000, 100, 10, 1, as well as 5000, 500, 50, and 25 copies per reaction) were prepared. Nine replicates of each dilution were run to determine the amplification efficiency and limit of detection defined as the lowest copies per reaction with >95% amplification success for each primer/probe set (Bustin et al., 2009).

In real-time PCR (qPCR), false positives can be caused by various sources of molecular biological noise. Establishing a false-positive threshold for the quantification of nucleic acids is essential for the performance of a robust and reliable qPCR assay. Establishing the analytical performance indicators of an assay, including the limit of blank (LoB) that can also be referred to as the Negative Control, and limit of detection (LoD), is fundamental and must be done according to precise procedures. The LoB is the upper limit of the target concentration that is considered acceptable in a blank sample. The LoB is then used to calculate the LoD, the target concentration limit above which the presence of the target can be asserted and quantified with a given statistical confidence. The method to calculate the LoB, LoD and LoQ is described in the Supplementary Data.

2.3 Sea sampling

2.3.1 Baltic Sea

In the Baltic Sea, 21 water samples (5L per sample) were collected from seven sampling stations during research survey No. 441 of FRV Walther Herwig III from 30.11. to 20.12.2020, conducted by the Thünen Institute of Fisheries Ecology in the frame of the North and Baltic Sea monitoring of environmental radioactivity. Water samples were collected from 3 different positions along the drag line of each bottom trawl station (start, middle and end of each drag line). The sampling was performed as described in the North Sea section.

In the Baltic Sea a 140 ft. bottom trawl gear with rock hoppers and a mesh size of 20 mm in the cod-end was used immediately after water sampling to collect Baltic cod. Catch time was 60 min each with a towing speed of 3.9- 4.1 kn during daytime. Atlantic cod catches varied between 0.00 and 9.25 kg/Ha.

2.3.2 North Sea

A total of 32 water samples consisting on 5L each were collected during research survey No. 428 of FRV Walther Herwig III in the North Sea in summer 2019, conducted by the Thünen Institute of Sea Fisheries (Figure 1). The procedure was performed before trawling to minimize contamination from trawl-derived DNA sources and possible disturbance due to the resuspension of sediment caused by the bottom trawler gear. Six water samples were collected along the drag line of each bottom trawl station (two replicates each at the start, the middle and at the end).

Demersal trawling was carried out immediately after water sampling to collect North Sea cod, according to International Bottom Trawl Survey (IBTS) standards (ICES, 2020). Catch time was up to 30 min each at a trawling speed of approximately 4 knots, Total weight of each trawl catch was recorded, before sorting by species. On predefined species, such as Atlantic cod (ICES, 2020), individual length and weight were measured, sex and maturity stage discrimination were performed and otoliths were removed for subsequent age determination in the laboratory.

A total of 32 5 L seawater samples were collected at 16 trawl positions (Figure 1) by Niskin bottles mounted on a stainless steel CTD frame at a depth of 4 m above the seafloor to match the trawl height of 5 m and to minimize the possibility of sampling eDNA from non-recent sediment sources. Immediately after sampling, the water was filtered through Sartorius™ PES membranes (pore size: 0.45 µm, diameter: 47 mm) with a vacuum pump. Filters were subsequently stored at -20°C until DNA extraction at the laboratory.

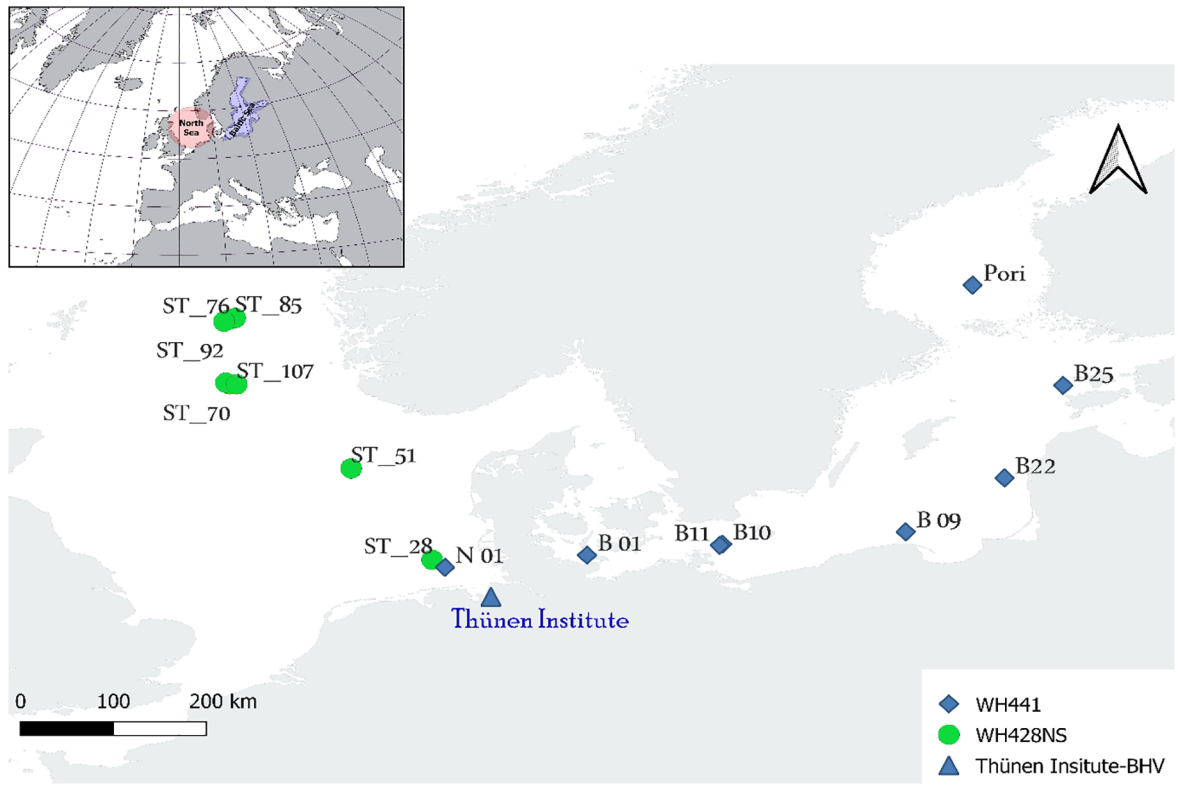


FIGURE 1
Sampling sites in the North Sea and Baltic Seas (Latitude: 2.00°-8.00° and Longitude: 52.00°-63.00°), WH441 and WH428 refer to the research missions aboard FFS Walther Herwig III conducted by the Thünen Institute in the Baltic Sea and North Sea, respectively. The map is constructed by QGIS.

2.3.3 Negative controls and contamination avoidance

Contamination avoidance was key throughout the analysis. Two negative control samples were taken from the ultrapure water which was used to rinse the Niskin bottles before each sampling. Field blanks consisted of 2.5 L of ultrapure water rinsed through the corresponding Niskin Bottles before water sample collection onboard the research vessel. After recovery of the CTD rosette, all Niskin sample bottles were thoroughly rinsed with fresh water on deck, removed from the sampling frame and transported to a CTD control laboratory isolated from the deck area where they were mounted on wall mounts for further processing. Prior to sub-sampling, the exterior of the Niskin bottles and the sampling nozzle were rinsed with a solution of sodium hypochlorite (10% commercial bleach), followed by ultrapure water. The on-board workbench area was covered with aluminium foil and rinsed with a 20% commercial bleach solution followed by ultrapure water. Each subsample bottle was rinsed three times with sample water and then filled to a 2.5 L mark. Negatives samples were treated as real samples and were filtered immediately. All further processing of the samples took place in a sterile environment in a molecular biology lab on land.

2.4 eDNA extraction

eDNA was extracted from the water according to the protocol of (Renshaw et al., 2015). In brief, the frozen PES membranes from each samples were placed in 2 ml tubes filled with 700 µL of CTAB-buffer (2% CTAB (w/v), 1.4 M NaCl, 100 mM Tris, 20 mM EDTA) and incubated at 65°C for 10 min while shaking at 650 rpm. Subsequently, 900 µl PCI phenol/chloroform/isoamyl alcohol (25:24:1) was added and vortexed for 5 s, prior to centrifugation at 15.000 x g for 5 min. 700 µl of supernatant were transferred to a new 2 ml tube, and 700 µl of chloroform was added, followed by centrifugation at 15.000 x g for 5 min. 500 µL of supernatant was transferred to a new 2 mL tube, and from here we proceeded with Monarch® Cleanup Kit (New England Biolabs, Frankfurt, Germany) from the step of adding 1000 µl of binding buffer. The final elution step was accomplished with 20 µl of elution buffer. The eDNA extracts were either stored at -20°C until qPCR or archived at -80°C. All steps of eDNA extractions were performed in a dedicated lab area under a chemical hood.

2.5 Atlantic cod qPCR from North Sea water samples

Quantitative Real-time PCR of Atlantic cod eDNA was performed using the best primer/probe couple developed in the first and second phase of this study. The primers Gm_Cytb_For2a (5'-TACACTATACCTCAGACATCGAGAC-3') and Gm_Cytb_Rev2b (5'-GGCAATGTGCATATAAAGACAAATG-3'), coupled with the LNA-based TaqMan probe GmCytb-LNA-P (5'-[FAM] A[+C]TA[+C]GGCTGA[+C]TAATTCG[+G]A[BHQ1]-3')

were used for the amplification of the CytB gene. A qTOWER³ Real-time PCR thermal cycler (Analytik Jena, Germany) was used for amplification. The 20 µL qPCR reaction mix contained 5.5 µL eDNA template, 10 µL 2 × TaqMan™ Environmental Master Mix 2.0 (Thermofisher, Germany), 0.4 µM of each primer, 0.2 µM of Atlantic cod probe (GmCytb-LNA-P probe), 0.1 µL of AmpErase™ Uracil N-Glycosylase (UNG) (ThermoFisher, Germany) and 2.4 µL of ultrapure water. qPCR reactions were performed under thermal cycler conditions of 10 min at 95°C followed by 45 cycles of 15 s at 95°C and 59 s at 62°C. Data collection was enabled at each combined annealing/extension step. eDNA samples were measured in triplicate reactions without dilution. Numerous qPCR assays were carried out to determine the optimal PCR condition, including primers/probe concentrations as well as the DNA amount to be added in each reaction.

During the DNA extraction and qPCR assays, negative controls for extractions and no template controls (NTCs) were included in all manipulations to ensure that no cross- contamination occurred. All controls and NTCs contained only nuclease-free water. A positive control consisting of Atlantic cod genomic DNA was used to prove the qPCR reaction performance. Extractions were performed identically. Field blanks consisted of 2.5 L of ultrapure water rinsed through the corresponding Niskin Bottles before water sample collection onboard the research vessel. Neither the extraction blanks nor the field sample blanks showed amplification with Atlantic cod qPCR primers.

2.6 Computational modeling of trawling and eDNA data

The relationship between Atlantic cod eDNA concentrations and trawl catches in the standard survey area was assessed by various combinations of covariates where eDNA levels were above LoQ. The response variable, eDNA level per fishing position, and all explanatory variables were log10 transformed with the exception noted in the comment column in section 3.3. The Shapiro-Wilk test was used to validate the dependent and independent variables before the regression model. All variables and residuals used in the regional regression models were characterized by Shapiro-Wilk p-values > 0.05 confirming normal distributions.

The correlation between the number of Atlantic cod gene copies obtained by qPCR and the catch per unit effort (CPUE) was tested by simple correlation regression, generalized least squares (GLS), forest-tree, nonlinear regression, and neural network. A model was developed based on the data obtained by Knudsen et al., 2019 and then readjusted according to the output of this study. The model with the best correlation degree between real and predictive data with a minimal RMSD and MRE was selected as the best model. The model was developed using R and MATLAB scripts.

3 Results

3.1 Atlantic cod primer-probes development

3.1.1 Primer screening

Four primer pairs targeting the mitochondrial genes ATPase6, ND5, COI, and CytB, were designed in silico to distinguish Atlantic cod from other species of the family *Gadidae*. By comparing the four designed primers with the ones published by Knudsen et al. (2019), ATPase6, ND5, and CytB showed more than 6 additional mutations (Table 1). A first screening for cross-amplification of other gadids *via* conventional PCR for ATPase6, ND5, COI, and CytB was negative in each of the three technical replicates (Tables 2, 3).

TABLE 1 The number of mutations observed *In-silico* between the primers designed or published compared to the reference *Gadidae* mitochondrial genomes.

Species	GenBank Acc. Nr.	*ATPase6_Cod*	*ND5_Cod*	*COI_Cod*	*GmCytb-LNAP*	(Knudsen et al., 2019)
Gadus morhua		0	0	0	0	0
Gadus macrocephalus	MK990531.1	6	9	2	4	1
Boreogadus saida	MG100545.1	7	9	2	7	2
Gadus ogac	LN908945.1	7	9	2	5	2
Gadus chalcogramma	DQ356946.1	9	8	3	4	2
Eleginus gracilis	MH061057.1	9	N	-	N	6
Theragra finnmarchica	AM489719.1	10	8	-	4	2
Pollachius virens	KP644330.1	10	15	4	N	6
Micromesistius australis	AB550326.1	11	N	4	N	6
Trisopterus minutus	KP644339.1	N	N	4	N	6
Microgadus proximus	DQ174066.1	10	N	4	N	6
Melanogrammus aeglefinus	KP644328.1	8	13	4	8	5

The presence of more than 5 mutations between the primer and the non-target species sequence and therefore a high distinction of Atlantic cod is highlighted in red. 4 to 5 primer mutations with the non-target species and therefore a medium risk of detection is highlighted in yellow, less than 3 mutations in green. The letter N means that no combination/similarity was observed between the primer and the untarget species.

10.3389/fmars.2023.1058354

TABLE 2 List of primers/probes pairs developed and tested, with the optimal concentrations and lengths of target fragments.

species	Primer/Probe name	Sequence	Optimal concentration [nM]	Fragment length (bp)
Gadus morhua	Gm_Cytb_For	TACACTATACCTCAGACATCGAGAC	400 nM	140
	Gm_Cytb_Rev	GGCAATGTGCATATAAAGACAAATG	400 nM	
	GmCytb-LNA-P	[FAM]A[+C]TA[+C]GGCTGA[+C]TAATTCG[+G]A[BHQ1]	200 nM	
	Target fragments for each assay			
	Sequence Cytb	TACACTATACCTCAGACATCGAGACAGCCTTCTCATCCGTAGTCCACATCTGTCGTG ATGTAAACTACGGCTGACTAATTCGGAATATACATGCTAATGGTGCCTCTTTCTTTT TCATTTGTCTTTATATGCACATTGCC		
	ND5-Cod_For	GCAAGAATTTGGACATAACTCTCCCTCTA	400 nM	250
	ND5_Cod_Rev	AATATAGTGGTTAAGGCTCCTAGACAGA	400 nM	
	ND5_Cod-P	[FAM] CCT AAT TCG GAT GAG CCC [MGBEQ]	200 nM	
	Target fragments for each assay			
	Sequence ND5	GCAAGAATTTGGACATAACTCTCCCTCTACTCGGTTTAATCTTGGCTGCCACTGGTAA ATCCGCCCAGTTTGGACTTCACCCATGACTACCAGCCGCAATAGAAGGTCCAACGCC AGTGTCTGCCCTACTTCATTCTAGCACAATAGTTGTAGCAGGAATTTTTCTCCTAATT CGGATGAGCCCTCTTATAGAAAATAATCAGACTGCACTAACTCTCTGTCTCTGTCTA GGAGCCTTAACCACTATATT		
	COI-Cod_For	TATTAATATGAAACCTCCGGCA	400 nM	98
	COI_Cod_Rev	CGGGGAGAGATAATAGTAGAA	400 nM	
	COI_Cod-P	[FAM]CCTATTTGTTTGAGCAGTACTAATTACAGCTGTG [BHQ1]	200 nM	
	Target fragments for each assay			
	Sequence COI	GCAAGAATTTGGACATAACTCTCCCTCTACTCGGTTTAATCTTGGCTGCCACTGGTA AATCCGCCCAGTTTGGACTTCACCCATGACTACCAGCCGCAATAGAAGGTCCAACG		
	ATPase6Cod-F	ACCCTGACTTTTAATTCCTACACCTAC	400 nM	
	ATPase6Cod-R	TAGGAGTGAAGATATATGGTATTAAGC	400 nM	209
	Target fragments for each assay			
	Sequence 12S	ACCCTGACTTTTAATTCCTACACCTACTTCCCGATGACTAAGCAATCGAGTTGTATCTCTA CAAGGATGGTTTATCGCCCGCTTTACTAATCAACTCTTTTTACCTCTAAATGTGGGAGGAC ACAAATGAGCTCCTCTTCTTGCCTCACTAATAATGTTTTTACTCACTCTAAATATGTTAGG CTTAATACCATATATCTTCACTCCTA		

-For, forward primer; -Rev, reverse primer; -P, Probe.

TABLE 3 Comparison of polymerase performance based on LoD and LoQ values.

Primer	Polymerase	LoD	LoQ	Slope
GmCytb-LNA	Environmental TF	12	16	-3.57
GmCytb-LNA	KAPA	16	22	-3.74
GmCytb-LNA	probe Luna	34	70	-3.37
ND5_Cod	probe Luna	26	70	-4.03
ND5_Cod	SybrGreen Luna	7.88	16	-3.85
Techne	Techne	770	781	-4.35
Knudsen et al. (2019)		660	660	–

However, when using SybrGreen PCR, some replicates showed unintended cross-amplification of untargeted gadid species, generally in the late cycles of the qPCR. The same bias was observed when using the probe, resulting in a strategy of defining specificity and sensitivity individually for each primer/probe pair with LoQ being a minimum accepted specificity value, and LoD the minimum sensitivity of a qPCR assay.

For ATPase, even both primers (forward and reverse) are located in hypervariable sites and should allow high specificity, it was problematic to determine probe regions for an unambiguous discrimination of Atlantic cod. Therefore, the ATPase-targeted primer pair is proposed for a qPCR assay coupled to SyberGreen. For this primer pair, the initial Ct value for DNA control was in the range of 12.02, while all investigated untargeted species would start from 32, with an equivalent LoQ Ct value of 31.

For the COI primers/probes, we obtained an initial amplification at Ct 18.3 for genomic DNA control, however, all untargeted species started to show some Ct values equal to 33.43. the same genomics DNA amount were used in all assays, the quantification of the genomics DNA were made by Qubit 4.0. The in-silico evaluation revealed a borderline number of mutations which allowed a distinction between Atlantic cod (*Gadus morhua*), the congeneric Pacific cod (*Gadus macrocephalus*), Greenland cod (*Gadus ogac*) and Alaska pollock (*Gadus chalcogramma*) as well as the Polar cod (*Boreogadus saida*).

The GmCytb-LNAP and ND5_Cod primers/probes were found to amplify only Cytb, and ND5 DNA of Atlantic cod as the target species. Cross-amplification tests on untargeted species collected during different missions to the North Sea were performed to test the specificity of the designed primers in qPCR. For the GmCytb-LNA primer pair, DNA extracted from cod tissue samples amplified at an initial Ct value of 18.8 ± 0.25, whereas for non-target species amplification was only proven after Ct 36.5, knowing that the LoQ value equals Ct=35.5, equivalent to 15 copies per reaction. The ND5_Cod primer/probe pair was able to reliably detect cod DNA at an initial Ct value of 11.56 corresponding to 8 x 10^7 copies, its LoQ value was 16.27 copies per reaction, equivalent to a Ct value of 29, while non-target species appeared after a Ct value of 31. To prevent the inclusion of false-positive test results, LoQ Ct values were established as the positive threshold for the test. The standard curve based on the genomic control had a slope of -3.57 and -3.85 for CytB and ND5, respectively.

3.1.2 Comparison of polymerase screening

The use of TaqMan™ Environmental Master Mix 2.0 (Thermofisher, Germany) increased the specificity and sensitivity of the assay compared to KAPA qPCR kit and Luna Universal probe qPCR mastermix kit, observed at the initial Ct level for the control, which was in the order of 16.80 and a final Ct of 36.1 for 10 copies per reaction, whereas non-target species only began to appear after a Ct of 36.01, with an LoQ value that went up to 16 copies per reaction when the Environmental Master Mix 2.0 kit was used (Figure 2). Similarly, the sensitivity also increased with the KAPA qPCR kit, but less significantly than with the Environmental Master mix. Using KAPA qPCR kit, the initial control Ct value was 16.45, and final Ct was 35.04 for 10 copies per reaction, while the non-target species started only after a Ct of 34.9, making the LoQ value equal to 22 copies per reaction (equivalent to a Ct of 33.8). Whereas for the Luna Universal probe qPCR mastermix kit, **untarget** species started to appear at a Ct of 34.5 which is equivalent to 57 copies per reaction, and therefore the LoQ value was 70 copies per reaction. Similarly, the ND5_Cod results had almost similar results concerning the sensitivity of the Environmental Taq polymerase compared to Luna Probe (Table 3). Techne has an LoQ value equivalent to 781 copies per reaction.

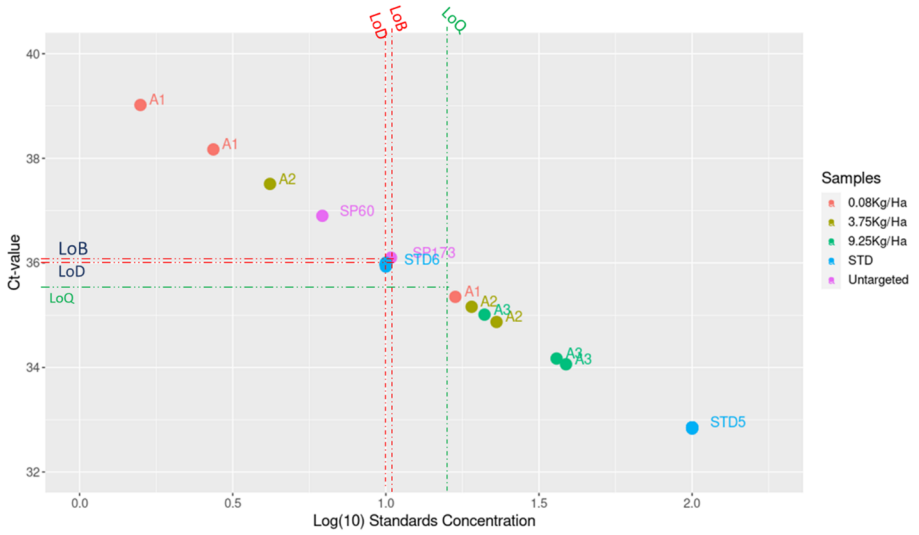


FIGURE 2
The results from different qPCR assays of *in situ* validation as well as the standard curve of threshold cycle number (Ct values) plotted against the log concentration (copy number). Dark green dots represent 9 replicates for each dilution. STD: the qPCR Standards. Positive control sample A1, A2, A3 with 0.08, 3.75, 9.25 kg/Ha of Atlantic cod in catch, respectively.

3.2 Artificial and *in situ* validation

Considering that the GmCytb-LNA assay performed best in the evaluation part (with a slope of about -3.5), we opted to proceed with *in situ* validation steps only with the primer/probe pair, setting the threshold of positive signal to a value of 1.7 for the GmCytb-LNA primer/probe pair.

For the Baltic sea samples, all technical replicates from stations with trawl catches of 9.25 and 3.75 kg/Ha of Atlantic cod were positive with eDNA signals between Ct 34.06 and 35.48, equivalent to copy numbers between 38.75 and 19.06 copies per reaction. For the stations with trawl catches of 0.08 kg/ha, only one of the three replicates was positive.

In the analysis of eDNA North Sea samples, the obtained Atlantic cod eDNA copy numbers were above the LOQ in 12 of the 17 samples, with a maximum copy number equivalent to 153 copies L^{-1}, clearly higher than those found in the Baltic Sea. Within single stations, water samples taken in the middle of each drag line were generally richer in Atlantic cod eDNA than those at the start and the end (Figure 3). The highest Atlantic cod biomass recorded by bottom trawling in this study was 9.5 kg/Ha in the Baltic Sea (Station B10:54°49,377N; 013°55,848E), which corresponded to a concentration of 134 eDNA copies per L. Conversely, with 153 copies per L, the highest eDNA concentration was found in the North Sea, corresponding to a trawl biomass of 1.25 kg/Ha.

Overall, the GmCytb-LNA qPCR analysis of 15 eDNA samples resulted in the detection of Atlantic cod at all sites where the species was found in the concomitant trawl catches. Additionally, at station ST51-2, where cod was absent in the trawl catch, eDNA analysis was positive with 20 copies per reaction while all other negative trawl stations tested also negative using qPCR. This means that our qPCR assay was able to detect *G. morhua* in 112% of eDNA samples. All the expected positive samples produced an amplification signal in all three qPCR replicates for each sample (Table 4). The qPCR assays had an efficiency of 94.04% +/- 1.0%, with a slope equal to -3.53 +/-0.27. The R2 values ranged from 0.99 +/-0.2%. All negative extraction and PCR controls tested were negative with no sign of contamination.

Overall, Atlantic cod was detected in 60% (9/16) of the analyzed trawl hauls at the North Sea.

3.3 Computational modeling of trawling and eDNA data

The correlation between eDNA copy number and the amount of Cod found by the simple models was insignificant and weak. Despite attempts to standardize the data by logarithmic, exceptional, (max-min)/max, (max-min)/(average-x) functions, the best regression obtained was 28% between eDNA copy number per L and biomass, for the abundance in CPUE/Ha the best correlation obtained with the Log of eDNA copies/L was 40%. The correlation between eDNA copy number and CPUE in ordinary least squares (OLS) models was less than 25%, while for biomass it was around 0. By adopting Gaussian process regression (GPR), Neural network and non-linear regression modelling, we observed an increase in the regression rate. However, the use of only one variable (eDNA/L copy number) achieved only 55%. The use of two variables (eDNA/L copy number and sampling depth) significantly increased the regression rates. This allowed us to achieve 96% regression between the trawl results and the model outputs by GPR and Neural Network method. At the same time, the error rate was significantly reduced to 0.002 (MAE) (Table 5).

The Gaussian process regression (GPR) model (GPR_4) with two explanatory variables was the best at explaining eDNA levels for the fishing position (GPR_4, R2 98%, MAE=0,002), compared to

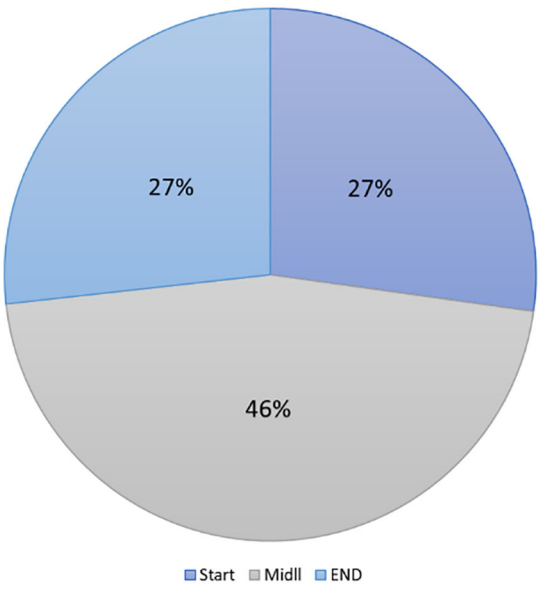


FIGURE 3
Percent of the number of copies per station, 46% of DNA copies obtained by qPCR were in middle stations of trawling. (p=0.004168*).

10.3389/fmars.2023.1058354

TABLE 4 The average measured eDNA (copies/L of seawater) in seawater samples from 17 stations is presented together with catches (kg), eDNA levels (copies/L of filtered water), and catches per unit effort (unit/ha).

Sampling Mission Number	Station Sampling Nr.	rep. pos	Kg/ha	eDNA (copies eDNA/L)
WH441	N01	1	0.1	16,5
	B09	3	3.8	76
	B10	3	9.3	134
	B 01	0	0,3	–
	B25	0	0,3	–
	B11	3	0,4	19.56
	B22	1	0,8	21.54
	Pori	0	0	0
WH428	ST28-1	0	0	–
	ST28-2	0	0	–
	ST28-3	0	0	–
	ST51-1	0	0	–
	ST51-2	1	0	20.36
	ST51-3	0	0	–
	ST70-1	3	1.25	110.32
	ST70-2	3	1.25	153.69
	ST70-3	3	1.25	109.82
	ST76-1	3	1.0	0
	ST76-2	3	1.0	64.24
	ST76-3	3	1.0	27.86
	ST98-1	3	1.0	67.46
	ST98-2	3	1.0	121.36
	ST98-3	3	1.0	49.69
	ST85	0	0	0

All water samples were analyzed by quantitative PCR in three replicates. The number of replicates with positive eDNA detection (eDNA level above the LOQ) is indicated for each sample (columns designated "pos."). If one or more replicates were above the level of quantification, LOQ = 16 copies of eDNA/L, an average was calculated using the positive replicates. Rep.pos, number of replicates positive.

other models with only one (eDNA/L copy number) or two (eDNA/L copy number and sampling depth) explanatory variables. In this model, we did not consider the fishing position and the latitude of the fishing position. The positive correlation between eDNA concentrations and catch per unit effort (CPUE) per Hectare on a station-by-station basis was statistically significant but considerably weaker than the regional comparisons.

4 Discussion

eDNA promises to provide reliable answers on marine diversity for marine and fisheries resource management studies, without posing a risk to dangerous species, due to the non-invasive approach. The objective of this study was to evaluate the performance of eDNA tools for revealing the abundance of *Gadus morhua* in the North and Baltic Sea, based on the qPCR approach.

The results of this research provide a basis for future studies to use this non-invasive tool. Significant differences were found by comparing the performance of the primers previously published and the ones designed in this study, and the type of primer used, as well as the computational model, used to predict the relationship between eDNA and the number of fishes per Ha. In addition, the three sampling positions at the same station were comparable, although the samples from the middle station were significantly better to provide more eDNA than both others positions, (according to Anova test, the f-ratio value is 6.5003. The p-value is.004168. The result is significant at $p < 0.05$). Also, the models used in this study were significantly different, with Gaussian process regression (GPR) applied to logarithmic data offering the best correlation between observed and predicted data.

TABLE 5 Modeling results with tests for various combinations of covariates on cod eDNA concentrations at fishing positions where eDNA levels were above the level of quantification.

Method Name	Model No	Variable	Error (MAE)	Regression (%) Predicted vs True	Comments
Gaussian process regression (GPR)	GPR_1	• Copies/L • kg/h	59.4	94%	Exponential, Model till 20 copies per reaction
Gaussian process regression (GPR)	GPR_2	• Copies eDNA/L • kg/h	74.5	92%	Square Exponential, Model till 20copies kg/
Gaussian process regression (GPR)	GPR_3	• Copies eDNA/L • kg/h	0.03	94.7%	Optimized GPR
Gaussian process regression (GPR)	GPR_4	• Copies eDNA/L • CPUE/ha	0.002	96%	Optimized GPR
Neural Network	NN_1	• Copies eDNA/L • kg/ha	6.04	89%	Wide NN
Neural Network	NN_2	• Copies eDNA/L • kg/ha	7,21	84.5%	Trilayered NN
Bayesian Fixed	Bay_1	• Copies eDNA/L • kg/ha		0.05	
GLS regression	GLS_1	• Log copies/L	0.5	92%	Log copies
OLS	OLS_1	• Log Copies/L	0.6	16%	Log CPUE
OLS	OLS_2	• Sqrt Copies/L	0.2	24%	Sqrt CPUE
R2	R2_1	• Log Copies eDNA/L		40%	Log10 CUPE/ha
R2	R2_2	• Log Copies eDNA/L		28%	Log kg/ha

GPR, Gaussian process regression; GLS regression, generalized least squares (GLS); OLS, ordinary least squares (OLS); R2, coefficient of determination, denoted R^2 or r^2 and pronounced "R squared"; NN, Neural Network; CPUE, Catch per unit effort; Ha, hectare; MAE, mean absolute error.
Predicted vs True: regression between the outputs of the developed model basing on eDNA (predicted) and the catch results by trawl bottom (True).

4.1 Atlantic cod qPCR test/primer development

Four primer pairs targeting the mitochondrial genes mt-ATPase6, mt-ND5, mt-COI, and mt-CytB were designed in silico and validated in-situ. All designed qPCR assays were found to be specific to *Gadus morhua* in the North and Baltic Sea regions, with high sensitivity and specificity. Although the primers may give a positive signal for other untargeted species, this occurs only after a Ct value higher than 36, which is equivalent to 16 copies per reaction or less, which means that every signal after Ct 36 is a false positive signal. This false positive signal may be only due to primer breaks or incomplete hybridization due to thermic exchange during PCR cycles.

The GmCytb-LNA and Cod_ND5 pair was found to be highly specific to *Gadus morhua* with the Environmental Thermo-Fisher polymerase/master mix and the KAPA master mix than with other polymerases. This highlights the importance of polymerase choice in eDNA studies. (Knudsen et al., 2019) determined the LoQ value of their *Gadus morhua* primers to be 669 copies per reaction, while the commercial qPCR (Techne, Bibby Scientific, United Kingdom) has an LoQ value of around 771 copies per reaction. With an LoQ value equivalent to 16 copies per reaction, the GmCytb-LNA pair, coupled with TaqMan™ Environmental Master Mix 2.0, has proven to provide more sensitive and reliable results than previously available or published *Gadus morhua* primers/kits. This means that environmental DNA studies are more accurate and possible with the GmCytb-LNA pair than with other primers.

4.2 eDNA and Stock assignment

A study of the variation in the number of eDNA copies was carried out to determine the best horizontal position to sample the water. Based on the data obtained (Figure 3), we observed that samples taken at the middle/center of the station would cover 46% of the total copy number obtained by qPCR, whereas samples taken at the beginning and end of the station would only cover 27%. The center of the station was, moreover, more consistent with the trawling data than the other two sampling points. We note here that only one station had more significant samples taken at the beginning of the station than at the middle or end of the station.

Despite the large number of studies that have dealt with the subject of eDNA, on different organisms, there are still crucial open questions that need further consideration, especially in the field of fish stock assessment in the sea. These not only include the quantitative aspect, but also the richness of mathematical models, which can offer solutions to accelerate the use of eDNA methods in the field of standard fish population assessment and monitoring, being a non-invasive method.

The comparison of eDNA data with trawling data showed that simple regression models were less effective in establishing a

significant relationship between the number of eDNA copies and the amount of fish caught or the CPUE number, where the regression percentage did not exceed 40%. On the contrary, the regression models based on neural networks and the Gaussian process allowed obtaining a high percentage of correlation between the predicted results of the model and those obtained from the trawling, especially when the predictor is the value of CPUE, coupled with a standardization by log10. Therefore, the GPR-based model (GPR_4), predicting the CPUE based on the copy number per liter, achieved a degree of correlation in the order of 95% with an error rate in the order of 0.002 (Figure 4). In the same line with our results, the outputs of modelling by (Lacoursière-Roussel et al., 2016), showed that a simple regression is only able to offer a maximum of 40% regression between the trawling data and the predictive. In addition, the works of (Tillotson et al., 2018; Moushomi et al., 2019), showed that generalized least squares (GLS) regression was able to provide a 90% regression, which is observed in our results, however, the error rate is of the order of 0.5 which may introduce a bias in the predictive data. The autoregressive integrated moving average (ARIMA) model invented by (Thalinger et al., 2019) also offers significant results between predicted and observed, however, this model focuses on studying the relationship between target eDNA concentration and fish numbers *via* time series modelling. Salter et al. (2019) found that their qPCR test could detect cod only at catch densities higher than 200 kg/h. Furthermore, data published by (Mahon et al., 2013) suggested that positive detection increased with the relative abundance of fish species in the Chicago area waterway system. In the same context, but in a controlled environment, (Doi et al., 2015) found a positive correlation between fish biomass and eDNA concentration in two experimental ponds.

In the present study, we were able to predict the stock abundance of cod in the North Sea and the Baltic Sea using eDNA tools from the samples. The results were similar to those

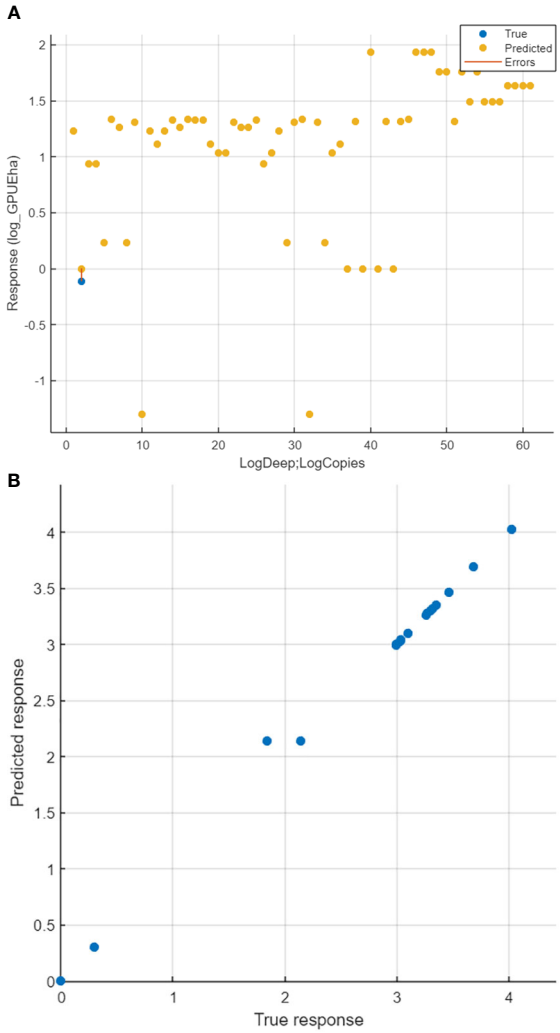


FIGURE 4
Comparison between output of Catch results and output of GPR model for eDNA copies. (R = 96%, Error MAE = 0,002). **(A)** co-Plot of the real data from trawl (blue) and the predicted basing on the eDNA model (yellow), the x-axis present predictive values of Log Depth (log Deep) of sampling and number of DNA copies (Log Copies); most of data were in predicted as it was observed. **(B)** The correlation between predicted data by the model (Predicted response) and the real data collected from trawl (True response).

obtained by trawling and CPUE for cod. Therefore, the compatibility between the trawl and eDNA data, supports also that eDNA can be used as a tool to reveal the cod stock situation in real time. In addition, as the half-life of eDNA in water samples is significantly short - around 48h- (Maruyama et al., 2014; Collins et al., 2018; Harrison et al., 2019), which makes the effects of eDNA accumulation in water through time negligible, and then eliminates the assumption that the amount of eDNA due to the long-term fishery repertoire of stations. But only for short times prior to trawling (maximum 48h). The issues of transport from other areas *via* currents are important to study, but the results from the three sampling positions for each station show that transport by current was not significant in our study. The sampling sites in the middle of the stations are counted as the most eDNA rich sites in *Gadus morhua* compared to the stations within the boundaries. The results obtained at the different stations show that the intermediate samples were richer in DNA copy numbers than those taken at the beginning or end of the station. To the best of our knowledge, there are currently no studies explaining these results. In this respect, we suspect that it is a question of random chance or that the ship created a secondary current along the flight path inside the station, causing an accumulation of DNA in the middle. However, this interpretation has been abandoned, as the amount of eDNA in the samples from the end of the station were often lower than in the center and the starting point, which contradicts this proposaland therefore negates it. The true reasons remain unknown, and subject to further studies in different locations and times of year, to examine their results and compare them to the current results, as well as to understand the impact of spatio-temporal variation on the validity of the current results.

5 Conclusions

Our results support the idea that eDNA can be used in the assessment of commercial fish stocks species to estimate the abundance of marine species, and elaborate a new non-invasive complementary method to the conventional methods currently used. Although a total replacement of the reference methods in the short term is not envisaged, a complementary integration of eDNA tools and especially of our eDNA method, will be a plus to put these technical tools on the right track of implementation and familiarization for monitoring and stock assessment teams in the short and medium term. This allows also to standardize and determine the effect of biotic and abiotic conditions on the performance and predictive capacity of fish stock assessment by eDNA from marine waters for other marine species.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. All data are available upon request to YK (yassine.kasmi@thuenen.de) or to RH (reinhold.hanel@thuenen.de).

Author contributions

BM and TB: Filtration and DNA Extraction. EE, TB, and YK: qPCR test development, optimization and Molecular Biology analysis. AE and YK: Bioinformatics and computational modeling. BM, EE, GD, MB, PN, SK, and TM, Water sampling of water from North Sea and Baltic Sea, Fisheries data. CS, EE, LC, RH, and YK: Conception of research and establish the methodology. LC, CS, and RH: Funding, Supervision, Orientation, and Coordination. RH: Correction of manuscript. YK: First draft manuscript. All authors contributed to the article and approved the submitted version.

Funding

The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union.

Acknowledgments

The authors gratefully acknowledge comments from the editor and the reviewers, which markedly improved the quality of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of CINEA or of the European Commission. The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments

using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. AE is a Rotating Program Director at the US National Science Foundation (NSF). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1058354/full#supplementary-material

References

Andruszkiewicz, E. A., Koseff, J. R., Fringer, O. B., Ouellette, N. T., Lowe, A. B., Edwards, C. A., et al. (2019). Modeling environmental DNA transport in the coastal ocean using Lagrangian particle tracking. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00477

Baudrier, J., Lefebvre, A., Galgani, F., Saraux, C., and Doray, M. (2018). Optimising French fisheries surveys for marine strategy framework directive integrated ecosystem monitoring. *Mar. Policy* 94, 10–19. doi: 10.1016/j.marpol.2018.04.024

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, 36–42. doi: 10.1093/nar/gks1195

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., and Genner, M. J. (2018). Persistence of environmental DNA in marine systems. *Commun. Biol.* 1, 1–11. doi: 10.1038/s42003-018-0192-6

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., et al. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Mol. Ecol. Resour.* 18, 940–952. doi: 10.1111/1755-0998.12907

Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H., and Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS One* 10, e0122763. doi: 10.1371/journal.pone.0122763

Elbrecht, V., and Leese, F. (2017). PrimerMiner: An r package for development and in silico validation of DNA metabarcoding primers. *Methods Ecol. Evol.* 8, 622–626. doi: 10.1111/2041-210X.12687

Groeneveld, J. C. (2000). Stock assessment, ecology and economics as criteria for choosing between trap and trawl fisheries for spiny lobster palinurus delagoae. *Fish. Res.* 48, 141–155. doi: 10.1016/S0165-7836(00)00178-8

Hall, T., Biosciences, I., and Carlsbad, C. (2011). BioEdit: An important software for molecular biology. *GERF Bull. Biosci.* 2, 60–61.

Hanel, R. (2019) *Aqugen database: The curated genetic reference database for FISH SPECIES*. Available at: http://www.aquagene.org/.

Harrison, J. B., Sunday, J. M., and Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proc. R Soc. B* 286, 20191409. doi: 10.1098/RSPB.2019.1409

ICES (2020)Manual for the north Sea international bottom trawl surveys. In: *Ser. ICES surv. protoc. SISP 10-IBTS 10*. Available at: https://ices-library.figshare.com/articles/_/19051361 (Accessed April 25, 2022).

Jerde, C. L. (2019). Can we manage fisheries with the inherent uncertainty from eDNA? *J. Fish Biol.* 98, jfb.14218. doi: 10.1111/jfb.14218

Jo, T., Takao, K., and Minamoto, T. (2021). Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. *Environ. DNA* 00, 1–13. doi: 10.1002/edn3.253

Jourdain, N. O. A. S., Breivik, O., Fuglebakk, E., Aanes, S., and Vølstad, J. H. (2020). Evaluation of sampling strategies for age determination of cod (Gadus morhua) sampled at the north Sea international bottom trawl survey. *ICES J. Mar. Sci.* 77, 859–869. doi: 10.1093/icesjms/fsaa013

Katoh, K., Rozewicki, J., and Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Knudsen, S. W., Ebert, R. B., Hesselsøe, M., Kuntke, F., Hassingboe, J., Mortensen, P. B., et al. (2019). Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. *J. Exp. Mar. Biol. Ecol.* 510, 31–45. doi: 10.1016/j.jembe.2018.09.004

Lacoursière-Roussel, A., Côté, G., Leclerc, V., and Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *J. Appl. Ecol.* 53, 1148–1157. doi: 10.1111/1365-2664.12598

Mahon, A. R., Jerde, C. L., Galaska, M., Bergner, J. L., Chadderton, W. L., Lodge, D. M., et al. (2013). Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS One* 8, e58316. doi: 10.1371/journal.pone.0058316

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., and Minamoto, T. (2014). The release rate of environmental DNA from juvenile and adult fish. *PLoS One* 9, e114639. doi: 10.1371/JOURNAL.PONE.0114639

Minamoto, T. (2022). Environmental DNA analysis for macro-organisms: Species distribution and more. *DNA Res.* 29, dsac018. doi: 10.1093/dnares/dsac018

Moushomi, R., Wilgar, G., Carvalho, G., Creer, S., and Seymour, M. (2019). Environmental DNA size sorting and degradation experiment indicates the state of daphnia magna mitochondrial and nuclear eDNA is subcellular. *Sci. Rep.* 9, 1–9. doi: 10.1038/s41598-019-48984-7

Renshaw, M. A., Olds, B. P., Jerde, C. L., Mcveigh, M. M., and Lodge, D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Mol. Ecol. Resour.* 15, 168–176. doi: 10.1111/1755-0998.12281

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., and Vestergaard, P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Commun. Biol.* 2, 1–9. doi: 10.1038/s42003-019-0696-8

Sevilla, R. G., Diez, A., Norén, M., Mouchel, O., Jérôme, M., Verrez-Bagnis, V., et al. (2007). Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. *Mol. Ecol. Notes* 7, 730–734. doi: 10.1111/j.1471-8286.2007.01863.x

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., and Thomsen, P. F. (2020). Population-level inferences from environmental DNA–current status and future perspectives. *Evol. Appl.* 13, 245–262. doi: 10.1111/eva.12882

Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z., Dunton, K. J., Hinks, G., and Vanmorter, S. M. (2021). Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal new Jersey, USA. *ICES J. Mar. Sci.* 78, 293–304. doi: 10.1093/icesjms/fsaa225

Sward, D., Monk, J., and Barrett, N. (2019). A systematic review of remotely operated vehicle surveys for visually assessing fish assemblages. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00134

Thalinger, B., Wolf, E., Traugott, M., and Wanzenböck, J. (2019). Monitoring spawning migrations of potamodromous fish species *via* eDNA. *Sci. Rep. 2019 91* 9, 1–11. doi: 10.1038/s41598-019-51398-0

Thompson, M. J., Putt, R. E., Gettleson, D. A., Hammer, R. M., and Stevens, R. C. (1982). "Utilization of remotely operated vehicles (Rovs) for fish survey standing stock assessment," in *Oceans Conference Record (IEEE)*. 82, 1288–1293. doi: 10.1109/oceans.1982.1151911

Tillotson, M. D., Kelly, R. P., Duda, J. J., Hoy, M., Kralj, J., and Quinn, T. P. (2018). Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biol. Conserv.* 220, 1–11. doi: 10.1016/j.biocon.2018.01.030

Trenkel, V. M., Vaz, S., Albouy, C., Amour, A. B., Duhamel, E., Laffargue, P., et al. (2019). We can reduce the impact of scientific trawling on marine ecosystems. *Mar. Ecol. Prog. Ser.* 609, 277–282. doi: 10.3354/meps12834

Tsuji, S., Ushio, M., Sakurai, S., Minamoto, T., and Yamanaka, H. (2017). Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. *PLoS One* 12, e0176608. doi: 10.1371/journal.pone.0176608

Wilcox, T. M., McKelvey, K. S., Young, M. K., Jane, S. F., Lowe, W. H., Whiteley, A. R., et al. (2013). Robust detection of rare species using environmental DNA: The importance of primer specificity. *PLoS One* 8, e59520. doi: 10.1371/journal.pone.0059520

Wright, E. S., Yilmaz, L. S., and Noguera, D. R. (2012). DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl. Environ. Microbiol.* 78, 717–725. doi: 10.1128/AEM.06516-11

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinf.* 13, 134. doi: 10.1186/1471-2105-13-134

Zhao, B., van Bodegom, P. M., and Trimbos, K. B. (2023). Environmental DNA methylation of lymnaea stagnalis varies with age and is hypermethylated compared to tissue DNA. *Mol. Ecol. Resour.* 23, 81–91. doi: 10.1111/1755-0998.13691

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 04 April 2023
DOI 10.3389/fmars.2023.1057206

OPEN ACCESS

EDITED BY
Fran Saborido-Rey,
Spanish National Research Council (CSIC), Spain

REVIEWED BY
Tim Regan,
University of Edinburgh, United Kingdom
Tereza Manousaki,
Hellenic Centre for Marine Research (HCMR), Greece

*CORRESPONDENCE
P. Martínez
paulino.martinez@usc.es

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 29 September 2022
ACCEPTED 21 March 2023
PUBLISHED 04 April 2023

CITATION
Pampín M, Casanova A, Fernández C, Blanco A, Hermida M, Vera M, Pardo BG, Coimbra RM, Cao A, Iglesias D, Carballal MJ, Villalba A and Martínez P (2023) Genetic markers associated with divergent selection against the parasite *Marteilia cochillia* in common cockle (*Cerastoderma edule*) using transcriptomics and population genomics data.
Front. Mar. Sci. 10:1057206.
doi: 10.3389/fmars.2023.1057206

Genetic markers associated with divergent selection against the parasite *Marteilia cochillia* in common cockle (*Cerastoderma edule*) using transcriptomics and population genomics data

M. Pampín[1], A. Casanova[1], C. Fernández[1], A. Blanco[1], M. Hermida[1], M. Vera[1], B. G. Pardo[1], R. M. Coimbra[1,2], A. Cao[3], D. Iglesias[3], M.J. Carballal[3], A. Villalba[3,4,5] and P. Martínez[1]*

[1]Departamento de Zoología, Genética y Antropología Física, Facultad de Veterinaria, Universidade de Santiago de Compostela, Lugo, Spain, [2]Departamento de Pesca e Aquicultura, Universidade Federal Rural de Pernambuco, Recife, Brazil, [3]Centro de Investigacións Mariñas (CIMA), Consellería do Mar da Xunta de Galicia, Vilanova de Arousa, Spain, [4]Departamento de Ciencias de la Vida, Universidad de Alcalá, Alcalá de Henares, Spain, [5]Research Centre for Experimental Marine Biology and Biotechnology (PIE), University of the Basque Country (UPV/EHU), Plentzia, Spain

The common cockle (Cerastoderma edule) plays an important role in marine ecosystems and represents a valuable socioeconomic resource for coastal communities. In 2012, the cockle beds from Ría de Arousa (Galicia, NW Spain) were seriously decimated by the protozoan Marteilia cochillia responsible for marteiliosis. We aimed to identify single nucleotide polymorphisms (SNP) markers potentially associated with resilience to marteiliosis to be used in marker-assisted selection programs for restoring affected cockle beds and recovering their production. For this, we carried out a population genomics approach using 2b-RADseq, where 38 naive samples (before the first detection of M. cochillia in 2012) from two beds of Ría de Arousa were compared with 39 affected samples collected in 2018/2019 (after several years of marteiliosis occurring in the area), collected either before (15 non-exposed samples) or during (24 exposed samples) the marteiliosis outbreak. Additionally, 767 differentially expressed genes (DEG) from a previous transcriptomic study addressed during the aforementioned 2018/19 marteiliosis outbreak, were evaluated to identify SNPs showing signals of selection. Using 2b-RADseq, 9,154 SNPs were genotyped and among them, 110 consistent outliers for divergent selection were identified. This set of SNPs was able to discriminate the samples according to their marteiliosis status (naive vs affected; exposed vs non-exposed), while another 123 SNPs were identified linked to DEGs associated with the level of infection across a temporal series. Finally, combining the population genomics and transcriptomics information, we selected the 60 most reliable SNPs associated with marteiliosis resilience. These SNPs were close to or within DEGs, and many of them were related to immune response (phagocytosis

and cell adhesion), defence, such as apoptosis, stress, and cellular cycle, among other functions. This set of SNPs will eventually be validated to develop a cost-effective genotyping tool for their application for obtaining cockle-resilient strains for marteiliosis.

KEYWORDS

SNP, bivalve, cockles, transcriptomics, population genomics, *Marteilia cochillia*, resilience

1 Introduction

The common cockle (*Cerastoderma edule*) is a marine bivalve species buried just below the surface of the sand or mud in intertidal and shallow subtidal areas of estuarine and marine coastal waters. It is widely distributed in the Northeast Atlantic, from the west coast of Africa in Senegal to the Barents Sea in Norway (Hayward and Ryland, 1995; Tyler-Walters, 2007). *C. edule* lives on average 2-4 years but can reach up to 10 years (Ponsero et al., 2009). It is a dioecious species that reaches sexual maturity at about one year of age (Maia et al., 2021). Like most bivalves, the common cockle releases gametes into the water, where external fertilization takes place (Moreira Sanmartín et al., 2016), and larvae remain in the water column for about 30 days (Creek, 1960), which allows their dispersal by marine currents (more than 100 km; Coscia et al., 2020). The spawning season runs from March to October, reaching the peak of activity between July and September when the water temperature is higher (~20°C; Maia et al., 2021).

Cockles are a highly valued shellfish species due to the range of ecosystem services that provide, e.g., the ability to reshape the seabed and alter the sediment properties (Ciutat et al., 2006; Neumeier et al., 2006; Andersen et al., 2010) depending on the substrate type (Rakotomalala et al., 2015; Eriksson et al., 2017). Like other bivalves, cockles contribute to improve water quality as filter feeders (Carmichael et al., 2012; van der Schatte Olivier et al., 2020) and aid to maintain biodiversity either indirectly, increasing the production of microphytobenthos (Swanberg, 1991) or as a direct source of food for different species (Sutherland, 1982; Beukema and Dekker, 2005). Furthermore, cockles are largely appreciated as a feeding resource for humans; Europe reported captures of 24,237 tonnes of *C. edule* in 2019, with Denmark, UK, Spain, and Portugal being the main producers (Food and Agriculture Organization of the United Nations [FAO], 2021). However, inter-annual production is unstable depending on different biotic and abiotic factors such as bacterial, viral, and parasitic infections (Lauckner, 1983; Bower et al., 1994), predation (Sutherland, 1982; Mascaró and Seed, 2000; Beukema and Dekker, 2005), food limitation (Bos et al., 2006), over-exploitation (Ferns et al., 2000), and environmental changes such as water temperature, salinity, pollution and in recent years climate change (Møhlenberg and Kiørboe, 1983; Ducrotoy et al., 1991; Beukema and Dekker, 2005; Parada and Molares, 2008; Burdon et al., 2014; Rowley et al., 2014).

Galicia (NW Spain) holds the most productive shellfisheries in Spain (Subsecretaría Subdirección General de Análisis Coordinacion y Estadística, 2020). In 2012, massive cockle losses were recorded in the Ría de Arousa, which houses one of the most productive shellfisheries on the Galician coast, associated with the protozoan parasite *Marteilia cochillia* (Villalba et al., 2014). In the following years, marteiliosis outbreaks spread to the southern Rías of Pontevedra and Vigo, almost depleting cockle production in the area. An annual pattern outbreak was recorded since the first 2012 outbreak; cases of infection are detected in newly recruited individuals in summer/early fall, followed by a progressive increase of prevalence and mortality until reaching almost 100% of cumulative mortality in the next months (Iglesias et al., 2023).

The parasite *M. cochillia* is a protistan parasite that colonizes the epithelium of the digestive gland of cockles destroying digestive diverticula and causing death due to starvation (Montaudouin et al., 2021). The complete life cycle of *M. cochillia* has not been yet disclosed. Darriba et al. (2020) observed parasitic forms (sporangia) being released through faeces into the environment. Intermediary hosts are hypothesized for its transmission, similarly to what is suspected to occur in the flat oyster (*Ostrea edulis*) infected by *M. refrigens*, where copepods of the genus *Paracartia*, i.e., *P. grani*, are involved (Audemard et al., 2002; Carrasco et al., 2015; Carballal et al., 2019). In this natural scenario, generating effective preventive measures against parasite infection is complex. Increasing resilience against *M. cochillia* through breeding programmes is an appealing approach to diminish the impact of the parasite in cockle beds, as has been demonstrated before in other bivalves (Ford and Haskin, 1987; Ragone Calvo et al., 2003; Kube et al., 2011; Proestou et al., 2016; Smits et al., 2020). This approach has been tested in natural environments, e.g. the breeding program for *Saccostrea glomerata* to obtain strains resistant to winter mortality and Qx disease caused by *Bonamia roughleyi* and *Marteilia sydneyi*, respectively (Nell et al., 2000), and in controlled conditions, such as the increased survival to ostreid herpesvirus 1 (OsHV-1) in *Crassostrea gigas* (up to 61.8%) after four generations of selection (Dégremont et al., 2015). On average, response to disease resistance selection in molluscs was higher than any other traits, such as growth (15% *vs* 10% per generation; see review of Hollenbeck and Johnston, 2018).

Genomic strategies are essential to understand the genetic basis of host-parasite interaction, for controlling marteiliosis and, eventually, for its application in breeding programs. For example, in *Crassostrea gigas* genomic prediction of OsHV-1 resistance was

more accurate (around 19%) than family-based prediction (Gutierrez et al., 2020). Genomic resources of common cockle have recently increased in the framework of the COCKLES Interreg (EAPA_458/2016) and the Scuba Cancers (ERC-2016-STG) projects, which ensured a robust genetic baseline for that purpose. A population genomics approach using 2b-RADseq along with the chromosome-level genome assembly of the species (Bruzos et al., 2022) was applied to disentangle the demographic and environmental factors underlying the common cockle structure in the Northeast Atlantic (Vera et al., 2022). Furthermore, RNAseq was applied to identify differentially expressed genes (DEG) in the digestive gland across the different infection stages. In this study, 767 DEGs, among the ~ 9000 annotated in the cockle's transcriptome, were identified when comparing samples of different infection levels across the outbreak 2018/19, many of which related to key immune pathways (Pardo et al., 2022).

The main goal of our study was to identify SNPs (Single Nucleotide Polymorphism), associated with genomic regions related to marteiliosis resilience in common cockle from Ría de Arousa for their eventual application in breeding programs and management of cockle beds. For this purpose, we followed two complementary approaches: i) identification of SNPs associated with divergent selection using groups of samples subjected to differential parasite pressure, and ii) detection of SNPs linked to the DEGs detected in response to *M. cochillia* outbreak by Pardo et al. (2022) showing significant genetic differentiation across groups with different level of infection. We used 2b-RAD and RNAseq data for genotyping anonymous and gene-linked SNPs, respectively, and further, we explored their involvement in the immune response that could explain the resilience to the parasite. A set of the most consistent SNPs were included thinking on its future validation for their potential application in breeding programs and management of common cockle beds.

2 Materials and methods

2.1 Population genomics approach

2.1.1 Sampling and DNA extraction

The sampling sites analysed in this study were selected according to relevant information on cockle marteiliosis epidemics. Marteiliosis was first detected in cockles from Galicia (NW Spain), namely from Ría de Arousa, in 2012 (Villalba et al., 2014); since then, outbreaks were recorded in this ría starting every summer/early fall, affecting each newly recruited annual cohort, and causing mass mortality (Iglesias et al., 2023). According to this information and the goals of the study, a total of 79 individuals were collected from two shellfish beds of Ría de Arousa (Table 1; Figure 1): i) 40 individuals in January 2012 before the first detection of *M. cochillia* (naive samples: NS) from Lombos do Ulla (SLO12) and O Sarrido (SSA12); and ii) 39 individuals from Lombos do Ulla in the 2018/19 period, after several generations of marteiliosis pressure ("affected"- samples, AS); among these, 15 juveniles were collected in spring 2018, before the annual marteiliosis outbreak (SLO18; non-exposed samples: NES), and 24 cockles mostly from a single cohort in spring 2019 (only four samples from September 2018) during the 2018/19 marteiliosis outbreak (SLO19; exposed samples: ES). A small portion of the gills was extracted from each individual and stored in 100% ethanol at 4°C. Genomic DNA was extracted from gills using the e.Z.N.A. E-96 mollusc DNA kit (OMEGA Bio-tech), following the manufacturer's recommendations. The quality and quantity of DNA were assessed with NanoDrop® ND-1000 (Nanodrop Technologies).

2.1.2 SNP calling and genotyping

Genotyping was performed following the 2b-RAD genotyping-by-sequencing (GBS) protocol (Wang et al., 2012). In brief, we obtained millions of 36 bp fragments in each sample produced by the digestion of genomic DNA with the A*lfI* IIb restriction enzyme (RE) (Thermo Fisher), which cut DNA at both sides of the RE site. 2b-RAD libraries were constructed at the Genomics Platform of Universidad de Santiago de Compostela (USC) and delivered to the FISABIO Platform (Valencia, Spain) for sequencing in a NextSeq 500 sequencer (Illumina). Then, reads from each individual were aligned to the common cockle genome (Bruzos et al., 2022) using Bowtie 1.1.2 (Langmead et al., 2009), and SNP calling was performed with Stacks 2.0 (Catchen et al., 2013; Rochette et al., 2019), following the parameters described by Vera et al. (2022). The RAD-tag SNP panel reported by Vera et al. (2022) mapped in the common cockle genome (Bruzos et al., 2022) was used as reference for genotyping to make feasible comparison with previous studies. Finally, some SNPs/RAD-tags or individuals were removed from the data using Plink 1.9 (Purcell et al., 2007) according to the following criteria: i) SNPs deviated from Hardy-Weinberg proportions ($p < 0.05$) in at least two sampling sites; ii) RAD-tags with more than 3 SNPs; iii) SNPs with missing data in > 50%

TABLE 1 *Cerastoderma edule* samples analysed from Ría de Arousa for the population genomics approach.

Sampling site	Date	Code	N	1st detection status	2018/19 outbreak status
Lombos do Ulla	Jan 2012	SLO12	20	NS	NES
O Sarrido	Jan 2012	SSA12	20*	NS	NES
Lombos do Ulla	May 2018	SLO18	15	AS	NES
Lombos do Ulla	April 2019	SLO19	24	AS	ES

NS and AS: naive (samples collected before 1st marteiliosis detection) and affected (samples collected after six years of marteiliosis outbreaks), respectively; NES and ES: non-exposed (samples collected before detection of the 2018/19 marteiliosis outbreak) and exposed (samples collected during the 2018/19 marteiliosis outbreak), respectively. *Two individuals were discarded from this sample after 2b-RAD filtering, totalling 18 individuals for analyses.

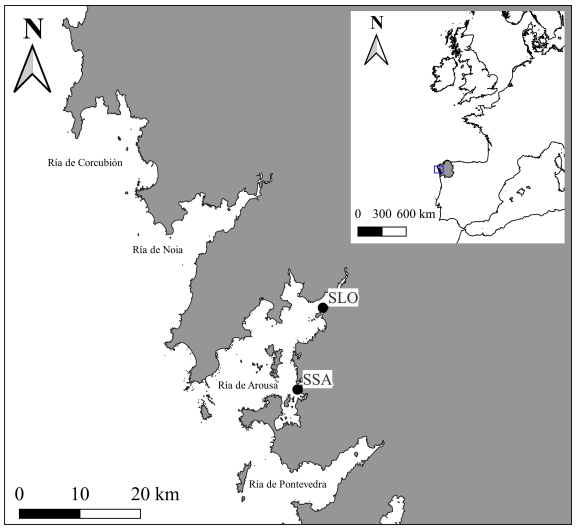


FIGURE 1
Study area showing the two sampled *Cerastoderma edule* beds in the Ría de Arousa (Galicia, NW Spain). Geographic Coordinate System – EPSG:25829 – SLO: Lombos do Ulla (518449 - 4719641); SSA: O Sarrido (514263 – 4706171).

individuals, and iv) individuals with < 30% of the SNP panel genotyped in any of the sampling sites.

2.1.3 Genetic diversity

The different subsets of SNPs used for analyses were extracted in Genepop format using the R package GENEPOPEDIT 1.0 package (Stanley et al., 2017). Allelic richness (Ar), observed (Ho) and expected (He) heterozygosity, and intrapopulation fixation index (F_{IS}) were calculated for each sample site to assess genetic diversity with the R package DiveRsity 1.9 (Keenan et al., 2013) (function "divBasic") with 1000 bootstraps. Departure from Hardy-Weinberg equilibrium (HWE) was estimated with exact tests using the enumeration method with GENEPOP 4.7.5 (Rousset, 2008).

2.1.4 Divergent selection for marteiliosis: Outlier detection

Two statistical approaches were performed to detect consistent outlier loci related to divergent selection against the neutral genomic background: i) the Bayesian F_{ST}-based method implemented in BAYESCAN v2.01 (Foll & Gaggiotti, 2008) was run using default parameters (i.e., 20 pilot runs; prior odds value of 10; burn-in of 50,000), 100,000 iterations and a sample size of 5,000; ii) the FDIST F_{ST} method implemented in ARLEQUIN v3.5 (Excoffier and Lischer, 2010) which uses a maximum likelihood approach (Beaumont & Nichols, 1996) was applied to incorporate *a priori* information regarding population structure with 100,000 simulations and 100 demes. For this purpose, we considered two scenarios where significant changes at specific genomic regions could hypothetically occur as a consequence of the differential selective marteiliosis pressure regarding the neutral background (Table 1): i) a temporal criterion (2012 *vs* 2018/19 outbreak), where naive samples (NS: SLO12, SSA12) constituted one group and affected samples (AS: SLO2018, SLO19) another; and ii) an exposure criterion, where samples from non-exposed cockles (NES: SLO12, SSA12 and SLO18) were grouped and compared to exposed samples (ES: SLO19). We used more strict parameters for ARLEQUIN (three technical replicates at $p < 0.01$), since this approach is more prone to false positives and a standard $p < 0.05$ for BAYESCAN, since it follows a more conservative approach (Narum and Hess, 2011).

All outliers detected were mapped on the *C. edule* genome (Bruzos et al., 2022), and those close to DEGs or another outlier (± 250 kb) were considered the most consistent ones (Population Genomic Candidates - PGCAND). Minor allele frequency (MAF) obtained with R package "adegenet" (Jombart, 2008) using "minorAllele" and "tab" functions respectively were calculated for further filtering steps.

2.1.5 Population structure

Global and pairwise relative coefficients of genetic differentiation (F_{ST}) were calculated between cockle sampling sites with GENEPOP 4.7.5 and R package StAMPP 1.6.2 with the 'stamppFst' function (Pembleton et al., 2013) using 10,000 bootstraps to calculate 95% confidence interval to test the null hypothesis ($F_{ST} = 0$). Genetic structure was additionally investigated through STRUCTURE 2.3.4 (Pritchard et al., 2000) using the R package ParallelStructure 1.0 (Besnier and Glover, 2013) to identify the most likely number of population units (K) in the samples. This program uses a Bayesian clustering approach to explore the population genetic units (clusters) using genotyping data. The program assigns the proportion of the genome that belongs to each of the clusters identified in each individual. Tests were performed without *a priori* information regarding the origin of samples, using an admixture model with correlated allele frequencies and burn-in of 100,000 iterations and 200,000 Markov Chain Monte Carlo steps. The number of K tested ranged from 1 to 5 (the number of sampling sites +1). For each K, ten replicates were performed to increase statistical confidence. The optimal number of K was estimated using the website program StructureSelector (Li and Liu, 2018) using different approaches: deltaK (Evanno et al., 2005), Mean LnP (K) (Pritchard et al., 2000) and those published by Puechmaille (2016). Graphical outputs were obtained with CLUMPAK (Kopelman et al., 2015). Further, a Discriminant Analysis Principal Component Analysis (DAPC) was performed with the R package adegenet to complement the STRUCTURE analysis. First, "find.cluster" function was used to assess the number of clusters in the population determining the optimal number of subpopulations with the Bayesian Information Criterion (BIC). Then, a cross-validation function was performed to detect the best number of principal components (PCs) given by the smallest mean square error (RMSE).

2.2 Transcriptomics approach

Pardo et al. (2022) identified 767 differentially expressed genes (DEG) in cockles collected before (July 2018) and at three different

times (November 2018, April 2019 and July 2019) during a natural outbreak of *M. cochillia* in 2018/2019 in Lombos do Ulla. Samples were classified histologically according to the level of infection as non-infected, mild, moderate and heavily infected. Then, DEGs were detected across a temporal series and according to the level of infection during the 2018/19 outbreak in Ría de Arousa. RNAseq data from these 767 differentially expressed genes (DEG) was used to call associated SNPs and estimate allele frequencies in each sample using SAMtools 1.9 (Li et al., 2009) with the following parameters: –skip-indels, –adjust-MQ 0, –max-depth 250. Then, SNPs were investigated for their association with the level of infection in exposed samples taken at three different times during a parasite outbreak (T1, T2 and T3). Samples were classified and pooled according to their level of infection across time using histopathology (Iglesias et al., 2023): I0: non-infected; I1: early infection; I2: moderate infection; I3: heavy infection; I4: final infection stage (Table 2). Allele frequency, missing data, expected heterozygosity and minimum allele frequency (MAF) were estimated per locus from the VCF file using all exposed samples (N = 50). SNPs that fitted the cut-off criteria of MAF > 0.05 and missing data < 30% were selected and mapped into the cockle's genome (Bruzos et al., 2022). Finally, the highest polymorphic SNP with the lowest missing data per DEG was chosen among those available.

Assuming the presence of genetic variation at DEGs related to marteiliosis response in Lombos do Ulla samples, we hypothesized that if divergent selection was occurring due to selective pressure, associated SNPs would show genetic differentiation between samples according to their level of infection, to say, on average exposed but non-infected samples would carry allelic variants related to resilience at a higher frequency, while heavily infected ones, would do for susceptibility variants. To increase statistical power, exposed samples were grouped into three sets according to their infection level: i) non-infected (15 individuals); ii) early/moderately infected (22 individuals), and iii) heavily infected/final stage of infection (13 individuals). Global F_{ST} was estimated for all selected SNPs using those three groups and their significance was estimated with exact tests ($p < 0.05$) using Genepop 4.7.5. When possible, the two most polymorphic SNPs were retrieved per DEG. Candidate SNPs were finally selected from the transcriptome approach (TCAND) from those showing the highest genetic differentiation ($p < 0.05$).

TABLE 2 *Cerastoderma edule* samples collected in Lombos do Ulla used for the transcriptomics approach classified by infection level.

Date	Code	I0	I1	I2	I3	I4
Nov-18	T1	5	5	5	5	4
Apr-19	T2	5	5	5	3	1
Jul-19	T3	5	0	2	0	0

I0: non-infected; I1: early infection; I2: moderate infection; I3: heavy infection; I4: final infection stage.

2.3 Final selection SNP panel

Once candidates from the population genomics approach (PGCAND) and transcriptomics (TCAND) were identified, a final set of SNPs was selected to design a cost-effective molecular tool including ~60 SNPs to be eventually used for obtaining resilient strains to marteiliosis. Significant SNPs from both approaches were first filtered by missing data < 30% and MAF > 0.05 and then by technical issues related to primer design and multiplexing. This preselected SNP set was next sorted according to the statistical confidence to be under divergent selection (*p*-value) and further prioritize for consistency using the following criteria: i) signals of selection in more than one approach (i.e., temporal and exposition among PGCAND); ii) more than one SNP detected in less than 250 kb; and iii) higher genetic diversity and lower missing data. Finally, when more than one outlier was found in the same gene or region (± 250 kb), only one of the markers was selected.

3 Results

3.1 Population genomics approach

Genetic diversity and structure were investigated with different sets of SNPs on several groups of samples related to the strategies followed to identify the most reliable set of outlier markers associated with divergent selection against marteiliosis (PGCAND outliers). Analyses were performed with i) the whole polymorphic SNP dataset; ii) the divergent outlier dataset; and iii) the neutral dataset, defined after removing outliers from the whole data.

3.1.1 Outlier detection and mapping

After filtering, 9,154 SNPs were retained (Table S1) from the SNP panel reported by Vera et al. (2022), and among them, 6,252 SNPs (68.3%) were polymorphic in our collection. The detection of outlier markers was performed using the 6,252 polymorphic SNPs, representing 1.6 SNPs/Mb according to the common cockle genome size (794 Mb; Bruzos et al., 2022), under the null hypothesis of neutrality across the whole genome. Thus, outliers with F_{ST} significantly above the neutral background were considered under divergent selection, while those with F_{ST} below the neutral background were considered under stabilizing selection. The two statistical methods implemented in BAYESCAN and ARLEQUIN programs, respectively, were applied and tested in the temporal (naive *vs* affected) and exposure (non-exposed *vs* exposed) scenarios (see Materials and Methods). BAYESCAN, the most conservative and sensitive to sampling error method, only detected one outlier under divergent selection in the temporal scenario, while ARLEQUIN detected a total of 213 consistent outliers ($p < 0.01$ in three technical replicates), 74 in the temporal (t) and 156 in the exposure (E) scenarios, respectively, 17 of them shared in both scenarios, including the one detected by BAYESCAN (Table S2). No outliers under stabilizing selection were detected with any of both methods. All the 213 consistent divergent outliers were mapped in the *C. edule* genome and additionally checked for their proximity to

other outliers (< 250 kb) (Table 3) or to any of the 767 DEGs reported by Pardo et al. (2022) (Table S3). A total of 110 SNPs (10 shared between both scenarios) met the criteria and were selected as the most reliable set of SNPs from the population genomics approach (PGCAND).

3.1.2 Genetic diversity

Genetic diversity using the whole 9,154 SNP dataset was slightly but significantly higher ($p < 0.05$) in samples from 2012 as compared to those from the 2018-19 period both for allelic richness (Ar: 1.362 *vs* 1.321) and expected heterozygosity (He: 0.085 *vs* 0.077). None of the samples showed global deviation from HWE ($p > 0.05$), although there was a significant deficit of heterozygotes in most samples using the confidence interval approach (Table 4A). Results were very similar when considering only the neutral SNPs (data not shown). However, when using the 110 PGCAND outlier panel (Table 4B), genetic diversity was lower in non-exposed *vs* exposed samples (NES *vs* ES: Ar: 1.330 *vs* 1.672, He: 0.100 *vs* 0.169, respectively; $p < 0.05$, Mann-Whitney tests) and further, a highly significant departure from HWE involving a remarkable heterozygote deficit was detected in the exposed sample (SLO19 F_{IS} = 0.364; $p < 0.001$). Interestingly, the high value detected in the exposed sample was mainly due to the 156 outliers of the exposure scenario (F_{IS} = 0.456; $p < 0.001$), while it remained significant but very similar across the four populations when the 74 outliers of the temporal scenario were compared (F_{IS} SLO12: 0.293; SSA12: 0.307; SLO18: 0.287 and SLO19: 0.290).

3.1.3 Genetic structure and differentiation

The global F_{ST} = 0.0032 showed low but significant ($p < 0.01$) genetic differentiation with the whole SNP dataset. Pairwise F_{ST} comparisons showed low and usually non-significant differentiation when using the neutral panel (6,006 SNPs), but highly significant differentiation when using the 110 most consistent outlier SNPs, especially when comparing naive (2012) and affected (2018/19) groups (average F_{ST} = 0.091), but also between non-exposed and exposed samples in the 2018/19 outbreak (F_{ST} = 0.033; Table 5).

The clustering method of STRUCTURE applied to the neutral dataset showed K = 1 as the optimal number of clusters according

TABLE 3 Genomic location of consistent SNPs detected in *Cerastoderma edule* following population genomics and transcriptomics approaches along with the differentially expressed genes reported by Pardo et al. (2022).

Chromosome	Size (bp)	Total PGCAND outliers	PGCAND selected	DEGs	TCAND
1	64609245	14 (8E, 6t)	5 (3E, 2t)	58	13
2	56319168	16 (11E, 5t)	10 (7E,3t)	58	10
3	55987847	20 (15E, 5t)	11 (9E, 2t)	59	10
4	52087795	14 (7E, 4t, 3S)	3 (1E, 2S)	44	4
5	50828891	21 (12E, 7t, 2S)	11 (6E, 5t)	55	8
6	40237005	9 (5E, 4t)	5 (2E, 3t)	47	9
7	39934596	7 (5E, 2t)	3 (3E)	40	8
8	39684391	11 (5E, 6t)	7 (3E, 4t)	41	6
9	39070162	4 (4E)	2 (2E)	37	9
10	38264924	15 (8E, 5t, 2S)	11 (5E,5t, 1S)	54	13
11	38197540	5 (4E, 1S)	2 (1E, 1S),	35	2
12	36327582	10 (6E, 1t, 3S)	6 (4E, 2S)	35	2
13	35955507	10 (9E, 1S)	5 (4E, 1S)	36	3
14	33816358	10 (8E, 2t)	5 (4E, 1t)	33	5
15	31726440	15 (9E, 3t, 3S)	6 (3E, 1t, 2S)	32	4
16	31510408	7 (5E, 1t, 1S)	3 (3E)	29	5
17	26587828	8 (6E, 2t)	6 (4E, 2t)	34	6
18	22603465	8 (6E, 1t, 1S)	6 (4E, 1t, 1S)	20	4
19	21711631	5 (3E, 2t)	3 (2E, 1t)	19	2
Minor scaffolds		4 (3E, 1t)	0	1	0
Total	755,460,783	213 (139E, 57t, 17S)	110 (70E, 30t, 10S)	767	123

Chromosome size according to Bruzos et al. (2022); DEGs, differentially expressed genes; PGCAND, candidates from population genetics approach: E (exposure), t (temporal) and S (shared in t and E); TCAND: candidates from transcriptomics approach.

10.3389/fmars.2023.1057206

TABLE 4 Genetic diversity in *Cerastoderma edule* samples from Ría de Arousa with: A) Whole dataset (9154 SNPs); B)110 PGCAND dataset.

A)						
		Complete SNP dataset				
Code	N	Ar	Ho	He	F_{IS}	CI
SLO12	20	1.367	0.080	0.085	**0.048**	0.021 – 0.068
SSA12	18	1.357	0.079	0.085	**0.073**	0.029 – 0.111
SLO18	15	1.281	0.072	0.073	0.013	-0.055 – 0.062
SLO19	24	1.361	0.075	0.081	**0.070**	0.050 – 0.084
B)						
		110 PGCAND				
Code	N	Ar	Ho	Code	N	Ar
SLO12	20	1.372	0.091	0.107	**0.149**	0.075 – 0.215
SSA12	18	1.369	0.107	0.120	0.113	-0.011 – 0.232
SLO18	15	1.248	0.061	0.074	**0.176**	0.022 – 0.288
SLO19	24	1.672	0.107	0.169	**0.364**	0.280 – 0.449

N, number of individuals for each location; Ar, allelic richness; Ho, observed heterozygosity; He, expected heterozygosity; F_{IS}, intrapopulation fixation index. CI, Confidence interval 95%.
In bold type significant values ($p < 0.05$).
Sample codes are shown in Table 1.

to all K estimators described by (Puechmaille, 2016) and K = 2 with ΔK and LnP (K) (Figure 2). Nevertheless, the second population unit would be spurious according to the criterion defined by Puechmaille (2016). Results with 110 PGCAND yielded different optimal Ks depending on the estimator used. According to LnP (K) (Pritchard et al., 2000), seven distinct groups were detected, while Puechamaille's estimators and ΔK reported optimal K = 3 differentiating naive (NS: SLO12/SSA12), non-exposed (NES: SL18) and exposed (ES: SLO19) samples. DAPC analyses yielded the lowest Bayesian information criterion for K = 1 with the neutral panel, and K = 2 (BIC K = 2; 161.754) with the 110 PGCAND, but K = 3 (162.722) rendered a slightly lower value showing a sample differentiation similar to that by STRUCTURE (Figure 3). All in all, two or three groups were identified, respectively, in the exposure (2 groups) and temporal (3 groups) scenarios, the non-exposed group (SLO18) being in-between the two more differentiated naive (SLO12, SSA12) and the exposed (SLO19) groups (Figure 3B).

3.2 Transcriptomics approach

More than five million SNPs were initially identified from the RNAseq reads aligned against the common cockle reference genome, which were reduced to > 950,000 high-quality SNPs after filtering with SAMtools. Among them, ~ 45,000 SNPs were inside the 767 DEGs reported by Pardo et al. (2022). After filtered by MAF (> 0.05) and missing data (< 30%), a total of 12,753 SNPs were obtained. Two SNPs with the highest MAF and the lowest missing data were retained per DEG (when available), thus constituting a total of 1,418 SNPs more manageable dataset (Tables S3, S4). Among the three groups of samples classified according to the level of infection (I0: no infection, I1: early/moderate, I3: heavy/final), 123 SNPs showed significant genetic differentiation ($p_{FST} < 0.05$) and constituted the set of candidate markers from the transcriptomics approach (TCAND) (Tables 5, S5). Among them, 41 SNPs showed a progressive increase (or decrease) in the frequency of the reference allele across the three

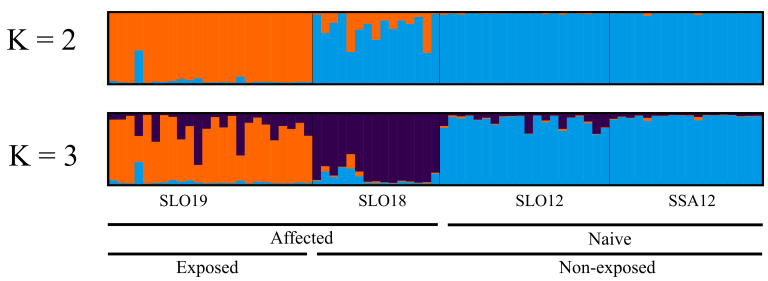


FIGURE 2
Population structure of *Cerastoderma edule* beds analysed with STRUCTURE with 110 PCAND dataset for K = 2 and K= 3). Each individual is represented by a vertical bar and its colour is proportional to the posterior probability assigned to each cluster.

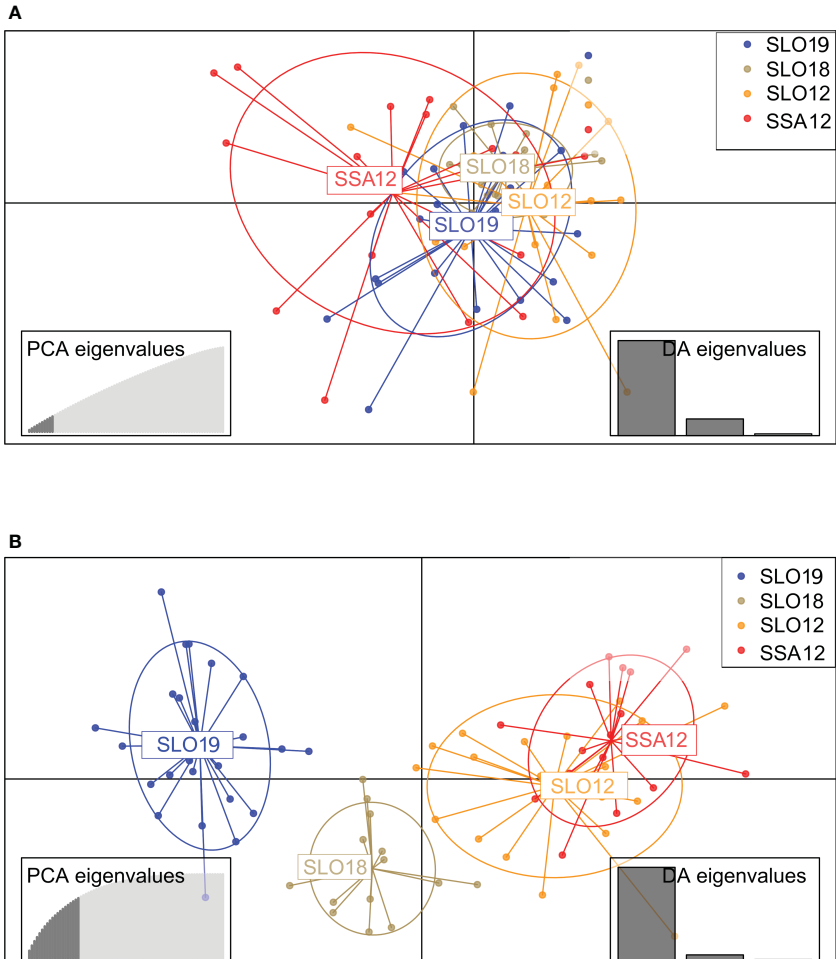


FIGURE 3
Discriminant Analysis of Principal Components (DAPC) plots of cockles *Cerastoderma edule* using: **(A)** neutral dataset (10 PCs; 18% of variance explained); **(B)** 110 PCAND dataset DAPC (20 PCs, 72% of explained variance). Codes are shown in Table 1. PCs retained according to the cross-validation method with the lowest RMSE are shown at the left bottom of each panel.

infection level groups (Table S5). Chromosomes (C) C1 and C10 housed the highest number of candidates (13 SNPs) (Table 3).

3.3 Final selection SNP panel

A final marker selection was made with the most suitable SNPs from the 110 PGCANDs and the 123 TCANDs to constitute a set of markers to be further validated by their technical and marteilioisis diagnosis usefulness (Table S6). PGCAND and TCAND were first preselected by MAF (> 0.05), missing data (< 30%) and technical criteria (± 100 bp flanking regions lacking additional polymorphisms), yielding a final set of 44 PGCAND and 38 TCAND. Then, they were ranked from the lowest to the highest genetic differentiation *p*-value for selection and additionally filtered to retain only one marker per genomic window (± 250 kb), and those with the highest F_{ST} and lowest missing data when more than one was available. A final panel of 60 SNPs was considered as a suitable set to define a cost-effective tool for validation (36 from PGCAND and 24 from TCAND approaches, respectively). These SNPs were placed in the cockle genome (Table 6). Chromosomes C5 and C10 housed the higher number of SNPs (5 and 8 SNPs respectively). Selected markers were related to catalytic functions and binding activities (17), such as the cathepsin family L, calmodulin, IgGFc-binding protein gene, Golgin subfamily A member, glutathione S-transferase sigma or proteasome related genes. Many of these genes have also been related to immune response (phagocytosis and cell adhesion), and defense, such as apoptosis, stress, and cellular cycle, among other functions (Table 6) (Niu et al., 2013; Nanut et al., 2014; Vigneron and Van den Eynde, 2014; Han et al., 2021).

4 Discussion

Cockle beds in Ría de Arousa experience annual outbreaks since 2012 (Villalba et al., 2014; Iglesias et al. 2023) due to the parasite *Marteilia cochillia*, which has collapsed its shellfishery. Although the parasite has only been recorded in restricted areas, namely Rías de Arousa, Pontevedra and Vigo in Galicia (Northwest Atlantic

TABLE 5 Pairwise F_{ST} matrix for *Cerastoderma edule* samples collected in 2012 (NS: naive) and in the 2018/19 period (AS: affected) in the Ría de Arousa, Galicia (NW Spain).

	SLO19	SLO18	SLO12	SSA12
SLO19	–	-0.001	-0.001	0.000
SLO18	**0.033**	–	0.001	**0.003**
SLO12	**0.080**	**0.065**	–	**0.002**
SSA12	**0.103**	**0.115**	0.002	–

Above diagonal: 6,006 neutral SNP dataset; below diagonal: 110 Candidate SNPs from population genomics approach (PGCAND). Significant values in bold ($p < 0.05$).

coast of Spain), Fangar and Alfacs bay in Catalonia (Northeast Mediterranean coast of Spain); Huelva (Southwest Atlantic coast of Spain); and Ría de Aveiro and Formosa Ras (North and South of Portugal, respectively) (Carrasco et al., 2013; Navas et al., 2018; Montaudouin et al., 2021; Iglesias et al., 2023), its presence threatens cockle production regardless that larval dispersal and connectivity between beds (Vera et al., 2022) could be aiding to recover recruitment every year (Iglesias et al. 2023).

Prevalence of marteiliosis and mortality of cockles have decreased since 2017 in Lombos do Ulla (Iglesias et al., 2023) and furthermore, batches from the naive shellfish bed of Noia (Ría de Noia, just north of Ría de Arousa) transplanted into Lombos do Ulla in 2017 and 2018 experienced much higher marteiliosis prevalence and mortality than those recruited in Lombos do Ulla during the same season. These observations led to hypothesize that Lombos do Ulla cockles had acquired certain resiliency to the parasite due to natural selection (Iglesias et al., 2019) and suggested that candidate genes and associated markers were probably underlying marteiliosis resilience. Thus, searching for those genetic markers was considered a main goal for recovering cockle production and preventing its expansion to other areas through breeding programs or appropriate shellfishery management. Similar approaches have been tackled for the identification of molecular markers for disease resilience in other mollusc species (de la Ballina et al., 2018; Ronza et al., 2018; Farhat et al., 2020; Leprêtre et al., 2021).

The immune system of molluscs lacks adaptative immunity and depends mostly on innate immune response, constituted by cellular and humoral responses (see review Allam and Raftos, 2015). Pardo et al. (2022) reported a set of DEGs associated with innate immune function, such as signal transduction, response to stimulus, cytoskeletal organization, pathogen recognition receptor (PRR), serine protease inhibition and antimicrobial response when analysing the transcriptomic response of cockles collected from Ría de Arousa during the same marteiliosis outbreak analysed in our study. Understanding the genetic basis of differential immune response may help to identify mechanisms conferring resilience or susceptibility to a particular disease. Moreover, detection of genetic markers associated with those differences, either responsible or not, but in linkage disequilibrium with the responsible variants, would be decisive for obtaining cockle strains resilient to the parasite.

We hypothesized that *M. cochillia* decreased prevalence and mortality rates recorded since 2017 in the inner area of the Ría de Arousa (Iglesias et al., 2019; Iglesias et al., 2023), are associated with selection of specific immune gene variants, which could eventually be identified using population genomics and transcriptomic approaches. Accordingly, outlier loci for divergent selection might be identified against the neutral background if appropriate cockle samples related to different marteiliosis pressures were compared. On the other hand, SNPs within DEGs showing genetic differentiation associated with the level of infection would be another relevant source of information for detecting consistent candidate genes holding allelic variants associated with resilience to the parasite. A similar approach has been recently reported in flat oyster *O. edulis*, where a major QTL related to resilience to *Bonamia ostreae* was identified (Vera et al., 2019; Sambade et al., 2022). Other selection models, such as overdominance, that have been reported for specific immune genes (Penn, 2002; Kekäläinen et al., 2009), could be operating and therefore our results would only explain part of the increased resilience. Furthermore, epigenetic imprinting is increasingly being claimed to be involved in immune memory in molluscs (Pradeu and Du Pasquier, 2018; Yao et al., 2021) and have also been suggested for bonamiosis resilience in flat oyster (Sambade et al., 2022).

Accordingly, we compared, on one hand, DNA samples preserved since 2012, corresponding to cockles recruited before the parasite's first detection (naive) *vs* those from the 2018/2019 period (affected), when the bed had been affected by marteiliosis for six years, the last ones including non-exposed (2018) and exposed (2019) samples to the parasite; and, on the other hand, RNAseq data from DEGs among samples with different levels of infection from the same outbreak (Pardo et al., 2022). The recent publication of the common cockle's genome (Bruzos et al., 2022) enabled to integrate all that information to look for more consistent signals of selection associated with the response to the parasite.

Using a previously validated 2b-RAD panel by Vera et al. (2022), we observed that genetic diversity in Arousa samples (He: 0.073 to 0.085) was similar to that reported in other studies in northwest Europe (He = 0.077 – 0.088, Vera et al., 2022), although the number of polymorphic loci was lower (6,252 in Ría de Arousa), an expected outcome considering the small geographic area studied. We detected a lower genetic diversity in the affected samples (SLO18/SLO19) than the non-affected ones from 2012, also expected considering the heavy mortalities and strong bottlenecks affecting cockle beds after consecutive marteiliosis outbreaks in the Ría de Arousa (average He: 0.077 *vs* 0.085; $p < 0.05$). Besides, we detected a slight heterozygote deficit ($F_{IS} > 0$) in all the samples with the whole and neutral SNP datasets, a usual observation in molluscs due to the presence of null alleles (see review Hollenbeck and Johnston, 2018). However, the heterozygote deficit was higher with

10.3389/fmars.2023.1057206

TABLE 6 List of the 60 SNPs selected from transcriptomic (TCAND) and population genomics (PGCAND) approaches positioned in the cockle *Cerastoderma edule* genome selected for resilience to *Marteilia cochillia*.

SNP_ID	Chr	Position	Ref	Alt	F_{ST}	*p*-value	MD	MAF	Inside/closest gene – distance to gene
1_18985767	1	18,985,767	T	C	0.078	0.031	0.120	0.409	Cyclic nucleotide-binding domain-containing protein 2-like – Inside
152630_16	1	21,765,580	A	G	0.106	0.003	0.247	0.069	Calmodulin - ~120 kb
1_25594873	1	25,594,873	A	C	0.121	0.006	0.000	0.490	Kazal-like domain-containing protein – Inside
172379_8	2	7,331,175	G	T	0.086	0.010	0.065	0.096	BRISC and BRCA1-A complex member 1 – Inside
2_31820629	2	31,820,629	G	A	0.157	0.001	0.000	0.470	Uncharacterized protein LOC111119482 – Inside
270375_0	2	32,784,443	C	T	0.086	0.006	0.169	0.063	Centrosome and spindle pole associated protein 1-like ~57 kb
2_44323669	2	44,323,669	A	G	0.136	0.003	0.220	0.321	Thimet oligopeptidase - Inside
209292_33	3	10,407,569	T	C	0.109	0.004	0.013	0.227	3-phosphoadenosine-5-phosphosulfate synthase ~21 kb
3_12504155	3	12,504,155	A	C	0.131	0.006	0.000	0.420	Lysosomal alpha-glucosidase-like - Inside
3_26188137	3	26,188,137	C	T	0.081	0.019	0.000	0.420	Putative zyg-1-like serine/threonine protein kinase (Fragment) – Inside
3_33568023	3	33,568,023	G	A	0.089	0.014	0.000	0.290	RNA helicase – Inside
185992_6	3	52,822,008	G	T	0.096	0.003	0.234	0.057	Peroxisomal membrane protein PMP34 ~34 kb
221361_4	4	29,246,635	T	C	0.194	0.000	0.156	0.052	Guanine nucleotide exchange factor for Rab-3A-like ~3.5 kb
19806_17	4	30,868,560	A	G	0.106	0.003	0.299	0.411	Testis-expressed sequence 11 protein ~5.4 kb
4_45423046	4	45,423,046	T	C	0.092	0.014	0.000	0.380	Apolipoprotein D-like – Inside
182698_17	5	7,332,544	T	C	0.130	0.001	0.026	0.091	Low-quality protein: la-related protein 1B-like – Inside
39442_34	5	23,254,047	C	T	0.108	0.004	0.156	0.097	Kinesin-like protein KIF1C – Inside
5_29151780	5	29,151,780	T	A	0.097	0.006	0.000	0.210	Hexosyltransferase – Inside
166822_3	5	34,008,325	G	T	0.139	0.001	0.221	0.121	Folate_rec domain-containing protein – Inside
270079_29	5	35,790,787	T	C	0.172	0.000	0.273	0.155	FH2 domain-containing protein 1 ~28.7 kb
5_48857211	5	48,857,211	T	G	0.107	0.011	0.100	0.389	Cytochrome P450 4F22 – Inside
259542_9	6	8,587,257	C	T	0.108	0.004	0.156	0.077	Monocarboxylate transporter 12-like – Inside
206963_8	6	9,943,132	A	T	0.116	0.002	0.169	0.068	Gamma-secretase subunit PEN-2-like ~23.0 kb
6_11581379	6	11,581,379	T	A	0.178	0.000	0.000	0.420	26S proteasome non-ATPase regulatory subunit 9-like – Inside
6_32304661	6	32,304,661	C	A	0.279	0.000	0.300	0.143	NPHS1 adhesion molecule, nephrin – Inside
184559_32	7	2,881,173	C	T	0.098	0.005	0.247	0.092	Vacuolar protein sorting-associated protein 4B-like ~5.1 kb
7_20566394	7	20,566,394	T	A	0.088	0.016	0.000	0.460	Serine/threonine-protein kinase/endoribonuclease IRE1-like - Inside
136636_15	8	21,903,346	T	C	0.080	0.007	0.104	0.051	Sulfotransferase family cytosolic 1B member 1-like ~19.3 kb
285008_5	8	23,212,049	A	T	0.095	0.004	0.117	0.196	Septin-2B-like – Inside
8_31342192	8	31,342,192	A	C	0.140	0.002	0.060	0.426	Sodium/glucose cotransporter 4 – Inside
13580_33	9	12,460,482	G	A	0.093	0.007	0.039	0.079	Protein PRRC1-A-like ~40.7 kb
9_34396132	9	34,396,132	A	T	0.082	0.020	0.000	0.460	Nuclear receptor 2C2-associated protein – Inside
9_34714048	9	34,714,048	T	A	0.075	0.013	0.000	0.230	Cyclic AMP-responsive element-binding protein 3-like protein 3 – Inside
37532_4	10	85,175	A	G	0.094	0.008	0.039	0.153	SRSF protein kinase 3 ~44.3 kb
10_1731406	10	1,731,406	C	T	0.106	0.012	0.000	0.440	Low-quality protein: heme-binding protein 2-like – Inside
233444_3	10	15,677,896	G	A	0.093	0.007	0.013	0.065	Glutathione S-transferase sigma class protein – Inside
142010_20	10	18,474,986	C	T	0.091	0.009	0.130	0.134	Integrator complex subunit 13 – Inside
255888_35	10	19,786,344	G	T	0.102	0.003	0.065	0.236	RIB43A-like with coiled-coils protein 2 – Inside

(Continued)

TABLE 6 Continued

SNP_ID	Chr	Position	Ref	Alt	F_{ST}	*p*-value	MD	MAF	Inside/closest gene – distance to gene
84100_31	10	20,829,424	C	A	0.114	0.003	0.052	0.133	Mitosis inhibitor protein kinase wee1– Inside
10_23286715	10	23,286,715	G	A	0.127	0.005	0.000	0.430	Seryl-tRNA synthetase – Inside
234122_25	10	37,058,575	T	A	0.100	0.005	0.208	0.095	Tyrosine-protein phosphatase non-receptor type 13-like – Inside
32590_17	11	5,177,924	G	C	0.118	0.002	0.208	0.103	Deleted in lung and esophageal cancer protein 1 – Inside
215174_5	12	18,315,989	A	G	0.284	0.000	0.273	0.310	Cathepsin L-like ~21.8 kb
126933_19	*12*	28,518,329	A	C	0.139	0.001	0.065	0.081	Poly [ADP-ribose] polymerase ~116.3 kb
268319_32	13	8,706,747	C	G	0.126	0.002	0.208	0.246	Synaptojanin-1-like – Inside
13_24547378	13	24,547,378	C	T	0.133	0.006	0.000	0.320	3-Ketoacyl-CoA thiolase, mitochondrial – Inside
149069_19	14	4,017,350	T	A	0.122	0.002	0.143	0.076	Cathepsin L1 – Inside
189805_32	14	20,522,861	A	C	0.102	0.005	0.182	0.054	Guanine nucleotide-binding protein subunit beta ~48.9 kb
146461_31	14	24,563,619	G	A	0.115	0.002	0.117	0.486	Receptor protein-tyrosine kinase – Inside
245986_32	15	19,221,254	G	C	0.102	0.004	0.208	0.452	Contactin-5 – Inside
15_30613820	15	30,613,820	T	A	0.147	0.001	0.000	0.130	Golgin subfamily A member 2 - Inside
16_14629719	16	14,629,719	A	G	0.114	0.012	0.000	0.350	Importin-5 – Inside
45581_8	17	1,657,437	A	G	0.101	0.003	0.039	0.059	NRIF2 (Fragment) – Inside
294497_25	17	4,573,773	T	C	0.109	0.003	0.221	0.444	Acetylcholine receptor subunit alpha-type acr-7-like ~29.6 kb
17_11850848	17	11,850,848	T	A	0.269	0.002	0.300	0.243	DUF4149 domain-containing protein – Inside
263423_32	18	16,930,883	T	C	0.107	0.003	0.182	0.069	Fumarylacetoacetase-like – Inside
258399_24	18	17,697,950	T	C	0.108	0.005	0.260	0.153	IgGFc-binding protein-like – Inside
18_20951750	18	20,951,750	T	C	0.084	0.021	0.000	0.490	Cathepsin L - Inside
19_13393947	19	13,393,947	A	G	0.118	0.009	0.100	0.367	Nuclear receptor subfamily 4 group A member 2 – Inside
83241_35	19	19,121,785	C	T	0.097	0.006	0.026	0.100	Dynein heavy chain 7, axonemal-like – Inside

Chr, Chromosome; Position, SNP mapping (bp) within the assembled chromosome/scaffold of C. edule. MAF, Minimum allele frequency; MD, Missing data; F_{ST}, genetic differentiation; Ref, allelic variant in the genome; Alt, alternative variant.

the 110 outlier subset, and especially with the 156 exposure outliers dataset in the *Marteilia*-exposed sample (SLO19). Additionally, this sample showed significantly higher genetic diversity with outlier loci than non-exposed samples (0.169 *vs* 0.100; $p < 0.05$). We cannot rule out other types of selection operating on these markers/genes, i.e. diversifying selection, but currently there is not a straightforward explanation for these observations. New data coming from an ongoing common garden experiment carried out in Ría de Arousa with this set of markers should shed some light on the pattern of genetic diversity observed.

Low genetic structure was detected in Ría de Arousa using the whole SNP dataset, as previously reported for small geographic areas with microsatellites (Martínez et al., 2013) and SNPs (Coscia et al., 2020) in the common cockle, and in other mollusc species with microsatellites (Diz and Presa, 2009; Vera et al., 2016). This observation points towards the high dispersal capacity of the larvae while they remain in the water column. However, we identified 156 and 74 consistent outlier loci when comparing exposed and non-exposed samples and when considering the period of collection, respectively. Among them, a total of 110 outliers were close to DEG reported by Pardo et al. (2022) and were selected as the most consistent ones from this approach. These SNPs were able to discriminate between naive *vs* affected samples and even between exposed and non-exposed samples from the 2018/19 outbreak, although with less statistical support. On the other hand, we identified 123 SNPs linked to DEGs detected in the same marteiliosis outbreak by Pardo et al. (2022), many of them related to key immune functions, which showed significant genetic differentiation among samples with different levels of infection. We speculate that these SNPs could be associated with allelic variants responsible for the differential expression and consequently under selection to marteiliosis pressure. This approach has been followed to identify markers within DEGs associated with resistance to pathologies or other traits in aquaculture species (Robledo et al., 2017; Robledo et al., 2020; Moraleda et al., 2021).

Finally, taking advantage of the chromosome-level genome assembly, we selected a final panel of the most consistent 60 SNPs, to design a cost-effective molecular tool putatively useful for the selection of resistant strains and management of cockle beds for their recovery. Validation of the "in silico" genotyping information of this SNP set with a robust genotyping tool is

being undertaken for its application in an ongoing common garden experiment in the Ría de Arousa, involving cockle stocks from naive and marteiliosis-affected shellfish beds to confirm their usefulness for discriminating resilient and susceptible cockles. This could be eventually used for the appropriate management and recovery of both cockle production and natural bed ecosystems in Galicia, which holds the most important shellfishery of this species in Spain.

Data availability statement

The transcriptomic data presented in this study are deposited in the NCBI repository (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA945848), accession number PRJNA945848. Genotype information from 2b-RADseq has been included as supplementary material in the text (Table S1).

Author contributions

PM performed the population genomics and transcriptomics analyses and selected the SNP panel in collaboration with AdC, VM and CR. FC carried out SNP calling and filtering from transcriptomics data. BA designed custom scripts and supervised the bioinformatic analysis. HM called and genotyped SNPs from 2b-RAD libraries. PB constructed 2b-RAD libraries. AsC and ID were involved in sampling in the field and histological evaluation of infection. VA and MP conceived the study, supervised the project and revised the manuscript. All authors collaborated in the manuscript and approved the final version.

Funding

This study was possible due to funds granted by the European Union through the project COCKLES within the INTERREG-AA programme (EAPA_458/2016) and the scholarship granted to M.R.M.Coimbra from CNPq (202015/2020-3). This research has been carried out under the framework of the Spain's Recovery and Resilience Plan, and more specifically under the investment line no.1 of its component number 17, where the complementary RTDI plan with the autonomous regions of Spain are foreseen, with one of those being the Complementary RTDI Plan for Marine Science which includes the Marine Science Programme for Galicia. This research in this paper corresponds to the Programme Work Package nº 6 and activity no. 6.3.A.2 about "Genetic architecture of marteiliosis resistance in common cockle" and has been funded by the Resilience and Recovery Funds).

Acknowledgments

We thank the bioinformatic support of the Centro de Supercomputación de Galicia (CESGA). M.J. Brianes Beiras, A.I. González Fontela, G. Martínez Verde, M.I. Meléndez Ramos, E. Penas Pampín, P. Rúa Santervas, P. Díaz Cedillo, J. Fernández González, I. López Maneiro, G. Pena Thomas, A. Pérez Caamaño and R. Viturro García provided technical assistance in field work and/or sample processing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1057206/full#supplementary-material

References

Allam, B., and Raftos, D. (2015). Immune responses to infectious diseases in bivalves. *J. Invertebr. Pathol.* 131, 121–136. doi: 10.1016/j.jip.2015.05.005

Andersen, T. J., Lanuru, M., van Bernem, C., Pejrup, M., and Riethmueller, R. (2010). Erodibility of a mixed mudflat dominated by microphytobenthos and *Cerastoderma edule*, East Frisian Wadden Sea, Germany. *Estuar. Coast. Shelf Sci.* 87, 197–206. doi: 10.1016/j.ecss.2009.10.014

Audemard, C., Le Roux, F., Barnaud, A., Collins, C., Sautour, B., Sauriau, P. G., et al. (2002). Needle in a haystack: Involvement of the copepod *Paracartia grani* in the life-cycle of the oyster pathogen *Marteilia refringens*. *Parasitology* 124, 315–323. doi: 10.1017/S0031182001001111

Beaumont, M. A., and Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc B Biol. Sci.* 263, 1619–1626. doi: 10.1098/rspb.1996.0237

Besnier, F., and Glover, K. A. (2013). ParallelStructure: A r package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PloS One* 8, e70651. doi: 10.1371/journal.pone.0070651

Beukema, J. J., and Dekker, R. (2005). Decline of recruitment success in cockles and other bivalves in the Wadden Sea: Possible role of climate change, predation on postlarvae and fisheries. *Mar. Ecol. Ser.* 287, 149–167. doi: 10.3354/meps287149

Bos, O. G., Hendriks, I. E., Strasser, M., Dolmer, P., and Kamermans, P. (2006). Estimation of food limitation of bivalve larvae in coastal waters of north-western Europe. *J. Sea Res.* 55, 191–206. doi: 10.1016/j.seares.2005.10.006

Bower, S. M., McGladdery, S. E., and Price, I. M. (1994). Synopsis of infectious diseases and parasites of commercially exploited shellfish. *Annu. Rev. Fish Dis.* 4, 1–199. doi: 10.1016/0959-8030(94)90028-0

Bruzos, A. L., Santamarina, M., García-souto, D., Díaz, S., Rocha, S., Zamora, J., et al. (2022). The evolution of two transmissible leukaemias colonizing the coasts of Europe. *bioRixiv*. doi: 10.1101/2022.08.06.503021

Burdon, D., Callaway, R., Elliott, M., Smith, T., and Wither, A. (2014). Mass mortalities in bivalve populations: A review of the edible cockle *Cerastoderma edule* (L.). *Estuar. Coast. Shelf Sci.* 150 (partB), 271–280. doi: 10.1016/j.ecss.2014.04.011

Carballal, M. J., Iglesias, D., Abollo, E., Cao, A., García, L., Ramilo, A., et al. (2019). "Looking for hosts of *Marteilia cochillia* in the zooplankton," in *19th International Conference on Diseases of Fish and Shellfish Abstract Book, European Association of Fish Pathologists*. (Porto - Portugal). 502.

Carmichael, R. H., Walton, W., and Clark, H. (2012). Bivalve-enhanced nitrogen removal from coastal estuaries. *Can. J. Fish. Aquat. Sci.* 69, 1131–1149. doi: 10.1139/F2012-057

Carrasco, N., Green, T., and Itoh, N. (2015). *Marteilia* spp. parasites in bivalves: A revision of recent studies. *J. Invertebr. Pathol.* 131, 43–57. doi: 10.1016/j.jip.2015.07.016

Carrasco, N., Hine, P. M., Durfort, M., Andree, K. B., Malchus, N., Lacuesta, B., et al. (2013). *Marteilia cochillia* sp. nov., a new *Marteilia* species affecting the edible cockle *Cerastoderma edule* in European waters. *Aquaculture* 412–413, 223–230. doi: 10.1016/j.aquaculture.2013.07.027

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Ciutat, A., Widdows, J., and Readman, J. W. (2006). Influence of cockle *Cerastoderma edule* bioturbation and tidal-current cycles on resuspension of sediment and polycyclic aromatic hydrocarbons. *Mar. Ecol. Prog. Ser.* 328, 51–64. doi: 10.3354/meps328051

Coscia, I., Wilmes, S. B., Ironside, J. E., Goward-Brown, A., O'Dea, E., Malham, S. K., et al. (2020). Fine-scale seascape genomics of an exploited marine species, the common cockle *Cerastoderma edule*, using a multimodelling approach. *Evol. Appl.* 13, 1854–1867. doi: 10.1111/eva.12932

Creek, G. A. (1960). The development of the lamellibranch *Cardium edule* l. *Proc. Zool. Soc London* 135, 243–260. doi: 10.1111/j.1469-7998.1960.tb05843.x

Darriba, S., Iglesias, D., and Carballal, M. J. (2020). *Marteilia cochillia* is released into seawater *via* cockle *Cerastoderma edule* faeces. *J. Invertebr. Pathol.* 172, 107364. doi: 10.1016/j.jip.2020.107364

Dégremont, L., Nourry, M., and Maurouard, E. (2015). Mass selection for survival and resistance to OsHV-1 infection in crassostrea gigas spat in field conditions: Response to selection after four generations. *Aquaculture* 446, 111–121. doi: 10.1016/j.aquaculture.2015.04.029

de la Ballina, N. R., Villalba, A., and Cao, A. (2018). Proteomic profile of the European flat oyster *Ostrea edulis* haemolymph in response to bonamiosis and identification of candidate proteins as markers of resistance. *Dis. Aquat. Organ.* 128, 127–145. doi: 10.3354/dao03220

Diz, A. P., and Presa, P. (2009). The genetic diversity pattern of *Mytilus galloprovincialis* in Galician rías (NW Iberian estuaries). *Aquaculture* 287, 278–285. doi: 10.1016/j.aquaculture.2008.10.029

Ducrotoy, J.-P., Rybarczyk, H., Souprayen, S., Bachelet, G., Beukema, J. J., Desprez, M., et al. (1991). "A comparison of the population dynamics of the cockle (Cerastoderma edule, l.) in north-Western Europe," In M. Elliot and J.-P. Ducrotoy Eds. *Estuaries and Coast: Spatial and Temporal Inrercomparisons*. (Olsen and Olsen, pp. 173–183.

Eriksson, B. K., Westra, J., Van Gerwen, I., Weerman, E., Vander Zee, E., Vander Heide, T., et al. (2017). Facilitation by ecosystem engineers enhances nutrient effects in an intertidal system. *Ecosphere* 8, e02051. doi: 10.1002/ecs2.2051

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Farhat, S., Tanguy, A., Pales, E., Guo, X., Boutet, I., Smolowitz, R., et al. (2020). Identification of variants associated with hard clam, *Mercenaria mercenaria*, resistance to quahog parasite unknown disease. *Genomics* 112, 4887–4896. doi: 10.1016/j.ygeno.2020.08.036

Ferns, P. N., Rostron, D. M., and Siman, H. Y. (2000). Effects of mechanical cockle harvesting on intertidal communities. *J. Appl. Ecol.* 37, 464–474. doi: 10.1046/j.1365-2664.2000.00509.x

Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 993, 977–993. doi: 10.1534/genetics.108.092221

Food and Agriculture Organization of the United Nations [FAO] (2021) *Fisheries and aquaculture. global capture production quantity, (1950 - 2020)*. Available at: https://www.fao.org/fishery/statistics-query/en/aquaculture/aquaculture_value (Accessed August 2, 2022).

Ford, S. E., and Haskin, H. H. (1987). Infection and mortality patterns in strains of oysters *Crassostrea virginica* selected for resistance to the parasite *Haplosporidium nelsoni* (MSX). *J. Parasitol.* 73, 368–376. doi: 10.2307/3282092

Gutierrez, A. P., Symonds, J., King, N., Steiner, K., Bean, T. P., and Houston, R. D. (2020). Potential of genomic selection for improvement of resistance to ostreid herpesvirus in pacific oyster (*Crassostrea gigas*). *Anim. Genet.* 51, 249–257. doi: 10.1111/age.12909

Han, Y., Tang, Y., Sun, S., Kim, T., Ju, K., Ri, S., et al. (2021). Modulatory function of calmodulin on phagocytosis and potential regulation mechanisms in the blood clam *Tegillarca granosa*. *Dev. Comp. Immunol.* 116, 103910. doi: 10.1016/j.dci.2020.103910

Hayward, P. J., and Ryland, J. S. (1995). *Handbook of the marine fauna of north-west europe. 2nd ed.* Eds. P. J. Hayward and J. S. Ryland (Oxford: Oxford University Press). doi: 10.1093/acprof:oso/9780199549443.001.0001

Hollenbeck, C. M., and Johnston, I. A. (2018). Genomic tools and selective breeding in molluscs. *Front. Genet.* 9. doi: 10.3389/fgene.2018.00253

Iglesias, D., Villalba, A., Cao, A., and Carballal, M. (2019). "Is natural selection enhancing resistance against marteiliosis in cockles recruited in the inner side of the ria of arousa?," in *19th international conference on diseases of fish and shellfish, Oporto, Portugal* (Porto, Portugal: European Association of Fish Pathologists), 184. Available at: https://eafp.org/wp-content/uploads/2020/01/2019-porto-abstract-book.pdf.

Iglesias, D., Villalba, A., Mariño, C., No, E., and Carballal, M. J. (2023). *Long-Term Survey Discloses a Shift in the Dynamics Pattern of an Emerging Disease of Cockles Cerastoderma Edule, Marteiliosis, And Raises Hypothesis to Explain it*. doi: 10.2139/ssrn.4381126

Jombart, T. (2008). Adegenet: A r package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Keenan, K., Mcginnity, P., Cross, T. F., Crozier, W. W., and Prodöhl, P. A. (2013). DiveRsity: An r package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* 4, 782–788. doi: 10.1111/2041-210X.12067

Kekäläinen, J., Vallunen, J. A., Primmer, C. R., Rättyä, J., and Taskinen, J. (2009). Signals of major histocompatibility complex overdominance in a wild salmonid population. *Proc. R. Soc B Biol. Sci.* 276, 3133–3140. doi: 10.1098/rspb.2009.0727

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi: 10.1111/1755-0998.12387

Kube, P. D., Cunningham, M., Dominik, S., Parkinson, S., Henshall, J., Finn, B., et al. (2011). *Enhancement of the pacific oyster selective breeding program*. FRDC and SeafoodCRC Final Report Project No. 2006/227. (Hobart, TAS, Australia: CSIRO Marine and Atmospheric Research).

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25

Lauckner, G. (1983). Diseases of Mollusca: Bivalvia. In Ed. O. Kinne *Diseases of Marine Animals* (Hamburgh: Biologische Anstalt Helgoland) II (ed. O. Kinne) 805–817.

Leprêtre, M., Faury, N., Segarra, A., Claverol, S., Degremont, L., Palos-ladeiro, M., et al. (2021). Comparative proteomics of ostreid herpesvirus 1 and pacific oyster interactions with two families exhibiting contrasted susceptibility to viral infection. *Front. Immunol.* 11. doi: 10.3389/fimmu.2020.621994

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Y., and Liu, J. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* 18, 176–177. doi: 10.1111/1755-0998.12719

Møhlenberg, F., and Kiørboe, T. (1983). Burrowing and avoidance behaviour in marine organisms exposed to pesticide-contaminated sediment. *Mar. pollut. Bull.* 14, 57–60. doi: 10.1016/0025-326X(83)90192-3

Maia, F., Barroso, C. M., and Gaspar, M. B. (2021). Biology of the common cockle *Cerastoderma edule* (Linnaeus 1758) in ria de aveiro (NW portugal): Implications for fisheries management. *J. Sea Res.* 171, 102024. doi: 10.1016/j.seares.2021.102024

Martínez, L., Méndez, J., Insua, A., Arias-Pérez, A., and Freire, R. (2013). Genetic diversity and population differentiation in the cockle *Cerastoderma edule* estimated by microsatellite markers. *Helgol. Mar. Res.* 67, 179–189. doi: 10.1007/s10152-012-0314-3

Mascaró, M., and Seed, R. (2000). Foraging behavior of *Carcinus maenas* (L.): Comparisons of size-selective predation on four species of bivalve prey. *J. Shellfish Res.* 19, 283–291. doi: 10.1007/s002270100677

Montaudouin, X., Arzul, I., Cao, A., Carballal, M. J., Chollet, B., Correia, S., et al. (2021). *Catalogue of parasites and diseases of the common cockle cerastoderma edule - COCKLES PROJECT* (Aveiro: UA Editora – Universidade de Aveiro). doi: 10.34624/9a9c-9j21

Moraleda, C. P., Robledo, D., Gutiérrez, A. P., Del-Pozo, J., Yáñez, J. M., and Houston, R. D. (2021). Investigating mechanisms underlying genetic resistance to

salmon rickettsial syndrome in Atlantic salmon using RNA sequencing. *BMC Genomics* 22, 156. doi: 10.1186/s12864-021-07443-2

Moreira Sanmartín, R., Roberts, S., and Figueras, A. (2016). "Molluscs," in *Genomics in aquaculture*. Eds. S. MacKenzie and S. Jentoft (London: Elsevier), 223–245. doi: 10.1016/B978-0-12-801418-9.00009-3

Nanut, M. P., Sabotič, J., Jewett, A., and Kos, J. (2014). Cysteine cathepsins as regulators of the cytotoxicity of NK and T cells. *Front. Immunol.* 5. doi: 10.3389/fimmu.2014.00616

Narum, S. R., and Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* 11, 184–194. doi: 10.1111/j.1755-0998.2011.02987.x

Navas, J. I., López-Sanmartín, M., Perez-Miguel, M., and Drake, P. (2018). "Detection of *Martelia cochillia* in *Cerastoderma edule* from Huelva coast (Andalusia Spain), Montpellier," in *AQUA 2018 - we r aquaculture*, 463.

Nell, J. A., Smith, I. R., and McPhee, C. C. (2000). The Sydney rock oyster *Saccostrea glomerata* (Gould 1850) breeding programme: progress and goals. *Aquac. Res.* 31, 45–49. doi: 10.1016/S0044-8486(03)00133-9

Neumeier, U., Lucas, C. H., and Collins, M. (2006). Erodibility and erosion patterns of mudflat sediments investigated using an annular flume. *Aquat. Ecol.* 40, 543–554. doi: 10.1007/s10452-004-0189-8

Niu, D., Jin, K., Wang, L., Feng, B., and Li, J. (2013). Molecular characterization and expression analysis of four cathepsin l genes in the razor clam, *Sinonovacula constricta*. *Fish Shellfish Immunol.* 35, 581–588. doi: 10.1016/j.fsi.2013.06.001

Parada, J. M., and Molares, J. (2008). Natural mortality of the cockle *Cerastoderma edule* (L.) from the Ría of Arousa (NW Spain) intertidal zone. *Rev. Biol. Mar. Oceanogr.* 43, 501–511. doi: 10.4067/S0718-19572008000300009

Pardo, B. G., Fernández, C., Pampín, M., Blanco, A., Iglesias, D., Cao, A., et al. (2022). Transcriptome characterization of the common cockle (*Cerastoderma edule*) after exposure to a *Marteilia cochillia* outbreak. *bioRixiv*. doi: 10.1101/2022.10.18.512677

Pembleton, L. W., Cogan, N. O. I., and Forster, J. W. (2013). StAMPP: An r package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 13, 946–952. doi: 10.1111/1755-0998.12129

Penn, D. J. (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. *Ethology* 108, 1–21. doi: 10.1046/j.1439-0310.2002.00768.x

Ponsero, A., Dabouineau, L., and Allain, J. (2009). Modelling of common European cockle *Cerastoderma edule* fishing grounds aimed at sustainable management of traditional harvesting. *Fish. Sci.* 75, 839–850. doi: 10.1007/s12562-009-0110-4

Pradeu, T., and Du Pasquier, L. (2018). Immunological memory: What's in a name? *Immunol. Rev.* 283, 7–20. doi: 10.1111/imr.12652

Pritchard, J., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Proestou, D. A., Vinyard, B. T., Corbett, R. J., Piesz, J., Allen, S. K., Small, J. M., et al. (2016). Performance of selectively-bred lines of eastern oyster, crassostrea virginica, across eastern US estuaries. *Aquaculture* 464, 17–27. doi: 10.1016/j.aquaculture.2016.06.012

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* 16, 608–627. doi: 10.1111/1755-0998.12512

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ragone Calvo, L. M., Calvo, G. W., and Burreson, E. M. (2003). Dual disease resistance in a selectively bred eastern oyster, *Crassostrea virginica*, strain tested in Chesapeake bay. *Aquaculture* 220, 69–87. doi: 10.1016/S0044-8486(02)00399-X

Rakotomalala, C., Grangeré, K., Ubertini, M., Forêt, M., and Orvain, F. (2015). Modelling the effect of *Cerastoderma edule* bioturbation on microphytobenthos resuspension towards the planktonic food web of estuarine ecosystem. *Ecol. Modell.* 316, 155–167. doi: 10.1016/j.ecolmodel.2015.08.010

Robledo, D., Hamilton, A., Gutiérrez, A. P., Bron, J. E., and Houston, R. D. (2020). Characterising the mechanisms underlying genetic resistance to amoebic gill disease in Atlantic salmon using RNA sequencing. *BMC Genomics* 21, 271. doi: 10.1186/s12864-020-6694-x

Robledo, D., Rubiolo, J. A., Cabaleiro, S., Martínez, P., and Bouza, C. (2017). Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*). *Sci. Rep.* 7, 12105. doi: 10.1038/s41598-017-12459-4

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28, 4737–4754. doi: 10.1111/mec.15253

Ronza, P., Cao, A., Robledo, D., Gómez-Tato, A., Álvarez-Dios, J. A., Hasanuzzaman, A. F. M., et al. (2018). Long-term affected flat oyster (*Ostrea edulis*) haemocytes show differential gene expression profiles from naïve oysters in response to *Bonamia ostreae*. *Genomics* 110, 390–398. doi: 10.1016/j.ygeno.2018.04.002

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Rowley, A. F., Cross, M. E., Culloty, S. C., Lynch, S. A., Mackenzie, C. L., Morgan, E., et al. (2014). The potential impact of climate change on the infectious diseases of commercially important shellfish populations in the Irish Sea–a review. *ICES J. Mar. Sci.* 71, 741–759. doi: 10.1093/icesjms/fst234

Sambade, I. M., Casanova, A., Blanco, A., Gundappa, M. K., Bean, T. P., Macqueen, D. J., et al. (2022). A single genomic region involving a putative chromosome rearrangement in flat oyster (*Ostrea edulis*) is associated with differential host resilience to the parasite *Bonamia ostreae*. *Evol. Appl.* 00, 1–15. doi: 10.1111/eva.13446

Smits, M., Enez, F., Ferraresso, S., Dalla Rovere, G., Vetois, E., Auvray, J. F., et al. (2020). Potential for genetic improvement of resistance to *Perkinsus olseni* in the Manila clam, *Ruditapes philippinarum*, using DNA parentage assignment and mass spawning. *Front. Vet. Sci.* 7. doi: 10.3389/fvets.2020.579840

Stanley, R. R. E. E., Jeffery, N. W., Wringe, B. F., DiBacco, C., and Bradbury, I. R. (2017). GENEPOPEDIT: a simple and flexible tool for manipulating multilocus molecular data in r. *Mol. Ecol. Resour.* 17, 12–18. doi: 10.1111/1755-0998.12569

Subsecretaría Subdirección General de Análisis Coordinacion y Estadística (2020) *Estadísticas pesqueras noviembre 2020 [Fishery statistics November 2020]* (Ministerio de Agricultura Pesca y Alimentación). Available at: https://www.mapa.gob.es/es/estadistica/temas/estadisticas-pesqueras/ (Accessed June 15, 2022).

Sutherland, W. J. (1982). Do oystercatchers select the most profitable cockles? *Anim. Behav.* 30, 853. doi: 10.1016/S0003-3472(82)80159-0

Swanberg, I. (1991). The influence of the filter-feeding bivalve *Cerastoderma edule* l. on microphytobenthos: a laboratory study. *J. Exp. Mar. Bio. Ecol.* 151, 93–111. doi: 10.1016/0022-0981(91)90018-R

Tyler-Walters, H. (2007). "Cerastoderma edule (Common cockle)," in *Mar. life inf. netw. biol. sensit. key inf. rev*. Eds. H. Tyler-Walters and K. Hiscock (United Kingdom: Plymouth Mar. Biol. Assoc), 1–23. doi: 10.17031/marlinsp.1384.1

van der Schatte Olivier, A., Jones, L., Vay, L., Christie, M., Wilson, J., and Malham, S. K. (2020). A global review of the ecosystem services provided by bivalve aquaculture. *Rev. Aquac.* 12, 3–25. doi: 10.1111/raq.12301

Vera, M., Carlsson, J., Carlsson, J., Cross, T., Lynch, S., Kamermans, P., et al. (2016). Current genetic status, temporal stability and structure of the remnant wild European flat oyster populations: conservation and restoring implications. *Mar. Biol.* 163, 239. doi: 10.1007/s00227-016-3012-x

Vera, M., Maroso, F., Wilmes, S. B., Hermida, M., Blanco, A., Fernández, C., et al. (2022). Genomic survey of edible cockle (*Cerastoderma edule*) in the northeast Atlantic: a baseline for sustainable management of its wild resources. *Evol. Appl.* 15, 262–285. doi: 10.1111/eva.13340

Vera, M., Pardo, B. G., Cao, A., Vilas, R., Fernández, C., Blanco, A., et al. (2019). Signatures of selection for bonamiosis resistance in European flat oyster (*Ostrea edulis*): New genomic tools for breeding programs and management of natural resources. *Evol. Appl.* 12, 1781–1796. doi: 10.1111/eva.12832

Vigneron, N., and Van den Eynde, B. J. (2014). Proteasome subtypes and regulators in the processing of antigenic peptides presented by class I molecules of the major histocompatibility complex. *Biomolecules* 4, 994–1025. doi: 10.3390/biom4040994

Villalba, A., Iglesias, D., Ramilo, A., Darriba, S., Parada, J. M., No, E., et al. (2014). Cockle *Cerastoderma edule* fishery collapse in the Ria de Arousa (Galicia, NW Spain) associated with the protistan parasite marteilia cochillia. *Dis. Aquat. Organ.* 109, 55–80. doi: 10.3354/dao02723

Wang, S., Meyer, E., Mckay, J. K., and Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* 9, 808–810. doi: 10.1038/nmeth.2023

Yao, T., Lu, J., Bai, C., Xie, Z., and Ye, L. (2021). The enhanced immune protection in small abalone *Haliotis diversicolor* against a secondary infection with. *Vibrio harveyi*. *Front. Immunol.* 12. doi: 10.3389/fimmu.2021.685896

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 18 April 2023
DOI 10.3389/fmars.2023.1061530

OPEN ACCESS

EDITED BY
Ana Rotter,
National Institute of Biology (NIB), Slovenia

REVIEWED BY
Luca Mirimin,
Galway-Mayo Institute of Technology, Ireland
Simon Goodman,
University of Leeds, United Kingdom

*CORRESPONDENCE
Natalia Petit-Marty
npetit@iim.csic.es
Laura Casas
lauracasas@iim.csic.es
Fran Saborido-Rey
fran@iim.csic.es

SPECIALTY SECTION
This article was submitted to Marine Fisheries, Aquaculture and Living Resources, a section of the journal Frontiers in Marine Science

RECEIVED 04 October 2022
ACCEPTED 29 March 2023
PUBLISHED 18 April 2023

CITATION
Petit-Marty N, Casas L and Saborido-Rey F (2023) State-of-the-art of data analyses in environmental DNA approaches towards its applicability to sustainable fisheries management.
Front. Mar. Sci. 10:1061530.
doi: 10.3389/fmars.2023.1061530

State-of-the-art of data analyses in environmental DNA approaches towards its applicability to sustainable fisheries management

Natalia Petit-Marty*, Laura Casas* and Fran Saborido-Rey*

Department of Fisheries Ecology, Institute of Marine Research-Spanish National Research Council (IIM-CSIC), Vigo, Spain

An increasing number of studies using marine environmental DNA (eDNA) approaches are showing its potential application in marine fisheries management by helping and simplifying some of the labor-intensive traditional surveys required to assess exploited populations and ecosystem status. eDNA approaches (i.e. metabarcoding and targeted) can support to ecosystem-based fisheries management by providing information on species composition; surveillance of invasive, rare and/or endangered species; and providing estimates of species abundance. Due to these potential uses in fisheries and conservation sciences, the number of studies applying eDNA approaches in marine habitats has expanded in the very last few years. However, a lack of consistency across studies when applying pipelines for data analyses, makes results difficult to compare among them. Such lack of consistency is partially caused by poor knowledge in the management of raw sequences data, and analytical methods allowing comparative results. Hence, we review here the essential steps of eDNA data processing and analyses to get sound, reproducible, and comparable results, providing a set of bioinformatics tools useful for each step. Altogether this review presents the state of the art of eDNA data analyses towards a comprehensive application in fisheries management promoting sustainability.

KEYWORDS

eDNA, bioinformatic pipeline, qPCR, dPCR, marine ecosystems, metabarcoding, vulnerable marine ecosystems

1 Introduction

Management of marine fisheries requires a large amount of data collection. Such data include monitoring of fish stock abundance, biomass, and several life history parameters. Fisheries-independent research surveys based on capture or visual census of the fish species of interest provide a significant proportion of the data to assess commercial fish stocks. These surveys are complemented by studies determining maturity, fecundity, age and sex

structure of the stocks. Moreover, fisheries-dependent data are also collected to monitor *in situ* fisheries accordingly to the authority's regulation (Bradley et al., 2019; Urban et al., 2023). Altogether, these methods give essential information to assess the status of fisheries resources and to provide scientific advice on the best management strategies to achieve a sustainable exploitation.

At present, the effects of climate change (Pinsky et al., 2020) and the detected loss of marine biodiversity (Payne et al., 2016) impose new levels of complexities in the assessments of the sustainability of marine fisheries. Climate change is altering fish species distributions and likely accelerating adaptation events, potentially impacting species abundances and ecosystems structure (Perry et al., 2005; Gallego et al., 2020; Pinsky et al., 2020; Palacios-Abrantes et al., 2022). Moreover, the inception of ecosystem-based fisheries management (EBFM, (Pikitch et al., 2004)), highlights as crucial the evaluation of the effect of fishing operations on marine ecosystems. EBFM should ensure long-term productive, healthy, and resilient ecosystems. Thus, it is of maximum priority to characterize and monitor the ecosystem status at places where fishing operates, especially in ecosystems defined as vulnerable by FAO (2009). Hence, data collection to ensure sustainable management of fisheries is scaling up in complexity requiring specialized training and high budgets.

As fisheries science is urgently needing to collect an increasing amount of data to assess the long-term sustainability of fish resources, environmental DNA approaches (eDNA, Taberlet et al., 2012) have become a promising technology in simplifying survey tasks. Hence, studies using eDNA to investigate marine species richness, distribution, and abundance expanded in the very last years (reviewed Gilbey et al., 2021; Jo et al., 2022; Miya, 2022; Ramírez-Amaro et al., 2022; Rourke et al., 2022). eDNA refers to DNA that can be extracted from environmental samples (such as soil, water, or air), without first isolating any target organisms. It is characterized by a complex mixture of genomic DNA from many different organisms and by possible degradation (Taberlet et al., 2012). Environmental DNA can be classified into two types, organismal DNA, and extra-organismal DNA. The source of organismal DNA is from whole individuals most probably alive at the time of sampling, whereas extra-organismal DNA originates from a variety of sources such as part of tissue replacement, metabolic waste, gametes, etc. (Rodriguez-Ezpeleta et al., 2021). Thus, eDNA samples are composed of a complex mixture of both types of DNA (i.e., organismal and extra-organismal) from various sources and in varying proportions (Taberlet et al., 2012). This heterogeneity of eDNA samples provides for raw data to potentially answer wide-rank of scientific questions, though it also brings uncertainties at the time of interpreting and comparing results across studies.

Coupled with sequencing and laboratory technologies, the analysis of eDNA samples can be used to identify species and to characterize species communities associated with the environment from where the DNA was extracted. Basically, eDNA approaches are based on the extraction, amplification, and quantification of the DNA present in environmental samples. eDNA approaches can be divided into two main groups: (1) eDNA metabarcoding or community approach, and (2) species-specific or targeted approach (Seymour, 2019; Tsuji et al., 2019). Both strategies share the key characteristic of non-invasive indirect sampling but differ in their purpose, methodology, and interpretations. The current state of the art of sequencing and quantification technologies makes both eDNA approaches very powerful in helping sustainable marine fisheries surveys and monitoring (Gilbey et al., 2021; Ramírez-Amaro et al., 2022; Rourke et al., 2022). On the one hand, the metabarcoding or community eDNA approach is intended to characterize the species community living in a given environment (Alberdi et al., 2017). Therefore, it can greatly facilitate environmental monitoring surveys by providing data on species assemblages and changes in ecosystems where fishing operates. On the other hand, the targeted species-specific approach can inform on population dynamics such as geographical displacements and provide for quantitative estimates of abundance. In addition, intermediate approaches between targeted and metabarcoding eDNA approaches, are also being designed and implemented in aquatic ecosystems (e.g. Wilcox et al., 2020). Nevertheless, there are gaps in the knowledge of eDNA dynamics in the sea, which should be accounted for in a comprehensive implementation of eDNA approaches in fisheries data surveys.

Since the sea is a highly variable environment, and applications of eDNA cover a wide spectrum of scientific questions (Barnes & Turner, 2016), it is noted disparity among results of eDNA analyses. Such incongruencies have raised concerns on the potential applicability of eDNA surveys to help fisheries sustainable management (Hansen et al., 2018; Ramírez-Amaro et al., 2022). It has been highlighted that part of the disparate results found among studies could be explained by the heterogeneity across pipelines and bioinformatics tools used for data processing and analyses (Creedy et al., 2022). eDNA data analyses entail the knowledge of the analyzing algorithms and prior assumptions made by these. For example, the choice of the algorithm to classify taxonomically the composition of an environmental DNA sample, as well as the completeness of the reference database used for it, can have a significant impact on diversity estimates and should be made according to the goals of the study (Mathon et al., 2021, Liu and Zhang, 2021, Miya, 2022). Additionally, it is imperative to understand what are the limitations of each particular eDNA study with respect to the power of detection (Burian et al., 2021), especially when reporting the presence/absence or abundance of target species.

Several software, pipelines, and scripts exist to manage and analyze the thousands of millions of sequences obtained in a typical eDNA study. User-friendly bioinformatics tools for the analysis of eDNA allow for obtaining results without bioinformatics or computational knowledge. Nevertheless, understanding of every step of the analytical pipeline is relevant for the reliable and efficient application of eDNA approaches. When correctly applied, algorithms and models included within eDNA data analysis pipelines allow infer and, many times solve common errors related to eDNA approaches. Therefore, to know the state of the art of data analysis for eDNA approaches is of great importance before planning an eDNA study. Thus, the aim of this review is to present and explain the main analytical approaches used in eDNA studies towards to facilitate its application in helping

fisheries' sustainable management. This review expands and complements the review made by Ramírez-Amaro et al. (2022) included in this Research Topic Issue. Nevertheless, whereas Ramírez-Amaro et al. (2022) center on sampling and laboratory steps, here we focus on the next step: the analyses of the eDNA data.

2 eDNA Data analyses pipelines

Several tasks related to fisheries sustainable management can be supported by eDNA approaches: (1) Surveillance of rare (i.e. endangered) and invasive species, (2) Describing and monitoring Vulnerable Marine Ecosystems, (3) Monitoring fish assemblages, and (4) Quantifying biomass and/or abundance of target economically important species. Case-study examples of the applicability of both eDNA approaches to these four tasks are presented in BOX 1. eDNA metabarcoding can be used to describe species richness (e.g. in Leray and Knowlton, 2015; McClenaghan et al., 2020; Boulanger et al., 2021; Liu and Zhang, 2021; Sato et al., 2021; Yu et al., 2021; Fonseca et al., 2022; Good et al., 2022; Keck et al., 2022), surveillance of the presence/absence of target species such as invasive, or endangered (e.g. in Jerde et al., 2011; Boussarie et al., 2018; Holman et al., 2019; Bonfil et al., 2021; Manfrin et al., 2022), and monitor changes in species composition (e.g. in Stat et al., 2017; Jeunen et al., 2019; Gallego et al., 2020; Afzali et al., 2021; Gold et al., 2021; Russo et al., 2021; Stoeckle et al., 2021; Valdivia-Carrillo et al., 2021; West et al., 2021; Maiello et al., 2022). eDNA targeted approach is better suited for quantification of the abundance of target species (e.g. in Takahara et al., 2012; Lacoursière-Roussel et al., 2016; Yamamoto et al., 2016; Knudsen et al., 2019; Salter et al., 2019; Brys et al., 2021; Fukaya et al., 2021; Shelton et al., 2022; Urban et al., 2023), and therefore it could be applied in monitoring fish stocks and vulnerable species.

Data processing for eDNA metabarcoding requires several steps which include decision-making on the best algorithms to be used accordingly to the previous knowledge of the species community and the goal of the study, while for eDNA targeted approach it is simplified. Available, open-access bioinformatics tools with potential use for eDNA approaches in fisheries are listed in Supplementary Table 1.

3 eDNA metabarcoding

Metabarcoding refers to the amplification and sequencing of a DNA region, known as barcoding, simultaneously for a species community or taxonomic group. Most of the widely used barcoding regions belong to the mitochondrial genome. It is because mitochondria are present in multiple copies within cells and therefore are most likely to be amplified from environmental samples which contain low DNA content by species. PCR products are then sequenced by High Throughput Sequencing (HTS), commonly in Illumina platforms, and the obtained sequences are subsequently clustered by similarity, and then taxonomically classified.

The selection of the barcoding region to be obtained from an eDNA sample depends on the goal of the study as well as on the group of species intended to be surveyed. Universal primers for barcoding regions are designed to amplify taxonomically related species. The taxonomical assignment is performed by sequence identity searches against reference sequence databases. The quality of the results of the metabarcoding approach depends on the correct taxonomical assignments of the species present in an eDNA sample. Therefore, the correct selection of the barcoding region to be used and the completeness of the reference database are the keys, and also the main limitations of the metabarcoding approach (Miya, 2022).

Bioinformatics pipelines are the result of executing linked instructions to process a large amount of sequence data (often several hundred million reads) using scripts, software, and databases. Computational requirements for an eDNA metabarcoding data analysis will depend on the sampling magnitude, number of replicates, and goals of the study, though hundreds of gigabytes of data are expected from metabarcoding sequencing. Data analyses are mostly performed on Unix platforms, and in high-performance computing systems. However, pipelines such as PEMA (Zafeiropoulos et al., 2020) can be installed from regular computers to cloud or HPC environments. Processing times of different pipelines have been compared in Mathon et al., 2021. Requirements of any particular pipeline used in eDNA metabarcoding can be found in the source webpage listed in Supplementary Table 1.

Data analysis protocols should be adapted to the survey design and the ecological question under study (Alberdi et al., 2017; Zinger et al., 2019). However, independently of the particularities of each study, the metabarcoding data analyses pipeline has a number of common steps that are essential and summarized in Figure 1.

Because the first application of eDNA metabarcoding was for the study of microorganisms, there are very well-established packages for the study of microbial communities based on the 16S metabarcoding region, such as MOTHUR (Schloss et al., 2009), Qiime2 (Bolyen et al., 2019), USEARCH (Edgar, 2010), and RDP pipeline (Cole et al., 2014). However, the extended use of eDNA in conservation and marine sciences has produced an increasing number of additional bioinformatics tools as well as reference databases for the analysis of eukaryote taxonomic groups. Therefore, the list of tools provided in Supplementary Table 1 does not intend to conduct a systematic review of all existing metabarcoding pipelines. Instead, we focus on those that have been proven useful for the study of marine biodiversity and fish communities, are open access, and are well-documented.

To date, for the study of fish communities, MitoFish, MitoAnnotator, and MiFish pipelines constitute a key platform (Sato et al., 2018), OBITools (Boyer et al., 2016) and ANACAPA (Curd et al., 2019) pipelines have been tested offering reliable results for fish communities, while mBrave (Ratnasingham, 2019) linked to BOLD system platform (Ratnasingham and Hebert, 2007), and Meta-Fish-Lib (Collins et al., 2021) are powerful tools to increase the knowledge in fish biodiversity surveys. Most of the pipelines presented in Supplementary Table 1 do not perform all steps needed to perform a complete eDNA metabarcoding analysis. However, most of them are modular allowing users to run different modules separately and choose better options according to their case study.

BOX 1 Case-study examples of four tasks required in sustainable management of marine fisheries

Task 1: Surveillance of rare species
Case-study: *Environmental DNA illuminates the dark diversity of sharks.*
Boussarie et al. (2018) used eDNA metabarcoding to detect shark species that are no longer detected by traditional surveys in habitats where they formerly occurred in New Caledonian. The authors performed eDNA sampling from a wide area encompassed by a gradient of human density. eDNA results were compared to visual censuses performed over more than 20 years, and data recorded by baited video over two years. An elasmobranch-specific COI primer set was used for the amplification of eDNA metabarcoding markers, and data analysis was performed with OBITools (Boyer et al., 2016), using a reference database with data retrieved from BOLD database (Hebert & Ratnasingham, 2007) for Elasmobranchii species. The reference database was enriched with homologous sequences from other non-elasmobranch taxa retrieved from the EMBL-EBI database by performing in silico PCR, to control for misidentification, given the short length of the amplified sequence (127bp). The Vegan R package (Dixon, 2003) was used for rarefaction analyses, followed by model fitting using the nls function in the stats package Models were fitted for the three methods independently (visual, video, and eDNA).
The results show that environmental DNA (eDNA) detected 44% more shark species than traditional underwater visual censuses and baited videos across the New Caledonian archipelago (south-western Pacific). eDNA analysis revealed the presence of previously unobserved shark species in human-impacted areas. The authors found that the main limitation of the method was the imperfect nature of currently available metabarcoding primers, which introduced a degree of uncertainty regarding the identification of certain species.

Task 2: Describing and monitoring Vulnerable Marine Ecosystems
Case-study: *Detection of community-wide impacts of bottom trawl fishing on deep-sea assemblages using environmental DNA metabarcoding.*
Good et al., 2022 assessed the impact of deep-sea trawling on open slope regions and marine canyons of the Mediterranean Sea for benthic meiofaunal species communities while assessing the power of eDNA metabarcoding in the detection of changes in species diversity richness and composition. The authors used different barcoding regions COI and 18S for targeting broad eukaryote diversity. eDNA samples from sediments were collected from five stations characterized by different levels of trawling impact, while meiofaunal samples were collected from two locations with different trawling activity. Data pre-processing and processing was performed using a combination of scripts and pipelines covering all standard steps. The taxonomic assignation of the Amplicon Sequence Variants (ASVs) was performed using the whole nucleotide database from NCBI, and the Vegan package (Dixon, 2003) was used to obtain rarefaction curves. Meiofaunal samples were classified by traditional methods.
The results indicated no effect of trawling on alpha diversity, but a significant effect on species composition by both methods (eDNA and meiofaunal surveys). Bryozoan taxa were only present at untrawled sites, suggesting this taxon could be used as trawling bioindicator. The main limitation of the study stems from a lack of reference databases. Therefore, the authors conclude that combining molecular and non-molecular methodologies remain the most holistic way to evaluate anthropogenic impacts, such as trawling, on benthic communities.

Task 3: Monitoring fish assemblages
Case-study: *Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA*
Patterns of diversity, seasonality, and abundance of marine fish species were analyzed by Stoeckle & collaborators (2021) by concurrently performing trawl survey and eDNA sampling. The authors took advantage of a bottom trawl survey of marine fisheries and evaluated the performance of eDNA metabarcoding in the estimates of fish richness, composition, seasonality, and relative abundance. eDNA metabarcoding sequencing was performed using the mitochondrial 12S gene for a total of 136 samples replicated across seasons. Amplicon sequence variants (ASVs) were obtained using DADA2 (Callahan et al., 2016). Taxon assignments were based on 100% of sequence similarity to a 12S reference sequence of a regional fish database.
Results found: (i) Agreement in seasonal abundance for 70 percent of the fish species. Inconsistent detection was due to rare taxa detected by one or another survey in single eDNA samples or tows. (ii) increased sensitivity in species diversity index of seven orders of magnitude with respect to trawl survey. (iii) Concordance in species composition between the two survey methods was about 75%, and close to 100% for abundant species. (iv) eDNA species reads (log-scaled) significantly correlated ($p<0.001$) with species biomass, and more strongly with an allometric index calculated from biomass which depends on body size ($R2 = 0.59$, and 0.66, respectively). The authors conclude that eDNA approaches have potential to improve the management of fisheries and MPAs.

Task 4: Quantifying biomass and/or abundance of target economically important species
Case-study: *Using eDNA to estimate biomass of bycatch in pelagic fisheries*
Collecting fisheries-dependent bycatch data is particularly challenging in large industrial fisheries. Urban et al. (2023) used eDNA sampling and qPCR to determine the biomass of Atlantic mackerel (*Scomber scombrus*) bycatch in herring (*Clupea harengus*) catch. Recording catch composition is required twice by the fisheries authority; at first, onboard the ship and then after landing in the factory. The water of holdings tanks of fisheries vessels is unchanged from when the fishing operation ends until the catch is landed at a factory, providing an excellent source of eDNA from catch composition. Species-specific assays targeting both, herring and mackerel mitochondrial cytochrome b sequences were used for DNA quantification. eDNA-to-biomass models were established with experimental data (shedding and decay experiments) and then used to predict the biomass of mackerel in the catch.
The results indicated that fractions and/or weights of mackerel estimated with eDNA analyses were comparable to routinely used visual-based estimation metrics. The variation in biomass of bycatch mackerel estimated from eDNA samples was lower than that found among the two visual assessments (i.e. onboard the ship, and at the factory), and it is within the 10% variation allowed by regulatory authorities. The authors concluded that the eDNA-based approach is more precise and consistent in estimating catch fractions than the currently used methods.

Independently of the analytical protocol, the steps described below are essential in any eDNA metabarcoding analyses.

3.1 Pre-processing: filtering raw data

The first steps of the bioinformatic pathway to analyze eDNA metabarcoding data are aimed to correct errors that can be introduced during DNA amplification and sequencing. The goal of pre-processing steps is to end up with a dataset composed of high-quality sequences that can be clustered and taxonomically assigned.

3.1.1 Checking the raw data quality profiles

HTS technologies produce massive amounts of data requiring multiple computationally intensive steps before performing an appropriate taxonomic assignment analysis. Commonly, raw files are generated in FASTQ format, a text file including the obtained DNA sequences (i.e. reads) and quality values encoded as ASCII characters associated with each base of the sequence. Commonly, reads are of 100-150 nucleotide length, and these can be single-end or paired-end depending on whether only one DNA strand or both are sequenced, respectively. For paired-end sequencing, two files per sample with the same number of reads are generated, corresponding to the forward (R1) and reverse (R2) reads.

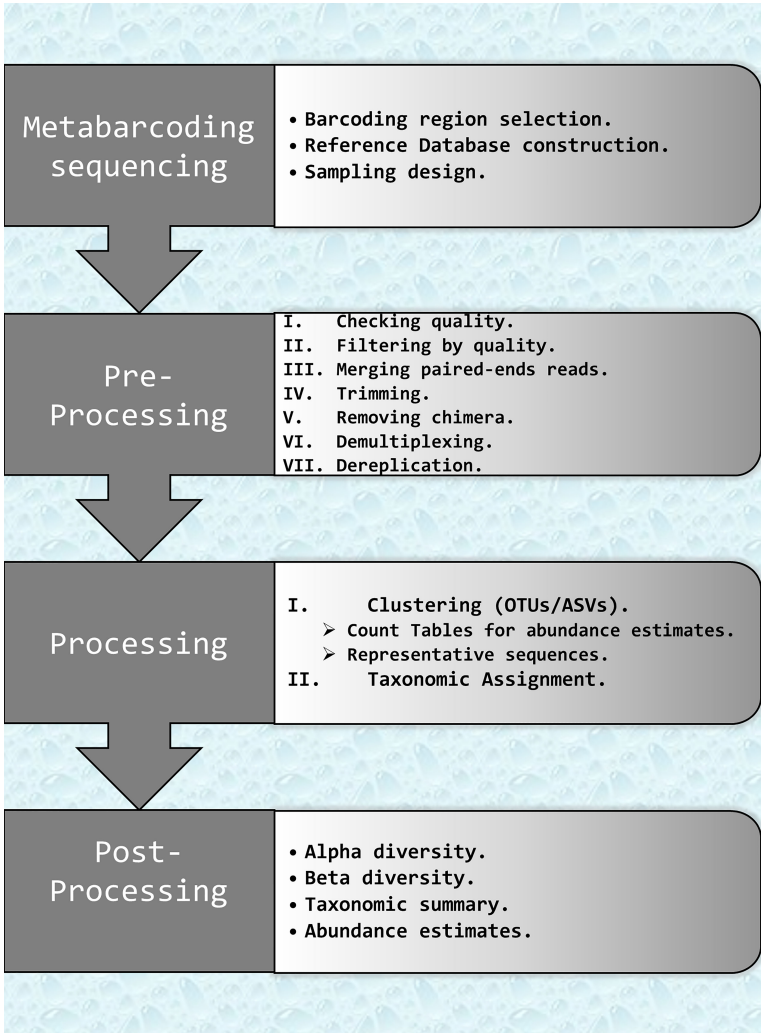


FIGURE 1
Flowchart describing the key bioinformatics steps required to analyze eDNA metabarcoding data. Steps are split in three parts: Pre-processing: filtering and correction raw data. Processing: clustering and taxonomic assignments; and Post-processing: statistical analysis and biodiversity index. OTUs, Organizational Taxonomic Units; ASVs, Amplicon Sequence Variants.

The reads-associated quality values are produced using the light intensity signal of each base call during the sequencing process. These quality scores are then translated into a Phred score using a modified version of the Phred algorithm (Ewing et al., 1998). Phred scores, are defined as the probability of error when calling a nucleotide base (A, C, T, or G) during the sequencing process, and are used for the control of the quality of the eDNA sequences. The quality of the reads can be obtained by using freely available software packages, such as FASTQC (Andrews, 2014) or FASTP (Chen et al., 2018). Both tools perform a primary check of the quality of all raw reads in a dataset through evaluation of different parameters such as the number of sequences, sequence length, GC content, presence of adaptors, ambiguous bases, overrepresented k-mers, and duplicated reads content. Visualizing and checking the quality scores from raw data is essential to set appropriate cut-off levels during further pre-processing steps and is fundamental for optimal downstream analysis.

3.1.2 Filtering by sequence quality

Typically, an average quality score (Q) of a read below 20 is considered low, and these sequences should be eliminated from the analysis. Moreover, sequencing accuracy decays with the sequence length and, accordingly, the 3' end of reads usually presents lower quality scores. Hence, it is recommended to trim the end of the forward and reverse reads based on their phred score to remove low-quality positions (typically Q20). This trimming can be performed by several programs, such as the trimmer tool of FASTQX-toolkit (https://hannonlab.cshl.edu/fastx_toolkit/), FastQC (Andrews, 2014), FastP (Chen et al., 2018), or Trimmomatic (Bolger et al., 2014).

3.1.3 Merging paired-end reads

This step should be performed if paired-end sequencing was obtained. Forward and reverse data files need to be combined into a single file to generate a complete amplicon sequence. Crucial parameters to perform merging of forward and reverse reads are

the choice of the algorithm, the presence of gaps in the alignment, or the minimum overlapping length (Taberlet et al., 2018). Algorithms for merging paired-end reads are implemented in different software packages, such as PEAR (Zhang et al., 2014), PANDAseq (Masella et al., 2012), FLASH (Magoč & Salzberg, 2011), or the *illuminapairedend* algorithm in OBITools (Boyer et al., 2016).

3.1.4 Trimming PCR primers and filtering by length

It is mandatory to remove the primers used in the eDNA metabarcoding experiment from every read. Primer sequences are usually removed using two approaches: (i) by allowing zero or a small number of mismatches between the sequence of the primers and the raw reads, with software packages such as Cutadapt (Martin, 2011) or (ii) by trimming the length of forward and reverse primers from the 5' -end and 3' -end of the sequences, respectively, using Trimmomatic (Bolger et al., 2014). Moreover, in most eDNA metabarcoding studies, only sequences within the expected length range of the amplified genomic region used are kept to reduce error rates in the dataset (Deiner et al., 2017). This can be done by setting a minimum and/or maximum length command with Cutadapt (Martin, 2011).

3.1.5 Removing chimeric sequences

Chimeric sequences are PCR artifacts made up of two or more sequences that have been erroneously combined during the extension step of the PCR amplification. The removal of chimeras has been identified as an essential quality control step to increase the diversity estimation accuracy (Alberdi et al., 2017; Deiner et al., 2017). Strategies for chimera detection are based on the comparison of all the sequences present in the dataset between themselves, in order to detect if any of them is derived from the 5' -end of one parent sequence and the 3' -end of another. Detection and removal of chimeras are implemented in several pipelines such as DADA2 (Callahan et al., 2016) and VSEARCH (Rognes et al., 2016), MIFISH (Sato et al., 2018), and BARQUE (Mathon et al., 2021). However, parameters used in the removal of chimeras during the bioinformatic pipeline could sometimes eliminate species that are actually present in the sample, so it is advisable to compare the results obtained with and without chimera removal.

3.1.6 Demultiplexing

Before pooling together eDNA samples for sequencing, a unique sequence named barcode is ligated to the genetic material of each individual sample. Then, the barcode information is used to allocate the sequence to their corresponding sample in a process known as demultiplexing. Different tools have been developed to this end as SABRE (https://github.com/najoshi/sabre), or FLEXBAR (Dodt et al., 2012), and it is also implemented in OBITools (*ngsfilter*, Boyer et al., 2016), eDNAFlow (Mousavi-Derazmahalleh et al., 2021), and SLIM (Dufresne et al., 2019).

3.1.7 Dereplication

Given that sequences obtained from eDNA samples contain different representations of species and/or individuals within the same species, it is needed to collapse strictly identical sequencing reads into unique sequences. This process is known as dereplication. It can be carried out in OBITools (Boyer et al., 2016) using the *obiuniq* command, or in VSEARCH (Rognes et al., 2016) using the *–derep_fulllength* or *–derep_prefix* scripts, and also in DADA2 (Callahan et al., 2016) and SWARM (Mahé et al., 2015). The abundance of identical sequences in each sample is tracked in the output file, information that is subsequently used to generate a count table after the clustering analyses.

3.2 Processing data: taxonomic assignment

The taxonomic assignment is key in the eDNA metabarcoding approach, and it is split into two steps: clustering and searching taxonomic identity in the reference databases.

3.2.1 Clustering

Clustering of the sequences by similarity is performed in order to split intra- from inter-specific genetic variability (Alberdi et al., 2017). However, a lack of knowledge on the expected spectrum of DNA sequence variation among and within species makes the selection of the clustering algorithm a critical step to get confident results. Different algorithms to perform the clustering have been proposed and evaluated, and the selection should be based on the study goals and the existing knowledge of the environment being analyzed (Xiong and Zhan, 2018).

Sequence clustering can be reference-based or *de novo*. Reference-based clustering is straightforward always that the reference database is complete enough, and the barcoding sequence has enough power to differentiate species. Clustering *de novo* entails grouping all sequences among themselves into clusters and assigning taxonomically the representative sequences.

The clustering algorithms most frequently used are based on generating clusters of sequences, named molecular Operational Taxonomic Units (OTUs). OTUs group sequences differing by less than a previously defined threshold, commonly set among 3 to 5% of sequence variation. Nevertheless, the imposed threshold of sequence similarity used is frequently arbitrary if the variability in DNA sequences of the target species community is unknown. Thus, it could produce misidentification of species if the threshold is set high or a lack of detection if it is too low. Therefore, other clustering methods have been developed without the need for a threshold of sequence similarity such as the Bayesian clustering algorithm of CROP (Hao et al., 2011) or the iterative growth process used in SWARM (Mahé et al., 2015b). These alternatives to OTUs, use Amplicon Sequence Variants (ASVs, Eren et al., 2013), also known as Exact Sequence Variants (ESVs), or zero-radius OTUs (ZOTUs).

Different from OTUs, ASVs are constructed by a *de novo* process grouping the sequences contained in the dataset by minimizing the number of differing nucleotides, frequently to one single difference. Therefore, all biological variation is captured in ASVs and results can be compared between different studies. ASVs algorithms are implemented in several software and pipelines such as DADA2 (Callahan et al., 2016), SLIM (Dufresne et al., 2019),

TABLE 1 Most common algorithms used for the clustering step in metabarcoding eDNA analyses.

Software/ Algorithm	Summary	Task performed	Reference
VSEARCH	VSEARCH is an open source tool for processing nucleotide sequence data. It performs global alignments between the query sequences and the potential target reference sequences. To identify similar sequences it uses a fast heuristic algorithm based on words shared by the query and target sequences.	• Clustering • Chimera detection • Dereplication • Sorting • Subsampling • FASTQ file processing (i.e. merging paired-ends reads)	Rognes et al. (2016) *PeerJ* 4 : e2584.
SWARM	SWARM is a single-linkage clustering method, it uses an iterative growth process and the sequence abundance values to delineate clusters. SWARM uses a local clustering threshold for alignments, instead of a global one such as the one used in Vsearch.	• Clustering • Dereplication • Outputs OTU representatives in fasta format	Mahé et al. (2015). *PeerJ* 3: e1420.
CROP	CROP (Clustering 16S rRNA for OTU Prediction) is an unsupervised Bayesian clustering method. It finds clusters without setting a hard cut-off threshold as required by hierarchical clustering methods. It uses a Gaussian mixture model to describe the data replacing the mean value of a Gaussian distribution by a 'center' sequence to characterize a specific cluster. Although originally designed for its use in 16S microorganism metabarcoding, it has been used with other markers and eukaryotes species.	• Clustering	Hao et al. (2011) *Bioinformatics* 27.5: 611-618.
DADA2	DADA (Divisive Amplicon Denoising Algorithm) is a model-based approach for correcting amplicon errors without constructing OTUs. DADA2 is reference free and applicable to any genetic locus. The core denoising algorithm is built on a model of the errors in Illumina-sequenced amplicon reads. A Poisson model for the number of repeated observations of a sequence parameterized by the error rate is then used to calculate the *p*-value of the null hypothesis that the number of amplicons reads of a sequence is consistent with the error model. These *p*-values are used as the division criteria for an iterative partitioning algorithm, which continues dividing sequencing reads until all partitions are consistent with being produced from their central sequence.	• Clustering • Filtering • Dereplication • Chimera identification • Merging of paired-end reads	Callahan et al. (2016). *Nature methods*, *13*(7), 581-583.

ANACAPA toolkit (Curd et al., 2019), eDNAFlow (Mousavi-Derazmahalleh et al., 2021), and PEMA (Zafeiropoulos et al., 2020).

An overview of the open-access bioinformatics tools implementing the four most common clustering algorithms: VSEARCH (Rognes et al., 2016), SWARM (Mahé et al., 2015), CROP (Hao et al., 2011) and DADA2 (Callahan et al., 2016) are presented in Table 1. Moreover, LULU (Frøslev et al., 2017), is an algorithm specially designed for removing erroneous OTUs improving the accuracy of species diversity statistics estimates based on similarity thresholds.

3.2.2 Taxonomic assignment

Several DNA regions are used as targets for HTS taxonomic identification using the eDNA metabarcoding approach in eukaryotes. Among the most widely used DNA regions are the ribosomal RNA genes (rRNA 12S/18S/28S), the internal transcribed spacer (ITS), and the Cytochrome oxidase subunit I (COI) mitochondrial gene. However, the use of one or another depends on the species groups under study, and the power of each one should be assessed when planning the study. Different studies recommend the use of more than one DNA region to increase the power of species detection (Kumar et al., 2022).

The taxonomic identification of OTUs/ASVs is one of the crucial steps in the pipeline and can be accomplished using different approaches: (i) the similarity-based method performs an alignment against all sequences of the reference database to assign taxonomy to a query sequence, and it is performed by BLAST (Altschul et al., 1990) or BOLD (for COI sequences, (Ratnasingham & Hebert, 2007)) algorithms; (ii) the phylogeny-based approach, implemented in pipelines such as SAP (Munch et al., 2008), pplacer (Matsen et al., 2010), or EPA-ng (Barbera et al., 2019), estimates the similarity between the query sequence and the reference sequences by analyzing the position of the query sequence in the phylogenetic tree generated for the reference database; and (iii) the composition-based method scores query sequences against the reference database and assigns the taxonomy based on the pattern of scores obtained, and it is implemented by the script ecotag in OBITools (Boyer et al., 2016)

As the goal of eDNA metabarcoding studies is to characterize the community of species living in a given environment, the quality and completeness of the reference databases used for taxonomic assignment is one of the crucial points to avoid false negative results (Miya, 2022). The best option is to build an *ad hoc* database containing the sequences of species inhabiting the area under study (e.g. in Boulanger et al., 2021; Gold et al., 2021), which requires an extra survey effort in marine environments where species composition is not fully characterized, or unknown. It is the most important for the description and surveillance of Marine Vulnerable Ecosystems, including invertebrate species which are poorly characterized molecularly (Miya, 2022).

For eDNA metabarcoding of fish species assemblages, reference databases are mostly complete if not at species levels, at least in high taxonomic categories such as genera or family. Moreover, for commercially important fish species monitored in fisheries management programs, biological material to build up sequences databases is available from research surveys, port landings, and markets. Thus, for fish species valued in markets, reference databases' incompleteness is not a limiting issue.

Universal primers for barcoding teleost fish are already available (Ivanova et al., 2007; Sato et al., 2018), as well as several databases specifically dedicated to teleosts, such as MitoFish (Sato et al., 2018) and Phylofish (Pasquier et al., 2016). To date, for fish species the two more frequently used barcoding regions are 12S and COI. Nowadays, it is possible to retrieve barcoding reference sequences of 25,924 fish species (i.e. 24,724 teleost and 1,200 Elasmobranchii) from BOLD systems (Ratnasingham and Hebert, 2007) and 8,523 from MIFISH database (Miya, 2022). Thus, using the COI mitochondrial gene potentially allows for the detection of approximately 70% out of the nearly 34,000 described fish species, though the length of the obtained COI sequences should be extended to avoid misidentifications (Collins et al., 2021). However, rRNA-related barcoding regions are preferred when different groups of species are intended to be surveyed from the eDNA sample (Miya, 2022), as these have the potential to differentiate wider taxonomic groups using the more common short reads metabarcoding sequencing. Additionally, bioinformatic tools for building up custom reference databases are available, allowing mining and retrieving barcoding sequences of species of interest from big sequences reference databases. The ANACAPA toolkit's (Curd et al., 2019) first module, CRUX, construct custom reference databases for user-defined primers by querying public databases such as NCBI's nucleotide database (Benson et al., 2013), and Meta-Fish-Lib (Collins et al., 2021) is a pipeline designed to retrieve mitochondrial DNA sequence data for a given list of fish species.

3.3 Post-processing and diversity statistics

Refining the distribution of sequences in the final count table is imperative to reduce the impact of false positives in further analyses. False positives are mainly due to external contamination, occurred in the laboratory, or internal contamination in the sequencing process (Taberlet et al., 2018).

External contaminations can be detected through the incorporation of controls during the processing of samples, such as negative DNA extractions (i.e. DNA extraction from storage/extraction buffers used in the field) or negative PCR controls. Internal contaminations are mainly due to the miss-assignment of indices during library preparation, sequencing, and/or demultiplexing steps, that causes the allocation of a low percentage of sequences of a sample to other samples. Additionally, due to miscalled bases during sequencing, OTUs/ASVs with a small number of sequences appear randomly spread through the count table. Therefore, error-correction in the count table (singletons and rare OTUs/ASVs) prior to starting biological interpretation is recommended in eDNA metabarcoding experiments. A summary of all potential source of errors in an eDNA processing workflow, as well as potential solutions to avoid these can be found in Zinger et al., 2019.

Different bioinformatics tools can also be used to minimize and correct false positive errors in eDNA metabarcoding analyses. DADA2 (Callahan et al., 2016) generates a parametric error model that is trained on each sequencing run and then applies that model to correct and collapse the sequence into ASVs. Deblur (Supplementary Table 1), computes error profiles to obtain putative error-free sequences from Illumina MiSeq and HiSeq. LULU (Frøslev et al, 2017) collapses erroneous OTUs/ASVs into their parents to remove them from the dataset. Finally, the R package microDecon (McKnight et al., 2019), uses the proportions of contaminant OTUs/ASVs in negative blank samples to identify and remove erroneous sequences from the eDNA metabarcoding datasets.

3.3.1 Alpha diversity patterns

Alpha-diversity analyses use the number of OTUs/ASVs sequences assigned to a taxonomic level at each sampling location. Different approaches are followed to accommodate differences in sampling effort and diversity coverages (McKnight et al., 2019). Rarefaction curves are used to check whether the sequencing depth obtained is sufficient to retrieve most of the species diversity present in the environmental samples. Rarefaction curves can be obtained by using the Vegan R package (Dixon, 2003). Alpha diversity indexes are frequently integrated within available eDNA metabarcoding pipelines (see Supplementary Table 1) or can be performed separately using TTT (Macher et al., 2021) or Vegan R package (Dixon, 2003).

3.3.2 Beta diversity patterns

Beta diversity can be defined as the variability in species composition among sampling units for a given area. Thus, it is different from alpha diversity as two groups can have identical species richness indexes, but different species compositions. Similar to alpha-diversity, beta-diversity analyses are frequently included in pipelines (See Supplementary Table 1) or can be obtained by using TTT (Macher et al., 2021) or Vegan R package (Dixon, 2003).

3.4 Quantitative monitoring of biodiversity

The quantification of the abundance of the detected species in an eDNA sample adds very valuable information to eDNA metabarcoding analyses. The quantitative relationship between species abundance and the amount of DNA present in the environment has been explored in seawater samples. It has been shown a significant positive correlation between trawl catches or visual census and estimated eDNA abundances (e.g. in Thomsen et al., 2016; Doi et al., 2017; Tillotson et al., 2018, Levi et al., 2019, Fukaya et al., 2021, Afzali et al., 2021; Russo et al., 2021; Sato et al., 2021; Stoeckle et al., 2021; Maiello et al., 2022). Nonetheless, many studies also report high uncertainty in these estimations, mainly due to the lack of knowledge on the dynamics of eDNA in marine environments (i.e. process-based models see point 3. e.g. Lacoursière-Roussel et al., 2016, Sepúlveda et al., 2021; Nakagawa et al., 2022), and the poor knowledge on the probability of detection in eDNA metabarcoding approaches (i.e. occupancy and process models, see

point 3, e.g. Doi et al., 2019; Fukaya et al., 2021). A modelling framework for eDNA metabarcoding data, allowing for all key sources of variation, error and noise in the data-generating process, has been proposed by Diana et al. (2022). Inference is performed using MCMC and can be used to estimate within-species biomass changes across sites and to link those changes to environmental covariates, while accounting for between-species and between-sites correlation. The modelling framework is available across the eDNAPLUS R script (Supplementary Table 1). However, currently, there is a consensus indicating that quantitative measures of species abundance are better achieved by using the targeted eDNA approach.

4 Targeted eDNA

The second category of eDNA approaches involves species-specific techniques that use assays tailored to target DNA fragments of particular species in an environmental sample. When designed stringently and after thorough validation, these are highly reliable and often effectively linked to the biomass and abundance of the target organism (e.g. in Takahara et al., 2012; Doi et al., 2015; Lacoursière-Roussel et al., 2016; Yamamoto et al., 2016; Capo et al., 2019; Knudsen et al., 2019; Salter et al., 2019; LeBlanc et al., 2020; Brys et al., 2021; Fukaya et al., 2021; Shelton et al., 2022; Kasmi et al., 2023; Urban et al., 2023). Nowadays, the main technique used in species-specific detection of environmental samples is real-time quantitative PCR (qPCR). However, other PCR relate-techniques such as digital PCR can also be used. The workflow of data analyses for the targeted eDNA approach is presented in Figure 2.

4.1 Real-time quantitative PCR

Real-time quantitative PCR or qPCR is a technique capable of detecting and quantifying tiny amounts of DNA present in a sample by contrasting the data obtained to a standard curve. Different from the original PCR method, qPCR measures DNA amplification by fluorescence signals as the reaction progresses in real-time. The quantification of the target DNA of a qPCR experiment is performed by measuring the emission of a fluorescent reporter dye that binds to the DNA in each amplification cycle. Thus, the initial amounts of DNA templates in a sample can be quantified by comparing the number of amplification cycles required to reach a particular threshold of fluorescence signal. Hence, the number of PCR cycles will be negatively related to the starting concentration of the target DNA (Kubista et al., 2006). Therefore, the most significant parameter derived from a qPCR experiment is the quantification cycle (Cq), which can be defined as a PCR cycle at which the accumulating PCR products' fluorescence reaches a pre-established threshold.

A standard curve is obtained by plotting Cq values versus DNA concentration for different dilutions of a target DNA sample of known concentration. The standard curves are used to calibrate the

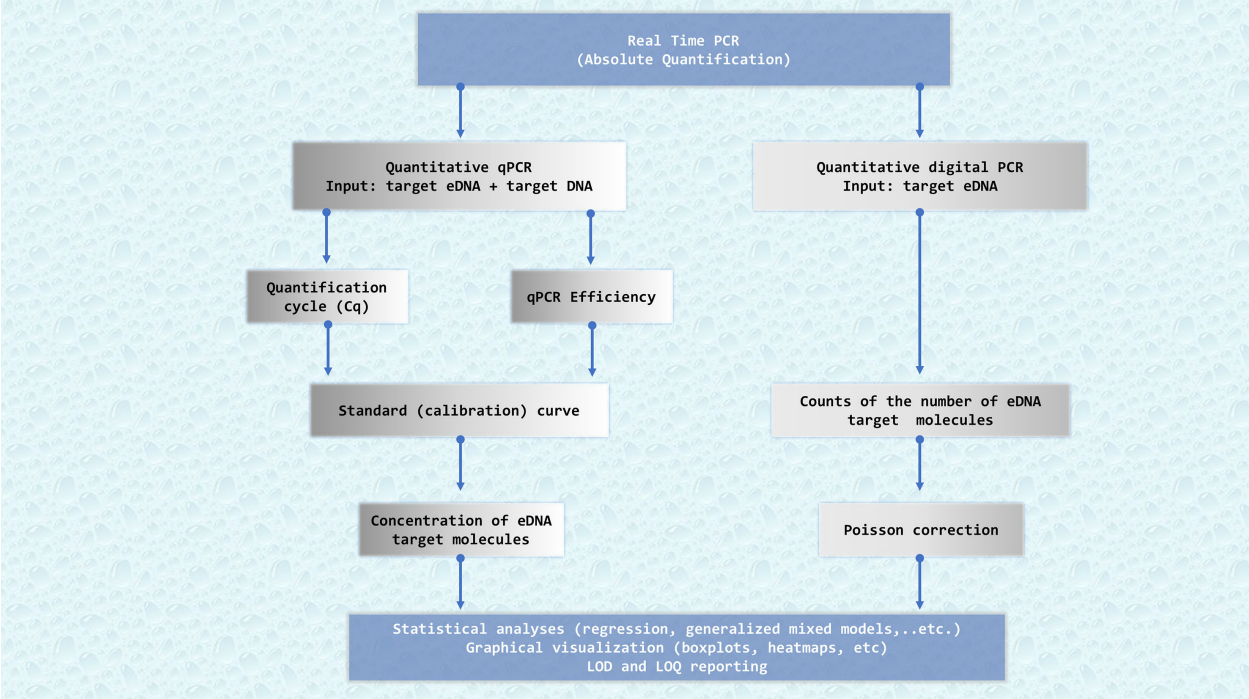


FIGURE 2
Flowchart describing the key bioinformatics steps required to get successful results quantifying eDNA from marine samples. qPCR: real-time quantitative PCR, dPCR: digital PCR. Quantification of eDNA samples by qPCR needs a standard calibration curve, whereas digital PCR directly counts eDNA target molecules splitting the PCR reaction in tiny volumes. Cq: quantification cycle, a PCR cycle at which the accumulating PCR products' fluorescence reaches a pre-established threshold. Efficiency, which measures the overall performance of the qPCR. A standard curve is obtained by plotting Cq values versus DNA concentration for different dilutions of a target DNA sample of known concentration. LOD, Limit of Detection. LOQ, Limit of Quantification.

qPCR and interpolate the data from the target samples with unknown concentrations collected in the environment.

The rate at which new amplicons are generated is the qPCR Efficiency, which measures the overall performance of the qPCR. An efficiency of 100% means that target DNA molecules double at each cycle. Typical efficiencies range from 90 to 110%. Efficiency values below or above the aforementioned range mean that the chosen qPCR conditions are not optimal (e.g. inefficient primer design, presence of PCR inhibitors, excessive amounts of starting DNA, or incorrect annealing temperature). Efficiency is calculated from the standard curve as $E= 10^{-1/slope} - 1$.

In order to get comparative results among qPCR studies it is necessary to report the relevant experimental conditions. Bustin et al. (2009) provide guidelines for the Minimum Information for the publication of Quantitative Real-Time PCR experiments (MIQE). Moreover, for eDNA quantitative results it is advised to report the assay Limit of Detection (LOD) and the assay Limit of Quantification (LOQ). LOD and LOQ are defined as the lowest standard concentration of template DNA that produced at least 95% positive replicates, and the lowest standard concentration that could be quantified with a CV value below 35%; respectively (Klymus et al. (2020); see Section 4.4 for more details).

4.2 Absolute quantification of target eDNA

The first step to getting absolute quantification is the generation of a standard curve for each run of qPCR. The standard curve is obtained using a DNA sample with a known concentration from the target species to be analysed (e.g DNA extracted from a small piece of tissue). These standard DNA samples are serially diluted to generate a minimum of five samples with decreasing DNA concentrations, and then added in triplicate to the qPCR assay and amplified jointly with the eDNA samples of unknown concentration (Yamamoto et al., 2016; Itakura et al., 2019). Usually, thermal cyclers can hold a plate containing up to 96 samples in a single run. Therefore, if more plates are needed for a given study, a unique standard curve should be generated for each run and also for each target species. Once the Cq values for the standard curve are obtained, these are plotted against the logarithm of the starting concentrations, which produce a linear relationship with a negative slope. With this standard curve, Cq data for the unknown eDNA samples are interpolated to obtain eDNA concentration of the target species. Most real-time thermal cyclers include proprietary software that performs these types of analyses automatically. Additionally, there is a wide range of software for the different raw data analyses that have been exhaustively reviewed by Pabinger et al., 2014.

4.3 Digital PCR techniques

Similar to qPCR, digital PCR (dPCR) allows for the quantification of minimal amounts of DNA, but it does not require a standard curve for quantification. Thus, quantification is possible even when a standard sample is not available. Moreover, eliminating calibration curves also eliminates a potential source of errors.

In a dPCR, the reaction mixture is split into a large number of separate tiny volumes, such that there is one or no target molecule present in any individual reaction. Currently, available commercial systems can generate up to ten thousand (Bio-Rad, Life Technologies), nearly fifty thousand (Standard BioTools), and ten million (RainDance) partitions per experiment (Pabinger et al., 2014). Therefore, each reaction becomes binary (i.e. presence/absence) and these discrete signals are counted. Obtaining target eDNA concentration values from a dPCR experiment is made directly by the software implemented in the machine. The data that this system generates is gathered from a binary signal that, after applying a Poisson correction to consider partitions with more than one molecule, can be used to directly count the number of targets eDNA molecules in the original sample.

Comparisons between qPCR and dPCR methods has shown the later to be more tolerant to PCR inhibition. Therefore, the obtained concentration estimates are less biased and, consequently, the correlation between eDNA concentration versus abundance or biomass should be stronger. There are few studies relying exclusively on dPCR for the estimation of target species abundance or biomass, likely as the technology is still expensive for its wide use (e.g. in Marx, 2014; Doi et al., 2015; Capo et al., 2019; Brys et al., 2021; Manfrin et al., 2022).

4.4 Getting standard measures for comparative studies: LOD and LOQ

eDNA samples contain low concentrations of target DNA, thus, the ability of an assay to detect and quantify these low concentrations of DNAs is of the most importance. Klymus et al. (2020) presented a simple method based on discrete thresholds for determining the LOD and LOQ for an eDNA qPCR assay. LOD determines how many copies of target DNA can be detected with 95% of confidence in a qPCR reaction. Thus, LOD describes the ability of an assay to detect target DNA, which is key in monitoring rare species based on detection/nondetection. While LOQ determines which is the minimum number of target DNA copies quantifiable in a qPCR reaction with defined precision, which is of the maximum importance when the goal of a study is quantifying biomass or abundance. LOD and LOQ can also be obtained by a curve-fitting modelling method, which is recommended to avoid rigorous testing of a large number of different DNA concentrations. Multiple models can be evaluated to select the best for each assessed dataset. LOD and LOQ can be estimated using the R script qPCR LOD CALC (Merkes et al., 2019, Supplementary Table 1)

4.5 Correlating eDNA concentration with abundance or biomass

Different studies focused on correlating abundances or biomass with eDNA concentration use diverse statistical methods for their analysis. Most studies find that the residual errors do not follow a normal distribution and need, therefore, to be first log-transformed in order to improve the homogeneity of the variance (Thomsen et al., 2016; Knudsen et al., 2019; Murakami et al., 2019). The statistical model used greatly depends on the experimental design and the data

obtained. For studies dealing with just two variables (abundance or biomass and eDNA concentration), the statistical method of choice is either type II linear regression (Takahara et al., 2012; Doi et al., 2015; Knudsen et al., 2019) or type I regression (Yamamoto et al., 2016). Studies also considering the different variables that may be having an influence on eDNA concentration (such as temperature or salinity) use linear mixed-effects models (Itakura et al., 2019) or generalized linear models (Thomsen et al., 2016; Tillotson et al., 2018; Knudsen et al., 2019). However, Kasmi et al. (2023) found differences in the relationship between eDNA copies and biomass depending on the statistical methods, getting better results when using Gaussian Process Regression (GPR), neural network and non-linear regression model, than with simple regression models.

High correspondence between abundance or biomass estimated from traditional and eDNA methods have been found (eg. Shelton et al., 2022; Kasmi et al., 2023; Urban et al., 2023). However, natural environmental conditions (e.g. Lacoursière-Roussel et al., 2016) and species-specific characteristics such as eDNA shed and decay (Urban et al., 2023), are source of uncertainty and need to be considered in results interpretation. Nevertheless, occupancy and process-based modelling can help to better interpret eDNA quantification results obtained from natural environments (See Section 5).

5 Results interpretation: occupancy and process-based models

When assessments of species distribution are carried out, it is imperative to understand what are the limitations of the survey methods used. Similarly to all survey methods, detection (or lack of detection) of species by eDNA sampling is not free of errors (see Burian et al., 2021 for a review of potential sources of error of eDNA-based methods). A particular species might not be detected even if it is present in the environment (i.e. false negative), or it can be erroneously detected even when absent (i.e. false positive). Therefore, it is most important to understand the level of error that the particular eDNA survey presents, to make sound and consistent interpretations.

As commented in section 3.3.3 several bioinformatics tools help to filter out false positive errors in eDNA metabarcoding studies. However, it is important to consider the rate of false negative results too, especially when the goal of the study is to monitor a group of target species (i.e. by metabarcoding) or single target species. False negative detection can be produced because enzymatic inhibition, or eDNA for a species was not collected within the sample, or it is in such a low concentration that became undetectable or because of the sensitivity of PCR (Ficetola et al., 2015). To avoid missing detection of taxa that are actually present (false negatives), multiple extractions and amplifications of the same samples are often performed (Ficetola et al., 2015; McClenaghan et al., 2020). Moreover, increasing sequencing depth can also improve the rates of detection (McClenaghan et al., 2020; Fukaya et al., 2021).

Currently, it is becoming common and advisable to interpret surveys based on eDNA within an occupancy modelling framework. Occupancy modelling allows estimating the probability of detection of a given species in a given environment (Griffin et al., 2020). Occupancy modelling can be applied to eDNA metabarcoding (i.e. Doi et al., 2019; McClenaghan et al., 2020, Valdivia-Carrillo et al., 2021), and in targeted eDNA approach (Dorazio and Erickson, 2018, Fukaya et al., 2021; Buxton et al., 2022). Appropriate data for fitting an Occupancy Site model consist of a series of samples obtained from S sampling sites, with K replicate samples obtained per site (Dorazio and Erickson, 2018). Models have been developed allowing for multiscale occupancy models to be applied accounting for both positive and negative errors at the field and laboratory stages (Griffin et al., 2020). Occupancy models successfully estimated true prevalence, detection probability, and false-positive rates, and their performance increased with the number of replicates (Ficetola et al., 2015). Thus, the occupancy modelling framework can be also applied to study the optimal conditions for sampling and laboratory stages (Doi et al., 2019). Modelling is frequently performed by Bayesian inference, and modeled parameters can include environmental and experimental covariates (Guillera-Arroita et al., 2017; McClenaghan et al., 2020; Burian et al., 2021; Buxton et al., 2022).

Other origins of uncertainty in the results of surveys of eDNA, especially for quantitative results, come from the dynamics of concentrations of eDNA at the sampling places (Burian et al., 2021). The amount of DNA shed into the environment by a particular species can be different from another, the biomass of an adult fish can be the same as hundred juvenile individuals of the same species; while environmental characteristics can make DNA rapidly degraded or transported by oceanic currents (Hansen et al., 2018). Process-based models are based on a mechanistic understanding of the dynamics of eDNA concentrations in the environment and can help in accounting for these sources of false positive and false negative results (Burian et al., 2021). eDNA studies have been favourably augmented in precision when accounting for eDNA transport shed and decay (e.g. Collins et al., 2021, Murakami et al., 2019; Kirtane et al., 2021; Urban et al., 2023). The experimental output, informs on the expected rates of detection or lack of it in the eDNA survey and can be integrated as prior information within the occupancy modelling framework.

Occupancy modelling can be performed using the eDNA SHINY APP (Diana et al., 2021), or the R script EDNAOCCUPANCY (Dorazio and Erickson, 2018).

6 Ten good practices in eDNA metabarcoding and targeted dPCR/qPCR data analyses

I. Make good planning of sampling (i.e. number of sites, replicates by sites, PCR replicates, amplicon sequencing depth), and use simulations to infer the probability of detection and number of replications.
II. Compile all the information available on the species or ecosystems to be surveyed, combining eDNA with traditional survey methodologies to get more confident results.
III. Follow strictly the pre-processing steps of eDNA metabarcoding data analyses pipelines.

IV. Assess the completeness of reference databases and choose the adequate clustering and taxonomic assignment algorithms accordingly to available data, and to the goal of the study.
V. Perform post-processing filtering of OTUS/ASVs to eliminate most of the possible sources of false positives produced in the laboratory.
VI. Estimate LOD and LOQ when performing qPCR assays, reporting these for comparative results among studies.
VII. Assess the probability of detection of the species of interest at both field and laboratory stages.
VIII. Research the eDNA dynamics (i.e. shed, decay, and transport) regarding your target species/ecosystem.
IX. Do not make conclusions about species absence without performing points V for the targeted eDNA approach, and VI and VII for both eDNA methods.
X. Conclusions on true species detection should be made only when using a robust protocol (e.g. including an appropriate number of positive replicates).

7 Final considerations

As sequencing and quantitative technologies grow, the applications of eDNA approaches also do. To date, automatic sequencing robots are already available to process eDNA *in situ* within aquatic environments (Sepulveda et al., 2020), and long-read sequencing of eDNA enhances taxonomic assignments opening also the possibility to estimate population genetics parameters at the same time (Sigsgaard et al., 2020; Tsuji et al., 2020). Bioinformatics tools to analyze the different data from eDNA sampling are growing along with the sequencing technologies, and are not a limiting issue in the application of eDNA approaches. The increasing number of eDNA studies published in the very last few years reflects both the simplicity of the method to potentially answer different scientific questions, and the need to simplify costly surveys traditionally used in ecological sciences. However, this diversity of studies with different analytical approaches makes it difficult to navigate the bibliography for a systematic implementation of eDNA approaches in the sustainable management of fisheries. Here, we show the state of the art of data analyses with the available bioinformatics tools proven to be useful for avoiding common mistakes in processing, analyzing, and interpreting results based on eDNA samples.

Across different studies referenced here, it seems clear that both eDNA approaches (metabarcoding and targeted) suffer uncertainty regarding detection probabilities. This uncertainty can be at least partially improved by: (i) designing eDNA sampling with enough numbers of sample replications, and sampling sites, and covering seasonal variance; (ii) getting PCR and amplicon sequencing replications, (iii) getting enough amplicon coverage, and (iv) by using occupancy and process-based modelling. This is particularly important in the application of eDNA approaches to surveillance of rare species (e.g. endangered), which are prone to suffer from uncertainties for the low probability of detection (Ficetola et al., 2015, but see Boussarie et al., 2018 in BOX 1-TASK 1). Major complexities are found in the application of eDNA approaches to surveillance of Marine Vulnerable Ecosystems, as many of them are poorly described, and therefore, reference databases are incomplete (e.g. Good et al., 2022, see BOX 1-TASK 2). In this regard, complementing and comparing eDNA sampling with traditional surveys such as visual census and video recording is necessary (e.g. Boussarie et al., 2018; Good et al., 2022). Nevertheless, the molecular characterization of the relevant invertebrate species of Vulnerable Ecosystems will be necessary for reference databases to be complete before monitoring can be systematically implemented. On this point, the use of different primer sets and different genomics regions gives more complete results (Good et al., 2022; Kumar et al., 2022). However, the use of eDNA approaches are giving promising results when monitoring Marine Protected Areas (MPA) (eg. Gold et al., 2021). Ecosystems within MPA are generally well described making the implementation of eDNA approaches augmented by prior knowledge. Similarly, barcoding regions have been obtained for thousands of fish species, and especially for economically important ones making that monitoring of fish communities does not present obvious methodological or analytical difficulties. Thus, the research on the limitations of the eDNA approaches for the study of target fish species (i.e. commercial) and fish communities are less challenging to be accomplished (Keck et al., 2022). Reference databases enriched with sequences of fish species locally known increase the likelihood of detection (e.g. Stoeckle et al., 2021, see BOX 1- TASK 3) and it is advisable when monitoring fish communities.Moreover, uncertainties about detection probabilities are decreased when DNA studies are performed in fisheries-related conditions (e.g. Russo et al., 2021; Stoeckle et al., 2021; Maiello et al., 2022; Kasmi et al., 2023, and Urban et al., 2023 presented in BOX 1-TASK 4).

In summary, eDNA approaches are showing realistic results for tasks related to fish abundance quantification and monitoring. Nonetheless, its use in ecosystem-based fisheries management could be delayed until reference databases of vulnerable marine ecosystems are completed. Several studies referenced here have demonstrated the feasibility of applying eDNA approaches to support tasks related to sustainable fisheries management. Given the urgent need of promoting sustainability in fisheries management, the extensive amount of data collection required for this, and the rapid scientific advances improving eDNA data analysis, it is likely that eDNA approaches can be incorporated within fisheries-related tasks in the nearby future.

Author's note

The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may

be held responsible for the use which may be made of the information contained therein.

Author contributions

NP-M, LC and FS-R wrote, edited and reviewed the manuscript. All authors approved the submitted version.

Funding

The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. NP-M is supported by MSCA-PF FishAdapt 101066785.

Acknowledgments

The authors would like to express their gratitude to Gary Carvalho, Laszlo Orban and Allan Tucker for their advice and expert revision of the State-of-the-Art review in which this manuscript is based on. We also acknowledge the good suggestions of the reviewers on the first version of this manuscript, which improved it in both content and quality.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1061530/full#supplementary-material

References

Afzali, S. F., Bourdages, H., Laporte, M., Mérot, C., Normandeau, E., Audet, C., et al. (2021). Comparing environmental metabarcoding and trawling survey of demersal fish communities in the gulf of st. Lawrence, Canada. *Environ. DNA* 3 (1), 22–42. doi: 10.1002/edn3.111

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., and Bohmann, K. (2017). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol. Evol.* 9, 134–147. doi: 10.1111/2041-210X.12849

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2014) *FastQC a quality-control tool for high-throughput sequence data*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., et al. (2019). EPA-Ng: massively parallel evolutionary placement of genetic sequences. *Systematic Biol.* 682), 365–369. doi: 10.1093/sysbio/syy054

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

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42. doi: 10.1093/nar/gks1195

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bonfil, R., Palacios-Barreto, P., Vargas, O. U. M., Ricaño-Soriano, M., and Díaz-Jaimes, P. (2021). Detection of critically endangered marine species with dwindling populations in the wild using eDNA gives hope for sawfishes. *Mar. Biol.* 168 (5), 1–12. doi: 10.1007/s00227-021-03862-7

Boulanger, E., Loiseau, N., Valentini, A., Arnal, V., Boissery, P., Dejean, T., et al. (2021). Environmental DNA metabarcoding reveals and unpacks a biodiversity conservation paradox in Mediterranean marine reserves. *Proc. R. Soc. B* 288 (1949), 20210112. doi: 10.1098/rspb.2021.0112

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J. B., et al. (2018). Environmental DNA illuminates the dark diversity of sharks. *Sci. Adv.* 4 (5), eaap9661. doi: 10.1126/sciadv.aap9661

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., and Coissac, E. (2016). Obitools: A unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* 161), 176–182. doi: 10.1111/1755-0998.12428

Bradley, D., Merrifield, M., Miller, K. M., Lomonico, S., Wilson, J. R., and Gleason, M. G. (2019). Opportunities to improve fisheries management through innovative technology and advanced data systems. *Fish Fisheries* 20 (3), 564–583. doi: 10.1111/faf.12361

Brys, R., Halfmaerten, D., Neyrinck, S., Mauvisseau, Q., Auwerx, J., Sweet, M., et al. (2021). Reliable eDNA detection and quantification of the European weather loach (Misgurnus fossilis). *J. Fish Biol.* 98 (2), 399–414. doi: 10.1111/jfb.14315

Burian, A., Mauvisseau, Q., Bulling, M., Domisch, S., Qian, S., and Sweet, M. (2021). Improving the reliability of eDNA data interpretation. *Mol. Ecol. Resour.* 21 (5), 1422–1433. doi: 10.1016/j.scitotenv.2020.144466

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Buxton, A., Diana, A., Matechou, E., Griffin, J., and Griffiths, R. A. (2022). Reliability of environmental DNA surveys to detect pond occupancy by newts at a national scale. *Sci. Rep.* 12 (1), 1295. doi: 10.1038/s41598-022-05442-1

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Capo, E., Spong, G., Norman, S., Königsson, H., Bartels, P., and Byström, P. (2019). Droplet digital PCR assays for the quantification of brown trout (Salmo trutta) and Arctic char (Salvelinus alpinus) from environmental DNA collected in the water of mountain lakes. *PloS One* 14 (12), e0226638. doi: 10.1371/journal.pone.0226638

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Collins, R. A., Trauzzi, G., Maltby, K. M., Gibson, T. I., Ratcliffe, F. C., Hallam, J., et al. (2021). Meta-Fish-Lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes. *J. Fish Biol.* 99 (4), 1446–1454. doi: 10.1111/jfb.14852

Creedy, T. J., Andujar, C., Meramveliotakis, E., Noguerales, V., Overcast, I., Papadopoulou, A., et al. (2022). Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Mol. Ecol. Resour.* 22 (3), 847–861. doi: 10.1111/1755-0998.13502

Curd, E. E., Gold, Z., Kandlikar, G. S., Gomer, J., Ogden, M., O'Connell, T., et al. (2019). Anacapa toolkit: An environmental DNA toolkit for processing multilocus metabarcode datasets. *Methods Ecol. Evol.* 109), 1469–1475. doi: 10.1111/2041-210X.13214

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., and Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods Ecol. Evol.* 8 (12), 1888–1898. doi: 10.1111/2041-210X.12836

Diana, A., Matechou, E., Griffin, J. E., Buxton, A. S., and Griffiths, R. A. (2021). An RShiny app for modelling environmental DNA data: accounting for false positive and false negative observation error. *Ecography* 44 (12), 1838–1844. doi: 10.1111/ecog.05718

Diana, A., Matechou, E., Griffin, J., Yu, D., Luo, M., Tosa, M., et al. (2022). eDNAPlus: A unifying modelling framework for DNA-based biodiversity monitoring. *arXiv preprint*. doi: 10.48550/arXiv.2211.12213

Dixon, P. (2003). VEGAN, a package of r functions for community ecology. *J. Vegetation Sci.* 146, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x

Dodt, M., Roehr, J. T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR–flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 13), 895–905. doi: 10.3390/biology1030895

Doi, H., Akamatsu, Y., Watanabe, Y., Goto, M., Inui, R., Katano, I., et al. (2017). Water sampling for environmental DNA surveys by using an unmanned aerial vehicle. *Limnology Oceanogr: Methods* 15 (11), 939–944. doi: 10.1002/lom3.10214

Doi, H., Fukaya, K., Oka, S. I., Sato, K., Kondoh, M., and Miya, M. (2019). Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Sci. Rep.* 9 (1), 3581. doi: 10.1038/s41598-019-40233-1

Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H., and Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PloS One* 10, e0122763. doi: 10.1371/journal.pone.0122763

Dorazio, R. M., and Erickson, R. A. (2018). Ednaoccupancy: An r package for multiscale occupancy modelling of environmental DNA data. *Mol. Ecol. Resour.* 18 (2), 368–380. doi: 10.1111/1755-0998.12735

Dufresne, Y., Lejzerowicz, F., Apotheloz Perret-Gentil, L., Pawlowski, J., and Cordier, T. (2019). SLIM : a flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinf.* 201, 88. doi: 10.1186/s12859-019-2663-2

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26 (19), 2460–2461. doi: 10.1093/bioinformatics/btq461

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., et al. (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* 4 (12), 1111–1119. doi: 10.1111/2041-210X.12114

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces usingPhred. i. accuracy assessment. *Genome Res.* 8 (3), 175–185. doi: 10.1101/gr.8.3.175

FAO (2009). *International guidelines for the management of deep-sea fisheries in the high seas* (Rome: Food And Agriculture Organization Of The United Nations), 73. Available at: https://www.fao.org/3/i0816t/i0816t00.htm.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., et al. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol. Ecol. Resour.* 15 (3), 543–556. doi: 10.1111/1755-0998.12338

Fonseca, V. G., Kirse, A., Giebner, H., Vause, B. J., Drago, T., Power, D. M., et al. (2022). Metabarcoding the Antarctic peninsula biodiversity using a multi-gene approach. *ISME Commun.* 2 (1), 1–11. doi: 10.1038/s43705-022-00118-3

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., et al. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* 8, 1188. doi: 10.1038/s41467-017-01312-x

Fukaya, K., Murakami, H., Yoon, S., Minami, K., Osada, Y., Yamamoto, S., et al. (2021). Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modelling. *Mol. Ecol.* 30 (13), 3057–3067. doi: 10.1111/mec.15530

Gallego, R., Jacobs-Palmer, E., Cribari, K., and Kelly, R. P. (2020). Environmental DNA metabarcoding reveals winners and losers of global change in coastal waters. *Proc. R. Soc. B* 287 (1940), 20202424. doi: 10.1098/rspb.2020.2424

Gilbey, J., Carvalho, G., Castilho, R., Coscia, I., Coulson, M. W., Dahle, G., et al. (2021). Life in a drop: Sampling environmental DNA for marine fishery management and ecosystem monitoring. *Mar. Policy* 124, 104331. doi: 10.1016/j.marpol.2020.104331

Gold, Z., Sprague, J., Kushner, D. J., Zerecero Marin, E., and Barber, P. H. (2021). eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PloS One* 16 (2), e0238557. doi: 10.1371/journal.pone.0238557

Good, E., Holman, L. E., Pusceddu, A., Russo, T., Rius, M., and Iacono, C. L. (2022). Detection of community-wide impacts of bottom trawl fishing on deep-sea assemblages using environmental DNA metabarcoding. *Mar. pollut. Bull.* 183, 114062. doi: 10.1016/j.marpolbul.2022.114062

Griffin, J. E., Matechou, E., Buxton, A. S., Bormpoudakis, D., and Griffiths, R. A. (2020). Modelling environmental DNA data; Bayesian variable selection accounting for false positive and false negative errors. *J. R. Stat. Soc.* 69 (2), 377–392. doi: 10.1111/rssc.12390

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., and Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods Ecol. Evol.* 8 (9), 1081–1091. doi: 10.1111/2041-210X.12743

Hansen, B. K., Bekkevold, D., Clausen, L. W., and Nielsen, E. E. (2018). The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish Fisheries* 19 (5), 751–768. doi: 10.1111/faf.12286

Hao, X., Jiang, R., and Chen, T. (2011). Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. *Bioinformatics* 27 (5), 611–618. doi: 10.1093/bioinformatics/btq725

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., and Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Sci. Rep.* 9 (1), 1–10. doi: 10.1038/s41598-019-47899-7

Itakura, H., Wakiya, R., Yamamoto, S., Kaifu, K., Sato, T., and Minamoto, T. (2019). Environmental DNA analysis reveals the spatial distribution, abundance and biomass of Japanese eels at the river basin scale. *Aquat. Conservation: Mar. Freshw. Ecosyst.* 29, 361–363. doi: 10.1002/aqc.3058

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., and Hebert, P. D. (2007). Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Notes* 7 (4), 544–548. doi: 10.1111/j.1471-8286.2007.01748.x

Jerde, C. L., Mahon, A. R., Chadderton, W. L., and Lodge, D. M. (2011). Sight-unseen detection of rare aquatic species using environmental DNA. *Conserv. Lett.* 4, 150–157. doi: 10.1111/j.1755-263X.2010.00158.x

Jeunen, G. J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., et al. (2019). Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Mol. Ecol. Resour.* 19 (2), 426–438. doi: 10.1111/1755-0998.12982

Jo, T., Takao, K., and Minamoto, T. (2022). Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. *Environ. DNA* 4 (2), 271–283. doi: 10.1002/edn3.253

Kasmi, Y., Blanke, T., Eschbach, E., Möckel, B., Casas, L., MatthiasBernreuther, M., et al. (2023). Atlantic Cod (Gadus morhua) assessment approaches in the north and Baltic Sea: a comparison of environmental DNA analysis versus bottom trawl sampling. *Front. Mar. Sci.* 10, 1058354. doi: 10.3389/fmars.2023.1058354

Keck, F., Blackman, R. C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., et al. (2022). Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. *Mol. Ecol.* 31 (6), 1820–1835. doi: 10.1111/mec.16364

Kirtane, A., Wieczorek, D., Noji, T., Baskin, L., Ober, C., Plosica, R., et al. (2021). Quantification of environmental DNA (eDNA) shedding and decay rates for three commercially harvested fish species and comparison between eDNA detection and trawl catches. *Environ. DNA* 3 (6), 1142–1155. doi: 10.1002/edn3.236

Klymus, K. E., Merkes, C. M., Allison, M. J., Goldberg, C. S., Helbing, C. C., Hunter, M. E., et al. (2020). Reporting the limits of detection and quantification for environmental DNA assays. *Environ. DNA* 2 (3), 271–282. doi: 10.1002/edn3.29

Knudsen, S. W., Ebert, R. B., Hesselsøe, M., Kuntke, F., Hassingboe, J., Mortensen, P. B., et al. (2019). Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. *J. Exp. Mar. Biol. Ecol.* 510, 31–45. doi: 10.1016/j.jembe.2018.09.004

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., et al. (2006). The real-time polymerase chain reaction. *Mol. aspects Med.* 27 (2-3), 95–125. doi: 10.1016/j.mam.2005.12.007

Kumar, G., Reaume, A. M., Farrell, E., and Gaither, M. R. (2022). Comparing eDNA metabarcoding primers for assessing fish communities in a biodiverse estuary. *PloS One* 17 (6), e0266720. doi: 10.1371/journal.pone.0266720

Lacoursière-Roussel, A., Rosabal, M., and Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Mol. Ecol. Resour.* 16 (6), 1401–1414. doi: 10.1111/1755-0998.12522

LeBlanc, F., Belliveau, V., Watson, E., Coomber, C., Simard, N., DiBacco, C., et al. (2020). Environmental DNA (eDNA) detection of marine aquatic invasive species (AIS) in Eastern Canada using a targeted species-specific qPCR approach. *Manage. Biol. Invasions* 11 (2), 201. doi: 10.3391/mbi.2020.11.2.03

Leray, M., and Knowlton, N. (2015). DNA Barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl. Acad. Sci.* 112 (7), 2076–2081. doi: 10.1073/pnas.1424997112

Levi, T., Allen, J. M., Bell, D., Joyce, J., Russell, J. R., Tallmon, D. A., et al. (2019). Environmental DNA for the enumeration and management of pacific salmon. *Mol. Ecol. Resour.* 19 (3), 597–608. doi: 10.1111/1755-0998.12987

Liu, J., and Zhang, H. (2021). Combining multiple markers in environmental DNA metabarcoding to assess deep-sea benthic biodiversity. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.684955

Macher, T. H., Beermann, A. J., and Leese, F. (2021). TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. *Mol. Ecol. Resour.* 21 (5), 1705–1714. doi: 10.1111/1755-0998.13358

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Mahé, F., Rognes, T., Quince, C., De Vargas, C., and Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3, e1420. doi: 10.7717/peerj.1420

Maiello, G., Talarico, L., Carpentieri, P., De Angelis, F., Franceschini, S., Harper, L. R., et al. (2022). Little samplers, big fleet: eDNA metabarcoding from commercial trawlers enhances ocean monitoring. *Fisheries Res.* 249, 106259. doi: 10.1016/j.fishres.2022.106259

Manfrin, C., Zanetti, M., Stanković, D., Fattori, U., Bertucci-Maresca, V., Giulianini, P. G., et al. (2022). Detection of the endangered stone crayfish austropotamobius torrentium (Schrank 1803) and its congeneric a. pallipes in its last Italian biotope by eDNA analysis. *Diversity* 14 (3), 205. doi: 10.1007/s10530-022-02761-2

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 171), 10–12. doi: 10.14806/ej.17.1.200

Marx, V. (2014). PCR: paths to sensitivity. *Nat. Methods* 11, 241–245. doi: 10.1038/nmeth.2849

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., and Neufeld, J. D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinf.* 13 (1), 1–7. doi: 10.1186/1471-2105-13-31

Mathon, L., Valentini, A., Guérin, P. E., Normandeau, E., Noel, C., Lionnet, C., et al. (2021). Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Mol. Ecol. Resour.* 21 (7), 2565–2579. doi: 10.1111/1755-0998.13430

Matsen, F. A., Kodner, R. B., and Armbrust, E. V. (2010). Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinf.* 11, 538. doi: 10.1186/1471-2105-11-538

McClenaghan, B., Compson, Z. G., and Hajibabaei, M. (2020). Validating metabarcoding-based biodiversity assessments with multi-species occupancy models: a case study using coastal marine eDNA. *PloS One* 15 (3), e0224119. doi: 10.1371/journal.pone.0224119

McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., and Zenger, K. R. (2019). A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environ. DNA* 1, 14–25. doi: 10.1002/edn3.11

Merkes, C., Klymus, K., Allison, M., Goldberg, C., Helbin, C., Hunter, M., et al. (2019). Reporting the limits of detection (LOD) and quantification (LOQ) for environmental DNA assays. *Data: U.S. Geological Survey Data release*. doi: 10.5066/P9AKHU1R

Miya, M. (2022). Environmental DNA metabarcoding: A novel method for biodiversity monitoring of marine fish communities. *Annu. Rev. Mar. Sci.* 14, 161–185. doi: 10.1146/annurev-marine-041421-082251

Mousavi-Derazmahalleh, M., Stott, A., Lines, R., Peverley, G., Nester, G., Simpson, T., et al. (2021). eDNAFlow, an automated, reproducible and scalable workflow for analysis of environmental DNA sequences exploiting nextflow and singularity. *Mol. Ecol. Resour.* 215), 1697–1704. doi: 10.1111/1755-0998.13356

Munch, K., Boomsma, W., Willerslev, E., and Nielsen, R. (2008). Fast phylogenetic DNA barcoding. *Philos. Translation R. Soc. B* 363, 3997–4002. doi: 10.1098/rstb.2008.0169

Murakami, H., Yoon, S., Kasai, A., Minamoto, T., Yamamoto, S., Sakata, M. K., et al. (2019). Dispersion and degradation of environmental DNA from caged fish in a marine environment. *Fisheries Sci.* 85, 327–337. doi: 10.1007/s12562-018-1282-6

Nakagawa, H., Fukushima, K., Sakai, M., Wu, L., and Minamoto, T. (2022). Relationships between the eDNA concentration obtained from metabarcoding and stream fish abundance estimated by the removal method under field conditions. *Environ. DNA* 4, 1369–1380. doi: 10.1002/edn3.346

Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., and Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomol Detect Quantif* 1 (1), 23–33. doi: 10.1016/j.bdq.2014.08.002

Palacios-Abrantes, J., Frölicher, T. L., Reygondeau, G., Sumaila, U. R., Tagliabue, A., Wabnitz, C. C., et al. (2022). Timing and magnitude of climate-driven range shifts in transboundary fish stocks challenge their management. *Global Change Biol.* 28 (7), 2312–2326. doi: 10.1111/gcb.16058

Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., et al. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics* 17 (1), 1–10. doi: 10.1186/s12864-016-2709-z

Payne, J. L., Bush, A. M., Heim, N. A., Knope, M. L., and McCauley, D. J. (2016). Ecological selectivity of the emerging mass extinction in the oceans. *Science* 353 (6305), 1284–1286. doi: 10.1126/science.aaf2416

Perry, A. L., Low, P. J., Ellis, J. R., and Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *science* 308 (5730), 1912–1915. doi: 10.1126/science.1111322

Pikitch, E. K., Santora, C., Babcock, E. A., Bakun, A., Bonfil, R., Conover, D. O., et al. (2004). Ecosystem-based fishery management. *Science* 305 (5682), 346–347. doi: 10.1126/science.1098222

Pinsky, M. L., Selden, R. L., and Kitchel, Z. J. (2020). Climate-driven shifts in marine species ranges: Scaling from organisms to communities. *Annu. Rev. Mar. Sci.* 12, 153–179. doi: 10.1146/annurev-marine-010419-010916

Ramírez-Amaro, S., Bassitta, M., Picornell, A., Ramon, C., and Terrasa, B. (2022). Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments. *Front. Mar. Sci.* 9, 1004674. doi: 10.3389/fmars.2022.1004674

Ratnasingham, S. (2019). mBRAVE: The multiplex barcode research and visualization environment. *Biodivers Inf Sci Stand.* 3, e37986. doi: 10.3897/biss.3.37986

Ratnasingham, S., and Hebert, P. D. N. (2007). BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7 (3), 355–364. doi: 10.1111/j.1471-8286.2007.01678.x

Rodriguez-Ezpeleta, N., Morissette, O., Bean, C. W., Manu, S., Banerjee, P., Lacoursière-Roussel, A., et al. (2021). Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: Comment on "Environmental DNA: What's behind the term?" by pawlowski et al., (2020). *Mol. Ecol.* 30 (19), 4601–4605. doi: 10.1111/mec.15942

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. doi: 10.7717/peerj.2584

Rourke, M. L., Fowler, A. M., Hughes, J. M., Broadhurst, M. K., DiBattista, J. D., Fielder, S., et al. (2022). Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. *Environ. DNA* 4 (1), 9–33. doi: 10.1002/edn3.185

Russo, T., Maiello, G., Talarico, L., Baillie, C., Colosimo, G., D'Andrea, L., et al. (2021). All is fish that comes to the net: metabarcoding for rapid fisheries catch assessment. *Ecol. Appl.* 31 (2), e02273. doi: 10.1002/eap.2273

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., and Vestergaard, P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Commun. Biol.* 2, 461. doi: 10.1038/s42003-019-0696-8

Sato, M., Inoue, N., Nambu, R., Furuichi, N., Imaizumi, T., and Ushio, M. (2021). Quantitative assessment of multiple fish species around artificial reefs combining environmental DNA metabarcoding and acoustic survey. *Sci. Rep.* 11, 19477. doi: 10.1038/s41598-021-98926-5

Sato, Y., Miya, M., Fukunaga, T., Sado, T., and Iwasaki, W. (2018). MitoFish and MiFish pipeline: a mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Mol. Biol. Evol.* 35, 1553–1555. doi: 10.1093/molbev/msy074

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environmental. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Sepúlveda, A. J., Al-Chokhachy, R., Laramie, M. B., Crapster, K., Knotek, L., Miller, B., et al. (2021). It's complicated … environmental DNA as a predictor of trout and char abundance in streams. *Can. J. Fisheries Aquat. Sci.* 78, 422–432. doi: 10.1139/cjfas-2020-0182

Sepulveda, A. J., Birch, J. M., Barnhart, E. P., Merkes, C. M., Yamahara, K. M., Marin, R., et al. (2020). Robotic environmental DNA bio-surveillance of freshwater health. *Sci. Rep.* 10 (1), 1–8. doi: 10.1038/s41598-020-71304-3

Seymour, M. (2019). Rapid progression and future of environmental DNA research. *Commun. Biol.* 2, 80. doi: 10.1038/s42003-019-0330-9

Shelton, A. O., Ramón-Laca, A., Wells, A., Clemons, J., Chu, D., Feist, B. E., et al. (2022). Environmental DNA provides quantitative estimates of pacific hake abundance and distribution in the open ocean. *Proc. R. Soc B.* 289 (1971), 20212613. doi: 10.1098/rspb.2021.2613

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., and Thomsen, P. F. (2020). Population-level inferences from environmental DNA–current status and future perspectives. *Evolutionary Appl.* 13 (2), 245–262. doi: 10.1111/eva.12882

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., et al. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci. Rep.* 7 (1), 1–11. doi: 10.1038/s41598-017-12501-5

Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z., Dunton, K. J., Hinks, G., and VanMorter, S. M. (2021). Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal new Jersey, USA. *ICES J. Mar. Sci.* 78 (1), 293–304. doi: 10.1093/icesjms/fsaa225

Taberlet, P., Bonin, A., Zinger, L., and Coissac, E. (2018). *Environmental DNA. for biodiversity research and monitoring* (UK: Oxford University Press). doi: 10.1093/050/9780198767220.001.0001

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050. doi: 10.1111/j.1365-294X.2012.05470.x

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., and Kawabata, Z. (2012). Estimation of fish biomass using environmental DNA. *PloS One* 7, e35868. doi: 10.1371/journal.pone.0035868

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., and Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PloS One* 11 (11), e0165252. doi: 10.1371/journal.pone.0165252

Tillotson, M. D., Kelly, R. P., Duda, J. J., Hoy, M., Kralj, J., and Quinn, T. P. (2018). Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biol. Conserv.* 220, 1–11. doi: 10.1016/j.biocon.2018.01.030

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., and Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environ. DNA* 2 (1), 42–52. doi: 10.1002/edn3.44

Tsuji, S., Takahara, T., Doi, H., Shibata, N., and Yamanaka, H. (2019). The detection of aquatic macroorganisms using environmental DNA analysis–a review of methods for collection, extraction, and detection. *Environ. DNA* 1 (2), 99–108. doi: 10.1002/edn3.21

Urban, P., Bekkevold, D., Hansen, B. K., Jacobsen, M. W., Nielsen, A., and Nielsen, E. E. (2023). Using eDNA to estimate biomass of bycatch in pelagic fisheries. *Environ. DNA* 00, 1–16. doi: 10.1002/edn3.377

Valdivia-Carrillo, T., Rocha-Olivares, A., Reyes-Bonilla, H., Domínguez-Contreras, J. F., and Munguia-Vega, A. (2021). Integrating eDNA metabarcoding and simultaneous underwater visual surveys to describe complex fish communities in a marine biodiversity hotspot. *Mol. Ecol. Resour.* 21 (5), 1558–1574. doi: 10.1111/1755-0998.13375

West, K., Travers, M. J., Stat, M., Harvey, E. S., Richards, Z. T., DiBattista, J. D., et al. (2021). Large-Scale eDNA metabarcoding survey reveals marine biogeographic break and transitions over tropical north-western Australia. *Diversity Distributions* 27 (10), 1942–1957. doi: 10.1111/ddi.13228

Wilcox, T. M., McKelvey, K. S., Young, M. K., Engkjer, C., Lance, R. F., Lahr, A., et al. (2020). Parallel, targeted analysis of environmental samples *via* high-throughput quantitative PCR. *Environ. DNA* 2 (4), 544–553. doi: 10.1002/edn3.80

Xiong, W., and Zhan, A. (2018). Testing clustering strategies for metabarcoding-based investigation of community–environment interactions. *Mol. Ecol. Resour.* 18 (6), 1326–1338. doi: 10.1111/1755-0998.12922

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., et al. (2016). Environmental DNA as a "snapshot" of fish distribution: a case study of Japanese jack mackerel in maizuru bay, Sea of Japan. *PloS One* 11, e0149786. doi: 10.1371/journal.pone.0149786

Yu, D., Shen, Z., Chang, T., Li, S., and Liu, H. (2021). Using environmental DNA methods to improve detectability in an endangered sturgeon (Acipenser sinensis) monitoring program. *BMC Ecol. Evol.* 21 (1), 216. doi: 10.1186/s12862-021-01948-w

Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., et al. (2020). PEMA: a flexible pipeline for environmental DNA metabarcoding analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience* 9 (3), giaa022. doi: 10.1093/gigascience/giaa150

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate illumina paired-end read merger. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., et al. (2019). DNA Metabarcoding. need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* 28, 1857–1862. doi: 10.1111/mec.15060

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 22 May 2023
DOI 10.3389/fmars.2023.1140538

OPEN ACCESS

EDITED BY
Reinhold Hanel,
Thünen Institute of Fisheries Ecology, Germany

REVIEWED BY
Simo Njabulo Maduna,
Norwegian Institute of Bioeconomy Research (NIBIO), Norway
Filippo Biscarini,
National Research Council (CNR), Italy

*CORRESPONDENCE
Daria A. Zelenina
d.zelenina@gmail.com

RECEIVED 09 January 2023
ACCEPTED 02 May 2023
PUBLISHED 22 May 2023

CITATION
Zelenina DA, Sergeev AA, Kositsina AI, Soshnina VA and Shpigalskaya NY (2023) SNP-based discrimination of pink salmon stocks of the Sea of Okhotsk basin: resolution of the approach and possible ways to increase it.
Front. Mar. Sci. 10:1140538.
doi: 10.3389/fmars.2023.1140538

SNP-based discrimination of pink salmon stocks of the Sea of Okhotsk basin: resolution of the approach and possible ways to increase it

Daria A. Zelenina[1*], Alexey A. Sergeev[1], Anna I. Kositsina[2], Valeria A. Soshnina[1,3] and Nina Yu. Shpigalskaya[2]

[1]Department of Molecular Genetics, Russian Federal Institute of Fisheries and Oceanography, Moscow, Russia, [2]Laboratory of Molecular Genetics, Kamchatka Branch of Russian Federal Institute of Fisheries and Oceanography, Petropavlovsk-Kamchatsky, Russia, [3]Department of Ichthyology, Moscow State University, Moscow, Russia

In this work, we studied the intraspecific polymorphism of pink salmon, *Oncorhynchus gorbuscha* (Walbaum, 1792), the least genetically explored species among Pacific salmon and one of the central fisheries objects in the Russian Far East. The most urgent task facing Russian fishery science is to determine the proportion of fish from the main reproduction areas in mixed stocks and, based on these data, predict the number of pink salmon returning to these regions for spawning. Due to the unique feature of the species, which evolved into two allochronous lineages, these lineages have been explored independently in parallel. We designed and used here two sets of outlier SNP markers, and this allowed us to reliably distinguish the most northern (Western Kamchatka and the Magadan coast) and the most southern (Iturup Island) regional stocks as well as intermediate stocks from Sakhalin Island and the Mainland coast in both even and odd lineages of pink salmon. In addition, in odd-year lineage, we discovered pronounced genetic differences between early-run and late-run spawners in Sakhalin Island and the proximity of this early spawning form to the mainland stocks. The created baseline covers the main areas of pink salmon reproduction in the Sea of Okhotsk basin and underlies the regional identification of pink salmon in mixed marine stocks.

KEYWORDS

pink salmon, stocks discrimination, genetic structure, SNP markers, KASP assay, fisheries management, the Sea of Okhotsk

1 Introduction

Pink salmon (*Oncorhynchus gorbuscha*, Walbaum, 1792) belongs to the Pacific salmon and therefore has many features typical for the genus *Oncorhynchus*. These traits include semelparity, anadromy, and philopatry (or homing), and a pronounced population structure due to homing behavior is a characteristic of this group of fish.

The range of pink salmon covers the northern part of the Pacific Rim and its coasts. It spawns in rivers from the Korean Peninsula to the Lena River in the Asian part of its range and from California to the Mackenzie River in North America (Heard, 1991). Pink salmon has a strictly two-year life cycle, the shortest among Pacific salmon. Because of this phenomenon, the species consists of two reproductively separate lineages: one of them spawns in even years and the other in odd years. According to various estimates, the species was divided into two allochronous groups from 0,9-1,1 million (Brykov et al., 1996) to 23,6 thousand (Churikov and Gharrett, 2002) years ago. During the years of independent evolution of these two lineages, various processes took place within each of them, and as a result, they markedly differ in some morphological (Glubokovskii and Zhivotovskii, 1986; Beacham et al., 1988), ecological (Gritsenko, 1981) and karyological (Gorshkova and Gorshkov, 1983) traits. According to mitochondrial (Sato and Urawa, 2017; Podlesnykh et al., 2020; Zelenina et al., 2022), nuclear (Salmenkova et al., 2006; Beacham et al., 2012) and genomic data (Tarpey et al., 2018; Christensen et al., 2021), the most significant intraspecific differences in pink salmon exist between the lineages. The latter statement gives grounds to conduct genetic assessments for each of them independently.

Studies of the population genetic structure of pink salmon began in the 1970s, and for a long time, they were based mainly on allozyme analysis (Aspinwall, 1974; Beacham et al., 1985; Hawkins et al., 2002). However, the development of molecular genetic techniques in recent decades brought them to a new level.

The studies were carried out using a wide range of DNA markers: various mtDNA fragments, microsatellites, genes of the major histocompatibility complex (Gordeeva, 2012), and mainly covered only limited areas of the species range. Thus, it was shown that the population structure of pink salmon is much weaker compared to other Pacific salmons, which is most likely due to the lowest level of homing in this species. A notable breakthrough in understanding the population structure of pink salmon and the pathways of the distribution of the species throughout its contemporary range has been achieved using next-generation sequencing approaches. First RAD sequencing of six North American pink salmon populations (three from each lineage) was performed in 2014 (Seeb et al., 2014), and three years later, eight Asian populations (four even and four odd) were added to this research (Tarpey et al., 2018). As a result, the second level of intra-species differences in hierarchical significance was revealed. Within each lineage, the populations were clustered by their origin, making it possible to trace phylogeographic patterns within the species. In each lineage, one of the groups was formed by populations originating from the Cascadian Pleistocene glacial refugium, while the second consisted of populations of Beringian origin (Figure 1).

As for this species' more detailed population organization, a generally accepted understanding of this issue is still required. Studies based on the mitochondrial and microsatellite polymorphism analysis in Asian pink salmon did not allow for reliable discrimination of populations and regional complexes. From this point of view, great hope is placed on genomic research, making it possible to obtain many SNP markers. Of course, the putatively adaptive markers used in this study cannot be considered an ideal unbiased tool in population studies. However, they will make it possible to trace some of the patterns that determine the population structure of pink salmon.

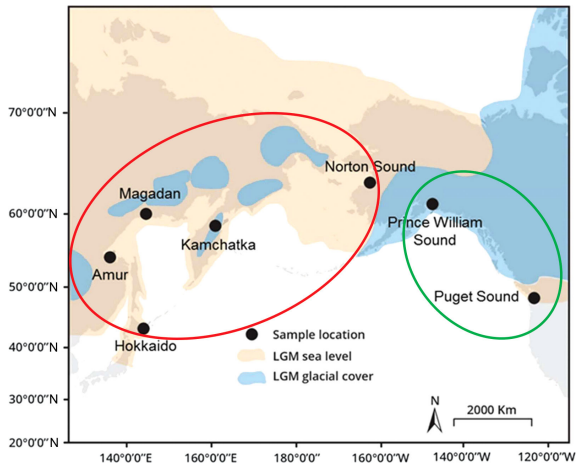


FIGURE 1
Map of the sampling localities from Tarpey et al. (2018). The description of the map is given in accordance with the paper mentioned above: "Map of the present-day coastline (grey) overlaid with the exposed land during lowered sea levels (120 to 130 m lower than present, orange) and glacial cover (blue) during the Last Glacial Maximum (LGM), approximately 18–26 thousand years ago". Populations originating from the Beringian Refugium are marked with a red oval, while those from the Cascadian Refugium – are with a green oval.

Pink salmon is the most abundant species of the genus Oncorhynchus (according to Ruggerone and Irvine (2018), 67% of total pacific salmon abundance and 48% of total biomass) and a key object of fishery in the Russian Far East. However, despite a rich history of research, this species is the most mysterious among all Pacific salmon and, from a practical point of view, the most difficult when you want to forecast the amount of fish returning for spawning to a particular area. Thus, one of the main challenges for Russian fishery science is correctly assessing the number of pink salmons migrating to a definite region.

Of course, geneticists cannot predict the total number of returning fish: the overall forecast is based mainly on traditional ichthyological approaches. Our task is to calculate the proportion of salmon from the main reproduction areas in the total population. From this point of view, the most problematic area in the Russian Far East is the Sea of Okhotsk basin. At the end of the marine feeding period, pink salmon migrate for spawning to one of five reproduction regions: 1) West Kamchatka, 2) the Magadan area, 3) the Mainland Coast and the Amur River, 4) the Sakhalin Island and 5) the South Kuriles Islands. The success of the pink salmon fishing season in these areas depends on the accuracy of forecasts: when fishermen know the estimated number of fish coming to spawn in a particular area, they can prepare vessels and fishing gear in advance. The genetic surveys are based on the mixed stocks analysis. Therefore, to successfully solve this problem, we need a reliable genetic baseline covering all the main stocks in the target region.

The main practical goal of our research was to develop the basis for a genetic tool that would allow us to reliably identify pink

salmon stocks in the Sea of Okhotsk basin, at least at the regional level. In this study, we designed two panels of SNP markers, one for an even-year lineage and one for an odd-year lineage, and assessed their ability to differentiate regional stocks. In addition, we expect to contribute to understanding the species' population structure and the processes that formed it.

2 Materials and methods

2.1 Study sites, sample collection and DNA extraction

The material for this work was selected in accordance with the purpose of the study and covers all the main regions of pink salmon reproduction in the Sea of Okhotsk basin: West Coast of Kamchatka Peninsula (hereinafter referred to as Western Kamchatka), the Sea of Okhotsk mainland coast above 59°N (Magadan Coast), the Sea of Okhotsk mainland coast below 59°N (Mainland Coast), Sakhalin Island (Sakhalin) and Iturup Island (South Kuriles) (Figure 2). In all figures, each region is marked with a certain color: Western Kamchatka (1) – green, Magadan coast (2) – blue, Mainland coast (3) – pink, Sakhalin (4) – yellow, South Kuriles (5) – red. All samples were taken in rivers from adult fish during the spawning run in July-September 2009-2020 (Table 1). Most of samples were collected by staff of Russian Federal Institute of Fisheries and Oceanography and specimen from the River Langery – by E.A. Kirillova (A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences).

Numbers of regions are given in accordance with Figure 2. Samples from Sakhalin Island are listed according to the date they were taken, starting with the earliest.

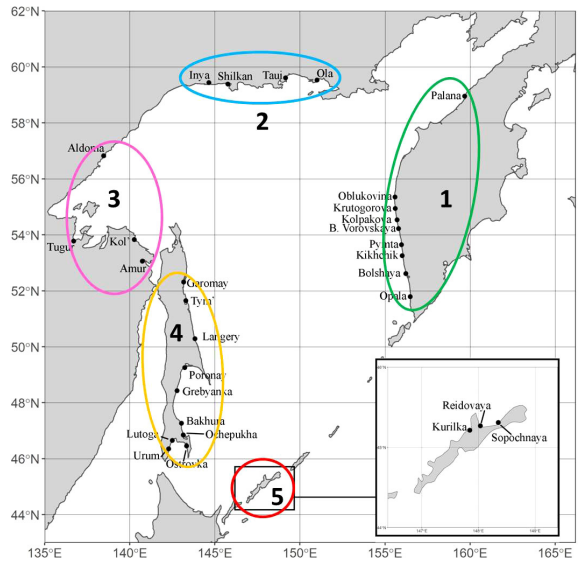


FIGURE 2
Sampling locations of *Oncorhynchus gorbuscha* in the basin of the Sea of Okhotsk. Sampling details for both lineages are given in Table 1.

Pectoral fin tissues were preserved in 96% ethanol and stored at -20°C until DNA isolation. All samples were deposited in the Collection of Genetic Materials of the Russian Federal Institute of Fisheries and Oceanography. Isolation and purification of total DNA was carried out by adsorption on AcroPrep™ 96 filter plate, 1 mL–1.0 µm, glass fiber media microcolumns (PALL, United States), as described earlier (Ivanova et al., 2006). We genotyped 47 fish from each population; thus, the total amount of the analyzed material consisted of 940 fish from the even-year spawning lineage and 1081 – from the odd-year spawning lineage.

2.2 SNP markers: search and development

The search for markers suitable for solving the tasks was based on genomic studies of pink salmon (Tarpey et al., 2018). The DNA sequences for the primer design, 50 for the even lineage pink salmon and 50 – for the odd lineage, were kindly provided by Dr. James Seeb, University of Washington. They contained SNPs, F_{st}-outliers, that supported the strongest differentiation among several populations of pink salmon, presumably originating from the Beringian refugium: one American, three Russian, and one Japanese (Figure 1).

We used Kompetitive Allele Specific PCR (KASP) for genotyping (He et al., 2014). The PCR primers were designed and synthesized by LGC Genomics, UK, and due to specific requirements for the provided DNA sequences, only half of them were suitable for primer design. Thus, at the validation stage, we had 26 loci for the even lineage and 24 for the odd one. Based on the results of preliminary testing, two panels of SNP markers were compiled, consisting of 8 loci for the "even" pink salmon and of 11 for the "odd" pink salmon. The names of the loci and the sequences of DNA-fragments, used for primer design, are available in the Supplementary Table 1.

2.3 SNP genotyping

The analysis was performed using KASP™ chemistry (LGC Genomics, UK), according to the manufacturer's standard protocol. The concentration of DNA samples was determined using Qubit 3.0 Fluorometer (Thermo Fisher Scientific, MA, USA), and then the samples were diluted to a concentration of approximately 50 ng/µL. KASP assay was carried out on The LightCycler® 480 Real-Time PCR System (Roche, USA) in 96-well plates in 10 µL reaction volume consisted of 5 µL KASP master mix, 0.14 µL KASP primer mix (both components were provided by LGC Genomics), 1 µL of DNA sample at approximate concentration 50 ng/µL and 4 µL of Milli-Q water. The cycling conditions were the following: initial stage at 94°C for 15 min, then 10 touchdown cycles at 94°C for 20 s, 61°C (decreasing by 0.6°C per cycle) for 60 s, and then 35 cycles of standard PCR at 94°C for 20 s, 55°C for 60 s.

2.4 Data analysis

2.4.1 Genetic diversity

To assess the discriminatory power of SNP loci the polymorphism information content (PIC) and average non-

10.3389/fmars.2023.1140538

TABLE 1 Sampling information. Regardless of the initial sample size, 47 fish from each locality/year were analyzed.

Region	Even-year spawning lineage			Odd-year spawning lineage		
	Locality (river)	Abbreviation	Date of collection	Locality (river)	Abbreviation	Date of collection
West Kamchatka (**1**)	Oblukovina	OBL16	Aug 2016	Palana	PAL15	July 2015
	Krutogorova	KRUT16	Aug 2016	Kolpakova	KOLP09	July 2009
	Kolpakova	KOLP16	Aug 2016	Kikhchik	KIKH15	July 2015
	B.Vorovskaya	BVOR16	Aug 2016	Opala	OPA13	July-Aug 2013
	Pymta	PYM20	Aug 2020			
	Bolshaya	BOL18	Aug 2018			
Magadan coast (**2**)	Ola	OLA16	July 2016	Taui	TAUI17	July 2017
	Taui	TAUI16	July 2016	Inya	INYA17	July 2017
	Shilkan	SHIL16	July 2016			
Mainland coast (**3**)	Aldoma	ALD16	Aug 2016	Aldoma	ALD17	July 2017
	Kol’	KOL16	July 2016	Tugur	TUG19	Aug 2019
	Amur (early run)	AMUR16E	June 2016	Kol’	KOL19	July 2019
	Amur (late run)	AMUR16L	July 2016	Amur	AMUR19	Aug 2019
Sakhalin Island (**4**)	Poronay	POR10	30.07.2010	Garomay	GAR09	02.08.2009
	Ostrovka	OSTR20	30.07.2020	Langery (early run)	LANG17E	16.06-05.07.2017
	Bakhura (early run)	BAH18E	06.08.2018	Grebyanka	GREB11	19.07.2011
				Poronay	POR09	31.07.2009
	Lutoga	LUT20	21.08.2020	Tym’	TYM09	26.08.2009
	Urum	URUM10	27.08.2010	Langery (late run)	LANG17L	01-24.09.2017
	Bakhura (late run)	BAH18L	01.09.2018	Urum	URUM09	01.09.2009
				Ochepukha	OCHE09	15.08.2009
South Kuriles, Iturup Island (**5**)	Sopochnaya (early run)	SOP16E	02.09.2016	Reidovaya (early run)	REJD13E	Aug 2013
	Sopochnaya (late run)	SOP16L	20.09.2016	Reidovaya (late run)	REJD13L	Sept 2013
	Kurilka	KUR12	Sept 2012			
	Reidovaya	REJD12	Sept 2012			

exclusion probability for identity of two unrelated individuals (NE-I) were estimated using CERVUS 3.0.7 (Kalinowski et al., 2007). For each locus average observed (Mean H_O) and expected (Mean H_E) heterozygosity across populations, as well as F-statistics parameters (F_{IS}, F_{IT} and F_{ST}) were calculated in GenAlEx 6 (Peakall and Smouse, 2006).

Genetic SNP diversity in populations was assessed in GenAlEx 6 by calculating several genetic parameters, namely the number of alleles (N), observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}).

2.4.2 Genetic structure and differentiation

To study the genetic structure, two statistical approaches were used: the clustering method implemented in STRUCTURE 2.3.4. (Pritchard et al., 2000) and principal coordinate analysis based on pairwise F_{ST} estimates performed in GenAlEx 6. The individual Bayesian population assignment test in STRUCTURE was performed with a model of admixed ancestry among populations, correlated allele frequencies and prior population information. We tested from 1 to 6 putative numbers of populations (K) with 10 iterations for each K value using 500,000 burn-in steps followed by 1,000,000 Markov Chain Monte Carlo (MCMC) generations. The number of genetic clusters was determined in StructureSelector (Li and Liu, 2018) based on the Evanno method (Evanno et al., 2005) and the Puechmaille method (Puechmaille, 2016). Calculation of pairwise F_{ST} estimates (Weir and Cockerham, 1984) and principal coordinates analysis (PCoA) based on their values were performed in GenAlEx 6.

To further assess genetic subdivision, a hierarchical analysis of molecular variance (AMOVA) was performed in Arlequin ver. 3.5 (Excoffier and Lischer, 2010). For the even-year lineage, the regional clustering was unambiguous, and the AMOVA analysis aimed to

confirm its validity. Thus, we formed putative population groups according to PCoA clusters corresponding to geographical regions. For "odd" pink salmon, the situation was not clear: the genetic clusters partly did not conform to regional ones. So we tried three analysis variants and tested i) three population groups conclusively with STRUCTURE results, ii) four groups based on geographical origin, and iii) four groups according to PCoA results. We assumed the model with the smallest intragroup differences to be the most feasible.

3 Results

3.1 Genetic diversity within loci and populations

The summary of genetic diversity for 8 "even-year" and 11 "odd-year" loci is given in the Table 2. For "even-year" loci observed heterozygosity ranged from 0.131 to 0.423 (mean: 0.224) and expected – from 0.123 to 0.403 (mean: 0.22); for "odd-year" loci H_O ranged from 0.102 to 0.455 (mean: 0.221) and H_E – from 0.111 to 0.445 (mean: 0.227). The polymorphism information content (PIC) varied from 0.123 to 0.342, with an average of 0.215, for "even-year" markers, and from 0.111 to 0.357, with an average of 0.21, for "odd-year" markers. The average non-exclusion probability (NE-I) for identity of two unrelated individuals in each locus was rather high, however the combined probability values across all loci (marked by asterisk) was satisfactory for even (0.02) and odd (0.004) lineages.

Full genetic diversity statistics by locus and population are detailed in the Supplementary Table 2.

3.2 Genetic structure and differentiation

3.2.1 Even-year spawning pink salmon

Figure 3A demonstrates the STRUCTURE analysis results of 23 populations from different regions of the Sea of Okhotsk. The proportions of individuals' ancestry to the genetic cluster were determined at K values from 2 to 5, and it needed to be made clear which of the values was optional. However, we observed some differences among geographical regions, with the strongest between the combined Northern region (I), which included the Magadan coast and Western Kamchatka, and South Kuriles (IV). To clarify

TABLE 2 Summary statistics (mean across populations) of SNP loci used to differentiate pink salmon populations in the Sea of Okhotsk basin.

	Locus name	Mean H_O	Mean H_E	F_{IS}	F_{IT}	F_{ST}	PIC	NE-I
Markers for even-year lineage	GOR-EV_55106_43	0.131	0.123	-0.06	0.008	0.064	0.123	0.762
	GOR-EV_41329_56	0.325	0.325	0.002	0.072	0.07	0.289	0.484
	GOR-EV_55959_29	0.157	0.156	-0.007	0.049	0.055	0.151	0.712
	GOR-EV_32728_58	0.423	0.403	-0.051	0.036	0.083	0.342	0.411
	GOR-EV_80318_35	0.156	0.155	-0.008	0.021	0.029	0.147	0.719
	GOR-EV_1611_62	0.334	0.337	0.007	0.123	0.117	0.309	0.456
	GOR-EV_71784_42	0.166	0.164	-0.014	0.037	0.050	0.157	0.7
	GOR-EV_41546_51	0.098	0.099	0.008	0.071	0.063	0.199	0.808
	Mean/*combined	0.224	0.22	-0.015	0.052	0.066	0.215	0.02*
Markers for odd-year lineage	GOR-OD_16062_57	0.382	0.376	-0.018	0.042	0.059	0.320	0.440
	GOR-OD_180_27	0.12	0.146	0.176	0.263	0.106	0.15	0.713
	GOR-OD_30678_51	0.146	0.144	-0.015	0.037	0.052	0.14	0.731
	GOR-OD_84997_61	0.146	0.162	0.095	0.136	0.046	0.156	0.703
	GOR-OD_26342_27	0.196	0.203	0.035	0.055	0.021	0.185	0.651
	GOR-OD_13427_28	0.455	0.445	-0.023	0.021	0.043	0.357	0.395
	GOR-OD_87107_52	0.182	0.183	0.002	0.08	0.078	0.179	0.662
	GOR-OD_52301_32	0.219	0.219	-0.002	0.064	0.065	0.206	0.614
	GOR-OD_6663_70	0.202	0.212	0.048	0.125	0.081	0.204	0.618
	GOR-OD_28942_40	0.286	0.292	0.023	0.245	0.228	0.307	0.458
	GOR-OD_33643_55	0.102	0.111	0.086	0.134	0.053	0.111	0.785
	Mean/*combined	0.221	0.227	0.037	0.109	0.075	0.21	0.004*

Mean H_O: average observed heterozygosity across all populations. Mean H_E: average expected heterozygosity across all populations. F_{IS}, F_{IT} and F_{ST}, fixation indexes; PIC, polymorphic information content; NE-I, average non-exclusion probability for identity of two unrelated individuals. *: combined non-exclusion probability across the entire set of loci.

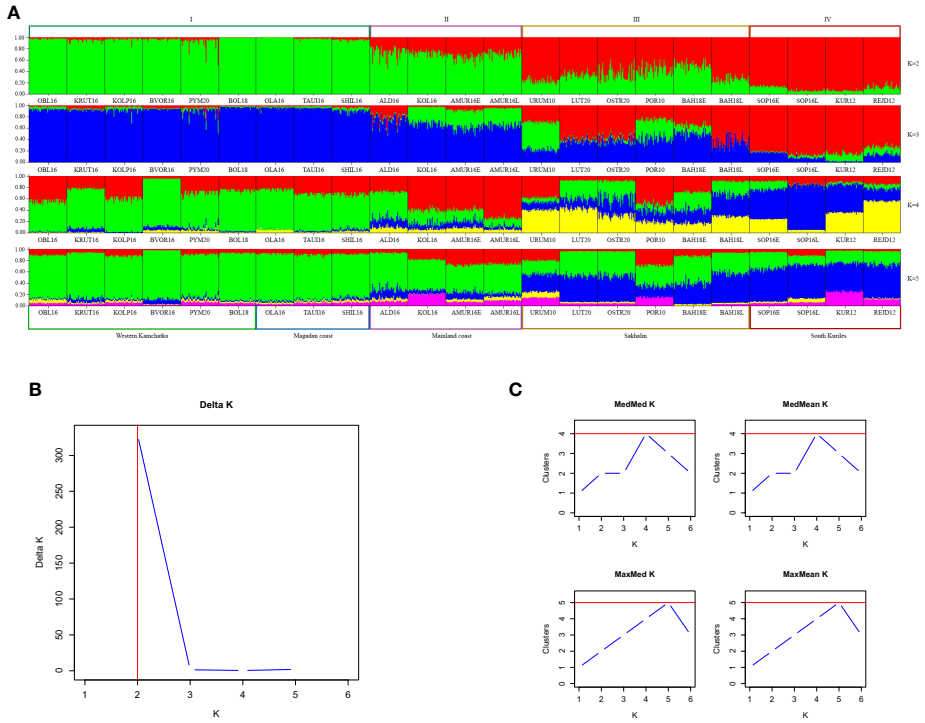


FIGURE 3
(A) Barplots reflecting the proportion of individual ancestry at K=2, 3, 4, 5 in the even-years lineage. K values are on the right. Sample designations are given in accordance with Table 1. Five regional groups are shown below the plots, and four putative clusters – above them. **(B, C)** Estimations of number of genetically distant groups within even-spawning lineage of pink salmon, based on ΔK according to (Evanno et al., 2005) **(B)** and on Puechmaille statistics (Puechmaille, 2016) **(C)**.

the most appropriate number of clusters, we applied two approaches: according to the Evanno method (Figure 3B), K appeared to be 2, whereas the Puechmaille method (Figure 3C) supported the existence of 4 clusters.

Principal coordinate analysis (PCoA) based on F_{st} genetic distances identified four population clusters formed on a regional principle (Figure 4A). This segregation was provided by the first axis, which accounted for nearly 80% of the molecular dispersion. The most distant were the Northern cluster (I), consisting of populations from Western Kamchatka and the Magadan coast, and the South Kuril cluster (IV), which included samples from Iturup Island. The Sakhalin cluster (III) and the cluster of the mainland coast of the Sea of Okhotsk (II) occupied intermediate positions.

The results of both types of analysis were generally in concordance. Therefore we considered it possible to divide the entire set of even-year spawning pink salmon populations into four groups: Northern (includes populations from West Kamchatka and Magadan area), Mainland, Sakhalin, and South Kuril (Figure 4A). The AMOVA discovered that strongly significant differences explained the majority of variations (8.46%) among these groups, and the interpopulation variability within the groups was almost 20 times less (0.45%) (Table 3); thus, our hypothesis was supported.

3.2.2 Odd-year spawning pink salmon

The STRUCTURE analysis of 20 odd-year populations (Figure 5A) revealed a pronounced clinal variability at K=2. As in the other lineage, all individuals from Western Kamchatka and the Magadan coast can be unambiguously assigned to one of the genetic clusters (group I), while the South Kuril pink salmon – to the opposite one (group IV). In contrast to the lineage of even years, there was no clear division into regions between the intermediate samples. Analysis performed at bigger K-values allowed us to find more noticeable visual differences. The optional K was 2 when we implemented the Evanno method (Figure 5B), and the Puechmaille approach (Figure 5C) found that the largest number of clusters (3) was at K=4 and 5.

However, PCoA results of odd-year spawning lineage (Figure 4B) revealed the existence of four clusters, and like in the even-year lineage Northern (I) and South Kuril clusters (IV) occupied the opposite sides of the PCoA plot. As for the two central clusters, the basis of their formation turned out to be different. All populations from the mainland coast formed a single cluster, with half of the samples from Sakhalin (II), while the other half of the Sakhalin samples formed a separate group (III). We have established a specific pattern that determines the location of each of the Sakhalin populations in a particular cluster, which is related to the timing of spawning. As seen from Table 1, all individuals sharing cluster II with pink salmon from the mainland coast came to spawn by the beginning of August and, therefore, can be considered early spawners. At the same time, pink salmon populations of late spawning dates form their separate cluster (III).

When conducting a hierarchical analysis for odd-year lineage (Table 4), we considered three models of population clustering:

10.3389/fmars.2023.1140538

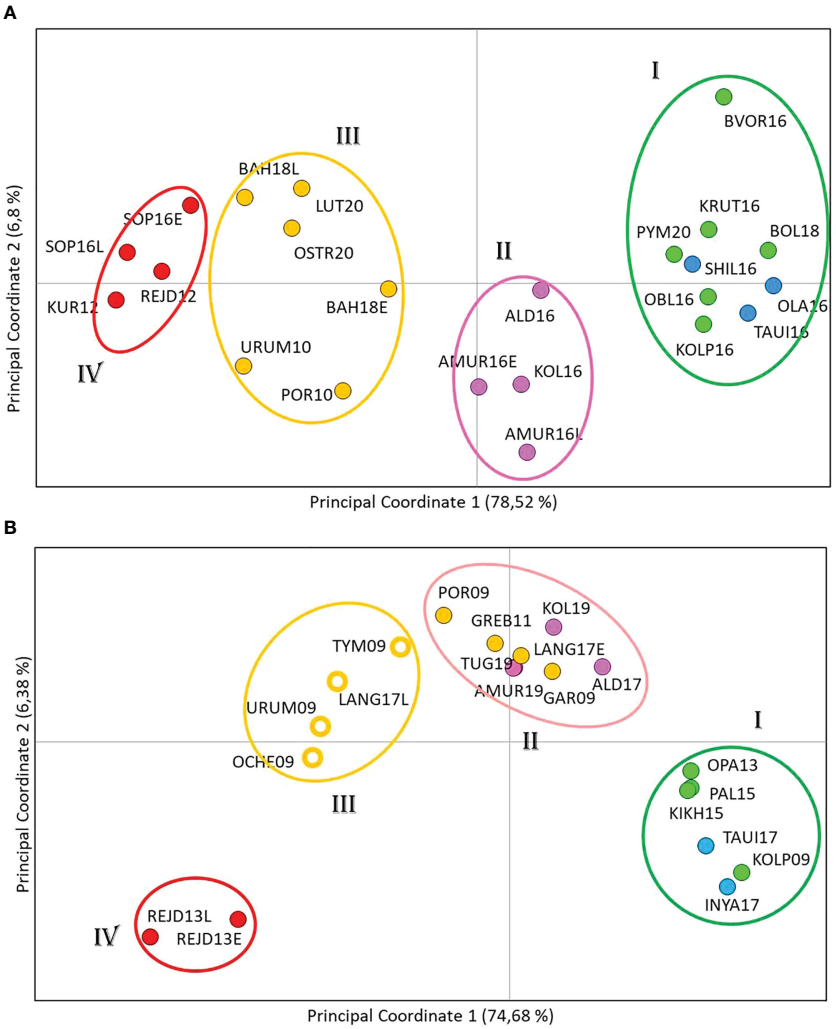


FIGURE 4
Principal coordinate analysis using the covariance matrix of pairwise F_{st} projected in a two-dimensional space for even- **(A)** and odd-spawning **(B)** lineages of pink salmon. Samples from different regions are marked with different colors.

1) three groups, that except Northern and Kuril clusters included one intermediate group (according to STRUCTURE results); 2) four groups formed on the regional principle; 3) four clusters in conformity with PCoA results. Near-zero P-values confirmed the significance of all AMOVA results. For each model, the major impact on the overall variability was provided by the differences among the groups and varied from 7.75% to 9.05%. We compared interpopulation variability within groups to determine the most likely clustering model. We found out that the lowest meaning (0.72%) was determined for the third model, suggesting the division of Sakhalin pink salmon into two groups depending on the timing of spawning.

We consider both SNP panels to be suitable for discrimination purposes. The genetic characteristics of populations were sufficient for forming stable regional clusters, both based on Bayesian estimates and the results of the analysis of principal coordinates. AMOVA outputs confirmed these assumptions.

TABLE 3 The results of AMOVA for even-years spawning pink salmon.

Source of variation	d.f.	Sum of squares	Variance composition	Percentage of variation	Fixation indexes
Among groups	3	130.517	0.08171	8.46	F_{CT}: 0,09465
Among populations within groups	19	24.444	0.00434	0.45	F_{SC}: 0,00491
Within populations	2135	1877.087	0.87920	91.09	F_{ST}: 0,08914
Total	2157	2032.048	0.96524		

All estimates were significantly different from zero ($P < 0.001$).

TABLE 4 The AMOVA analysis of odd-years spawning pink salmon based on 11 SNP-loci.

Source of variation	d.f.	Sum of squares	Variance composition	Percentage of variation	Fixation indexes
3 GROUPS_WK-MAG_KHAB-SAKH_KUR (MODEL 1)					
Among groups	2	135.893	0.12643	9.03	F_{CT}: 0,09032
Among populations within groups	17	66.484	0.02842	2.03	F_{SC}: 0,02232
Within populations	1856	2310.757	1.24502	88.94	F_{ST}: 0,11062
Total	1875	2513.134	1.39987		
4 GROUPS_WK-MAG_KHAB_SAKH_KUR (MODEL 2)					
Among groups	3	149.621	0.10637	7.75	F_{CT}: 0,07746
Among populations within groups	16	52.756	0.02188	1.59	F_{SC}: 0,01727
Within populations	1856	2310.757	1.24502	90.66	F_{ST}: 0,09339
Total	1875	2513.134	1.37327		
4 GROUPS_WK-MAG_KHAB-SAKHE_SAKHL_KUR (MODEL 3)					
Among groups	3	167.640	0.12274	8.91	F_{CT}: 0,08909
Among populations within groups	16	34.736	0.00987	0.72	F_{SC}: 0,00787
Within populations	1856	2310.757	1.24502	90.37	F_{ST}: 0,09626
Total	1875	2513.134	1.37763		

For all comparisons P-values were strongly significant ($P < 0.001$).

4 Discussion

Pacific salmon is an abundant, commercially valuable group of anadromous fish inhabiting the Pacific Ocean's northern part. The conservation of these species is ensured by the North Pacific Anadromous Fish Commission (NPAFC), the international intergovernmental organization that includes five member countries: Canada, Japan, the Republic of Korea, the Russian Federation, and the United States of America. According to the NPAFC anadromous fish Conservation Measures, the fishery of Pacific salmon is prohibited in the high seas. Almost all Pacific salmon originate in the waters of NPAFC member countries, and these countries are responsible for the harvest and conservation of their anadromous stocks. The Pacific salmon fishery occurs mainly in coastal areas or on the routes of spawning migrations within the 200-mile zones (exclusive economic zones) of the coastal states.

In many cases, delineating salmon stocks within a particular area is essential, as this allows for solving different tasks concerning their management and conservation, and fisheries genetics is responsible for answering such vital questions. For Russian salmon fisheries, one of the most urgent problems is the discrimination of pink salmon stocks in the basin of the Sea of Okhotsk. Many attempts were made to solve this problem using different kinds of standard ichthyological approaches, but their resolutions needed to be revised.

This paper presents a powerful tool that distinguishes the main regional pink salmon stocks in the Sea of Okhotsk basin. Our approach is based on advances in high-throughput sequencing: the results of the RAD-sequencing project (Tarpey et al., 2018) provided an opportunity to develop our panels of SNP markers.

Today, SNP loci are essential in fish and fisheries genetics as a convenient tool for population identification and individual assignment. Over the past two decades, the use of SNPs for fisheries management purposes has steadily increased. They first became popular because of their biallelic nature and hence the categorical information obtained (presence or absence of alleles) and reproducibility, regardless of the method of analysis (Sobrino et al., 2005; Zelenina et al., 2005). Then it became clear that the rapid development of genomic techniques promotes the further expansion of the application of SNP markers in fisheries and aquaculture (Zelenina et al., 2011).

Over the past decade, SNP markers have been widely used in mixed stock analysis and management of several species of Pacific salmon: chinook salmon (Larson et al., 2014; Beacham et al., 2022), sockeye salmon (Ackerman et al., 2011; Dann et al., 2013), coho salmon (Beacham et al., 2020; Deeg et al., 2022) and chum salmon (Kitada and Kishino, 2021; McKinney et al., 2022). However, there needed to be more information about the most abundant species of this genus, pink salmon.

Unlike most known Pacific salmon panels, our sets for genetic discrimination of pink salmon stocks consist of only a few SNP markers: 8 for lineage spawning in even years and 11 – for that from odd years. Such a small number of loci make these tools suitable for use close to fishing areas, in fisheries institutions in the Russian Far East, and the only equipment required is a real-time PCR machine.

Comparing the results obtained for both lineages, we stated that we managed to better solve the problem for the "even" pink salmon problem. All samples are reliably distributed over four genetic clusters corresponding to four main regions. However, the situation in the odd-year lineage is not so definite: samples from

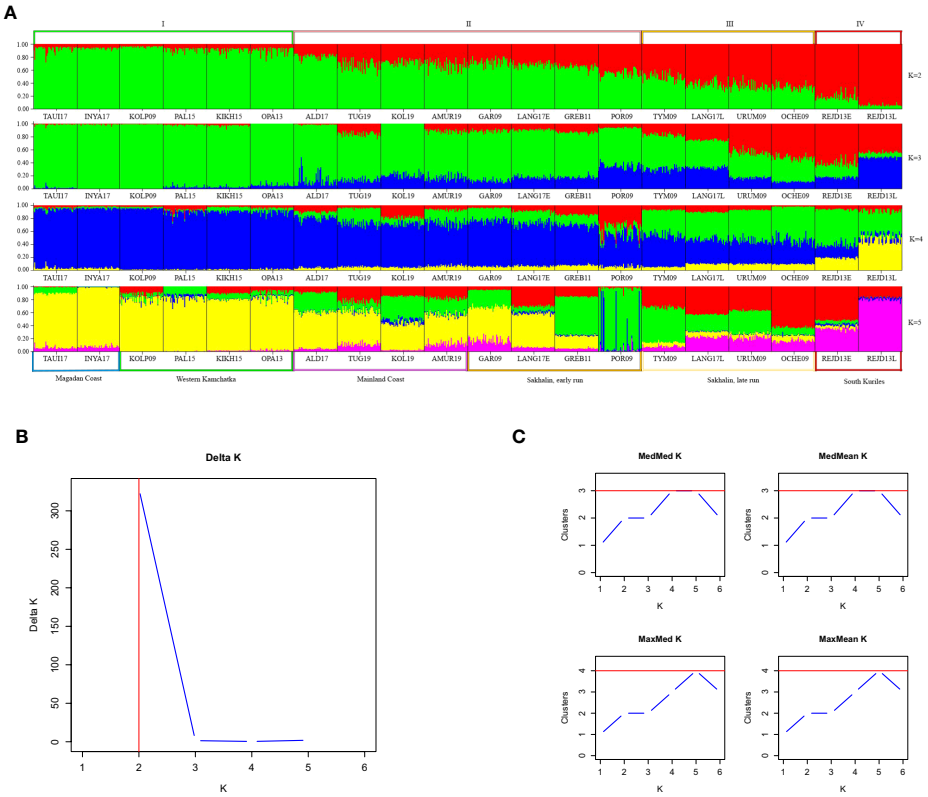


FIGURE 5
(A) Barplots reflecting the proportion of individual ancestry at K=2, 3, 4, 5 in the odd-years lineage. K values are on the right. Sample designations are given in accordance with Table 1. Five regional groups are shown below the plots, and four putative clusters – above them. **(B, C)** Estimations of number of genetically distant groups within odd-spawning lineage of pink salmon, based on ΔK according to (Evanno et al., 2005) **(B)** and on Puechmaille statistics (Puechmaille, 2016) **(C)**.

one of the main fishing areas, Sakhalin Island, were divided into two groups and distributed between two clusters, probably according to the timing of their spawning run.

We consider this phenomenon not to be accidental, and the insufficient discriminatory ability of our set of genetic markers cannot explain it. Instead, the most reasonable explanation lies in the evolutionary histories of the early and late spawning pink salmon on Sakhalin Island. In addition, data on different migration routes of these forms during their return to spawning after oceanic feeding support our assumptions (N. V. Kolpakov, E. A. Shevlyakov, V. D. Nikitin – personal communication).

Considering that the primary goal of the pink salmon forecast is to correctly predict the ratio of northern (West Kamchatka plus the Magadan coast), southern (Sakhalin plus the Mainland coast), and South Kuriles stocks, we conclude that we have reached it. Last year our "even" panel was for the first time tested by mixed stock analysis of the young-of-the-year pink salmon in the Sea of Okhotsk and verified during the fishing season, and the prediction for the northern stocks corresponded to the result (Kositsyna et al., 2022).

It should be noted that our markers were developed based on a comparison of only a few pink salmon populations, and we believe that an additional genomic study (ddRAD) of numerous samples from the rivers of the Sea of Okhotsk basin will allow us to create new sets of SNP markers with better discrimination ability.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval of the animal study were not required, as the pink salmon individuals for this study were sampled partly in the course of ichthyologic scientific surveys, partly during commercial fishing. In the first case, genetic samples (fin clips) were taken without harm to live fish; in case of commercial catches, tissues of already dead animals were used.

Author contributions

DZ wrote the manuscript draft, generated the study's main idea, designed the experiments and analyzed the data. AS, AK and VS performed the experiments. DZ and VS prepared figures and tables. NS managed the project and organized sampling work in the field.

Funding

This research was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1084) and by funding of State scientific task for the Russian Federal Institute of Fisheries and Oceanography No. 076-00007-22-01. The publication fee was covered by OOO "MAXIM MEDICAL", Arkhangelskoe, Russian Federation.

Acknowledgments

We thank researchers from the Russian Federal Institute of Fisheries and Oceanography and its Kamchatka, Magadan, Sakhalin and Khabarovsk branches, who collected genetic samples. The authors are grateful to Dr. Nikolai Mugue for editing the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1140538/full#supplementary-material

References

Ackerman, M. V., Habicht, C., and Seeb, L. W. (2011). Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the copper river, Alaska. *Trans. Am. Fish. Soc.* 140, 865–881. doi: 10.1080/00028487.2011.588137

Aspinwall, N. (1974). Genetic analysis of north American populations of pink salmon, *Oncorhynchus gorbuscha*, possible evidence for the neutral mutation–random drift hypothesis. *Evolution* 28, 295—305. doi: 10.1111/j.1558-5646.1974.tb00749.x

Beacham, T. D., Jonsen, K., Sutherland, B. J. G., Ramshaw, B., and Rondeau, E. B. (2022). Parentage-based tagging and genetic stock identification applied to assessment of mixed-stock fisheries and hatchery broodstocks for Chinook salmon in British Columbia, Canada. *Fish. Res.* 253, 106369. doi: 10.1016/j.fishres.2022.106369

Beacham, T. D., McIntosh, B., MacConnachie, C., Spilsted, B., and White, B. A. (2012). Population structure of pink salmon (*Oncorhynchus gorbuscha*) in British Columbia and Washington, determined with microsatellites. *Fish. Bull.* 110, 242–256.

Beacham, T. D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J. R., Rondeau, E. B., et al. (2020). Accurate estimation of conservation unit contribution to coho salmon mixed-stock fisheries in British Columbia, Canada, using direct DNA sequencing for single nucleotide polymorphisms. *Canad. J. Fish. Aquat. Sci.* 77, 1302–1315. doi: 10.1139/cjfas-2019-0339

Beacham, T. D., Withler, R. E., and Gould, A. P. (1985). Biochemical genetic stock identification of pink salmon (*Oncorhynchus gorbuscha*) in southern British Columbia and puget sound. *Canad. J. Fish. Aquat. Sci.* 42, 1474—1483. doi: 10.1139/f85-185

Beacham, T. D., Withler, R. E., Murray, C. B., and Barner, L. W. (1988). Variation in body size, morphology, egg size, and biochemical genetics of pink salmon in British Columbia. *Trans. Am. Fish. Soc* 117, 109–126. doi: 10.1577/1548-8659(1988)117<0109:VIBSME>2.3.CO;2

Brykov, V. A., Polyakova, N., Skurikhina, L. A., and Kukhlevsky., A. D. (1996). Geographical and temporal mitochondrial DNA variability in populations of pink salmon. *J. Fish Biol.* 48, 899–909. doi: 10.1111/j.1095-8649.1996.tb01485.x

Christensen, K. A., Rondeau, E. B., Sakhrani, D., Biagi, C. A., Johnson, H., Joshi, J., et al. (2021). The pink salmon genome: uncovering the genomic consequences of a two-year life cycle. *PloS One* 16 (12), e0255752. doi: 10.1371/journal.pone.0255752

Churikov, D., and Gharrett, A. J. (2002). Comparative phylogeography of the two pink salmon broodlines: an analysis based on a mitochondrial DNA genealogy. *Mol. Ecol.* 11, 1077–1011. doi: 10.1046/j.1365-294x.2002.01506.x

Dann, T. H., Habicht, C., Baker, T. T., and Seeb, J. E. (2013). Exploiting genetic diversity to balance conservation and harvest of migratory salmon. *Can. J. Fish. Aquat. Sci.* 70, 785–793. doi: 10.1139/cjfas-2012-0449

Deeg, C. M., Sutherland, B. J. G., Ming, T. J., Wallace, C., Jonsen, K., Flynn, K. L., et al. (2022). In-field genetic stock identification of overwintering coho salmon in the gulf of Alaska: evaluation of nanopore sequencing for remote real-time deployment. *Mol. Ecol. Res.* 22, 1824–1835. doi: 10.1111/1755-0998.13595

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Mol. Ecol. Res.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Glubokovskii, M. K., and Zhivotovskii, L. A. (1986). Population structure of pink salmon: a system of fluctuating stocks. *Biol. Mor.* 12, 39–44.

Gordeeva, N. V. (2012). High estimates of differentiation between pink salmon, *Oncorhynchus gorbuscha*, populations at locus of major histocompatibility complex MHC-I A1 support the "local stock" hypothesis. *J. Ichthyol.* 52, 68–76. doi: 10.1134/S0032945212010043

Gorshkova, G. V., and Gorshkov, S. A. (1983). Chromosomal polymorphism in even-year and odd-year pink salmon. *Dokl. Akad. Nauk SSSR.* 273, 1023–1024.

Gritsenko, O. F. (1981). Population structure of the pink salmon *Oncorhynchus gorbuscha* (Walbaum). *Vopr. Ikhtiol.* 21, 787—799.

Hawkins, S. L., Varnavskaya, N. V., Matzak, E. A., Efremov, V. V., Guthrie, C. M.III, Wilmot, R. L., et al. (2002). Population structure of odd-broodline Asian pink salmon and its contrast to the even-broodline structure. *J. Fish Biol.* 60, 370–388. doi: 10.1006/jfbi.2001.1850

He, C., Holme, J., and Anthony, J. (2014). SNP genotyping: the KASP assay. *Crop Breeding. Methods Mol. Biol.* 1145, 75–86. doi: 10.1007/978-1-4939-0446-4_7

Heard, W. R. (1991). "Life history of pink salmon (Oncorhynchus gorbuscha)," in *Pacific salmon life histories* (Vancouver: University of British Columbia Press), 119–230.

Ivanova, N. V., deWaard, J., and Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol. Ecol. Notes.* 6, 998–1002. doi: 10.1111/j.1471-8286.2006.01428.x

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106. doi: 10.1111/j.1365-294x.2007.03089.x

Kitada, S., and Kishino, H. (2021). Population structure of chum salmon and selection on the markers collected for stock identification. *Ecol. Evol.* 11, 13972–13985. doi: 10.1002/ece3.8102

Kositsyna, A. I., Shpigalskaya, N., Sergeev, A. A., Soshnina, V. A., Savenkov, V. V., Denisenko, A. D., et al. (2022). Genetic identification of juvenile pink salmon *Oncorhynchus gorbuscha* (Walbaum) in the Okhotsk sea basin based on the results of restriction analysis of mitochondrial DNA and analysis of single-nucleotide polymorphism. *Res. Aquat. Biol. Resour. Kamchatka North-West Part Pacific Ocean.* 66, 52–67. doi: 10.15853/2072-8212.2022.66.52-67

Larson, W. A., Seeb, J. E., Pascal, C. E., Templin, W. D., and Seeb, L. W. (2014). Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (Oncorhynchus tshawytscha) from western Alaska. *Can. J. Fish. Aquat. Sci.* 71, 1–11. doi: 10.1139/cjfas-2013-0502

Li, Y. L., and Liu, J. X. (2018). StructureSelector: a web based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Res.* 18, 176–177. doi: 10.1111/1755-0998.12719

McKinney, G. J., Barry, P. D., Pascal, C., Seeb, J. E., Seeb, L. W., and McPhee, M. V. (2022). A new genotyping-in-Thousands-by-Sequencing single nucleotide polymorphism panel for mixed-stock analysis of chum salmon from coastal Western Alaska. *North Am. J. Fish. Manag.* 42, 1134–1143. doi: 10.1002/nafm.10805

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in excel, population genetic software for teaching and research. *Mol. Ecol.* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Podlesnykh, A. V., Kukhlevsky, A. D., and Brykov, V. A. (2020). A comparative analysis of mitochondrial DNA genetic variation and demographic history in populations of even and odd-year broodline pink salmon, *Oncorhynchus gorbuscha* (Walbaum 1792), from Sakhalin island. *Environ. Biol. Fish.* 103, 1553–1564. doi: 10.1007/s10641-020-01040-0

Pritchard, J. K., Stefens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Puechmaille, S. J. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol. Ecol. Res.* 16, 608–627. doi: 10.1111/1755-0998.12512

Ruggerone, G. T., and Irvine, J. R. (2018). Numbers and biomass of natural- and hatchery-origin pink salmon, chum salmon, and sockeye salmon in the north pacific ocean 1925–2015. *Mar. Coast. Fish.* 10, 152–168. doi: 10.1002/mcf2.10023

Salmenkova, E. A., Gordeeva, N. V., Omel'chenko, V. T., Altukhov, I., Afanas'ev, K. I., Rubtsova, G. A., et al. (2006). Genetic differentiation of pink salmon *Oncorhynchus gorbuscha* walbaum in the Asian part of the range. *Genetika* 42, 1371–1387.

Sato, S., and Urawa, S. (2017). Genetic variation of Japanese pink salmon populations inferred from nucleotide sequence analysis of the mitochondrial DNA control region. *Environ. Biol. Fish.* 100, 1355–1372. doi: 10.1007/s10641-017-0648-4

Seeb, L. W., Waples, R. K., Limborg, M. T., Warheit, K. I., Pascal, C. E., and Seeb, J. E. (2014). Parallel signatures of selection in temporally isolated lineages of pink salmon. *Mol. Ecol.* 23, 2473–2485. doi: 10.1111/mec.12769

Sobrino, O. B., Brioan, M., and Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci. Int.* 154, 181–194. doi: 10.1016/j.forsciint.2004.10.020

Tarpey, C. M., Seeb, J. E., McKinney, G. J., and Seeb, L. W. (2018). Single-nucleotide polymorphism data describe contemporary population structure and diversity in allochronic lineages of pink salmon (*Oncorhynchus gorbuscha*). *Can. J. Fish. Aq. Sci.* 75, 987–997. doi: 10.1139/cjfas-2017-0023

Weir, B. S., and Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x

Zelenina, D. A., Khrustaleva, A. M., Volkov, A. A., Habicht, C., Smith, C., and Seeb, J. E. (2005). A case study of two genetic markers for Inter_Laboratory collaboration: SNPs provide transportability without standardization. *NPAFC Doc.* 913, 14.

Zelenina, D. A., Martinsohn, J. T., Ogden, R., Volkov, A. A., Zelenina, I. A., and Carvalho, G. R. (2011). Advanced approaches to studying the population diversity of marine fishes: new opportunities for fisheries control and management. *Russ. J. Genet.* 47, 1444–1455. doi: 10.1134/S1022795411120179

Zelenina, D. A., Zhivotovsky, L. A., Soshnina, V. A., Vilkova, O., and Yu., Glubokovsky, M. K. (2022). Intraspecific differentiation of Asian pink salmon inferred from the mitochondrial cytb gene sequences. *Rus. J. Genet.* 58, 1323–1333. doi: 10.1134/S1022795422110138

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 29 May 2023
DOI 10.3389/fmars.2023.1095171

OPEN ACCESS

EDITED BY
Jonathan Sandoval,
Flinders University, Australia

REVIEWED BY
Fausto Valenzuela,
Centro de Investigación Biológica del Noroeste (CIBNOR), Mexico
Tommaso Russo,
University of Rome Tor Vergata, Italy

*CORRESPONDENCE
Rebeca Rodríguez-Mendoza
rrodriguez@iim.csic.es

RECEIVED 10 November 2022
ACCEPTED 07 April 2023
PUBLISHED 29 May 2023

CITATION
Rodríguez-Mendoza R and Saborido-Rey F (2023) The potential use of genomic methods in bottom trawl surveys to improve stock assessments in Europe. *Front. Mar. Sci.* 10:1095171. doi: 10.3389/fmars.2023.1095171

The potential use of genomic methods in bottom trawl surveys to improve stock assessments in Europe

Rebeca Rodríguez-Mendoza* and Fran Saborido-Rey

Department of Marine Ecology and Resources, Instituto de Investigaciones Marinas, Consejo Superior de Investigaciones Científicas (IIM-CSIC), Vigo, Spain

In Europe, research surveys carried out by state governments provide the great majority of fishery-independent data. Member States (MS) in the European Union (EU) regularly conduct research surveys to provide the necessary data to assess the status of exploited fish stocks and to monitor the general condition of the marine ecosystem. In the surveys, samplings are carried out throughout the distribution range of the targeted fish species using standardized gears (e.g., trawls and seines) and other methods (e.g., hydroacoustics and underwater cameras). In the context of fish stock assessment, survey data are important because they provide indices that help tuning the stock assessment models (e.g., the index of fish abundance) and key information about the size and age distributions of the stock, the size-age relationships, the proportion of fish mature at each age, and information on reproductive performance of the stocks. However, research surveys have a number of shortcomings that include, for example, a high economic cost coupled with complex logistics and a long time required for processing the collected data. In addition, some of the parameters that are needed in stock assessment cannot be estimated from survey data for certain commercially important species. For instance, age is usually determined using hard structures (such as otoliths) in fish target species. However, for European hake, age cannot be determined accurately because there are many difficulties in interpreting the ring patterns of the otoliths. This highlights the need to look for alternative methodologies such as genomics, that have the potential of improving the data obtained from research surveys and hence, improve fish stock assessments. Considering this, we carried out a review of the bottom trawl research surveys in the EU with the purpose of: 1) identifying the current approaches for monitoring fishery resources and the ecosystem and 2) determining how genomic techniques can be used to improve survey data, taking into account the needs of current and future stock assessment in Europe.

KEYWORDS

research surveys, bottom trawl, data collection, genomic methods, fish stock assessment

1 Introduction

The status of marine fish stocks needs to be evaluated regularly to ensure fishing practices are kept at sustainable levels, through a process called "stock assessment". This process involves collecting, analyzing, and reporting demographic information to determine changes in the abundance of fishery stocks in response to fishing and, to the extent possible, predict future trends of stock abundance (NMFS, 2001). It is based on different data types, for example, fishery catches, landings, biological information recorded by onboard observers and fishery-independent data coming from research surveys. In the European Union (EU), the Common Fisheries Policy (CFP) established in Regulation (EU) No. 1380/2013 is supported by the fisheries data collected by Member States (MS). This regulation sets out the rules for managing European fishing fleets in a sustainable way and protecting marine resources exploited in European fisheries. According to the CFP, MS shall "collect biological, environmental, technical, and socio-economic data necessary for fisheries management", enabling "the assessment of: (a) the state of exploited marine biological resources; (b) the level of fishing and the impact that fishing activities have on the marine biological resources and on the marine ecosystems; and (c) the socio-economic performance of the fisheries, aquaculture and processing sectors within and outside Union waters".

To do this, a Data Collection Framework (DCF) was established in the EU and is currently set out in Regulation (EU) 2017/1004 and the Commission Delegated Decision (EU) 2021/1167, that establishes the multiannual Union programme for the collection and management of biological, environmental, technical and socioeconomic data in the fisheries and aquaculture sectors from 2022 (EU MAP). The DCF provides for the EU MAP that details the requirements for data collection by MS and a list mandatory scientific surveys at sea. Under EU MAP, a research survey at sea is defined as: "trips carried out on a research vessel, or a vessel dedicated to scientific research for stock and ecosystem monitoring, and designated for this task by the body in charge of the implementation of the national work plan established in accordance with Article 21 of Regulation (EU) No 508/2014" (Commission Delegated Decision (EU) 2019/910). The current list of mandatory surveys is established in Commission Implementing Decision (EU) 2021/1168, and includes 51 surveys that are carried out in the Baltic Sea, the North Sea and eastern Arctic, the North Atlantic and the Mediterranean and Black Sea.

In research surveys, samples are taken across the distribution range of the targeted fish species employing standardized fishing gears (e.g., trawls and seines), hydroacoustics and other devices (remotely operated vehicles, towed cameras, etc.). Survey data are important in stock assessment because they provide indices that help tuning the stock assessment models (e.g., fish abundance indices, usually the number or weight of fish caught per unit of effort). For example, the Baltic International Trawl Survey (BITS) provides two indices (one from the BITS-Q1 and another from the BITS-Q4) used in the stock assessment of cod in the eastern Baltic Sea. Surveys also provide key information about fish stocks such as the size and age distributions, the size-age relationships, the proportion of fish mature at each age, and information on reproductive performance of the stocks. Species' diet and trophic relationships can also be determined by sampling stomach contents in research surveys (Cooper, 2006). However, the sampling programs in the surveys also provide information on various ecosystem components such as hydrography, geochemistry of seawater and sediment, benthic epifauna and infauna, zooplankton, phytoplankton, and other aspects such as marine litter (ICES, 2019). In fact, the sampling design used in surveys takes into account the distribution of many different species, the overall community and the environmental characteristics. In the recent years, budget and time-dependent efforts have enabled the development of a full ecosystem monitoring programme without disrupting the fisheries time-series (ICES, 2019).

Nevertheless, traditional methods to assess the state of fish stocks through scientific surveys have experienced a very slow progress and present recognized shortcomings (Maunder and Piner, 2015). Research surveys involve complex logistics and are costly. The consequence of this is that the data collected are sparse in space and time. In addition, a long time is needed to process and analyze the collected data (Stamatopoulos, 2002). Moreover, some important parameters in fish stock assessment cannot be estimated using traditional methodologies at present. For example, the sex of juveniles cannot be determined using traditional methods (e.g., observation of the gonads), despite the importance of sex ratios to evaluate the status of exploited stocks. Similarly, difficulties have been faced when determining the age of individuals of some species, an essential parameter for growth estimates, population dynamic studies and for optimizing the harvesting time (Gursoy et al., 2005). Otoliths of some commercially important fish species such as European hake (*Merluccius merluccius*) or cod (*Gadus morhua*), for instance, have proved unreliable due to the presence of false rings, lack of definition of rings or deposition at irregular intervals (Morales-Nin et al., 1998; Hüssy et al., 2010; Ligas et al., 2011). In this context, High-throughput sequencing (HTS) genomic methods can offer the possibility to resolve some of these hurdles and enhance the data collected for fisheries assessment. In recent years, several genomic methods have been developed and applied to study different aspects of marine organisms and biodiversity. For example, Close-Kin Mark-Recapture (CKMR) is a method that allows estimating abundance and other demographic parameters (e.g., mortality rates and connectivity) from kinship relationships determined from genetic samples (Bravington et al., 2016a). Until now, it has been used for few fish species, such as bluefin tuna (Bravington et al. 2016b), white sharks (Hillary et al., 2018), brook trout (Ruzzante et al., 2019) and thornback ray (Trenkel et al., 2022), but CKMR is being considered for several more species (e.g., Maunder et al., 2021). Another genomic method that could potentially provide an indicator of stock abundance and/or biomass is environmental DNA (eDNA). eDNA is DNA that is collected from an environmental sample (e.g., water, sediment and air) rather than directly from an organism. eDNA can originate in cells from the body or waste products of organisms (Ficetola et al., 2008; Taberlet et al., 2012). This genomic tool has been mostly used to determine the presence and distribution of a species, but recent

evidence suggests the concentration of eDNA could also be used to estimate abundance and/or biomass for stock assessments in a fast and cost-effective way (Rourke et al., 2022 and references therein). Regarding age estimation, epigenetic age determination (Horvath and Raj, 2018) could provide an accurate alternative method for aging fish and overcome the limitations of the current methods (e.g., otolith readings). This genomic method analyses changes in DNA methylation associated with ageing and allows constructing epigenetic clocks to predict age, such as the one recently developed for sea bass (*Dicentrarchus labrax*) (Anastasiadi and Piferrer, 2020). Genomic methods could be used as well to improve knowledge on stock structure and substructure, which is also relevant to stock assessment. For example, restriction site-associated DNA sequencing (RAD-seq) could be used to analyze genome-wide diversity of fish populations using thousands of single nucleotide polymorphisms (SNPs). Then, for example, differences of RAD-Seq-derived SNP frequencies within stocks could be analyzed to define their substructure (e.g., Ceballos et al., 2021); or screening of SNPs could be used to search for sex markers that would allow sexing fish individuals and determining the sex structure of the stock (e.g., Palaiokostas et al., 2013; Gamble, 2016; Feron et al., 2021).

Considering this, we carried out a review of the bottom trawl research surveys in the EU with the purpose of: 1) identifying the current approaches for monitoring fishery resources and the ecosystem and 2) determining how genomic techniques can be used to improve survey data, taking into account the needs of current and future stock assessment in Europe.

2 Overview of bottom trawl surveys in Europe

Currently, there are 19 mandatory bottom trawl surveys that are carried out by EU MS and other European countries (see tables in Supplementary Materials S1, S2). These surveys are carried out in several regions: The Baltic Sea, the North Sea (ICES areas 1 and 2), the North Atlantic (ICES Areas 5-14 and the NAFO areas) and the Mediterranean and Black Sea (Figure 1). Thus, different organizations are responsible for coordinating and standardizing these surveys. The main management bodies involved in planning and coordinating research surveys in Europe are the following:

- The International Council for the Exploration of the Sea (ICES). Three groups are involved in the coordination and standardization of bottom trawl surveys in European waters: i) The ICES Working Group on Beam Trawl Surveys (WGBEAM) is in charge of planning, coordinating and implementing European inshore and offshore beam trawl surveys, ii) The ICES International Bottom Trawl Survey Working Group (IBTSWG) coordinates bottom-trawl research surveys within the ICES area and iii) The ICES Baltic International Fish Survey Working Group (WGBIFS) plans, coordinates, and implements bottom-trawl and hydroacoustic research surveys in the Baltic Sea.
- The Northwest Atlantic Fisheries Organization (NAFO): Three of the bottom trawl surveys in this review are carried

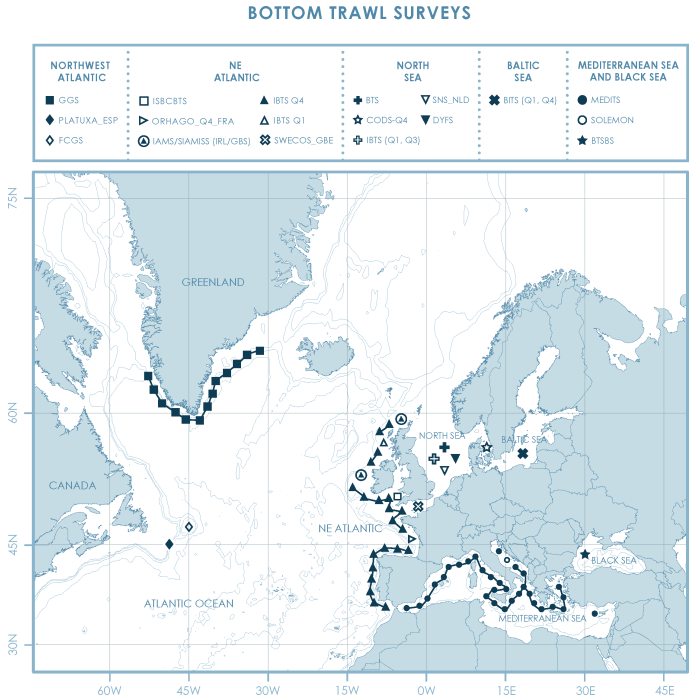


FIGURE 1
Overview of mandatory bottom trawl surveys as established in Commission Implementing Decision (EU) 2021/1168. Bottom trawl surveys are carried out in several regions: the North Atlantic (ICES Areas 5-14 and the NAFO areas), the North Sea (ICES areas 1 and 2), the Baltic Sea and the Mediterranean and Black Sea.

out in the NAFO convention area, i) The Flemish Cap Survey (3M), ii) the 3LNO Survey (Flemish Pass and Grand Bank) and iii) the Greenland Groundfish Survey (around Greenland).

- The General Fisheries Commission for the Mediterranean (GFCM), which coordinates three surveys: i) the Bottom Trawl Survey in the Black Sea, the International Bottom Trawl Survey in the Mediterranean (MEDITS) and the Beam Trawl Survey in GSA 17 (SOLEMON).

Mandatory bottom trawl surveys are diverse. Primarily, this is a consequence of the different regions covered by each survey, which means that different ecosystems with different species of marine organisms are surveyed. Regarding methodological aspects, bottom trawl surveys have different characteristics related to the fishing gear used, the survey duration, the distance from the shore, and the international participation (i.e., number of countries contributing to a particular survey). There are two types of fishing gears used in bottom trawl surveys in Europe, the otter trawl (e.g., Northeast Atlantic IBTS) and the beam trawl (e.g., North Sea BTS). Otter trawls are generally used in surveys targeting a wide variety of demersal species and that cover large areas, many of them located off-shore. Beam trawls are used mostly in surveys targeting flatfish species. In beam trawl surveys, there is no standardized gear across surveys, as the width and rigging of the beam trawls depends on the local circumstances and the vessel's capacity. The setup of the gear is mainly determined by seafloor conditions. For instance, in the beam trawl survey carried out by the Netherlands (part of the BTS survey) the gear is rigged with a flip-up rope in the central and western North Sea to avoid rocks from entering the net. In the south-eastern North Sea, no flip-up rope is used because there are no rocks in that area (ICES, 2019). Regarding distance, surveys can be divided into in-shore surveys and off-shore surveys. In-shore surveys such as the DYFS and SNS_NLD 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. 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). In addition to these differences, some surveys are focused in only a few species (for example, the Sole Net Survey in the North Sea, that targets 0-4 group sole, plaice and turbot) while others target dozens of species (e.g., MEDITS or western IBTS) (see table in Supplementary Material S1). Regarding international participation, the number of countries contributing to a particular survey also differs among surveys. About half of the surveys are carried out by one or two countries but there are surveys, such as MEDITS, the North Sea IBTS and the western IBTS, that involve many countries.

Regarding data collection, biological parameters such as length, weight, sex, maturity and age are determined in all surveys for target species. For the rest of species, taxonomic identification is carried out and, in many cases, 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 and marine litter data are recorded as well. The type of data that are collected in each bottom trawl survey is shown in the table in Supplementary Material S3.

3 Data used for stock assessment

A stock assessment is "the process of collecting, analyzing, and reporting demographic information to determine changes in the abundance of fishery stocks in response to fishing and, to the extent possible, predict future trends of stock abundance" (NMFS, 2001). Figure 2 shows the general process of stock assessment, including in

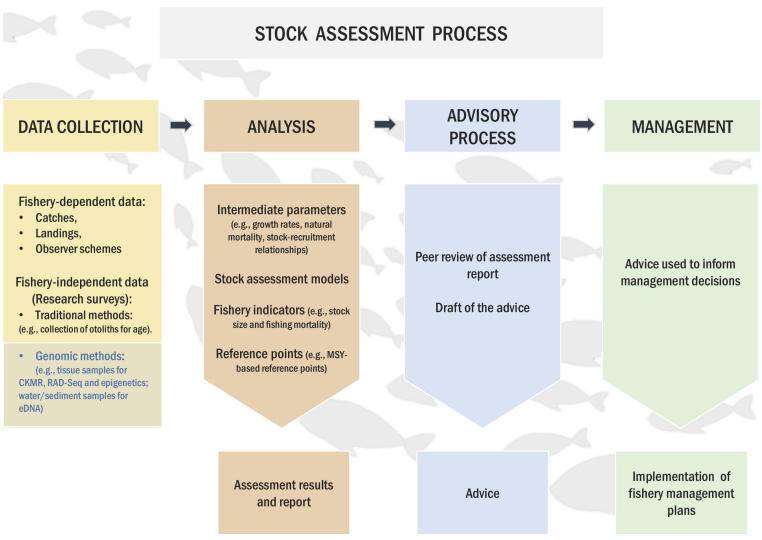


FIGURE 2
Overview of the stock assessment process.

this case, sampling for genomic methods within "Data Collection". Stock assessments are used to provide scientific advice to resource managers and policy makers about the present health and expected trends of a particular fish stock. Certainly, stock assessments are the technical basis for establishing annual fishery harvest levels (e.g., catch limits and quotas) and other management measures for the fishery (NOAA, 2012). These assessments are usually carried out by fitting mathematical models to the available information to produce simplified representations of population and fishery dynamics. Stock assessment methods have progressed over time, from descriptive models to complex statistical models with many estimated parameters and formal approaches to evaluating uncertainty (Cadrin and Dickey-Collas, 2015).

Stock assessment models require different sources of information depending on their assumptions. In general, the data used in assessments includes commercial catches and landings, effort, biological information recorded by onboard observers (e.g., age and length composition data) and fishery-independent data coming from research surveys. As already mentioned, survey data are important in stock assessment because they provide indices that help tuning the stock assessment models (e.g., fish abundance indices, usually the number or weight of fish caught per unit of effort). For example, survey abundance indices are used in analytical stock assessment models such as Stock synthesis models (Methot and Wetzel, 2013), in State-space assessment models (SAM, Nielsen and Berg, 2014; Berg and Nielsen, 2016) or in Stochastic surplus production model in continuous time (SPiCT, Pedersen and Berg, 2017). Equally important, surveys provide information about fish stocks such as the size and age distributions, the size-age relationships, the proportion of fish mature at each age, and information on their reproductive performance. For instance, the eastern Baltic cod (cod in ICES subdivisions 24-32) is currently assessed using an age-length based Stock synthesis model, where abundance indices from the Baltic International Trawl Surveys (BITS Q1 and Q4) are used. In this model, age-length keys are based on otolith readings from these surveys as well, and maturity and weight-length data are provided by the BITS Q1 survey (ICES, 2022b). In the case of the northern European hake (ICES subareas 4,6 and 7, and in divisions 3a, 8ab) the assessment is done using a length-based and sex-disaggregated Stock Synthesis model. In this model, several survey indices are used, including those from the western IBTS survey (i.e., from the French Southern Atlantic Bottom trawl survey (EVHOE-WIBTS-Q4), the Spanish Porcupine Bottom Trawl Survey (SpPGFS-WIBTS-Q3), the Irish Groundfish Survey (IGFS-WIBTS-Q4), the French surveys in the Bay of Biscay (FR-RESSGACQ)) and from the Irish Anglerfish and Megrim Survey (IAMS_IRL) (ICES, 2022c). For other species (e.g., data-limited species such as stocks in the ICES stock category 3), management advice relies solely on a survey trend-based assessment. Such assessments are "restricted to qualitative advice about whether the stock is trending up, down or is stable, and on whether the stock is approaching a possible trigger for management action (e.g., the lowest point in the abundance index time series)". For this reason, these assessments cannot be used to provide advice on the absolute level of the fish stock or the direct effect of fishing on it. (ICES, 2012). For example, for the beaked redfish in Division 14b, the assessment is based only on the Greenland Groundfish Survey (GGS) abundance indices (ICES, 2022a).

The assessments are carried out by several expert working groups within ICES, NAFO, GFCM and the European Commission's Scientific Technical and Economic Committee on Fisheries (STECF). A list of these groups can be found in Table 1.

4 Genomic data applications for stock assessment

At present, there are several genomic methods that could be used to estimate key parameters of fish populations, and possibly resolve some of the current difficulties and enhance the data collected for fisheries assessment. In this section, an overview of the most relevant methods is presented, namely, CKMR, eDNA, epigenetic age determination and other approaches, such as RAD-Seq.

TABLE 1 Expert working groups involved in stock assessments that are related to bottom trawl surveys in the European Union.

Organization	Acronym	Expert Working Group
GFCM	SGSABS	Subregional Group for Black Sea Stock Assessment
GFCM	WGSAD	Working Group on Stock Assessment of Demersal Species
ICES	HAWG	Herring Assessment Working Group for the Area South of 62°N
ICES	NWWG	Northwestern Working Group
ICES	WGBFAS	Baltic Fisheries Assessment Working Group
ICES	WGBIE	Working Group for the Bay of Biscay and the Iberian Waters Ecoregion
ICES	WGCSE	Working Group for the Celtic Seas Ecoregion
ICES	WGEF	Working Group on Elasmobranch Fishes
ICES	WGNSSK	Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak
NAFO	SC-STACFIS	Scientific Council - Standing Committee on Fisheries Science

4.1 CKMR

Close-Kin Mark-Recapture (CKMR) is a method that was developed to estimate abundance and other demographic parameters (e.g., mortality rates and connectivity) from kinship relationships determined from genetic samples (Bravington et al., 2016a). This method is based in the traditional mark-recapture framework, where population size and other parameters can be estimated from marking experiments adequately designed for this purpose. In CKMR, however, DNA tags are used instead of traditional physical tags. With DNA tags, an individual is marked by its presence in the sample, and "recaptured" if one or more close relatives are also present (e.g., parents and offspring or siblings). The idea behind this is that finding relatives is less likely to occur in bigger populations, so the number of "recaptures" provides information on adult abundance (Bravington et al., 2016a)

Regarding trawl abundance indices, a fundamental problem is the unknown detection probability or catchability, which prevents from obtaining absolute abundance indices. Therefore, CKMR method represents a major advance that allows estimating absolute abundance, leading to a change in paradigm (Trenkel et al., 2022 and references therein). Until now, CKMR has allowed determining absolute abundance of several fish populations around the world, such as southern bluefin tuna (Bravington et al., 2016b), white sharks in eastern Australia and New Zealand (Hillary et al., 2018), brook trout in Nova Scotia (Ruzzante et al., 2019) and thornback ray in the Bay of Biscay (Trenkel et al., 2022). CKMR is also being considered to estimate abundance of other species like silky shark and bigeye tuna in the eastern Pacific Ocean (Maunder et al., 2021).

At present, CKMR has already produced significant progress in the stock assessment of some species. For example, in the case of southern bluefin tuna (*Thunnus maccoyii*), the stock assessment relies on CPUE from fisheries as a primary index of (relative) abundance (Bravington et al., 2016a). For this species, CKMR has demonstrated its ability for estimating three of the key parameters of fisheries stock assessments, namely, the absolute abundance of adults and total adult mortality and selectivity of the southern bluefin tuna stock. In this case, CKMR has been used as a stand-alone assessment framework that is independent of the catch and effort of southern bluefin tuna fisheries, avoiding the large uncertainties associated with the interpretation of the CPUE series derived from them. The results of CKMR for southern bluefin tuna have been already reviewed and accepted by the Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). This method has been adopted by the CCSBT as a fisheries-independent method to monitor spawner abundance (Davies et al., 2020).

Another key parameter in fisheries stock assessment and management is natural mortality (M), because it directly affects estimates of stock productivity and reference points. This parameter is very difficult to estimate and has a large uncertainty associated, mainly because there is a lack of information and it is difficult to obtain unbiased data (e.g., age-composition in the absence of fishing). In practice, there are several approaches for estimating M (e.g., methods based on life history theory; empirical relationships; analysis of tagging data; analysis of catch-at-age data or estimation within integrated population models) (Maunder et al., 2023). At present, mark-recapture is considered one of the most reliable methods for estimating M, in spite of limitations such as such as non-reporting of tags, tag loss and tag-related behavioral changes and mortality. In this context, CKMR, is probably the most promising direct method to estimate M for stocks, because it avoids many of the problems associated with conventional tagging mentioned before (Maunder et al., 2023).

For further information about CKMR and its use for estimating fish population parameters, the reader is referred to the relevant articles that are also part of this Research Topic "Prospects and Challenges for the Implementation of HTS Genetic Methods in Fisheries Research Surveys and Stock Assessment".

4.2 eDNA

eDNA refers to DNA that can be extracted from environmental samples (e.g., soil, water or air), without first isolating any target organisms. eDNA is actually made up by a complex mixture of genomic DNA from many different organisms, and may be partly degraded (i.e., formed by small fragments of DNA molecules) (Taberlet et al., 2012). In recent years, eDNA has been increasingly applied to assess marine fish diversity (e.g., Fraija-Fernández et al., 2020; Zhou et al., 2022), as it is a non-invasive and cost-effective approach that yields good results (Miya, 2022), In addition to this, recent evidence suggests the concentration of eDNA could also be used to provide a rapid, cost-effective indicator of abundance and/or biomass for fisheries stock assessments (Rourke et al., 2022 and references therein). However, more research is needed before this tool can be routinely applied in the context of research surveys, because factors affecting eDNA concentrations in seawater (or sediment) need to be better understood. For example, it is known that biotic factors such as intraspecific variation in DNA production, shedding rates among individuals, metabolic rate and size of individuals affect the quantity of DNA that is released to the environment. Also, abiotic factors such as underwater currents or water temperature can affect DNA concentrations as well (Rourke et al., 2022). Genomic methods, such as eDNA, could help overcome some situations where survey approaches fail and may produce more reliable data than research surveys in such cases. For instance, in trawl surveys, the catch may not be representative of the true abundance and biomass of fish in a determined area (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 behavior of fish ahead of the trawl gear (some are herded into the net by the otter boards while others show net-avoidance behavior), 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 genomic methods could provide more accurate information than traditional surveys. For example, Thomsen et al.

(2016) showed that the estimated abundance of Greenland sharks in the continental slope in Southwest Greenland using eDNA was much higher than the one calculated from the trawling data. Given that Greenland sharks are considered to be highly abundant in the surveyed area (Nielsen et al., 2014), the authors suggest that the ability of these sharks for escaping from the trawling net must be affecting the estimated abundance from the trawl survey. So, in this case, the results from using eDNA methodology appeared to reflect shark abundance more accurately. Moreover, it is difficult to obtain trawl-based abundance estimates for those stocks closer to the shore, because rocky coasts or shallow waters are not accessible to trawling. For these situations, eDNA could improve existing fish stock monitoring programs (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 evidences the limited ability of trawl surveys to capture taxa in certain types of habitats, or fish at different life stages, sizes, and behaviors, whereas eDNA can in theory detect fish in any type of habitat, regardless of their swimming behaviors and sizes once the eDNA metabarcoding protocols are well-established (Afzali et al., 2021). Nevertheless, there are still insufficient studies where the efficacy of traditional methods (surveys) versus genomic methods has been formally compared. To the present, most of these studies have been focused on eDNA metabarcoding (i.e., biodiversity approach). For example, the study carried out by Thomsen et al. (2016) demonstrated that eDNA results are equivalent to catch data obtained from trawling: In total, 26 fish families were detected by both methods, while three families were only detected using eDNA and two families only by trawling. While eDNA was able to identify species that are not commonly observed in the trawling nets, trawling detected other taxa that were not recognized at the species level by eDNA. As a matter of fact, different studies suggest that traditional surveys and eDNA metabarcoding are complementary, because they offer a broader picture of marine biodiversity (Deiner et al., 2017; Evans et al., 2017; Gillet et al., 2018; Stat et al., 2019).

A detailed explanation of how eDNA works and its application for fisheries assessment in marine environments is available in the article by Ramírez-Amaro et al. (2022) and other relevant articles in this Research Topic: "Prospects and Challenges for the Implementation of HTS Genetic Methods in Fisheries Research Surveys and Stock Assessment".

4.3 Epigenetic age determination

Epigenetic age determination is based on the clock-like patterns of change in DNA methylation that occur at particular cytosine-guanine dinucleotides in the genome (i.e., CpG sites). At these sites, the proportion of methylated copies of the genome in the cells from a particular tissue sample either progressively increases or decreases across the life span (Guevara and Lawler, 2018). Chronological age predictors built from DNA methylation are termed 'epigenetic clocks' (Zhang et al., 2019). Regarding age estimation, epigenetic clocks could provide an accurate alternative method for aging fish and overcome the limitations of the current methods (e.g., otolith readings). At present, epigenetic clocks have been built for a few marine species (e.g., seabass, Anastasiadi and Piferrer, 2020) but further evidence is needed to validate its use across different fish species. In the future, development of universal epigenetic clocks that are evolutionary conserved across a broad range of fish species should be explored, as this would provide a fast, reader-independent tool for ageing fishes that are sampled in bottom trawl surveys. Epigenetic age determination may offer a solution for obtaining accurate age structure of monitored fish stock in cases where traditional otolith reads is challenging. It must be considered that many teleost fish species do not show otolith growth increments or other phenotypic features related to age, complicating monitoring of the population dynamics for those species (Mayne et al., 2020 and references therein). Also, extracting otoliths for age estimation is a lethal procedure and is undesirable in the case of endangered, threatened and protected species. In addition, age estimates that are based on counting otolith increments may be affected by large biases and uncertainties due to the combination of processing and interpretation errors. Such errors are then carried on into the growth and mortality estimates and other demographic rates required for population dynamics models (Dortel et al., 2013). This is the case of the Atlantic cod (*Gadus morhua)* in the eastern Baltic Sea, for which ageing uncertainty has led to failed analytical stock assessment, greatly affecting the management of the stock 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., European hake – *Merluccius merluccius*, or monkfish – *Lophius* spp.) (ICES, 2021). Moreover, epigenetic age determination is non-lethal, which makes it very attractive in the case of threatened species, such as sharks. A detailed explanation of the method to determine age using epigenetics and its applications in fisheries management and conservation biology is available in Piferrer and Anastasiadi (2023).

4.4 Other applications of genomic techniques

In fisheries management, genomics has been successfully used to define fish stocks and quantify the extent of adaptive divergence and connectivity between them, also allowing performing mixed-stock analysis with substantially increased resolution (Bernatchez et al., 2017). Genomic high-throughput methods are now enabling the discovery and genotyping of thousands of genetic markers for all kinds of species at affordable costs, including non-model organisms. Because of this, these methods are revolutionizing ecological, evolutionary and conservation genetics (Andrews et al., 2016). The availability of thousands of molecular markers to representing the genome, has greatly expanded the number of characters available for stock identification, improving also our

understanding of genetic geographical variation. In addition, more variable genetic markers have been developed offering a greater sensitivity for detecting genetic differences among groups. For example, genetic differences have been found within many coastal and oceanic species that were initially considered to be genetically homogeneous (Cadrin, 2020 and references therein).

In this context, restriction site-associated DNA sequencing (RAD-seq) has been increasingly applied to analyze genome-wide diversity of fish populations using thousands of genetic markers called single nucleotide polymorphisms (SNPs). RAD-Seq is a fractional genome sequencing strategy that is designed to interrogate anywhere from 0.1 to 10% of a selected genome, instead of analyzing the whole genome. RAD-Seq works by first fragmenting the target genome using a restriction enzyme. After digestion, DNA is transformed into a pool of DNA of fragments (i.e., library) suitable for use on a sequencing platform (e.g., Illumina). Sequence data are then analyzed to identify and score genetic variations in the samples or population of interest. Parallel screening of the thousands of markers derived from RAD-Seq allow then researchers to map natural variation and induced mutations (Floragenex, 2015).

In the context of fisheries, differences of RAD-Seq-derived SNP frequencies can be analyzed to define stock structure. For instance, the population structure of European hake was analyzed in samples from the Mediterranean Sea, the northwestern Iberian Peninsula (southern stock), the eastern Bay of Biscay (northern stock), and the Norwegian Sea (northern stock) using RAD-seq (Leone et al., 2019). The study indicated that hake in the Norwegian Sea is genetically different from that of the rest of the locations under study and confirmed differentiation of Mediterranean and northeast Atlantic locations. However, samples from the eastern Bay of Biscay and the northwestern Iberian Peninsula were not found to be genetically different. These results imply that samples from the northern stock belong to different genetic populations, and that samples belonging to locations included in the northern and southern stocks are part of a single genetically homogeneous population. This information can help improving management of European hake by defining more meaningful management units. In the Northeast Atlantic, assessment and management of the European hake is currently done independently for two stocks (southern and northern), separated by the Capbreton Canyon. However, as shown by the results of that study, the southern stock and the southern part of the northern stock should be assessed jointly.

Thus, RAD-Seq and similar approaches can help improve stock assessments mainly by providing better information of stock identity, their spatial boundaries and connectivity between different stocks. Generally, including spatial structure in assessments when the available evidence indicates stock heterogeneity leads to less bias in estimates of management interest, even if resulting in less precision. In addition, failure to include spatial and stock structure in the management of fisheries can lead to overexploitation of local populations (Punt et al., 2020 and references therein). Nevertheless, incorporating spatial and stock structure into stock assessments is still challenging. Using spatial models leads to more complex population dynamics models, requiring additional parameters that describe movement, the spatial allocation of recruitment and, potentially, spatial variation in biological parameters (especially growth rates) (Punt, 2019). Moreover, including movement among spatial areas introduces another dimension of flexibility that will be confounded with natural mortality, growth, selectivity, and recruitment. Thus, future stock assessment packages will need to provide tools to diagnose and control this confounding to provide accurate assessments (Punt et al., 2020).

Screening of SNPs could be also used to search for sex markers that would allow sexing fish individuals. Such genomic-based sex markers would be useful for developing simple sex identification assays for species or developmental stages (e.g., eggs and larvae) where sex of the individuals is not identifiable. In addition, the sex of the younger juveniles cannot be assigned using traditional methods in those species that lack sexual dimorphism. Sex determination using RAD-Seq would allow us to determine the sex of such individuals, and for larger specimens as well, with the advantage of being non-lethal (important for protected, endangered or threatened species, for example). At present, genomic markers for sex have been identified in a variety of fish species. For example, in the Atlantic halibut (*Hippoglossus hippoglossus*), a major sex determining locus has been identified and assays for 10 SNPs with significant association with phenotypic sex have been used to distinguish males and females successfully (Palaiokostas et al., 2013). RAD-Seq has also been used to identify a male specific genomic region in Atlantic cod and to identify positions in the genome that displayed significant differences in read depth between males and females (Kirubakaran et al., 2019). A diagnostic test to determine gender, using a simple PCR reaction, was also developed in that same study.

So far, RAD-seq has been the most popular approach for population genomics of non-model organisms. However, one of the main limitations of this method is that large stretches of the genome between markers remain unsampled and signatures of selection and adaptive divergence that are highly localized in the genome can be missed (Lou et al., 2021 and references therein). Thus, whole genome sequencing approaches are increasingly being used to overcome this limitation. For example, a recent study used whole-genome screening to investigate the population structure of Greenland halibut (*Reinhardtius hippoglossoides*) in the NW Atlantic. This approach allowed identifying a weak but significant divergence between Greenland Halibut from the Gulf of Saint Lawrence and those from the rest of locations analyzed in the Northwest Atlantic (Ferchaud et al., 2022). In this case, whole-genome sequencing allowed finding all the SNPs that were differentiated, including those in the very restricted genomic regions that explained most of the differentiation between the two stocks. The authors of the study highlight that a restricted approach (such as RAD-Seq) would have likely missed such narrow regions, potentially leading to a lower average differentiation.

5 Final considerations

Stock assessments are continuously improving, mainly due to methodological and computational advances (Punt et al., 2020), but

also due to improvements in data collection that allow better estimates of stock parameters. In this context, the information coming from genomic methods, such as abundance estimates (CKMR, eDNA), mortality rates (CKMR), age (epigenetics) and sex (RAD-Seq) determination and stock substructure (RAD-Seq) has the potential to improve stock assessments. In general, the use of genomic approaches can improve stock assessments by allowing, facilitating or improving estimation of key parameters and by providing insights of stock structure and connectivity. However, this does not necessarily mean that these approaches will end up substituting traditional data collection. The set of traditional methods is the outcome of a long process of adaptation to the goals and needs of stock assessments, while genomic methodologies are scientific developments which still need to follow further innovation steps for fitting them with the stock assessment specific needs. It is expected that genomic 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 genomic methods regarding traditional approaches is yet to be demonstrated in a variety of scenarios. Data obtained from genomic 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 genomic accuracy and precision in comparison with traditional methods is required for each of the stocks where the genomic methods can be expected to be implemented.

Also, the type of data that can be collected by using the different approaches 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 genomic method is capable of determining size structure because the length of a fish cannot be determined by genomic analysis. When comparing the type of data that can be obtained using traditional methodology and genomic methods (see Table 2), it is evident that genomic methods do not provide all the parameters that traditional methods are able to provide for the stock assessment of targeted species. Traditional surveys, in addition, provide information for monitoring the general conditions of the marine environment (e.g., marine litter and pollutants). Most notably, genomic methods do not provide information on size structure and maturity, and there are difficulties for estimating abundance-at-age.

It is clear that traditional methods show a number of difficulties in relation to data collection and stock assessment that genomic methods may help to overcome. But at present, substituting traditional methods with genomic methods would lead to a loss of information, and both methodologies seem to be rather complementary than substitutes. For instance, the few assessments based on the CKMR methodology (e.g., Bravington et al., 2016a; Ruzzante et al., 2019; Marcy-Quay et al., 2020) also required traditional parameters (e.g., length-based estimated age or sex).

From the above, it stands out that data collection, in terms of quantity and quality, would benefit from complementing traditional surveys with genomic methods. The next question that arises is: How can genomic methods be implemented into the already existing traditional surveys? To answer this, two options could be considered: the first would be to implement directly genomic methods into existing surveys. This would mean adding the collection of tissue and water/sediment samples required by genomic methods on top of the already existing activities in the survey, but without altering the survey program (i.e., survey duration, number of sampled stations, etc.). The second option would be to re-structure the surveys, for example, by reducing sampling effort using traditional methodology and giving genomic methods their own space in the survey.

The first option can be considered for surveys with sufficient resources —mainly, scientific staff and storage capacity, given that the extraction of samples for the genomic analysis does not require significant additional equipment and/or consumables onboard. In these cases, integrating the tissue sample collection for genetic analysis into the regular survey work can provide a solution to this without increasing costs significantly. However, even if costs do not increase during the survey itself, the processing and analysis of samples using genomic methods is still costly and would add up to the already expensive survey. The second option (restructuring the survey) could be considered when data collection for genomic methods cannot be directly implemented on an existing survey or if cost reduction is necessary. In that case, the survey would need to be restructured, so that dedicated resources (e.g., scientific/technical staff time) can be reorganized to gather samples for both traditional and genomic methods. For example, in the case of a particular survey which could be reduced from 30 days to 20 days, the first 5 days could be dedicated to collect eDNA samples and the remaining 15 days to collect biological samples (for genomics, size structure, age, sex and maturity) but from fewer stations (due to fewer days available). This could reduce the total survey cost, but not without several consequences regarding the quantity and quality of the data obtained. According to ICES (2020), reductions in survey effort (i.e., number of sampled stations, tow duration, survey frequency) can have consequences on many aspects of the information produced from surveys. These consequences not only may affect stock assessments, but would also have an impact on fisheries management, ecosystem indicators, and fisheries research (e.g., loss of data on ecosystem indicators, loss of non-target fish and invertebrates' abundance data, loss of food habits information needed for multispecies ecosystem models, loss of platforms for novel studies, etc.).

Another possibility worth exploring is complementing traditional surveys with genomic data obtained using commercial fishing vessels. For example, Russo et al. (2021) have investigated the possibility of assessing catch composition of single hauls carried out by trawlers by applying eDNA metabarcoding to the dense water draining from fishing nets just after the end of hauling operations (i.e., slush). In that study, the authors demonstrate that the fish assemblages identified using eDNA in the slush reflected those determined by visual inspection of net content (approx. 71% of species and 86% of families of fish) and detected a strong relationship between read counts and species abundances in the catch. Thus, this approach could be upscaled to serve as a powerful source of information on the structure of demersal

TABLE 2 Information provided by the traditional fisheries surveys, compared with the potential information provided by genomic methods.

Type of data	Traditional methods Y/N; comment	Genomic methods 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. However, in a number of situations 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); it has been used for some fish species, such as bluefin tuna (Bravington et al., 2016b), white sharks (Hillary et al., 2018), brook trout (Ruzzante et al., 2019) and thornback ray (Trenkel et al., 2022), and is being considered for several more species (e.g., Maunder et al., 2021). A noteworthy aspect is that sample size for CKMR depends on the expected population size, so the method is likely not applicable to species with large population sizes, i.e., systems numbering in the tens of millions of individuals or larger (Ruzzante et al., 2019).
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).	Potentially yes (eDNA – using for example, quantitative PCR, qPCR); Current evidence and studies demonstrate positive correlations between detectable DNA in the environment and abundance/biomass of the species of interest (Rourke et al., 2022). 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 become available for the species of interest (e.g., sea bass, see Anastasiadi and Piferrer, 2020, northern red snapper and red grouper, see Weber et al., 2022).
Sex	Yes; sex is usually determined by visual inspection of the animals.	Yes; (RAD-seq); When/if sex markers are available (e.g., for Atlantic halibut, see Palaiokostas et al., 2013; Nile tilapia, see Palaiokostas et al., 2013; for icefish, see Xing et al., 2021).
Maturity	Yes; maturity is determined by visual inspection or histological examination of the gonads.	No
Diet	Yes; diet is determined by analyzing stomach contents.	No
Other data		
Marine litter	Yes	No
Biodiversity	Yes; although with some limitations.	Yes (eDNA)
Stock structure	Could be possible (e.g., using stock identification methods such as analysis of parasites or using morphometric characters).	Yes (RAD-Seq) (e.g., for European hake, see Leone et al., 2019; for rockfish, see Longo et al., 2022; for Antarctic toothfish, see Ceballos et al., 2021).
Oceanographic data	Yes; Oceanographic data include seawater temperature and salinity, for example.	No

assemblages and the impact of fisheries. Following that line, Maiello et al. (2022) designed a customized low-cost 3D-printed plastic probe that, placed inside the trawl net, serves as a container for rolls of gauze that are positioned to capture DNA from the during fishing operations. Their results strengthen the idea that eDNA-based biomonitoring can become embedded in fishery-dependent surveys, at negligible additional cost and effort, to study catch composition and the broader faunal features of the ecosystems that sustain commercial fishing.

As already mentioned, fish stock assessments have been evolving over time due to several reasons, for example, due to improvements in computational approaches, advances in methods for fitting models to data, and the need to not only provide best estimates of model parameters and outputs but also to quantify the uncertainties associated with the estimates (Punt et al., 2020). However, the incorporation of genomic data into these methods will necessarily mean that assessment methods must be also adapted to be able to handle these data. In fact, some experts already consider that one of the necessary improvements for a next-generation stock assessment package in the next 5-15 years is to ensure its ability for handling genetic data, and in particular, to be able to use close-kin mark-recapture data (Punt et al., 2020).

In conclusion, in the short and medium-term, it does not seem feasible to implement directly genomic tools into data collection and stock assessment, given that stock assessment accuracy requires a certain length and stability of the data time series. However, given the growing evidence of the utility of genomic tools in the stock assessment context, it is necessary to devise a roadmap to implement such tools in the future. The guideline for future implementation could be based on the evolution of the cost-efficiency and on further evidence of precision and accuracy gains.

Author contributions

RR-M and FS-R: conceived the article. RR-M: wrote the article. FS-R: reviewed the article. All authors contributed to the article and approved the submitted version.

Funding

The information and views set out in this publication are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union.

Acknowledgments

The information and views set out in this publication are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. RR-M would like to thank Carmen Blanco Tilve for her help editing some parts of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1095171/full#supplementary-material

References

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

Anastasiadi, D., and Piferrer, F. (2020). A clockwork fish: age prediction using DNA methylation-based biomarkers in the European seabass. *Mol. Ecol. Resour.* 20, 287–397. doi: 10.1111/1755-0998.13111

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17 (2), 81–92. doi: 10.1038/nrg.2015.28

Berg, C. W., and Nielsen, A. (2016). Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* 73 (7), 1788–1797. doi: 10.1093/icesjms/fsw046

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32 (9), 665–680. doi: 10.1016/j.tree.2017.06.010

Bravington, M. V., Grewe, P. M., and Davies, C. R. (2016b). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat. Commun.* 7, 13162. doi: 10.1038/ncomms13162

Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016a). Close-kin mark-recapture. *Stat. Sci.* 31 (2), 259–274. doi: 10.1214/16-STS552

Cadrin, S. X. (2020). Defining spatial structure for fishery stock assessment. *Fish. Res.* 221, 105397. doi: 10.1016/j.fishres.2019.105397

Cadrin, S. X., and Dickey-Collas, M. (2015). Stock assessment methods for sustainable fisheries. *ICES J. Mar. Sci.* 72 (1), 1–6. doi: 10.1093/icesjms/fsu228

Ceballos, S. G., Papetti, C., Babbucci, M., Fernández, D. A., Schiavon, L., and Cheng, C. H. C. (2021). Genome-wide analysis reveals striking lack of genetic differentiation over long distances for the Antarctic toothfish *Dissostichus mawsoni*: high genetic connectivity or shared spawning grounds? *Fish. Res.* 243, 106074. doi: 10.1016/j.fishres.2021.106074

Cooper, A. B. (2006). *A guide to fisheries stock assessment: from data to recommendations* (Durham, USA: University of New Hampshire/New Hampshire Sea Grant College Program - NOAA), 44.

Davies, C., Bravington, M., Eveson, P., Lansdell, M., Aulich, J., and Grewe, P. (2020). *Next-generation close-kin mark recapture: using SNPs to identify half-sibling pairs in southern bluefin tuna and estimate abundance, mortality and selectivity. technical report* Vol. 52 (Hobart, Australia: Fisheries research and Development Corporation and CSIRO, Australia).

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., et al. (2017). Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* 26 (21), 5872–5895. doi: 10.1111/mec.14350

Dortel, E., Massiot-Granier, F., Rivot, E., Million, J., Hallier, J. P., Morize, E., et al. (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. doi: 10.1371/journal.pone.0060886

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., et al. (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Can. J. Fish. Aquat. Sci.* 74 (9), 1362–1374. doi: 10.1139/cjfas-2016-0306

Ferchaud, A. L., Normandeau, E., Babin, C., Præbel, K., Hedeholm, R., Audet, C., et al. (2022). A cold-water fish striving in a warming ocean: insights from whole-genome sequencing of the Greenland halibut in the Northwest Atlantic. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.992504

Feron, R., Pan, Q., Wen, M., Imarazene, B., Jouanno, E., Anderson, J., et al. (2021). RADSex: a computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol. Ecol. Resour.* 21 (5), 1715–1731. doi: 10.1111/1755-0998.13360

Ficetola, G. F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biol. Lett.* 4 (4), 423–425. doi: 10.1098/rsbl.2008.0118

Floragenex (2015). *Technical brief: RAD-sequencing* (Beaverton, Oregon, USA: Floragenex Inc). Available at: https://static1.squarespace.com/static/542f1ecee4b0a7afa8674662/t/555b7a68e4b014a961e564b3/1432058472478/2015+RAD-Seq+Technical+Whitepaper+v1.0.pdf.

Fraija-Fernández, N., Bouquieaux, M. C., Rey, A., Mendibil, I., Cotano, U., Irigoien, X., et al. (2020). Marine water environmental DNA metabarcoding provides a

comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. *Ecol. Evol.* 10 (14), 7560–7584. doi: 10.1002/ece3.6482

Fraser, H. M., Greenstreet, S. P. R., and Piet, G. J. (2007). Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass. *ICES J. Mar. Sci.* 64, 1800–1819. doi: 10.1093/icesjms/fsm145

Gamble, T. (2016). Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Mol. Ecol.* 25, 2114–2116. doi: 10.1111/mec.13648

Gillet, B., Cottet, M., Destanque, T., Kue, K., Descloux, S., Chanudet, V., et al. (2018). Direct fishing and eDNA metabarcoding for biomonitoring during a 3-year survey significantly improves number of fish detected around a South East Asian reservoir. *PloS One* 13 (12). doi: 10.1371/journal.pone.0208592

Guevara, E. E., and Lawler, R. R. (2018). Epigenetic clocks. Evolutionary anthropology: issues, news, and reviews. *Evol. Anthropol.* 27, 6, 256–260. doi: 10.1002/evan.21745

Gursoy, C., Tarkan, A. S., Acipinar, H., and Gaygusuz, O. (2005). Validation of the scales for age determination in a cyprinid, *Scardinius erythrophthalmus* (L. 1758). *Int. J. Zoological Res.* 1 (1), 33–36. doi: 10.3923/ijzr.2005.33.36

Heimbrand, Y., Limburg, K. E., Hüssy, K., Casini, M., Sjöberg, R., Palmén Bratt, A. M., et al. (2020). Seeking the true time: exploring otolith chemistry as an age-determination tool. *J. Fish Biol.* 97 (2), 552–565. doi: 10.1111/jfb.14422

Hillary, R. M., Bravington, M. V., Patterson, T. A., Grewe, P., Bradford, R., Feutry, P., et al. (2018). Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. *Sci. Rep.* 8, 2661. doi: 10.1038/s41598-018-20593-w

Horvath, S., and Raj, K. (2018). DNA Methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19 (6), 371–384. doi: 10.1038/s41576-018-0004-3

Hüssy, K., Hinrichsen, H. H., Fey, D. P., Walther, Y., and Velasco, A. (2010). The use of otolith microstructure to estimate age in adult Atlantic cod, *Gadus morhua*. *J. Fish Biol.* 76 (7), 1640–1654. doi: 10.1111/j.1095-8649.2010.02602.x

ICES (2012). Report on the classification of stock assessment methods developed by SISAM. *ICES CM 2012/ACOM/SCICOM:01*, 15.

ICES (2019). *SISP 14 - Manual for the Offshore Beam Trawl Surveys (WGBEAM). Series of ICES Survey Protocols (2012–2020).* Report. 54. doi: 10.17895/ices.pub.5353

ICES (2020). ICES Workshop on Unavoidable Survey Effort Reduction (WKUSER). *ICES Sci. Rep.* 2 (72), 92. doi: 10.17895/ices.pub.7453

ICES (2021). Workshop on tools and development of stock assessment models using a4a and stock synthesis (WKTADSA). *ICES Sci. Rep.* 3 (33), 197. doi: 10.17895/ices.pub.8004

ICES (2022a). "Beaked redfish (*Sebastes mentella*) in division 14.b, demersal (Southeast Greenland)," in *In report of the ICES Advisory Committe. ICES advice 2022* (Copenhagen, Denmark: ICES), reb.27.14b. Available at: https://www.ices.dk/community/groups/Pages/WGBIE.aspx.

ICES (2022b). *Cod (Gadus morhua) in subdivisions 24–32, eastern Baltic stock (eastern Baltic sea). In report of the ICES Advisory Committe. ICES advice 2022* (Copenhagen, Denmark: ICES), cod.27.24–32. doi: 10.17895/ices.advice.19447874

ICES (2022c). *Hake (Merluccius merluccius) in subareas 4, 6, and 7, and in divisions 3.a, 8.a–b, and 8.d, northern stock (Greater North Sea, Celtic seas, and the Northern Bay of Biscay). in report of the ICES Advisory Committe. ICES advice 2022*, hke.27.3a46-8abd. (Copenhagen, Denmark: ICES). doi: 10.17895/ices.advice.19448012

Kirubakaran, T. G., Andersen, Ø., De Rosa, M. C., Andersstuen, T., Hallan, K., Kent, M. P., et al. (2019). Characterization of a male specific region containing a candidate sex determining gene in Atlantic cod. *Sci. Rep.* 9 (1), 1–9. doi: 10.1038/s41598-018-36748-8

Knudsen, S. W., Ebert, R. B., Hesselsøe, M., Kuntke, F., Hassingboe, J., Mortensen, P. B., et al. (2019). Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. *J. Exp. Mar. Biol. Ecol.* 510, 31–45. doi: 10.1016/j.jembe.2018.09.004

Leone, A., Álvarez, P., García, D., Saborido-Rey, F., and Rodriguez-Ezpeleta, N. (2019). Genome-wide SNP based population structure in European hake reveals the need for harmonizing biological and management units. *ICES J. Mar. Sci.* 76 (7), 2260–2266. doi: 10.1093/icesjms/fsz161

Ligas, A., Pierattini, C., Viva, C., Bertolini, D., and Belcari, P. (2011). Age estimation and growth of European hake, *Merluccius merluccius* (Linnaeus 1758), in the northern Tyrrhenian Sea. *Atti Della Societa Toscana Di Sci. Naturali Memorie Serie B.* 118, 9–14. doi: 10.2424/ASTSN.M.2011.16

Longo, G. C., Harms, J., Hyde, J. R., Craig, M. T., Ramón-Laca, A., and Nichols, K. M. (2022). Genome-wide markers reveal differentiation between and within the cryptic sister species, sunset and vermilion rockfish. *Conserv. Genet.* 23 (1), 75–89. doi: 10.1007/s10592-021-01397-4

Lou, R. N., Jacobs, A., Wilder, A. P., and Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol. Ecol.* 30 (23), 5966–5993. doi: 10.1111/mec.16077

Maiello, G., Talarico, L., Carpentieri, P., De Angelis, F., Franceschini, S., Harper, L. R., et al. (2022). Little samplers, big fleet: eDNA metabarcoding from commercial trawlers enhances ocean monitoring. *Fish. Res.* 249, 106259. doi: 10.1016/j.fishres.2022.106259

Marcy-Quay, B., Sethi, S. A., Therkildsen, N. O., and Kraft, C. E. (2020). Expanding the feasibility of fish and wildlife assessments with close-kin mark–recapture. *Ecosphere Freshw. Ecol.* 11 (10), e03259. doi: 10.1002/ecs2.3259

Maunder, M. N., Hamel, O. S., Lee, H. H., Piner, K. R., Cope, J. M., Punt, A. E., et al. (2023). A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. *Fish. Res.* 257, 106489. doi: 10.1016/j.fishres.2022.106489

Maunder, M. N., and Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. *ICES J. Mar. Sci.* 72 (1), 7–18. doi: 10.1093/icesjms/fsu015

Maunder, M. N., Xu, H., Schaefer, K. M., and Fuller, D. W. (2021). *Inter-American tropical tuna commission scientific Advisory Committee 12th meeting* (Inter-American Tropical Tuna Commission – Scientific Advisory Committee), 41.

Mayne, B., Korbie, D., Kenchington, L., Ezzy, B., Berry, O., and Jarman, S. (2020). A DNA methylation age predictor for zebrafish. *Aging* 12, 24817–24835. doi: 10.18632/aging.202400

Methot, R. D. Jr., and Wetzel, C. R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Res.* 142, 86–99. doi: 10.1016/j.fishres.2012.10.012

Miya, M. (2022). Environmental DNA metabarcoding: a novel method for biodiversity monitoring of marine fish communities. *Annu. Rev. Mar. Sci.* 14 (1), 161–185. doi: 10.1146/annurev-marine-041421-082251

Morales-Nin, B., Tores, G. J., Lombarte, A., and Recasens, L. (1998). Otolith growth and age estimation in the European hake. *J. Fish Biol.* 53 (6), 1155–1168. doi: 10.1111/j.1095-8649.1998.tb00239.x

Nielsen, A., and Berg, C. W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* 158, 96–101. doi: 10.1016/j.fishres.2014.01.014

Nielsen, J., Hedeholm, R. B., Simon, M., and Steffensen, J. F. (2014). Distribution and feeding ecology of the Greenland shark (*Somniosus microcephalus*) in Greenland waters. *Polar Biol.* 37, 37–46. doi: 10.1007/s00300-013-1408-3

NMFS (2001) *Marine fisheries stock assessment improvement plan. Report of the National Marine Fisheries Service - National Task Force for Improving fish stock assessments. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-56, 69 p., 25 appendices*. Available at: https://spo.nmfs.noaa.gov/sites/default/files/tm56.pdf.

NOAA (2012) *Population assessments: fish stocks*. Available at: https://www.fisheries.noaa.gov/topic/population-assessments/fish-stocks.

Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M. E., Oral, M., Taggart, J. B., et al. (2013). Mapping the sex determination locus in the Atlantic halibut (*Hipoglossus hippoglossus*) using RAD sequencing. *BMC Genomics* 14 (1), 1–12. doi: 10.1186/1471-2164-14-566

Piferrer, F., and Anastasiadi, D. (2023). Age estimation in fishes using epigenetic clocks: applications to fisheries management and conservation biology. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1062151

Pedersen, M. W., and Berg, C. W. (2017). A stochastic surplus production model in continuous time. *Fish Fish.* 18 (2), 226–243. doi: 10.1111/faf.12174

Punt, A. E. (2019). Spatial stock assessment methods: a viewpoint on current issues and assumptions. *Fish. Res.* 213, 132–143. doi: 10.1016/j.fishres.2019.01.014

Punt, A. E., Dunn, A., Elvarsson, B. Þ., Hampton, J., Hoyle, S. D., Maunder, M. N., et al. (2020). Essential features of the next-generation integrated fisheries stock assessment package: a perspective. *Fish. Res.* 229, 105617. doi: 10.1016/j.fishres.2020.105617

Ramírez-Amaro, S., Bassitta, M., Picornell, A., Ramon, C., and Terrasa, B. (2022). Environmental DNA: state-of-the-art of its application for fisheries assessment in marine environments. *Front. Mar. Sci.*, 2348. doi: 10.3389/fmars.2022.1004674

Rourke, M. L., Fowler, A. M., Hughes, J. M., Broadhurst, M. K., DiBattista, J. D., Fielder, S., et al. (2022). Environmental DNA (eDNA) as a tool for assessing fish biomass: a review of approaches and future considerations for resource surveys. *Environ. DNA* 4 (1), 9–33. doi: 10.1002/edn3.185

Russo, T., Maiello, G., Talarico, L., Baillie, C., Colosimo, G., D'Andrea, L., et al. (2021). All is fish that comes to the net: metabarcoding for rapid fisheries catch assessment. *Ecol. Appl.* 31 (2), e02273. doi: 10.1002/eap.2273

Ruzzante, D. E., McCracken, G. R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., et al. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. *Methods Ecol. Evol.* 10, 1445–1453. doi: 10.1111/2041-210X.13243

Spedicato, M. T., Massutí, E., Mérigot, B., Tserpes, G., Jadaud, A., and Relini, G. (2019) *The MEDITS trawl survey specifications in an ecosystem approach to fishery management*. Available at: https://scientiamarina.revistas.csic.es/index.php/scientiamarina/article/view/1829.

Stamatopoulos, C. (2002). "Sample-based fishery surveys. a technical handbook," in *FAO fishery technical paper*, vol. 425. (Rome: FAO), 1–132.

Stat, M., John, J., DiBattista, J. D., Newman, S. J., Bunce, M., and Harvey, E. S. (2019). Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. *Conserv. Biol.* 33 (1), 196–205. doi: 10.1111/cobi.13183

Taberlet, P., Coissac, E., Hajibabaei, M., and Rieseberg, L. H. (2012). Environmental DNA. *Mol. Ecol.* 21 (8), 1789–1793. doi: 10.1111/j.1365-294X.2012.05542.x

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., and Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PloS One* 11 (11), e0165252. doi: 10.1371/journal.pone.0165252

Trenkel, V. M., Charrier, G., Lorance, P., and Bravington, M. V. (2022). Close-kin mark–recapture abundance estimation: practical insights and lessons learned. *ICES J. Mar. Sci.* 79 (2), 413–422. doi: 10.1093/icesjms/fsac002

Weber, D. N., Fields, A. T., Patterson, W. F. III, Barnett, B. K., Hollenbeck, C. M., and Portnoy, D. S. (2022). Novel epigenetic age estimation in wild-caught gulf of Mexico reef fishes. *Can. J. Fish. Aquat. Sci.* 79 (1), 1–5. doi: 10.1139/cjfas-2021-0240

Xing, T. F., Li, Y. L., and Liu, J. X. (2021). Female-specific genomic regions and molecular sex identification of the clearhead icefish *(Protosalanx hyalocranius*). *BMC Genomics* 22 (1), 1–9. doi: 10.1186/s12864-021-07830-9

Zhang, Q., Vallerga., C. L., Walker, R. M., Lin, T., Henders, A. K., Montgomery, G. W., et al. (2019). Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* 11 (54). doi: 10.1186/s13073-019-0667-1

Zhou, S., Fan, C. R., Zhang, J., Yang, W., Ji, D. J., Wang, L., et al. (2022). Combined use of eDNA metabarcoding and bottom trawling for the assessment of fish biodiversity in the zhoushan Sea. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.809703

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 07 July 2023
DOI 10.3389/fmars.2023.1113870

OPEN ACCESS

EDITED BY
Montse Pérez,
Spanish Institute of Oceanography, Spain

REVIEWED BY
Simo Njabulo Maduna,
Norwegian Institute of Bioeconomy Research (NIBIO), Norway
Anthony Lee Dellinger,
Kepley BioSystems, Inc., United States
Dan Ovando,
University of Washington, United States

*CORRESPONDENCE
Laura Casas
lauracasas@iim.csic.es
Fran Saborido-Rey
fran@iim.csic.es

RECEIVED 01 December 2022
ACCEPTED 12 May 2023
PUBLISHED 07 July 2023

CITATION
Casas L and Saborido-Rey F (2023) A review of genomics methods and bioinformatics tools for the analysis of close-kin mark-recapture. *Front. Mar. Sci.* 10:1113870. doi: 10.3389/fmars.2023.1113870

A review of genomics methods and bioinformatics tools for the analysis of close-kin mark-recapture

Laura Casas* and Fran Saborido-Rey*

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

Stock assessments serve to monitor the condition of fish stocks and exploit them sustainably but require accurate data such as growth and mortality rates as input parameters. Most species fished worldwide lack the data needed to assess their status and even those closely assessed are often based on parameters that are known to contain uncertainty. This has resulted in an increased share of overfished stocks over the last half century, demanding urgently innovative methodologies that can provide novel means to reduce uncertainty of fish stocks assessments and expand the range of assessed species. CKMR has emerged recently attracting a great interest due to its potential to provide accurate demographic parameters of interest in stock assessments. The method is at the crossroads between fisheries science and genomics, requiring specialized knowledge that is usually outside of the experience of fisheries scientist and modellers, complicating the application of the method and its uptake in regular fisheries assessments. In this review, we provide useful information to perform the genomics and bioinformatics steps required to complete successfully a CKMR study. We discuss the most suitable genomics assays, considering the amount of information they provide, their easiness of use and cost of genotyping accurately the large number of individuals needed to assess most fish stocks. We provide an overview of methods of analysis and statistical methodologies that can be used to infer kinship with the accuracy required in a large population setting with sparse sampling, where most individuals are unrelated, determining a low probability of finding closely related individuals. We analyse potential sources of biases and errors and provide recommendations to facilitate the application of CKMR to a wider range of fish stocks.

KEYWORDS

CKMR, SNP markers, genotyping, kinship analysis, fisheries assessments

1 Introduction

Close-kin mark-recapture (CKMR) provides a fisheries-independent method for estimating demographic parameters (e.g. abundance, population trend, survival rates) of fish stocks, based on the frequency and degree of kinship determined from genetic samples (Bravington et al., 2016b).

The underlying principle is analogous to classical mark recapture, but uses the genomic information to determine if any given individual ("capture") has one or more close relatives ("recapture/s") in a sample. The probability of finding relatives ("recaptures") diminishes as the size of the population increases, providing information on adult abundance. The inverse relationship between abundance and probability of recapture holds if individuals represent a random sample from the population of interest (Bravington et al., 2016b).

CKMR relies on a mark-recapture analysis framework (Bravington and Grewe, 2007) and can be divided into five different stages (Figure 1); design, sampling, marker discovery and genotyping, kin finding and demographic/statistical modelling of the population.

The design stage involves several steps, starting with the compilation of existing biological knowledge, catch data (if available) and estimated stock parameters, including approximate population size, if the population is assessed. This information is needed to perform a preliminary evaluation of the potential feasibility of CKMR to assess the population of interest, which consists of an estimation of the number of samples and ancillary information required to infer population parameters with accuracy, the potential complications and the technical considerations for the design of the CKMR study. If the study is deemed viable, the next step is the design of sampling, considering the relevant biology (e.g., the presence of population substructure with differentiated adult and juvenile habitats (nursery grounds) or with biased sex-ratio in spawning-grounds) and the estimated samples sizes required for the target level of precision in the population estimates (Maunder et al., 2021).

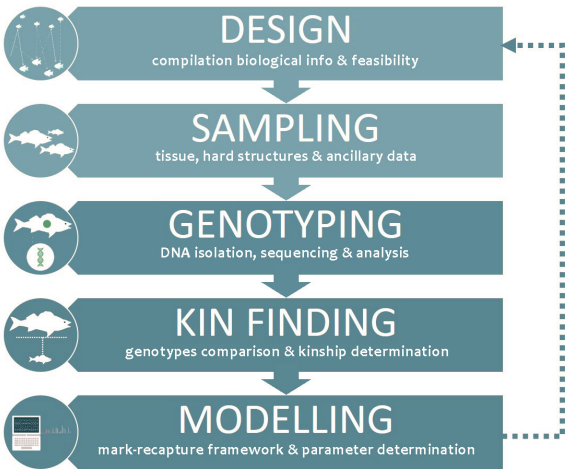


FIGURE 1
Flowchart displaying the five components of a CKMR study. The arrows indicate the succession of stages, including a connection between the final stage and the initial design stage, indicating an iterative process.

The second stage is the sampling, which involves the collection of samples, including tissue and hard structures (otoliths, scales or vertebrae, among others) as well as ancillary data, such as length, sex, and maturation stage.

The next two stages involve the use of genomics and bioinformatics tools. The term "genomics" refers to an organism's complete collection of heritable information stored in its DNA. Modern genomic technologies provide the means to study this information and uncover differences in genome content that provide insights into individuals, populations and species. Such differences, also known as DNA polymorphisms (in biology, polymorphism describes the existence of multiple forms) are the source of genetic diversity and can be used as molecular markers (Del Giacco and Cattaneo, 2012). Bioinformatics, on the other hand, can be defined as the application of tools of computation and analysis to capture and interpret biological data (Bayat, 2002). Both disciplines are involved in the third stage, the genotyping, which entails the determination of the DNA sequence at polymorphic positions within the genome of an individual. It requires the isolation of DNA from the tissues collected for each specimen, the amplification and sequencing of this DNA through a selected method, and searching for polymorphic regions across their genomes to produce unique genotypes. The fourth stage addresses the search of close relatives (equivalent to "recaptures") among the specimens and requires a bioinformatics workflow to perform a kinship analysis among the unique genotypes aimed at detecting highly related individuals. In the last stage, variants of capture-recapture modelling are applied to the kinship data to produce direct estimates of parental population abundance and other demographic parameters such as mortality or fecundity, depending on the type of kinship relationships analysed. The results obtained serve to feed back the design stage in an iterative process to refine and optimize the CKMR model (Delaval et al., 2022).

The CKMR method is at the crossroads between fisheries science and genomics, requiring specialized knowledge that is usually outside of the experience of fisheries scientist and modellers, complicating the application of the method and its uptake in regular fisheries assessments (Davies et al., 2015). This review expands and complements the review made by Casas and Saborido-Rey, (2023) within this Research Topic focused on close-kin mark–recapture (CKMR) as an emerging tool to estimate population parameters, focused on the theory behind the method, the stages of design and sampling (one and two) and existing case studies. The present manuscript addresses stages three and four, which involve the use of genomic and bioinformatics tools. The rapid advances in genomic technologies and the plethora of software and analysis pipelines represent a notable challenge and can be overwhelming when approaching a CKMR project. A simple terminology is used across the manuscript to reach potential users of the method, such as fisheries managers and scientist, with no expertise in genomics. Additionally, a brief glossary for key genomics and bioinformatics terms is included. We provide an overview of molecular marker types that have been applied in CKMR studies to date, their characteristics

and give recommendations for future CKMR studies. We discuss the genomic methods with higher potential for CKMR applications, considering the amount of information they provide, their easiness of use and cost. We provide an overview of methods of analysis and statistical methodologies that can be used to infer kinship, together with the potential sources of error and biases. Finally, we provide recommendations and important considerations to carry out close kin studies to assess fish populations, to facilitate bridging the gap between geneticists and fisheries assessment scientists and promote the use of genomic tools in fisheries science.

2 Genotyping stage

The genotyping stage consists of several steps that involve laboratory work to isolate, amplify and sequence each specimen´s DNA, followed by a bioinformatics workflow to analyse the resulting sequences (Figure 2). This analysis aims at detecting and scoring variable regions among the individuals that can be used as markers to produce unique genotypes.

2.1 Isolation of high quality DNA

The first step is the isolation of DNA and the importance of obtaining the highest possible quality cannot be overemphasized, as it is essential to later produce the best possible sequencing reads. The collection of tissue samples for CKMR studies commonly relies on a biopsy of tissue, such as a fin clip or a small piece of skin, which does not require the sacrifice of the specimen. Alternatively, a range of soft tissues can be targeted, including spleen, heart, blood, kidney or muscle, if specimens are sacrificed. A suitable storage of the tissues collected is essential to guarantee the integrity of the DNA. An adequate preserving solution (e.g. 90% ethanol, DNAzol, DMSO-EDTA, DNA/RNA Shield™, RNAlater) should be used, and is also important to minimize the time between collection and storage to prevent tissue degradation (Mulcahy et al., 2016; Oosting et al., 2020; Dahn et al., 2022). If the DNA is not isolated promptly, samples should be kept at -20°C. Handling procedures should minimize the risk of cross-contamination between different individuals. This is especially pertinent to studies involving the collection of tissue samples at sea, on board of research or commercial vessels, where access to sterile tools and clean workspaces is often limited (Anderson et al., 2023). Care is also needed in subsequent steps, as the risk of contamination remains along the sample processing. Sample contamination and mixing can seriously impact downstream results, causing erroneous inferences and is a common problem in large-scale studies (Zajac et al., 2019; Francois et al., 2020; Anderson et al., 2023).

Several DNA isolation methods can be used, from the classic phenol–chloroform DNA extraction method to any of the multiple high purity DNA isolation commercial kits available, as long as they

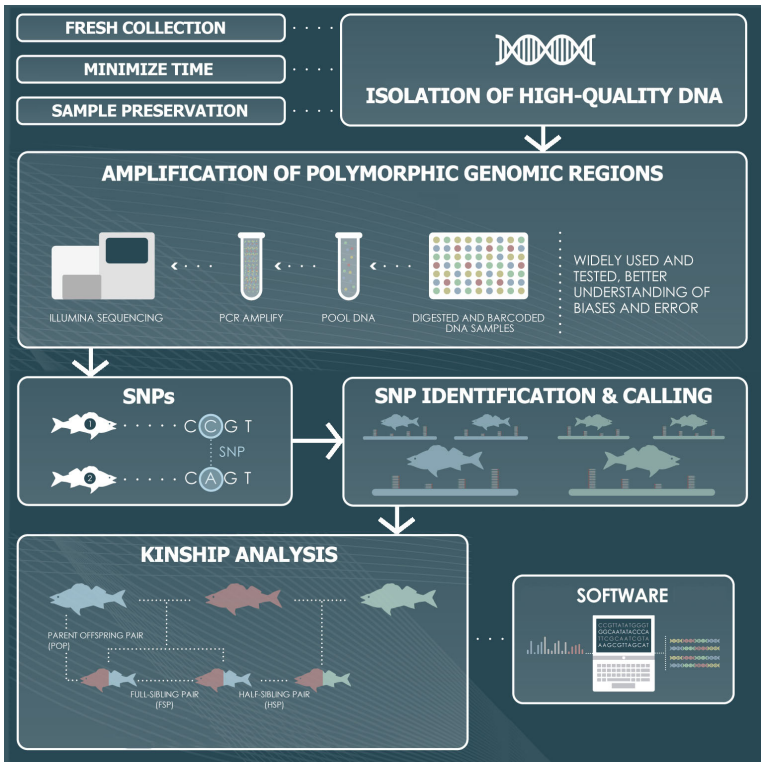


FIGURE 2
Representation of the genomics and bioinformatics steps involved in a CKMR study. The workflow entails the isolation of DNA, the amplification of polymorphic regions across the genome of the species of interest, the identification of molecular markers and the genotyping of the specimens, followed by an analysis of the kinship relationships among them to infer parent-offspring and half-sibling relationships.

produce high molecular weight genomic DNA. A large number of comparative studies have tested the efficiency of different methods in terms of DNA quality, quantity and purity (i.e. rate of fragmentation, concentration and rate of contaminants, respectively) showing that it is highly dependent on the tissue processed, the storage conditions and the species [e.g. (Silva et al., 2019; Martincová and Aghová, 2020; Lutz et al., 2023)]. Martincová and Aghová (2020), tested 12 different DNA extraction methods from eight manufacturers, including some the most widely used commercial silica membrane-based kits. The authors assessed the quality of the DNA obtained from four tissues of two vertebrate species and found that the highest DNA yields were consistently obtained with one of the kits but different ones produced a better DNA quality, in terms of purity and fragmentation. In the tests carried out by Lutz et al. (2023), three extraction methods were tested on several fish tissues and the results are in contrast with those found by Shuttleworth and Oosthuizen (2022) and Oduoye et al. (2020), who also tested several isolation methods in fishes from different species. These marked differences highlight the need of a thorough testing and optimization of the DNA preservation and isolation method on every CKMR study. Additionally, when selecting a method is also important to consider the cost, the processing time and the technical requirements of the different methodologies, as they differ widely (Silva et al., 2019; Martincová and Aghová, 2020; Oduoye et al., 2020).

The DNA quality, quantity and purity strongly affects downstream molecular analysis, conditioning the subsequent choices of markers and methods, as well as the bioinformatics workflow but poor quality DNA is highly unlikely to produce the accurate data need in CKMR studies, regardless of the choices.

2.2 Amplification of polymorphic genomic regions

After obtaining high-quality genomic DNA for each individual, the next step consists on the amplification and sequencing of this material to discover polymorphic regions across the genome of the species under study. These variable regions are heritable and, therefore, can be used as genetic markers, enabling the identification of closely related specimens (Stage 3) needed in CKMR studies. Accurate kinship analysis requires high resolution markers, which can be either highly polymorphic or very numerous to tackle the inherent challenges of studying wild populations, often characterized by large sizes, small numbers of true closely related individuals sampled and a high number of comparisons (Städele and Vigilant, 2016). All CKMR studies published to date have relied on either microsatellites (STRs) (Bravington et al., 2016a; Ruzzante et al., 2019; Marcy-Quay et al., 2020; Prystupa et al., 2021) or, more recently, on single nucleotide polymorphisms (SNPs) (Hillary et al., 2018; Wacker et al., 2021; Delaval et al., 2022; Patterson et al., 2022; Trenkel et al., 2022). The former consist of tandem repeats of short (one to six base pairs) genetic elements, in which differences between alleles are primarily in the number of repeats (Webster and Reichart, 2005). The latter, constitute the most common form of variation in a genome and are characterized by the substitution of a single nucleotide at a specific location (van Dijk et al., 2014). Both types have advantages and drawbacks and the optimal choice depends on several factors, reviewed below.

2.2.1 Molecular marker types

The first CKMR study that demonstrated the ability of the method to estimate abundance of a fish population was based on STRs and several others followed suit (Bravington et al., 2016a; Ruzzante et al., 2019; Marcy-Quay et al., 2020; Prystupa et al., 2021). The use of STRs involves a costly and time-consuming investment in their isolation and characterization, although the emergence of high capacity sequencing technologies have facilitated this process in recent years (Ellis et al., 2011; De Barba et al., 2017). High-throughput sequencing techniques (HTS) enable the simultaneous sequencing of large numbers of DNA fragments, sensibly reducing costs and processing times compared to previous technologies (Reuter et al., 2015). Several tools for identifying STRs from high-throughput data exist [e.g.: SSR pipeline, (Miller et al., 2013)], although false positive results and limited quality is still an issue, requiring often further experiments to identify and validate the polymorphic STRs (Guang et al., 2019). Moreover, current techniques present difficulties regarding amplification calibration and the choice of informative STRs with high specificity (Pimentel et al., 2018). The analysis of STRs from high-throughput sequences is hampered by their high propensity to vary in size during both the PCR amplification and the sequencing reaction itself. Another disadvantage of STRs, due to their PCR based nature, is that mutations in primer regions can lead to non-amplifying "null alleles" that can pose problems for kinship assignments (Paetkau and Strobeck, 1995; Ishibashi et al., 1996). Nonetheless, STRs also carry a number of significant advantages, being the most important, in the context of CKMR studies, a series of characteristics that make them especially useful for estimating kinship and relatedness. They are codominant markers, highly polymorphic due to high mutation rates, with a high power for paternity analyses (≈6× that of SNPs) and low requirements in terms of DNA amount and quality. In fact, the ability of STRs to accurately assign parentage from highly degraded DNA samples has been validated (De Barba et al., 2017). Therefore, they might be a suitable choice for CKMR studies that have to rely on the analysis of samples that do not reach the recommended DNA quality standards, if identification of first-degree relatives alone is sufficient. Moreover, STR genotyping is less expensive than SNP genotyping (Puckett, 2017; Lemopoulos et al., 2019), a very relevant factor in any CKMR study involving natural populations of interest in fisheries, as these mostly require the analysis of very large numbers of individuals [e.g (Bravington et al., 2016a; Trenkel et al., 2022)].

Single nucleotide polymorphisms (SNPs), on the other hand, have become increasingly popular in population genetic studies (von Thaden et al., 2017; Torrado et al., 2020; Wenne et al., 2020; Crespel et al., 2021) and, more specifically, in CKMR studies (Hillary et al., 2018; Wacker et al., 2021; Delaval et al., 2022; Trenkel et al., 2022). SNP markers show several practical improvements over STRs to conduct kinship analysis, including their higher abundance in the genomes, a lower and predictable

mutation rate and their easier automation that results in a higher reproducibility (Amorim and Pereira, 2005; Anderson and Garza, 2006; Fisher et al., 2009). Their main constraint is the limited genetic information they provide per locus, but given a sufficient number of markers, their collective strength can resolve almost any parentage or close kin relationship. Several studies have compared the power of SNPs and STRs in a parentage context and in virtually every case, the studies concluded that SNPs are at least as powerful as STR markers [for review see (Flanagan and Jones, 2019)].

Considering their characteristic, the latest CKMR studies and the increasing efficiency and affordability of genotyping genomic techniques, we argue that SNP markers are the sensible choice for any project initiating today, unless budget constraints or DNA quality advises otherwise. Thus, hereafter, we focus on methodologies and bioinformatics pipelines solely based on SNPs.

2.2.2 Methods for high throughput marker discovery

In spite of the outstanding developments in sequencing technologies and bioinformatics tools in the last decades and the worldwide initiatives to improve genomic resources across species [e.g.: the Vertebrate Genomes Project, (Rhie et al., 2021)], these are still scarce in non-model species (Christiansen et al., 2021). This is particularly true for fish (Fan et al., 2020) and thus, in a fisheries context, CKMR studies must target, mostly, wild populations of thousands to millions of individuals for which few prior genomic resources are available. Several methodologies can be used for the simultaneous discovery of thousands of genomic regions containing SNPs across genomes, and they can be broadly grouped into three categories (Table 1); i) restriction enzyme-based methods, ii) combined enzyme-based hybridization capture methods, iii) whole-genome sequencing methods.

The first two categories measure polymorphisms in a subset of genomic regions, a more economical approach than sequencing the whole genome, especially in studies involving a large number of individuals. They can assess accurately a wide array of biological questions and have been the choice in CKMR studies based on SNPs published to date (Hillary et al., 2018; Wacker et al., 2021; Delaval et al., 2022; Trenkel et al., 2022).

Restriction-enzyme-based methods were specifically developed to reduce the proportion of the genome targeted for sequencing. They define a large group of HTS methodologies that involve the digestion of genomic DNA with restriction endonucleases and the sequencing of the resulting restriction fragments. They encompass several classes of methodologies, including genotyping-by-sequencing [GBS (Elshire et al., 2011)], reduced-representation libraries [RRLs, (Van Tassell et al., 2008)], complexity reduction of polymorphic sequences [CroPS, (van Orsouw et al., 2007)] and restriction-site-associated DNA sequencing [RADseq (Miller et al., 2007; Baird et al., 2008)] that share key steps but also have substantial differences [reviewed in (Davey et al., 2011)]. The latter are especially useful for CKMR studies that often target organisms lacking a well-assembled reference genome, as they can provide high genome-wide marker densities scored with high accuracy (Davey et al., 2011). The term RADseq is used today to refer not only to the original (single digest) RADseq protocol but also to a number of variants that were developed to suit specific experimental needs. These include ddRADseq (Peterson et al., 2012), 2bRAD (Wang et al., 2012), ezRAD (Toonen et al., 2013), 3RAD (Graham et al., 2015), nextRAD (Fu et al., 2017) and quaddRAD (Franchini et al., 2017), among others (reviewed by Andrews et al., 2016; Campbell et al., 2018). However, with the notable exception of double-digest RADseq (ddRAD), the majority of these derivatives consist of only minor and subtle modifications of the parent protocol and have only been marginally used and tested, preventing their application in CKMR studies. The 2RAD and 3RAD methods (Bayona-Vásquez et al., 2019; Glenn et al., 2019) are also noteworthy as they overcome some of the technical challenges of RADseq-based methods, providing an efficient, flexible, and low-cost system to analyse large numbers of individuals. The 2RAD/3RAD methods have a lower startup cost and a higher capacity for sample multiplexing, as well as a simplified workflow that facilitates their implementation (Bayona-Vásquez et al., 2019). Nonetheless, it is important to note that technical differences among the methods lead to important considerations for the types of bias and error inherent in the resulting data and these are much better understood in sequences generated by sdRADseq and ddRADseq techniques (Andrews et al., 2016).

TABLE 1 Comparison of the main properties of restriction enzyme-based methods, combined enzyme-based hybridization capture methods and whole-genome sequencing methods for the discovery of SNP markers across genomes.

Category	Restriction enzyme-based methods (RADseq)	Combined enzyme-based hybridization capture methods	Whole-genome sequencing
Expertise required	++	++	+++
Number of markers	++	+	+++
Number of individuals	++	+++	+
Variant-calling and genotyping	Intermediate coverage, ++ genotype accuracy	Higher coverage, +++ genotype accuracy	Lower coverage, + genotype accuracy
Information content	intermediate overall information	more information per locus	more overall information
Practical considerations	intermediate cost per individual, faster	lower cost per individual, requires lower quality DNA	higher cost per individual, more information

Scores indicate (+) low, (++) medium, (+++) high.

The original single digest Restriction site Associated DNA is arguably the most popular reduced representation sequencing technique and has a number of advantages for identifying kinship-informative SNPs in non-model organisms. This technique sequences short regions surrounding essentially all restriction sites for a given restriction endonuclease (assuming a sufficient sequencing depth). Restriction fragments are randomly sheared to a length suitable for the sequencing platform of choice, and selective PCR is normally used to amplify for sequencing only those fragments containing a restriction site, generating a data set of RAD tags (sequences downstream of restriction sites) that derive from a much-reduced part of the original genome. The most popular derivative, ddRADseq differs from sdRADseq in two principal aspects. First, it eliminates the need for a sonicator, a specialized instrument not necessarily available in a standard molecular laboratory, using instead a double restriction enzyme digest (i.e., a restriction digest with two enzymes simultaneously). Second, it introduces a precise selection for genomic fragments by size since it relies on the distance between cut sites to determine the length of DNA that is sampled (Peterson et al., 2012).

Both techniques are flexible in the number of loci they can target as the choice of restriction enzyme(s) determines the number of resultant SNP markers. Considering that in CKMR studies it is critical to obtain a sufficient number to resolve kinship accurately, it is highly recommended to perform a prospective data simulation to model accurately the number and distribution of expected RAD loci before initiating a study. Simulations allow testing the behaviour of different molecular protocols in the system of interest, as well as assessing the magnitude of data recovered given variable experimental conditions. This can be performed by several tools, including simRAD (Lepais and Weir, 2014), ddRADseqtools (Mora-Márquez et al., 2017), RADinitio (Rivera-Colón et al., 2021) and PredRAD (Vendrami et al., 2019).

Restriction enzyme techniques are suitable to analyse sample sizes of a few thousand individuals or less, the range of most teleost CKMR studies published to date (Hillary et al., 2018; Ruzzante et al., 2019; Marcy-Quay et al., 2020; Prystupa et al., 2021; Wacker et al., 2021; Delaval et al., 2022). For larger sample sizes, a better strategy, in terms of costs and time, is the use of RADseq on a subset of samples for SNP discovery first, and subsequently using this information to design a custom panel of SNPs (called SNP chip or SNP array) for genotyping of the remaining samples (Trenkel et al., 2022). This approach has further advantages as it ensures consistent sequencing of the same genomic regions and significantly simplifies the analysis. SNP chips serve as a black-box presence-absence for each allele at each locus and a computer reports which alleles are present. The current generation of microarrays can accommodate hundreds of thousands or millions of DNA fragments (oligonucleotides) and the genotyping in parallel of hundreds of individuals (Adler et al., 2013). SNP chips can also serve as a workaround when is not possible to consistently obtain high-quality DNA samples, although the design of a SNP chip inevitably requires a small number of high-quality samples (Maunder et al., 2021).

This is precisely the basis of the second category of assays "combined enzyme-based hybridization capture methods", which use a restriction-enzyme-based method to identify candidate SNP loci for capture bait design and subsequently employs custom capture baits to enrich candidate SNP loci before sequencing (Hoffberg et al., 2016). The coupling of these two strategies improves the consistency of genotype data compared to stand alone restriction enzyme methods as it produces higher sequence read coverage of a refined set of loci, improving confidence in genotype calls. Additionally, it allows the multiplexing of a larger number of samples within a fixed sequencing effort, substantially reducing the cost per individual (Andrews et al., 2014; Ali et al., 2016). This group of "enriched" methods encompass several approaches, including Rapture (Restriction-site associated DNA capture; RAD capture) (Ali et al., 2016), Hybridization Capture Using RAD Probes (hyRAD), RADcap (Hoffberg et al., 2016) or HyRADX (Schmid et al., 2017).

The third category that could be used for the discovery of SNPs across a given genome are whole-genome sequencing (WGS) methods, which are rapidly becoming popular in ecological studies although have never been tested in a CKMR approach (Taylor et al., 2021). WGS methodologies provide significantly more information compared to the previous two categories, as they can theoretically unveil *all* the polymorphisms in a genome. Although they have, in principle, more resolution than needed for CKMR studies and a significantly higher cost, the application of WGS has the potential to boost the power of close kin analyses. Current studies are based on close kinship involving first and second order relatives; parent-offspring and half-sibling pairs. Nonetheless, WGS in species with well-assembled genomes could provide sufficient information to reliably identify one or two orders more distant pairwise relationships (e.g. half-first cousins or great uncles) requiring the analysis of a smaller percentage of the population´s individuals to accurately estimate parameters of interest (Anderson, 2022b).

2.3 SNP identification and calling

Once the sequences have been obtained, the next step involves the use of bioinformatics tools to convert this raw genetic data into a final set of SNP and genotype calls consisting of an inferred allele (i.e. AA, BB, AB) at each SNP locus for each individual analysed. Although analytic strategies vary across different high-throughput technologies, they all require critical validation to ensure precise and unbiased interpretation (Shafer et al., 2017). The number of SNP loci required to ensure a suitable statistical power for kinship analyses will vary across study systems based on genetic diversity, mating system, and the number of individuals sampled (Kopps et al., 2015).

The steps of the bioinformatics pipeline used to produce the genotypes need to be tailored to the methodology used to generate the libraries and the sequencing technology; however, all bioinformatics workflows share some common goals. For example, they must take into account the moderate genotyping error rate inherent in HTS data, identify and remove SNPs in paralogous and other repetitive genomic regions, and generate a set of unlinked loci. There are some basic steps that are similar among

all pipelines, starting with the demultiplexing step that serves to assign each sequence to its individual of origin [e.g. (Torkamaneh et al., 2017)]. This is followed by a pre-processing of the reads that includes a quality control and filtering steps to eliminate poor-quality or suspected artifactual SNP loci. The reads are subsequently assembled or aligned. If a reference genome is available, sequence reads are aligned to the reference using an alignment software such as Bowtie2, TopHat2, BWA or STAR, among others [for a comparison of their performance see (Musich et al., 2021)]. Alternatively, loci can be assembled *de novo* by clustering similar sequence reads together and assuming that variation among reads at a locus represents either sequencing error or true allelic variation. The final step consists on the discovery of polymorphic loci and the inference of the genotypes at these loci for each individual. Bi-allelic SNPs are identified for each individual sample and a filtering step is used to remove uninformative and unreliable loci to keep only high quality, error free genotypes. Retaining only reliably scored SNPs is essential in CKMR studies since the genotypes are subsequently used to infer kinship, thus locus appearing inconsistent with the assumptions underlying Hardy-Weinberg Equilibrium, showing linkage disequilibrium, and with low call frequencies must be filtered out (Trenkel et al., 2022).

There are a number of bioinformatics software packages that have been developed specifically to aid the workflow analysis of reduced-representation sequencing data. Several of these platforms utilize the same tools and algorithms commonly applied to whole-genome sequence data, while others utilize specifically developed algorithms. For RADseq, the most popular software analysis is Stacks v2 (Rochette et al., 2019) and its previous version, Stacks v1 (Catchen et al., 2013). This program is designed modularly to perform sequentially cleaning and filtering of raw sequence data, building loci, creating a catalog of loci, and matching samples back against the catalogue, transposing the data, adding paired-end reads to the analysis and calling genotypes.

Stacks employs a de Bruijn graph assembler to build contigs from paired-end reads and overlap those contigs with the corresponding single-end loci. This enables a Bayesian genotype caller to provide precise SNPs, and a robust algorithm to phase those SNPs into long haplotypes, generating RAD loci spanning several hundred base pairs (Rivera-Colón et al., 2021). Stacks implements several alternative models to call SNPs and genotypes and then converts SNPs into phased haplotypes using a graph-based algorithm that relies on sequence data, specifically on co-observations of alleles within a read pair. Despite its wide adoption, generating a reliable set of loci for downstream analysis requires appropriate use of the software and this implies the non-trivial task of selecting some parameters throughout the pipeline. Such parameters depend on key features of the RADseq dataset under analysis and enforces to explore the parameter space and assess how the analysis software interacts with the biological signal (Paris et al., 2017). Although these complex genomic analyses remain a daunting task for many researchers, very detailed road maps for a correct use of Stacks and robust SNP calling are available (Paris et al., 2017; Rochette and Catchen, 2017; Rochette et al., 2019; Rivera-Colón and Catchen, 2022).

Other alternatives include Ipyrad (Eaton and Overcast, 2020) which allows for the inclusion of indel variation and requires selecting a set of parameters that will affect SNP calling, analogous to Stacks v2. The pipeline dDocent (Puritz et al., 2014a), depends largely on other bioinformatics software packages and performs SNP calling using a Bayesian statistical framework with FreeBayes (Garrison and Marth, 2012). There are also a number of software that use genotype likelihoods and probabilities rather than explicit genotype calls. Analysis of Next Generation Sequencing Data (ANGSD) (Korneliussen et al., 2014) and polyRAD (Clark et al., 2019) estimate a posterior probability from the priors and likelihoods for each individual and allele using Bayes' theorem (Wang et al., 2019), facilitating the incorporation of statistical uncertainty regarding genotypes.

2.3.1 Sources of bias and error

Genotype data produced by high throughput sequencing and SNP arrays are imperfect due to missing (errors of omission) and erroneous (errors of commission) genotypes (Faria et al., 2011; Carroll et al., 2018). These errors strongly affect genotype-based analyses, such as inferences of identity, relatedness and relationship, resulting in incorrect assignments (Wang, 2010; Gomez-Raya et al., 2022). CKMR studies often target natural populations where most individuals are unrelated. In this context, ignoring or underestimating genotyping errors during SNP-based kinship inference can cause the exclusion of true relatives or false-positive assignments, having a great impact on the subsequent estimation of population parameters. Thus, in CKMR studies it is essential to minimize genotyping errors by avoiding artefacts that have the potential to bias allele frequencies and cause false alleles. Allelic dropout causes the masking of some alleles and occurs when there is a failure during the amplification of one or both alleles of a diploid individual (Sommer et al., 2013). If only one allele drops out, the other is revealed alone causing the misinterpretation of the individual as homozygous at the concerned locus. If drop out affects both alleles, it causes missing genotypes. In enzyme-based methods, allelic dropout manifests when the restriction enzyme recognition site contains a polymorphism, resulting in a failure to cut the genomic DNA at that location (Andrews et al., 2016). It has been shown that allele dropout increases with overall levels of polymorphism and has a greater impact on data generated by ddRAD than the original sdRAD, because loci depend on the presence of two cut sites rather than one (Arnold et al., 2013a; Gautier et al., 2013a). In capture-targeted assays, variability in regions surrounding the targeted SNP sites can interfere with hybridization introducing dropout-like effects (Gershoni et al., 2022).

Allelic dropout is generally caused by random effect and strongly correlated with three well know artefacts that can be introduced at various stages of the genomic workflow; poor DNA quality and quantity, low sequencing coverage and PCR duplicates (Nielsen et al., 2011; Puritz et al., 2014b).

Poor DNA quality strongly compromises the accuracy of genotype data. Restriction enzyme methods are highly susceptible to degraded DNA, since it reduces dramatically the percentage of

identical regions amplified among samples, strongly affecting the ability to identify SNPs. Additionally, the sequences produced from low quality DNA suffer from low quality scores or high uncertainty of base calls, resulting in high error rates and low genotyping call rates (Graham et al., 2015). Low amounts of DNA, on the other hand, require more cycles in protocols that include a PCR enrichment step to produce enough DNA for sequencing, introducing further biases and additional sequencing errors due to PCR amplification (Davey et al., 2013; Cumer et al., 2021). However, newer protocols like 3RAD have an improved performance, compared to traditional RADseq methods, that makes them better suited to low input DNA concentrations (Bayona-Vásquez et al., 2019).

To a less extent, targeted sequencing assays that select regions of interest through PCR amplification (amplicon-based approaches) or hybridization enrichment (bait hybridization) are also affected by poor DNA quality and quantity. The former requires annealing to the locations flanking the regions of interest while in the second the DNA hybridizes to a bait oligonucleotide. In both cases, the use of degraded DNA affects the uniformity of coverage across genomic targets and increases the likelihood of capturing off-target regions (So et al., 2018). With small amounts of DNA, extensive PCR amplification is needed, to generate a sufficient number of sequencing library molecules, exacerbating biases associated to this process.

Coverage (or depth) in DNA sequencing refers to the number of reads that align to a specific locus in a given genome. Since high throughput sequencing has an inherent error rate that compares unfavourably to Sanger sequencing, it normally requires a minimal number of reads to ensure accuracy of the bases detected (Huang and Knowles, 2016). A high coverage permits the calling algorithms to assess SNPs with a higher likelihood, resulting in a larger percentage of true loci (Paris et al., 2017). Loci below the coverage threshold has, in contrast, a high uncertainty and should be filtered out during the bioinformatics analysis, but when the starting coverage is deficient, this may result in the removal of all or most of the loci. Insufficient coverage is a common mistake in restriction-enzyme based studies and the main reason behind large genotyping error rates in SNPs (Fountain et al., 2016). This is especially relevant in CKMR studies that aim at determining kinship relationships. As a general rule, studies using restriction enzyme methods should aim at a minimum depth coverage of 25X to minimize genotyping error rates (Paris et al., 2017). Moreover, before embarking in a RADseq experiment, an estimation of the coverage should always be performed before library construction, based on the number of cut sites and the number of multiplexed samples (Rivera-Colón et al., 2021).

Finally, high throughput sequencing generally involves the preparation of libraries that mostly include a PCR step that generates copies ("PCR duplicates") of the original DNA fragments ("parent fragments"). Stochastic effects during PCR can cause uneven amplification of heterozygous alleles, causing two undesired phenomena. First, heterozygotes would appear as homozygotes and additionally, alleles containing PCR errors can appear as true alleles because PCR duplicates spuriously increase confidence in their calling (Andrews et al., 2014; Puritz et al., 2014b). Failure to remove bioinformatically PCR duplicates can potentially lead to downstream errors in genotyping (Tin et al., 2015; Flanagan and Jones, 2019) although this phenomenon is still being intensely debated ((Euclide et al., 2020). Thus, an important point to consider when selecting the genotyping method in a CKMR study is whether the assay allows the removal of PCR duplicates (Andrews et al., 2016). For instance, both sd- and ddRAD methods have a PCR step in their protocols, but only the first allows their removal. This is possible because the sdRADseq protocol has a mechanical fragmentation step, absent in ddRADseq, which generates fragments of slightly different sizes. Removal of PCR duplicates is only possible with ddRADseq if the protocol is modified to incorporate random oligo-nucleotides into the barcodes of the molecular library (Rochette et al., 2019). Thus, dual-digest techniques like quaddRAD and 2RAD/3RAD are suitable for CKMR studies, as they include molecular ID tags that allow detection and removal of PCR duplicates (Franchini et al., 2017; Bayona-Vásquez et al., 2019). Another alternative would be the use of a PCR-free protocol with any RADseq-based method but this approach has important drawbacks, as it is costly and has substantial technical limitations, especially with regard to the starting DNA amount (Toonen et al., 2013; Rochette et al., 2023).

3 Kinship analysis

The last step of the bioinformatics pipeline in any CKMR study consists on a precise detection of the close inter-familial relationships between the individuals present in a sample, based on their genotypes, using statistical methods that need to be adapted to the nature of the data (Bravington et al., 2016b; Ruzzante et al., 2019). The identification of related specimens is possible because of inheritance, as different types of kinship share different degrees of genetic relatedness (Städele and Vigilant, 2016). When two specimens reproduce, they each pass, on average, half of their DNA on to their progeny, in diploid organisms (i.e. those that have two complete sets of chromosomes, one from each parent; includes most animals). Thus, parent-offspring share 50% of their genome, approximately the same percentage as full-siblings although each of these relationships has a distinct chromosomal sharing pattern. Second-degree relatives (half-siblings) share, on average, 25% of their genomes and are also informative in CKMR studies (Waples et al., 2018; Delaval et al., 2022; Patterson et al., 2022). Despite the distinctiveness of the shared patterns, assigning individuals into discrete kinship categories such as "full-siblings" or "half-siblings" is difficult because the percentage of the genome shared can vary considerably due to stochastic processes that occur during cell division (Städele and Vigilant, 2016). Kinship analysis in wild marine fish populations is, moreover, a unique challenge due to the necessity to make large numbers of pairwise comparisons and the low percentage of true kinship pairs, requiring large panels of genomic markers to infer close relationships with accuracy [e.g. (Marcy-Quay et al., 2020; Delaval et al., 2022; Trenkel et al., 2022)].

Statistical methods used to assess kin relationships from molecular data can be grouped into three broad categories: 1) exclusion methods; 2) relatedness-based methods; and 3)

likelihood-based methods, which are of increasing power, but have substantial computational costs as a trade-off (Huisman, 2017).

Exclusion methods are qualitatively based on Mendel's laws of inheritance, excluding a relationship among a set of individuals if their genotypes are incompatible given the relationship under the laws. They are very fast and simple in concept and implementation, but suffer from several weaknesses, including the difficulty to incorporate genotyping errors and mutations that can invalidate true kinship pairs (Wang, 2012).

Relatedness-based methods estimate pairwise relatedness or kinship coefficients between individuals, and use these to categorize the data into first-degree relatives, second-degree relatives and unrelated. These coefficients quantify the amount of genetic sharing between pairs of individuals reflecting the actual level of shared ancestry between two individuals based on their DNA (Goudet et al., 2018). In simple systems, with non-overlapping generations and no inbreeding, a simple measure of relatedness, defined as the probability that a pair of randomly sampled homologous alleles are identical by descent (IBD), might be sufficient to assign kinship (Huisman, 2017). Nonetheless, most marine fish populations are characterized by overlapping generations, requiring a more precise description to differentiate between kinship types. Table 2 reflects the probability of sharing 0, 1 or 2 alleles that are IBD (kinship coefficients κ_0, κ_1 and κ_2), for some common relationships, although neither pairwise measure can distinguish between half-siblings, grandparents and full aunts/uncles (all k=0.25).

Likelihood methods are, in comparison, more powerful, accurate, and robust but computationally more demanding and thus, more suitable to achieve the accuracy needed in CKMR studies. Methods in this third category consider the relationships among all individuals in a sample to assign kinship so for large data sets with many individuals and markers, this approach can be computationally daunting (Wang, 2012). In such cases, is common to reduce computational cost by considering only pairwise likelihoods, ignoring all other individuals related or unrelated to the pair (Huisman, 2017). This group of methods work upon genotype likelihoods or posterior probabilities, allowing the incorporation of the uncertainty of genotype calls (Herzig et al., 2022).

Accuracy of kinship inference can be affected by two issues, false-positives that happen when an unrelated pair share enough alleles by chance to look as a related pair, and false-negatives that arise when a pair appears not to share alleles that are, in fact, present in both. The chance of false positives decreases with the increment in number of loci used and needs to be assessed in advance using the allele frequencies (and the per-locus exclusion criterion) to ensure the use of enough loci (Harrison et al., 2013a; Harrison et al., 2013b). False negatives cannot be predicted in advance and appear due to null alleles and genotyping errors that produce the incorrect recording of the true alleles, leading to the rejection of true kinship pairs (O'Leary et al., 2018). Incorporating both errors in kinship assignment is paramount in CKMR studies of marine teleost since they mostly target large populations. In a pairwise analysis of kinship, the expected number of related pairs is only a very small fraction of the total number of comparisons [e.g. (Bravington et al., 2016a; Trenkel et al., 2022)]. The large-scale sparse kinship nature of such applications determines a large effect of the inclusion of spurious kin or the exclusion of real kin pairs on subsequent estimates of population parameters.

Published CKMR studies targeting marine populations to date have all been based on the detection of Parent-Offspring-Pairs (POPs) and/or Half-Sibling-Pairs (HSP) (Bravington et al., 2016a; Hillary et al., 2018; Ruzzante et al., 2019; Marcy-Quay et al., 2020; Prystupa et al., 2021; Wacker et al., 2021; Delaval et al., 2022; Trenkel et al., 2022). They have mostly relied on the use of specifically designed algorithms to incorporate false positive and negative rates (see (Bravington et al., 2016a; Bravington et al., 2016b; Hillary et al., 2018) for full details of developing a likelihood-ratio kin identification statistic). Two statistics known as WPSEX (Weighted PSeudo-EXclusion) and PLOD (Pseudo log-odds) scores have been used in CKMR studies to calculate the probability of POPs and HSPs, respectively, between pairs of individuals (Bravington et al., 2016b; Hillary et al., 2018; Trenkel et al., 2022). WPSEX is designed to robustly identify parent-offspring pairs from biallelic SNP data characterized by many loci that may have (heritable) null alleles, and occasional genotyping errors that may prevent the (non-heritable) detection of alleles. The frequency of null alleles per locus is estimated in advance by maximum-likelihood, assuming Hardy-Weinberg equilibrium and

TABLE 2 Kinship categories (*pairwise kinship*) with their corresponding kinship coefficient (φ) and probability of sharing zero, one or two alleleles identitcal-by-descent (ibd) (κ_0, κ_1 and κ_2).

Pairwise kinship	φ	*ibd* probability		
		κ_0	κ_1	κ_2
Self, Monozygous twin (MZ)	1/2	0	0	1
Parent-offspring (PO)	1/4	0	1	0
Full sibling (FS)	1/4	1/4	1/2	1/4
Half-sibling (HS)	1/8	1/2	1/2	0
Grandparent-grandchild	1/8	1/2	1/2	0
Aunt-niece	1/8	1/2	1/2	0
Unrelated (U)	0	1	0	0

does not require estimates of genotyping error rate (Trenkel et al., 2022). The PLOD score provides the pseudo-likelihood that a pair of animals are HSPs and results from summing the log-likelihood per locus. A higher WPSEX or PLOD value indicates a greater likelihood that the pair are a POP/HSP, respectively. These scores are ultimately used to identify threshold values for the robust classification of kinship categories (POP, HSP, or UP (Unrelated Pair; all more-distant kinship categories)) (Hillary et al., 2018; Patterson et al., 2022).

An alternative approach followed by CKMR studies targeting solely POPs in populations with simple structure and life-histories [e.g., brown trout (Ruzzante et al., 2019)] was the use of widely available kinship inference software. COLONY (Jones and Wang, 2010) has been often the program of choice (Ruzzante et al., 2019; Marcy-Quay et al., 2020; Wacker et al., 2021) although a bewildering variety of software packages can be used to infer kinship (e.g. SNPRelate (Zheng et al., 2012); NGSRelate (Hanghøj et al., 2019); Sequoia (Huisman, 2017), among many others). Nonetheless, is not always clear how (or indeed whether) some methods differ from others and no systematic comparison studies have been performed, even less so in a CKMR framework. Moreover, such full-pedigree methods that attempt to address the huge complexities of family-reconstruction might be computationally too demanding to analyse many teleosts populations, due to their large sizes that imply huge number of possible pairwise relationships (Bravington et al., 2017).

The R package CKMRsim has been recently released and was specifically developed to compute likelihood ratios for different relationships between all pairs of individuals in a data set for close kin mark recapture studies. The software allows the assessment of false positive and false negative rates through Monte Carlo methods (Anderson, 2022a). CKMRsim simulates the genotypes of related pairs of individuals from the estimated allele frequencies and then calculates the probabilities of those genotype pairs to compute a log-likelihood ratio of the true relationship vs. the hypothesis of no relationship. Similarly, genotypes of unrelated pairs are also simulated and their log-likelihood ratios computed. The comparison of observed likelihood ratio values of related kin pairs with the distribution of simulated pairwise values is used to select the threshold values for classifying a pair into a given relationship, minimizing wrong assignments.

Two CKMR studies have compared CKMRsim with other kinship inference software. In blue skate, ML-relate identified a slightly higher number of kinship pairs than CKMRsim (27 vs. 19), all HSPs with lower relatedness values than those of HSPs detected by both methods (Delaval et al., 2022). COLONY unveiled, on average (5 runs) 11 more POPs than CKMRsim in Arctic Grayling (37.67 vs 26) (Prystupa et al., 2021). In both cases, all the related pairs detected with CKMRsim were also inferred by the other software's, but the former was more conservative. These differences are very relevant in CKMR studies, having large effects due to low numbers of kin pairs normally detected. Thus, if a software is selected to perform kinship inference, it is essential to determine whether the package of choice provides sufficient control over false-positive and false-negative error rates to assess confidence in kinship inference.

3.1 The value of non-autosomal markers and ancillary information

In addition to genetic markers, complementary biological data (ideally age, otherwise length can be used, sex) should be used as a check point of the feasibility of the inferred kin relationships and identify false-positive kinship assignments. For example, ancillary data can serve to identify pairs that cannot have the purported relationship due to their relative ages. Depending on the life history of the species, age may also be useful to differentiate relationships that are indistinguishable otherwise. For example, half siblings may be discriminated from grandparent–grandoffspring by the age difference between the individuals in species in which the reproductive life span is shorter than roughly twice the age at first reproduction. This information is also essential to separate within and among-cohort relationships, since only the later are useful in CKMR studies (Hillary et al., 2018; Davies et al., 2020; Maunder et al., 2021; Waples and Feutry, 2022).

Besides nuclear markers, CKMR studies can strongly benefit from the information provided by mitochondrial DNA (mtDNA) (Bravington et al., 2017). While the former are bi-parentally inherited, mtDNA is acquired only from the mother in most animals, including fishes (Breton and Stewart, 2015). This inheritance pattern can be extremely useful in kinship analysis as it can be used to discriminate between half siblings that are related through the father (paternal HSP) and the mother (maternal HSP) (Thompson et al., 2020). This maternally inherited marker is also useful for identifying false-positive assignments of kinship with Mendelian incompatibilities as well as for reducing the misclassification rates (Kopps et al., 2015). In addition, mtDNA of identified kin pairs can be used to uncover differences in the reproductive dynamics of females and males and obtain information on sex ratios, sexually dimorphic mortality, mating strategies or spatial reproductive structure (Mace et al., 2020).

4 Discussion

Most species fished worldwide lack the data needed to assess their status, despite the increasing trend of overfished stocks in the last four decades (FAO, 2022). Only a low number of fish stocks, among those targeted by fisheries, are subjected to detailed assessments and all inhabit the waters of developed regions (as in Europe, the USA, Canada or Australia) (Palomares et al., 2020). Even these detailed assessments suffer from recognized shortcomings as they utilize parameters that are known to contain uncertainty (Kokkalis et al., 2017), having a great impact on the quality of the scientific advice provided to management bodies and hence on fishery activity.

There is a crucial need for innovative methodologies that can provide novel means to reduce uncertainty of fish stocks assessments and expand the range of assessed species. Close kin mark recapture is an emerging fisheries independent approach to estimate population parameters with potential to improve fisheries assessments (Bravington et al., 2016b). The method is grounded in genomics and its application requires expertise and knowledge in

two distinct fields that have largely evolved separately, complicating its uptake and implementation by fisheries scientist and modellers. In spite of the value of genomic based methods to inform fisheries management being acknowledged, they have seldom been incorporated and remain underutilized to these days (Bernatchez et al., 2017).

The rapid advances and extraordinary number of sequencing and computational technologies certainly represent a major challenge to those outside the genomics field. Moreover, to date CKMR has only been applied to a handful of species, mainly characterized by smaller population sizes than most exploited fish stocks (Bravington et al., 2016a; Hillary et al., 2018; Ruzzante et al., 2019; Marcy-Quay et al., 2020; Prystupa et al., 2021; Wacker et al., 2021; Delaval et al., 2022; Trenkel et al., 2022).

The nature of CKMR studies in a large population setting with sparse sampling, where most individuals are unrelated and the probability of finding closely related individuals is low, determines a large effect of small deviations from true kin numbers on subsequent estimates of population parameters.

Achieving accurate kinship determinations requires strict quality controls at every stage, starting with the isolation of high-quality DNA samples, which is contingent upon an adequate sample preservation. Genetic markers should be selected in terms of the quality and amount of information they provide and we argue that SNPs should be preferred to STRs in contemporary and future CKMR projects. While studies using STRs can often confidently identify only first-order kin relations (parent–offspring or full-sibling), the use of high-density, genome-wide SNP markers can enable reasonably accurate assignment of individuals to second-order (e.g. half-siblings), thereby sensibly reducing the sample size needed in CKMR studies, a requirement that has possibly prevented more widespread adoption of kinship-based methodologies.

A plethora of methods can be used for marker discovery and genotyping but enzyme-based, particularly RADseq methods together with hybridization-based methodologies are particularly suited for CKMR applications. For RADseq, many different protocols are available and each has its own trade-offs but the original sdRADseq has been more intensively tested across multiple systems and the sources of bias in the resulting data are better known (Andrews et al., 2014; Flanagan and Jones, 2019; Rochette et al., 2019). Other variants can also be used as long as they allow identification and removal of PCR duplicates, including the 2RAD/3RAD protocols, which provide a streamlined workflow at a lower cost and can have an increased utility with low-concentration DNA samples (Bayona-Vásquez et al., 2019; Glenn et al., 2019). Undeniably, none of these methods are characterized by the simplicity of their protocols and they require an enormous amount of post-processing, but their advantages clearly outweigh these drawbacks. In parallel, a large number of bioinformatics software, dependent on the genomic method selected, can be applied for the analysis but their use is not always straightforward since it needs to be tailored to the user´s data characteristics. Nonetheless, these steps do not need to be carried out necessarily in-house as they can be outsourced to an ever growing number of companies worldwide.

Inference of kinship using a specific relatedness software is challenging due to the lack of systematic comparison studies among them and because not all provide a sufficient control over false-positive and false-negative error rates to assess confidence in kinship inference. To avoid this problem, the statistics WPSEX and PLOD can be used, alternatively, to identify reliable threshold values for the robust classification of kinship categories (Bravington et al., 2016b; Bradford et al., 2018; Hillary et al., 2018; Thompson et al., 2020; Trenkel et al., 2022). The package CKMRsim was specifically developed to compute likelihood ratios for different relationships between pairs of individuals in a CKMR framework. It has proved more conservative than other software's in a few studies (Prystupa et al., 2021; Delaval et al., 2022) but its accuracy has not been tested with individuals of known pedigree, to the best of our knowledge.

The use of ancillary data (age/length, sex) and non-autosomal markers (mtDNA) is pivotal to check the feasibility of the inferred kin relationships and identify false-positive kinship assignments but also to discriminate among types of a degree of kinship (e.g. grandparent–grandoffspring and half-siblings).

Although here we provide useful information to complete the genomics and bioinformatics steps required in a CKMR study today, it is very important to highlight that the application of the method is still very marginal. Considering this fact and the fast pace with which technology in this field progresses, we foresee the incorporation of newer methodologies with the capacity to boost the power of close kin analyses, and facilitate its application to a wider range of fish stocks.

Author contributions

All authors contributed to the article and approved the submitted version.

Funding

The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union.

Acknowledgments

The authors would like to express their gratitude to Gary Carvalho, Laszlo Orban and Allan Tucker for their advice and expert revision of the State of the Art review in which this manuscript is based on.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adler, A. J., Wiley, G. B., and Gaffney, P. M. (2013). Infinium assay for Large-scale SNP genotyping applications. *J. Vis. Exp* (81), e50683. doi: 10.3791/50683

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., et al. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202, 389–400. doi: 10.1534/genetics.115.183665

Amorim, A., and Pereira, L. (2005). Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci. Int.* 150, 17–21. doi: 10.1016/j.forsciint.2004.06.018

Anderson, E. (2022a). *CKMRsim: inference of pairwise relationships using likelihood ratios* (R package version). https://eriqande.github.io/CKMRsim/authors.html#citation

Anderson, E. C. (2022b). CKMRpop: forward-in-time simulation and tabulation of pairwise kin relationships in age-structured populations. *Mol. Ecol. Resour* 22, 1190–1199. doi: 10.1111/1755-0998.13513

Anderson, E. C., and Garza, J. C. (2006). The power of single-nucleotide polymorphisms for Large-scale parentage inference. *Genetics* 172, 2567–2582. doi: 10.1534/genetics.105.048074

Anderson, G., Macdonald, J. I., Potts, J., Feutry, P., Grewe, P. M., Boutigny, M., et al. (2023). Evaluating DNA cross-contamination risk using different tissue sampling procedures on board fishing and research vessels. *ICES J. Mar. Sci* 80 (4), 728–738. doi: 10.1093/icesjms/fsad004

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28

Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., and Luikart, G. (2014). Trade-offs and utility of alternative RADseq methods. *Mol. Ecol.* 23, 5943–5946. doi: 10.1111/mec.12964

Arnold, B., Corbett-Detig, R. B., Hartl, D., and Bomblies, K. (2013a). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol. Ecol.* 22, 3179–3190. doi: 10.1111/mec.12276

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One* 3, e3376. doi: 10.1371/journal.pone.0003376

Bayat, A. (2002). Science, medicine, and the future: bioinformatics. *BMJ* 324, 1018–1022. doi: 10.1136/bmj.324.7344.1018

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., et al. (2019). Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ* 7, e7724. doi: 10.7717/peerj.7724

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32, 665–680. doi: 10.1016/j.tree.2017.06.010

Bradford, R. W., Thomson, R., Bravington, M., Foote, D., Gunasekera, R., Bruce, B. D., et al. (2018). *A close-kin mark-recapture estimate the population size and trend of east coast grey nurse shark* (Tasmania: Rep to Natl Environ Sci Program, Mar Biodivers Hub CSIRO Ocean Atmos Hobart). Available at: https://www.nespmarine.edu.au/document/close-kin-mark-recapture-estimate-population-size-and-trend-east-coast-grey-nurse-shark.

Bravington, M. V., and Grewe, J. P. (2007). *A method for estimating the absolute spawning stock size of SBT, using close-kin genetics* (Hobart, Australia: CSIRO). Available at: https://www.ccsbt.org/ja/system/files/resource/ja/4d93fe45b53f2/18.pdf.

Bravington, M. V., Grewe, P. M., and Davies, C. R. (2016a). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat. Commun.* 7, 13162. doi: 10.1038/ncomms13162

Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016b). Close-kin mark-recapture. *Stat. Sci.* 31, 259–274. doi: 10.1214/16-STS552

Bravington, M., Thomson, R., and Davies, C. (2017) *Genotyping issues for CKMR on Atlantic bluefin tuna*. Available at: https://www.iccat.int/GBYP/Docs/Close_Kin_Mark_Recapture_Phase_6_Genotyping_Issues.pdf.

Breton, S., and Stewart, D. T. (2015). Atypical mitochondrial inheritance patterns in eukaryotes. *Genome* 58, 423–431. doi: 10.1139/gen-2015-0090

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

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., et al. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evol. Appl.* 11, 1094–1119. doi: 10.1111/eva.12600

Casas, L., and Saborido-Rey, F. (2023). A review of an emerging tool to estimate population parameters: the close-kin mark–recapture method. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1087027

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22 (11), 3124–3140. doi: 10.1111/mec.12354

Christiansen, H., Heindler, F. M., Hellemans, B., Jossart, Q., Pasotti, F., Robert, H., et al. (2021). Facilitating population genomics of non-model organisms through optimized experimental design for reduced representation sequencing. *BMC Genomics* 22, 625. doi: 10.1186/s12864-021-07917-3

Clark, L. V., Lipka, A. E., and Sacks, E. J. (2019). polyRAD: genotype calling with uncertainty from sequencing data in polyploids and diploids. *G3 (Bethesda)* 9 (3), 663–673. doi: 10.1534/g3.118.200913

Crespel, A., Schneider, K., Miller, T., Rácz, A., Jacobs, A., Lindström, J., et al. (2021). Genomic basis of fishing-associated selection varies with population density. *Proc. Natl. Acad. Sci.* 118 (51), e2020833118. doi: 10.1073/pnas.2020833118

Cumer, T., Pouchon, C., Boyer, F., Yannic, G., Rioux, D., Bonin, A., et al. (2021). Double-digest RAD-sequencing: do pre- and post-sequencing protocol parameters impact biological results? *Mol. Genet. Genomics* 296, 457–471. doi: 10.1007/s00438-020-01756-9

Dahn, H. A., Mountcastle, J., Balacco, J., Winkler, S., Bista, I., Schmitt, A. D., et al. (2022). Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing. *Gigascience* 11, giac68. doi: 10.1093/gigascience/giac068

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., and Blaxter, M. L. (2013). Special features of RAD sequencing data: implications for genotyping. *Mol. Ecol.* 22, 3151–3164. doi: 10.1111/mec.12084

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012

Davies, C. R., Bravington, M. V., Eveson, J. P., Lansdell, M., Aulich, J., and Grewe, P. M. (2020) *Next-generation close-kin mark recapture: using SNPs to identify half- sibling pairs in southern bluefin tuna and estimate abundance, mortality and selectivity*. Available at: https://www.frdc.com.au/sites/default/files/products/2016-044-DLD.pdf.

Davies, C. R., Bravington, M. V., and Thompson, R. (2015). Advice on close-kin mark-recapture for estimating abundance of eastern Atlantic blue tuna:a scoping study. *ICAAT GBYP07* 20, 34. Available at: https://www.iccat.int/GBYP/Docs/Close_Kin_Mark_Recapture_Phase_6_Scoping_Study.pdf.

De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., and Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Mol. Ecol. Resour* 17, 492–507. doi: 10.1111/1755-0998.12594

Delaval, A., Bendall, V., Hetherington, S. J., Skaug, H. J., Frost, M., Jones, C. S., et al. (2022). Evaluating the suitability of close-kin mark-recapture as a demographic modelling tool for a critically endangered elasmobranch population. *Evol. Appl* 16 (2), 461–473. doi: 10.1111/eva.13474

Del Giacco, L., and Cattaneo, C. (2012). Introduction to genomics. *Methods Mol Biol.* 823, 79–88. doi: 10.1007/978-1-60327-216-2_6

Eaton, D. A. R., and Overcast, I. (2020). Ipyrad: interactive assembly and analysis of RADseq data sets. *Bioinformatics* 36 (8), 2592–2594. doi: 10.1093/bioinformatics/btz966

Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., et al. (2011). Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (Salmo salar l.). *Genetica* 139, 353–367. doi: 10.1007/s10709-011-9554-4

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-Sequencing (GBS) approach for high diversity species. *PloS One* 6, e19379. doi: 10.1371/journal.pone.0019379

Euclide, P. T., McKinney, G. J., Bootsma, M., Tarsa, C., Meek, M. H., and Larson, W. A. (2020). Attack of the PCR clones: rates of clonality have little effect on RAD-seq genotype calls. *Mol. Ecol. Resour* 20, 66–78. doi: 10.1111/1755-0998.13087

Fan, G., Song, Y., Yang, L., Huang, X., Zhang, S., Zhang, M., et al. (2020). Initial data release and announcement of the 10,000 fish genomes project (Fish10K). *Gigascience* 9 (8), giaa080. doi: 10.1093/gigascience/giaa080

FAO (2022). *The state of world fisheries and aquaculture* (Rome: FAO). doi: 10.4060/cc0461en

Faria, P. J., Kavembe, G. D., Jung'a, J. O., Kimwele, C. N., Estes, L. D., Reillo, P. R., et al. (2011). The use of non-invasive molecular techniques to confirm the presence of mountain bongo tragelaphus eurycerus isaaci populations in Kenya and preliminary inference of their mitochondrial genetic variation. *Conserv. Genet.* 12, 745–751. doi: 10.1007/s10592-011-0181-5

Fisher, P. J., Malthus, B., Walker, M. C., Corbett, G., and Spelman, R. J. (2009). The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. *J. Dairy Sci.* 92, 369–374. doi: 10.3168/jds.2008-1086

Flanagan, S. P., and Jones, A. G. (2019). The future of parentage analysis: from microsatellites to SNPs and beyond. *Mol. Ecol.* 28, 544–567. doi: 10.1111/mec.14988

Fountain, E. D., Pauli, J. N., Reid, B. N., Palsbøll, P. J., and Peery, M. Z. (2016). Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Mol. Ecol. Resour* 16, 966–978. doi: 10.1111/1755-0998.12519

Franchini, P., Monné Parera, D., Kautt, A. F., and Meyer, A. (2017). quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Mol. Ecol.* 26, 2783–2795. doi: 10.1111/mec.14077

Francois, C. M., Durand, F., Figuet, E., and Galtier, N. (2020). Prevalence and implications of contamination in public genomic resources: a case study of 43 reference arthropod assemblies. *G3 Genes|Genomes|Genetics* 10, 721–730. doi: 10.1534/g3.119.400758

Fu, Z., Epstein, B., Kelley, J. L., Zheng, Q., Bergland, A. O., Castillo Carrillo, C. I., et al. (2017). Using NextRAD sequencing to infer movement of herbivores among host plants. *PloS One* 12 (5), e0177742. doi: 10.1371/journal.pone.0177742

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907*. doi: 10.48550/arXiv.1207.3907

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., et al. (2013a). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Mol. Ecol.* 22 (11), 3165–3178. doi: 10.1111/mec.12089

Gershoni, M., Shirak, A., Raz, R., and Seroussi, E. (2022). Comparing BeadChip and WGS genotyping: non-technical failed calling is attributable to additional variation within the probe target sequence. *Genes (Basel)* 13, 485. doi: 10.3390/genes13030485

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., et al. (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed illumina libraries (iTru & iNext). *PeerJ* 7, e7755. doi: 10.7717/peerj.7755

Gomez-Raya, L., Gómez Izquierdo, E., Garcia-Ruiz, F., and Rauw, W. M. (2022). First-degree relationships and genotyping errors deciphered by a high-density SNP array in a duroc × Iberian pig cross. *BMC Genomic Data* 23, 14. doi: 10.1186/s12863-022-01025-1

Goudet, J., Kay, T., and Weir, B. S. (2018). How to estimate kinship. *Mol. Ecol.* 27, 4121–4135. doi: 10.1111/mec.14833

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., et al. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Mol. Ecol. Resour* 15, 1304–1315. doi: 10.1111/1755-0998.12404

Guang, X.-M., Xia, J.-Q., Lin, J.-Q., Yu, J., Wan, Q.-H., and Fang, S.-G. (2019). IDSSR: an efficient pipeline for identifying polymorphic microsatellites from a single genome sequence. *Int. J. Mol. Sci.* 20, 3497. doi: 10.3390/ijms20143497

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., and Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* 8 (5), giz034. doi: 10.1093/gigascience/giz034

Harrison, H. B., Saenz-Agudelo, P., Planes, S., Jones, G. P., and Berumen, M. L. (2013a). On minimizing assignment errors and the trade-off between false positives and negatives in parentage analysis. *Mol. Ecol.* 22, 5738–5742. doi: 10.1111/mec.12527

Harrison, H. B., Saenz-Agudelo, P., Planes, S., Jones, G. P., and Berumen, M. L. (2013b). Relative accuracy of three common methods of parentage analysis in natural populations. *Mol. Ecol.* 22, 1158–1170. doi: 10.1111/mec.12138

Herzig, A. F., Ciullo, M., Deleuze, J.-F., Génin, E., Redon, R., Adjou, C., et al. (2022). Moment estimators of relatedness from low-depth whole-genome sequencing data. *BMC Bioinf.* 23, 254. doi: 10.1186/s12859-022-04795-8

Hillary, R. M., Bravington, M. V., Patterson, T. A., Grewe, P., Bradford, R., Feutry, P., et al. (2018). Genetic relatedness reveals total population size of white sharks in eastern Australia and new Zealand. *Sci. Rep.* 8, 2661. doi: 10.1038/s41598-018-20593-w

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., et al. (2016). RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Mol. Ecol. Resour* 16, 1264–1278. doi: 10.1111/1755-0998.12566

Huang, H., and Knowles, L. L. (2016). Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Syst. Biol.* 65, 357–365. doi: 10.1093/sysbio/syu046

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour* 17, 1009–1024. doi: 10.1111/1755-0998.12665

Ishibashi, Y., Saitoh, T., Abe, S., and Yoshida, M. C. (1996). Null microsatellite alleles due to nucleotide sequence variation in the grey-sided vole clethrionomys rufocanus. *[Internet]. . Mol. Ecol.* 5, 589–590. doi: 10.1111/j.1365-294X.1996.tb00351.x

Jones, O. R., and Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour* 10, 551–555. doi: 10.1111/j.1755-0998.2009.02787.x

Kokkalis, A., Eikeset, A. M., Thygesen, U. H., Steingrund, P., and Andersen, K. H. (2017). Estimating uncertainty of data limited stock assessments. *ICES J. Mar. Sci.* 74, 69–77. doi: 10.1093/icesjms/fsw145

Kopps, A. M., Kang, J., Sherwin, W. B., and 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? *G3: Genes|Genomes|Genetics* 5, 1815–1826. doi: 10.1534/g3.115.019323

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* 15 (1), 356. doi: 10.1186/s12859-014-0356-4

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., et al. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness {[/amp]]mdash; implications for brown trout conservation. *Ecol. Evol.* 9, 2106–2120. doi: 10.1002/ece3.4905

Lepais, O., and Weir, J. T. (2014). SimRAD: an r package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Mol. Ecol. Resour* 14, 1314–1321. doi: 10.1111/1755-0998.12273

Lutz, Í, Miranda, J., Santana, P., Martins, T., Ferreira, C., Sampaio, I., et al. (2023). Quality analysis of genomic DNA and authentication of fisheries products based on distinct methods of DNA extraction. *PloS One* 18, e0282369. doi: 10.1371/journal.pone.0282369

Mace, P., Ritchie, P., Wellenreuther, M., McKenzie, J., Hupman, K., Hillary, R., et al. (2020) *Report of the workshop on the utility of genetic analyses for addressing new Zealand fisheries questions*. Available at: https://www.mpi.govt.nz/dmsdocument/39668/direct.

Marcy-Quay, B., Sethi, S. A., Therkildsen, N. O., and Kraft, C. E. (2020). Expanding the feasibility of fish and wildlife assessments with close-kin mark–recapture. *Ecosphere* 11 (10), e03259. doi: 10.1002/ecs2.3259

Martincová, I., and Aghová, T. (2020). Comparison of 12 DNA extraction kits for vertebrate samples. *Anim. Biodivers Conserv.* 43 (1), 67–77. doi: 10.32800/abc.2020.43.0067

Maunder, M. N., Lennert-Cody, C. E., -da-Silva, A., Alexandre, A., and Xu, H. (2021) *Considerations for conducting close kin mark recapture of stocks managed by the IATTC (document SAC-12-14)*. Available at: https://www.iattc.org/getattachment/ddc8410e-4de7-401c-87a9-73fb33090f23/SAC-12-14_Considerations-for-conducting-Close-Kin-Mark-Recapture-of-stocks-managed-by-IATTC.pdf.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17, 240–248. doi: 10.1101/gr.5681207

Miller, M. P., Knaus, B. J., Mullins, T. D., and Haig, S. M. (2013). SSR_pipeline: a bioinformatic infrastructure for identifying microsatellites from paired-end illumina high-throughput DNA sequencing data. *J. Hered* 104, 881–885. doi: 10.1093/jhered/est056

Mora-Márquez, F., García-Olivares, V., Emerson, B. C., and López de Heredia, U. (2017). Ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. *Mol. Ecol. Resour* 17, 230–246. doi: 10.1111/1755-0998.12550

Mulcahy, D. G., Macdonald, K. S., Brady, S. G., Meyer, C., Barker, K. B., and Coddington, J. (2016). Greater than X kb: a quantitative assessment of preservation conditions on genomic DNA quality, and a proposed standard for genome-quality DNA. *PeerJ* 4, e2528. doi: 10.7717/peerj.2528

Musich, R., Cadle-Davidson, L., and Osier, M. V. (2021). Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.657240

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986

Oduoye, O., Ajayi, D., Oluwasanya, O., Hassan, T., Olajire, O., Coker, D., et al. (2020). COMPARATIVE ANALYSIS OF DNA ISOLATED FROM FISH SAMPLES BY THREE DIFFERENT EXTRACTION PROTOCOLS. *J. Exp. Biol. Agric. Sci.* 8, 644–653. doi: 10.18006/2020.8(5).644.653

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., and Portnoy, D. S. (2018). These aren't the loci you'e looking for: principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* 27, 3193–3206. doi: 10.1111/mec.14792

Oosting, T., Hilario, E., Wellenreuther, M., and Ritchie, P. A. (2020). DNA Degradation in fish: practical solutions and guidelines to improve DNA preservation for genomic research. *Ecol. Evol.* 10, 8643–8651. doi: 10.1002/ece3.6558

Paetkau, D., and Strobeck, C. (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol. Ecol.* 4, 519–520. doi: 10.1111/j.1365-294X.1995.tb00248.x

Palomares, M. L. D., Froese, R., Derrick, B., Meeuwig, J. J., Nöel, S.-L., Tsui, G., et al. (2020). Fishery biomass trends of exploited fish populations in marine ecoregions, climatic zones and ocean basins. *Estuar. Coast. Shelf Sci.* 243, 106896. doi: 10.1016/j.ecss.2020.106896

Paris, J. R., Stevens, J. R., and Catchen, J. M. (2017). Lost in parameter space: a road map for STACKS. *Methods Ecol. Evol.* 8, 1360–1373. doi: 10.1111/2041-210X.12775

Patterson, T. A., Hillary, R. M., Kyne, P. M., Pillans, R. D., Gunasekera, R. M., Marthick, J. R., et al. (2022). Rapid assessment of adult abundance and demographic connectivity from juvenile kin pairs in a critically endangered species. *Sci. Adv.* 8 (51), eadd1679. doi: 10.1126/sciadv.add1679

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *De novo* SNP discovery and genotyping in model and non-model species. *PloS One* 7, e37135. doi: 10.1371/journal.pone.0037135

Pimentel, J. S. M., Carmo, A. O., Rosse, I. C., Martins, A. P. V., Ludwig, S., Facchin, S., et al. (2018). High-throughput sequencing strategy for microsatellite genotyping using Neotropical fish as a model. *Front. Genet.* 9, 73. doi: 10.3389/fgene.2018.00073

Prystupa, S., McCracken, G. R., Perry, R., and Ruzzante, D. E. (2021). Population abundance in arctic grayling using genetics and close-kin mark-recapture. *Ecol. Evol.* 11, 4763–4773. doi: 10.1002/ece3.7378

Puckett, E. E. (2017). Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions. *Conserv. Genet. Resour* 9, 289–304. doi: 10.1007/s12686-016-0643-7

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014a). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2, e431. doi: 10.7717/peerj.431

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., and Bird, C. E. (2014b). Demystifying the RAD fad. *Mol. Ecol.* 23, 5937–5942. doi: 10.1111/mec.12965

Reuter, J. A., Spacek, D. V., and Snyder, M. P. (2015). High-throughput sequencing technologies. *Mol. Cell* 58, 586–597. doi: 10.1016/j.molcel.2015.05.004

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592, 737–746. doi: 10.1038/s41586-021-03451-0

Rivera-Colón, A. G., and Catchen, J. (2022). "Population genomics analysis with RAD, reprised: stacks 2," in *Methods in molecular biology* 2498, 99–149. doi: 10.1007/978-1-0716-2313-8_7

Rivera-Colón, A. G., Rochette, N. C., and Catchen, J. M. (2021). Simulation with RADinitio improves RADseq experimental design and sheds light on sources of missing data. *Mol. Ecol. Resour* 21, 363–378. doi: 10.1111/1755-0998.13163

Rochette, N. C., and Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using stacks. *Nat. Protoc.* 12, 2640–2659. doi: 10.1038/nprot.2017.123

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28, 4737–4754. doi: 10.1111/mec.15253

Rochette, N. C., Rivera-Colón, A. G., Walsh, J., Sanger, T. J., Campbell-Staton, S. C., and Catchen, J. M. (2023). On the causes, consequences, and avoidance of PCR duplicates: towards a theory of library complexity. *Mol. Ecol. Resour*. doi: 10.1111/1755-0998.13800

Ruzzante, D. E., McCracken, G. R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., et al. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. *Methods Ecol. Evol.* 10, 1445–1453. doi: 10.1111/2041-210X.13243

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., et al. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* 8, 907–917. doi: 10.1111/2041-210X.12700

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., et al. (2017). hyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods Ecol. Evol.* 8, 1374– 1388. doi: 10.1111/2041-210X.12785

Shuttleworth, L., and Oosthuizen, C. J. (2022). Comparing DNA yield from fish scales following different extraction protocols. *Sci. Rep.* 12, 2836. doi: 10.1038/s41598-022-06889-y

Silva, P. C., Malabarba, M. C., Vari, R., and Malabarba, L. R. (2019). Comparison and optimization for DNA extraction of archived fish specimens. *MethodsX* 6, 1433–1442. doi: 10.1016/j.mex.2019.06.001

So, A. P., Vilborg, A., Bouhlal, Y., Koehler, R. T., Grimes, S. M., Pouliot, Y., et al. (2018). A robust targeted sequencing approach for low input and variable quality DNA from clinical samples. *NPJ Genomic Med.* 3, 2. doi: 10.1038/s41525-017-0041-4

Sommer, S., Courtiol, A., and Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics* 14, 542. doi: 10.1186/1471-2164-14-542

Städele, V., and Vigilant, L. (2016). Strategies for determining kinship in wild populations using genetic data. *Ecol. Evol.* 6, 6107–6120. doi: 10.1002/ece3.2346

Taylor, R. S., Jensen, E. L., Coltman, D. W., Foote, A. D., and Lamichhaney, S. (2021). Seeing the whole picture: what molecular ecology is gaining from whole genomes. *Mol. Ecol.* 30, 5917–5922. doi: 10.1111/mec.16282

Thompson, R. B., Bravington, M. V., Feutry, P., Gunasekera, R., and Grewe, P. (2020) *Close kin mark recapture for school shark in the SESSF. FRDC 2014/024.* Available at: https://www.frdc.com.au/sites/default/files/products/2014-024-DLD.pdf.

Tin, M. M. Y., Rheindt, F. E., Cros, E., and Mikheyev, A. S. (2015). Degenerate adaptor sequences for detecting PCR duplicates in reduced representation sequencing data improve genotype calling accuracy. *Mol. Ecol. Resour* 15, 329–336. doi: 10.1111/1755-0998.12314

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., et al. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* 1, e203. doi: 10.7717/peerj.203

Torkamaneh, D., Laroche, J., Bastien, M., Abed, A., and Belzile, F. (2017). Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinf.* 18, 5. doi: 10.1186/s12859-016-1431-9

Torrado, H., Carreras, C., Raventos, N., Macpherson, E., and Pascual, M. (2020). Individual-based population genomics reveal different drivers of adaptation in sympatric fish. *Sci. Rep.* 10, 12683. doi: 10.1038/s41598-020-69160-2

Trenkel, V. M., Charrier, G., Lorance, P., and Bravington M, V. (2022). Close-kin mark–recapture abundance estimation: practical insights and lessons learned. *ICES J. Mar. Sci.* 79, 413–422. doi: 10.1093/icesjms/fsac002

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001

van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., et al. (2007). Complexity reduction of polymorphic sequences (CRoPS^{TM}): a novel approach for Large-scale polymorphism discovery in complex genomes. *PloS One* 2, e1172. doi: 10.1371/journal.pone.0001172

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., et al. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* 5, 247–252. doi: 10.1038/nmeth.1185

Vendrami, D. L. J., Forcada, J., and Hoffman, J. I. (2019). Experimental validation of in silico predicted RAD locus frequencies using genomic resources and short read data from a model marine mammal. *BMC Genomics* 20, 72. doi: 10.1186/s12864-019-5440-8

von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., et al. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Sci. Rep.* 7, 10768. doi: 10.1038/s41598-017-10647-w

Wacker, S., Skaug, H. J., Forseth, T., Solem, Ø, Ulvan, E. M., Fiske, P., et al. (2021). Considering sampling bias in close-kin mark–recapture abundance estimates of Atlantic salmon. *Ecol. Evol.* 11, 3917–3932. doi: 10.1002/ece3.7279

Wang, J. (2010). Effects of genotyping errors on parentage exclusion analysis. *Mol. Ecol.* 19, 5061–5078. doi: 10.1111/j.1365-294X.2010.04865.x

Wang, J. (2012). Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics* 191, 183–194. doi: 10.1534/genetics.111.138149

Wang, S., Meyer, E., McKay, J. K., and Matz M, V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* 9, 808–810. doi: 10.1038/nmeth.2023

Wang, X., Lund, M. S., Ma, P., Janss, L., Kadarmideen, H. N., and Su, G. (2019). Improving genomic predictions by correction of genotypes from genotyping by sequencing in livestock populations. *J. Anim. Sci. Biotechnol.* 10, 8. doi: 10.1186/s40104-019-0315-z

Waples, R. S., and Feutry, P. (2022). Close-kin methods to estimate census size and effective population size. *Fish Fish* 23, 273–293. doi: 10.1111/faf.12615

Waples, R. S., Grewe, P. M., Bravington, M. W., Hillary, R., and Feutry, P. (2018). Robust estimates of a high n e / n ratio in a top marine predator, southern bluefin tuna. *Sci. Adv.* 4, eaar7759. doi: 10.1126/sciadv.aar7759

Webster, M. S., and Reichart, L. (2005). Use of microsatellites for parentage and kinship analyses in animals. *Methods Enzymol.* 222–238. 10.1016/S0076-6879(05)95014-3

Wenne, R., Bernaś, R., Kijewska, A., Poćwierz-Kotus, A., Strand, J., Petereit, C., et al. (2020). SNP genotyping reveals substructuring in weakly differentiated populations of Atlantic cod (Gadus morhua) from diverse environments in the Baltic Sea. *Sci. Rep.* 10, 9738. doi: 10.1038/s41598-020-66518-4

Zajac, G. J. M., Fritsche, L. G., Weinstock, J. S., Dagenais, S. L., Lyons, R. H., Brummett, C. M., et al. (2019). Estimation of DNA contamination and its sources in genotyped samples. *Genet. Epidemiol.* 43, 980–995. doi: 10.1002/gepi.22257

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

Glossary

Allele	Alternative form of a DNA sequence (a single base or a segment of bases) at a given genomic location; a single allele for each locus is inherited from each parent.
Allelic dropout (ADO)	It is a common phenomenon caused by a partial amplification failure of the DNA, which results in the loss of one of the alleles, causing the heterozygous individuals to appear as false homozygous.
Codominant markers	Markers for which both alleles are expressed when co-occurring in an individual, allowing the discrimination between heterozygotes and homozygotes.
Contig	(as related to genomic studies; derived from the word "contiguous") is a contiguous sequence of DNA created by assembling overlapping sequenced fragments.
Demultiplexing	Is the process by which sequencing reads are assigned to their sample of origin based on the sequence of their unique molecular tag or barcode. This step is required when multiple samples are pooled (multiplexed) before sequencing, in order to increase sample throughput and reduce costs.
Diploid	Refers to the presence of two complete sets of chromosomes in an organism's cells, with each parent contributing a chromosome to each pair. Most animals and plants are diploids except for their sex cells or gametes that are haploid.
DNA isolation	Is the process of extracting DNA from the cells of an organism, typically using a sample of blood, saliva or tissue.
DNA amplification	Any process that increases the number of copies of a specific DNA fragment. See also PCR.
Genotyping	Is the process of determining the DNA sequence, called a genotype, at polymorphic positions within the genome of an individual.
Haplotype	A set of closely linked genetic markers or DNA variations on a chromosome that tend to be inherited together.
High Throughput Sequencing (HTS)	Also known as next-generation sequencing (NGS) and massively parallel sequencing, refers to a collection of methods and technologies that can sequence thousands/millions of DNA fragments at a time. This is in contrast to older technologies that can produce a limited number of fragments.
Homozygous	Indicates two alleles on homologous chromosomes that are identical for a given locus.
Heterozygous	Indicates two alleles on homologous chromosomes that are different for a given locus.
Hybridization	The pairing of a single-stranded, labeled probe (usually DNA) to its complementary sequence.
Identity by descent (IBD)	It is a term used in genetic genealogy to describe a matching segment of DNA shared by two or more individuals that has been inherited from a common ancestor in the absence of recombination. Estimating the proportion of IBD segments is useful to determine relatedness.
Kinship coefficient	Probability that two homologous alleles drawn from each of two individuals are identical by descent (IBD), is a classic measurement of relatedness.

(Continued)

Continued

Linkage disequilibrium (LD)	Refers to the non-random association of alleles at neighboring loci that result from their close physical proximity, which makes recombination (crossing over) between them highly unlikely.
Locus	Specific physical location on the genome where a DNA sequence is located. The plural is loci. The size of the region (from a single base up to thousands of bases) depends on the context in which the term is being used.
Microsatellite	Also known as single-tandem repeats (STRs), consist of repetitive segments of DNA that present high variability in repeat number between individuals.
Mutation rate	Is the frequency of mutations in a locus or organism over time.
Non-autosomal markers	Markers located on the sex chromosomes and the mitochondrial DNA. The autosomes are the chromosomes other than the sex chromosomes.
Oligonucleotides	Short polymers of the nucleotide building blocks of nucleic acids.
Paired-end reads	Refers to reads produced by sequencing both ends of the same molecule. When the sequenced DNA fragments are shorter than two times the read length (determined by the sequencing technology), the paired reads overlap and can be merged into a longer read.
PCR	Polymerase Chain reaction. A method for amplifying a DNA based sequence using repeated cycles of replication by a heat-stable polymerase and two oligonucleotides called primers, one complementary to the (+) strand at one end of the sequence to be amplified and one complementary to the (-) strand at the other.
Polymorphism	As related to genomics, refers to the presence of two or more variant forms of a specific DNA sequence that can occur among different individuals or populations. The most common type of polymorphism involves variation at a single nucleotide (also called a single-nucleotide polymorphism, or SNP).
RAD tags	Restriction site associated DNA (RAD) markers.
Restriction endonucleases	Also called restriction enzymes are enzymes that recognize a specific DNA sequence called a restriction site, and cleave the DNA within or adjacent to that site.
Sequencing coverage	Also called sequencing depth, refers to the number of times a nucleotide is read during sequencing. The higher the depth of read coverage, the higher confidence in the resulting consensus sequence.
Single-nucleotide polymorphism (SNP)	DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) is replaced in the genome sequence. Is the most common form of variation in the genome and it is used widely to study genetic differentiation among individuals or populations.
SNP chip	Also called SNP array, is a type of DNA microarray used to detect SNP polymorphisms. It contains designed probes flaking the SNPs of interest for which the specific alleles are determined by hybridization.
Whole-genome sequencing	Also known as full genome sequencing, is the process of determining the entirety, or nearly the entirety, of the DNA sequence of an organism's genome at a single time.

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 11 July 2023
DOI 10.3389/fmars.2023.1087027

OPEN ACCESS

EDITED BY
Francisco Vargas-Albores,
National Council of Science and Technology (CONACYT), Mexico

REVIEWED BY
Einar Eg Nielsen,
Technical University of Denmark, Denmark
Tetsuya Akita,
Japan Fisheries Research and Education Agency, Japan
Aurelien Delaval,
Norwegian Institute of Marine Research (IMR), Norway

*CORRESPONDENCE
Laura Casas
lauracasas@iim.csic.es
Fran Saborido-Rey
fran@iim.csic.es

RECEIVED 01 November 2022
ACCEPTED 15 May 2023
PUBLISHED 11 July 2023

CITATION
Casas L and Saborido-Rey F (2023) A review of an emerging tool to estimate population parameters: the close-kin mark-recapture method.
Front. Mar. Sci. 10:1087027.
doi: 10.3389/fmars.2023.1087027

A review of an emerging tool to estimate population parameters: the close-kin mark-recapture method

Laura Casas* and Fran Saborido-Rey*

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

Knowing the number of individuals in a population is fundamental for the sustainable management of exploited marine resources but estimating this parameter is often extremely challenging, especially in large, highly mobile and dispersed populations. Abundance estimation traditionally relies on multiple data types that include the relationship between fishery catches and effort (Catch Per Unit Effort or CPUE), scientific research surveys and demographic models that are developed to estimate past and current stock dynamics, but uncertainty is often high. Close-kin mark-recapture (CKMR) is an alternative method for estimating abundance and other demographic parameters (e.g. population trend, survival rates, connectivity), using kinship relationships determined from genetic samples. This methodology is based on a simple concept - the larger the population the less likely to find relatives and vice versa - and was proposed two decades ago although regained considerable attention recently. Refinements in the statistical methodology and advances in high throughput sequencing methods have boosted the efficiency of genomic analysis, promising to revolutionize the field of fisheries stock assessments. In theory, it can be applied to almost any species, provided that there is sufficient information about the life-history/biology of the organism and that the populations are not so small as to be almost extinct or so large that finding relatives becomes extremely difficult. Thus, it has the potential to provide baseline data for the many exploited fish stocks that remain largely unassessed and to reduce uncertainty in those that are regularly evaluated. Here, we provide an overview of the method in the context of fisheries assessments, analyze the advances and synthetize the field studies published in the last five years. Moreover, we evaluate the readiness, viability and maturity of the method to infer demographic parameters in species spanning diverse life histories. We evaluate technical considerations and requirements for a successful application and analyze the main challenges and limitations preventing a broader implementation.

KEYWORDS

CKMR, kinship, abundance, mortality, fisheries assessments, fisheries stock benchmarking

1 Introduction

The status of marine fish stocks needs to be assessed to ensure fishing practices that exploit the populations at sustainable levels. Stock assessments are based on multiple data types that include catch data, monitoring of fishery landings, biological observers and scientific research surveys. The latter provide critical fishery-independent information of fish stocks that is subsequently utilized for estimating key population demographic parameters such as biomass, abundance, fecundity, recruitment and mortality. These parameters are not only crucial to guarantee the effective management of the stocks but also to understand their recent evolutionary history (Swain, 2011; Kindsvater et al., 2016).

Despite their critical importance, scientific research surveys used to assess fish stocks present recognized shortfalls (Maunder and Piner, 2015). These include a slow progress coupled with high economic costs and complex logistics, that often results in sparse data with limited coverage in space and time (Stamatopoulos, 2002, Pennino et al., 2016). Moreover, the data obtained is not always accurate and the elevated costs prevent the assessment of many exploited stocks that remain data limited or, directly, unassessed, representing a major conservation concern.

The emergence of novel genomic techniques together with the plummeting sequencing costs have provided novel means to improve the cost-efficiency of fisheries research surveys, reduce bias and uncertainty of fish stocks assessments and expand the range of assessed species.

In this review, we focus particularly on the method of close-kin mark-recapture (CKMR), based on an idea founded nearly two decades ago (Nielsen et al., 2001; Skaug, 2001) that resurfaced recently as a promising tool to estimate key demographic parameters, through genotyping and the identification of close-kin using modern genomic methods (Bravington et al., 2016b). CKMR can be used to estimate abundance or biomass among other demographic parameters, which are especially challenging to infer in many marine species, characterized by large, highly mobile and dispersed populations.

This methodology has gathered considerable attention as it has important advantages over classical mark-recapture (MR), which requires capture, physical marking and recapture of individuals, who must remain alive, in multiple sampling events. In contrast, CKMR can be applied to samples collected during a unique sampling event as well as to dead specimens (Wacker et al., 2021). Besides, CKMR overcomes many of the challenges inherent to traditional MR modelling, as it does not suffer from the effects that can occur after initial capture, such as trap shyness or tag-loss (Marcy-Quay et al., 2020). Advances in high-throughput sequencing technologies, enable today the fast and accurate genotyping of large numbers of samples across many loci to identify close-kin with precision.

Despite its potential, the successful application of the method is yet very restricted, as it has been used to assess only a handful of species (Delaval et al., 2022). The broad applicability of CKMR to fish populations and, therefore, its usefulness for fisheries assessments has yet to be demonstrated. Here, we provide an overview of the method, together with guidance for its application, which includes crucial information that is scattered across grey-literature (mostly non-peer reviewed reports, including technical, workshop and project reports). Bravington et al. (2016b) lays the foundation of the method detailing the mathematical and statistical framework, assumptions and conditions needed for its application whereas Waples and Feutry (2022) compare the methodology with kin-based methods to estimate effective population size using analytical models. In contrast, here, we provide a deliberately oversimplified description of the statistical framework and use an accessible terminology to reach potential users alike – fisheries scientist and managers with no expertise in genomics and geneticists unfamiliar with modelling, respectively. We review the methodology in the context of fisheries assessments focusing on practical aspects and technical considerations. We analyze all CKMR studies applied to fish species published to date, discuss the readiness, viability, maturity and limitations of the method and issue recommendations for its uptake in stock assessments.

2 The basics of close-kin mark-recapture

2.1 What is CKMR?

CKMR is a method for estimating abundance and other demographic parameters (e.g. survival rates, fecundity, selectivity) from kinship relationships determined from genetic samples. It uses modern genetics to identify close relatives amongst large sample sizes of fish, and then makes demographic inferences about the adult stock from the number and pattern of pairs found (Bravington et al., 2015; Bravington et al., 2016a). The likelihood of a specimen being a close relative of the rest of the individuals can be calculated by comparing the genetic make-up of fish from a population and accounting for their life-history information (e.g. year of birth). CKMR offers a direct methodology to assess wild fish stocks that is fishery-independent of commercial catch per unit effort (CPUE) and total catch data so crucially, does not suffer from errors in catch reporting and other potential sources of bias associated with more traditional fishery-dependent data. This methodology can provide estimates of key population parameters from relatively short studies (over a few years, e.g. Wacker et al., 2021) and therefore, has the potential to be widely deployed for routine assessments of fisheries resources (Rodriguez-Ezpeleta et al., 2020).

It has been nearly two decades since the use of individual genotypes as genetic tags was first proposed in the literature (Nielsen et al., 2001; Skaug, 2001). At the time, genetic data were in short supply, and it was necessary to make allowances to handle the substantial uncertainty regarding inferred kin relationships. Today, advances in sequencing technologies, especially their increasing throughput, have made CKMR projects feasible in a number of different contexts, as several degrees of kinship can be determined with enough accuracy.

CKMR is based on the principles of DNA-based kinship as it uses the unique DNA profiles of the individuals to determine if they are related or not. An specimen is considered “tagged” by its presence in a sample, and “recaptured” by the occurrence of one or more close relatives in the sample (Bravington et al., 2016b; Seber

and Schofield, 2019). Molecular tags pass automatically from generation to generation, as a consequence of inheritance, and the probability of detecting relatives is directly linked to population size. Naturally, finding relatives is more likely to occur in smaller populations, so the number of "recaptures" can be used to infer abundance (Bravington et al., 2016b).

The first applications of CKMR were based solely on the identification of Parent-Offspring Pairs (POPs) (Figure 1) (e.g. (Bravington et al., 2016a, southern bluefin tuna) although novel high throughput sequencing methods allow also inferring more distant kinship relationship accurately and thus, second-order (e.g. half-sibling) relatives have also been incorporated in posterior CKMR studies. This extended kinship methodology is especially relevant in the context of fisheries management for two reasons. First, it widens the scope of applicability of CKMR studies enabling the study of species characterized by ontogenetic shifts in habitat use without implying complex sampling designs. Many fish species occupy coastal shallow areas as juveniles and move to deeper, more diverse habitats as adults (Cheminée et al., 2021). Collecting both components complicates sampling logistics and can be extremely challenging for large, solitary species like sharks, where adults are mostly found dispersed across huge areas of deep offshore open ocean waters (Ramírez-Macías et al., 2017; Hoffmayer et al., 2021). Second, the use of second-order kinship allows the estimation of additional population parameters, such adult mortality, which cannot be inferred using solely first-order kinship (Davies et al., 2015; Maunder et al., 2021). Estimates of kinship derived from genetic data, together with life-history traits such as maturation schedules and reproductive output (fecundity, reproductive behaviour, egg/larval quality, offspring survival rate), along with stock structure information (age, growth and length-weight relationships, all potentially estimated by sex), are modelled to obtain abundance and other demographic parameters. These can be integrated into stock assessment models to assist management procedures.

2.2 Theory and assumptions behind close-kin abundance estimation

For an in-depth view of the statistical underpinnings of CKMR, the interested reader is referred to (Bravington et al., 2016a; Bravington et al., 2016b). Here, we aim to provide a simplified overview into the basic theoretical foundation behind close-kin abundance estimation. In its simplest version, the method is based on POPs and rests on two notions: i) each juvenile has two parents and ii) genomic information allows establishing accurately if two fish constitute a POP. Thus, each juvenile is an offspring that genetically "tags" its two adult parents among the adult population of size N_{adult}. The probability of a sampled juvenile being the offspring of a randomly sampled adult is $2/N_{adult.}$ It is important to note here that N_{adult} is the number of adults alive when the juveniles were born, and thus, the method serves to estimate adult abundance retroactively. This parameter, N_{adult}, can be calculated based on the genotypes of sampled adults and juveniles (m_A and m_J), by examining all possible pairwise comparisons, and counting the number of POPs found (P), using the following formula (Bravington and Grewe, 2007; Bravington et al., 2016a):

$$\hat{N}_{adult} = m_J m_A \times 2/P \quad (1)$$

Therefore, the observed total of POPs provides a natural estimate of absolute abundance of the adult population that is directly analogous to a Lincoln–Petersen abundance estimate in standard mark-release recapture (Bravington et al., 2016a).

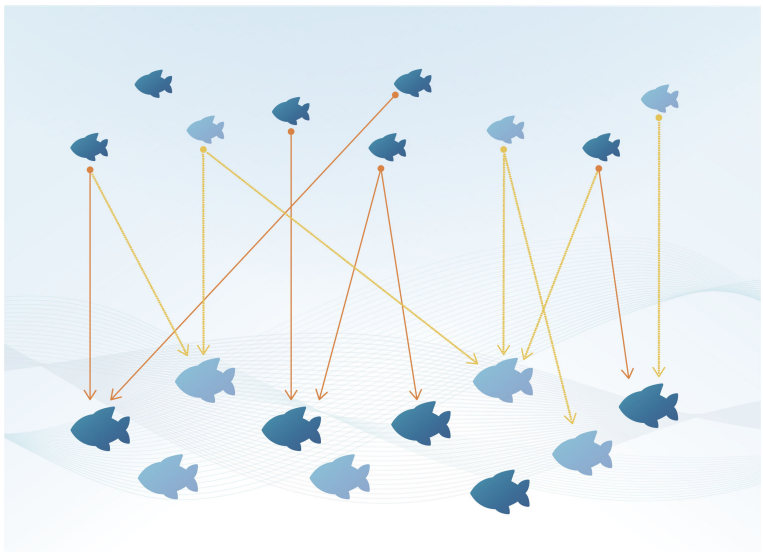


FIGURE 1
Illustration of the CKMR principle; adults (big fish) and juveniles (small fish) are sampled (dark blue) from the total population (dark and light blue). Each juvenile "tags" two fish: each of its parents (orange and yellow lines) in the adult population; but only sampled fish provide kinship information – POPs (orange 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), using the formula: (2x5x6)/6 = 10 (see section 2.2); Figure redrawn from (Bravington et al., 2016b).

The derivation of an unbiased estimate of $\hat{N}_{adult}$ requires the random sampling of adults and collecting the juveniles independently of adults (e.g. the method cannot be applied to adults and offspring sampled concurrently when parental care is ongoing) (Bravington and Grewe, 2007). Failing to comply with these two requirements would increment artificially the probability of finding a POPs, resulting in a negatively biased estimation of adult abundance.

The alternative version of CKMR based on half-siblings conforms to three principles: i) given any pair of juveniles, each of them has one mother and one father; ii) assuming a balanced adult sex ratio, the chance that they share the same mother is 1 over the number of female adults, i.e. $2/N_{adult}$ and iii) the chance that they have the same father is $2/N_{adult.}$ Thus, for a given pair of juveniles, the overall chance that they constitute an HSP is $4/N_{adult}$. Making all m_J pairwise comparisons, gives $m_J^2/2$ "non-double-counted" comparisons, and can be used, with the number of half-siblings detected (H) to estimate N_{adult} using the following equation (Bravington, 2014):

$$\hat{N}_{adult} = m_J^2 \times 2/H \quad (2)$$

The observed number of HSPs provides an estimate that is independent of the POP-based estimation and reflects the effective breeding population abundance, or the effective number of breeders in one year, N_b (Davies et al., 2020; Waples and Feutry, 2022). If a substantial proportion of the adults do not contribute to produce offspring, they are invisible to a sibling-based approach. Skip-spawning, maternal effects, variation in reproductive potential and other mechanisms are some of the causes producing different reproductive success among individuals (Lowerre-Barbieri, 2009). Thus, the use of POPs (Eq. 1) provides the total number of adults in the population, an essential parameter in fisheries stock assessment. The use of HSPs (Eq. 2) estimates, instead, the number of adults that have effectively contributed to produce offspring, a parameter that is not currently used in stock assessment and fisheries management. If all or most of the adults have equal reproductive success, and there is an even sex ratio in the population, both estimators yield similar results. However, most of the exploited teleost species show very large fecundity and variation in reproductive success (Wright and Trippel, 2009). Nonetheless, for simplicity, we hereafter use the term "adult abundance" for both approaches (POP and HSP-based)). The differences and commonalities between them are reviewed by Waples and Feutry (2022).

2.3 Adjusting the simple estimators to the specific biology of the species under study

Both simple estimators provide accurate inferences if strong assumptions are made, including the absence of temporal variation in life history traits (fecundity, survival rate, migration), the random sampling of a balanced number of adults and/or juveniles showing a homogeneous distribution during a unique sampling event, and other standard population genetics assumptions, i.e. Hardy-Weinberg equilibrium, linkage equilibrium and random mating. However, such assumptions cannot be made for the vast majority of the fish populations as they display variations in maturation, fecundity, spawning dynamics, survival rate, migration and other factors that can affect the probability of finding close-kin matches, potentially causing a strong bias of the simple estimators (Bravington and Grewe, 2007; Bravington et al., 2016b). Below, we review the most important parameters that can affect the probability and therefore, require attention when applying CKMR.

2.3.1 Population substructure

In the presence of population substructure, the basic method remains unbiased if sampling is proportional to abundance across either the sub-population of adults or the sub-population of juveniles. Planning such sampling strategy requires previous knowledge, however, genetic data generated for CKMR can be used to detect substructure and determine the adequacy of sampling *a posteriori*. Spatial patterns in close-kin distribution can inform about population subdivision and even an exploratory rough sampling design should be enough to reveal substructuring when is strong enough to affect CKMR estimates (Wang, 2014; Feutry et al., 2017; Conn et al., 2020; Waples and Feutry, 2022).

2.3.2 Sex-specific life history characteristics

Many species show biased sex-ratios, some across the whole adult population while others display a balanced number of males and females that becomes skewed during certain periods of the year (e.g. during breeding or feeding season). This situation would, generally, require independent estimates of the adult male and female abundances ($\hat{N}_{adult-male}$ and $\hat{N}_{adult-female}$), using the male and female adult samples, respectively. Nonetheless, if adults have unequal sex ratios but the sampling holds the same sex bias, the POP-based estimator would be unaffected (Bravington and Grewe, 2007). Demographic modelling should consider the existence of strong sexual dimorphism in growth rates, fecundity-at-age, hermaphroditism, and different catchability between sexes (gear selectivity pattern) among other particularities (Trenkel et al., 2022).

2.3.3 Sampling delays

The abundance estimate derived from the CKMR method is retrospective, i.e., $\hat{N}_{adult}$ is the number of adults that were alive when the juveniles were born, rather than when the juveniles were sampled (Bravington and Grewe, 2007; Bravington et al., 2012). Thus, comparing each offspring only against their "parental-cohort-group" would require knowing the age of all the fish and that the fish would mature at the same length/age. In realistic CKMR settings, this is almost never the case. Most studies are based on samples collected years after the birth of juveniles, and thus, comparisons between juvenile and adults to estimate kinship necessarily involve a high proportion of "impostor" adults that could not have been parents as they were immature at the time (Figure 2) (Bravington et al., 2014). Ignoring these time lags and the differential probability of an adult being the parent of a juvenile in each comparison would lead to a high bias of the

simple estimator for POPs. Sibling-based analysis also require to assign accurately the cohort year of each specimen as not all types of siblings are useful in CKMR studies (see section 3.2.2). This problem can be tackled by recording the age of all the individuals included in the analysis (Bravington and Grewe, 2007). The preferred source is always a direct and precise age estimation based on hard structures such as otoliths, scales or vertebrae, among others. Nonetheless, this is not always possible as these methods often require the sacrifice of the specimens and are not always reliable (e.g. otolith-based ageing in European hake (de Pontual et al., 2006)). In the absence of direct age determination, a length cut-off can be used to assign the birth year through growth curves but this alternative requires knowledge on growth rates and associated uncertainty in the age estimate. For species displaying considerable variability in length-at-age, this uncertainty needs to be accounted for in the CKMR model and might even prevent the use of a considerable percentage of informative kinship pairs in a sample (Rodriguez-Ezpeleta et al., 2020; Trenkel et al., 2022). Additionally, other stock parameters such as rates of adult mortality, age at maturity and knowledge of whether the reproductive output is dependent on size need to be also considered as they can heavily affect the probability of finding kinship matches of both, POPs and HSPs.

2.3.4 Multi-year sampling

A similar problem arises if the experimental design implies multi-year samplings, since for a given cohort of juveniles, their potential parents will be sampled across several years, rather than in one year (Bravington et al., 2014). Therefore, the sample will contain multiple cohorts of juveniles. Moreover, for many species there might be a general delay of several years before the potential parents of a given juvenile cohort are sampled, during which some of the parents will die. Again, it is essential to consider both, the birth-year of the juveniles and the sampling-year of the adults to restrict the pairwise comparisons to those individuals that were mature in the birth-year. Also, when comparing offspring from different cohorts it is paramount to account these delays as they can greatly affect the probability of kinship matches (Waples and Feutry, 2022). Although certain species can benefit from the use of pre-existing archived DNA samples, very old specimens may not be relevant for the estimation of the current stock parameters as the cohorts that produced them will have died out. Since any CKMR analysis reflects the abundance of the adult stock in the birth year of the collected specimens, in case of very long intervening periods or of late-maturing species, the stock abundance might have change considerably (Rodriguez-Ezpeleta et al., 2020).

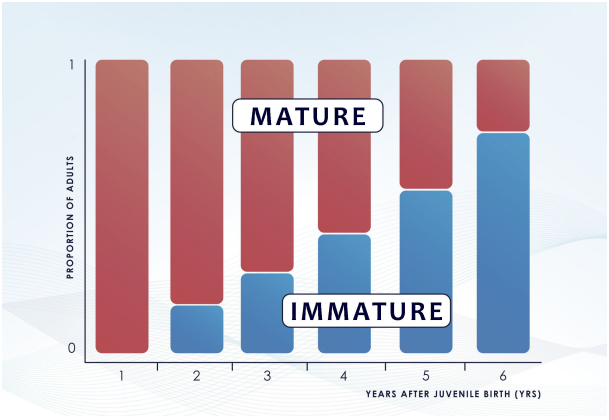


FIGURE 2
Effect of the incoming recruitment on the original parent-cohort group. The time gap between juvenile birth and sampling determines the number of immature specimens that enters the "adult" group but could not have been parents at the time the juveniles were born as they were immature. Figure redrawn from (Bravington et al., 2014).

2.3.5 Reproductive variability

The basic method can also produce skewed estimates if variation in reproductive potential is not considered. Most fish species show a gradual maturation (not all fish mature at the same length or age) and an increase in fecundity and offspring quality with age/size (Saborido-Rey and Kjesbu, 2005). Accordingly, large/old fish produce more offspring with increased viability, i.e. more "tags" per capita. If adult sampling is strongly selective towards large specimens, each tag would be more likely to be "recaptured". This can lead to an underestimation of N_{adult} as each comparison would be more likely to yield a POP than would a comparison with a randomly chosen adult (Figure 3, right panel). Likewise, there are temporal variations in maturation schedules and fecundity that are also important to consider in CKMR analyses. If only POPs are analyzed, fluctuations in reproductive output among years due to variation in reproductive potential (which in teleosts it is well known to occur), do not bias the simple estimator $\hat{N}_{adult}$ but affect its precision (Figure 3) (Bravington, 2014). In contrast, random within year fluctuations (due to e.g. sweepstake reproduction) would produce large numbers of within-year full-sibs and half-sibs that, if accounted for, would affect independence of samples leading to overdispersion in the close-kin data (Maunder et al., 2021). Thus, variation in the reproductive potential is even more relevant in HSP-based studies as it could directly affect the number of pairs found within each sampled cohort of juveniles, requiring the exclusion of within-cohort comparisons (see section 3.2.2 for further information).

3 Planning a CKMR study

3.1 Considerations for experimental design and sampling

Before embarking on a CKMR project, a careful evaluation of the sampling, sequencing and modelling needs, together with their associated costs is indispensable to assess the feasibility of the method for our case study. Proceeding without a clear understanding of requirements is likely to lead to either wrong conclusions or unsuccessful studies (Rodriguez-Ezpeleta et al., 2020). An essential requirement for the successful implementation of CKMR is prior knowledge on the biology of the species. Although CKMR can be applied to a variety of species and life histories, its actual

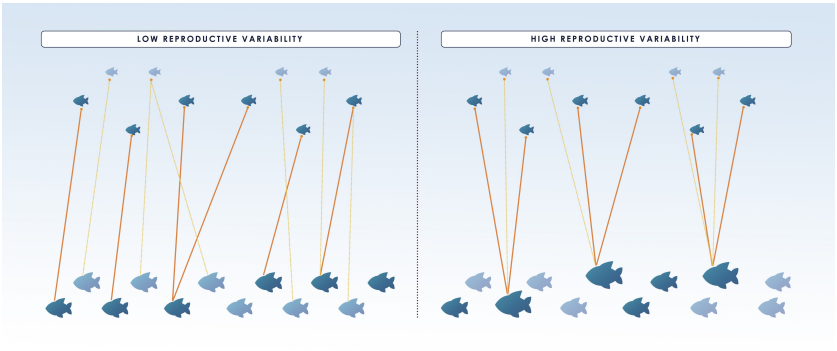


FIGURE 3
Illustration displaying the effect of reproductive variability (high variability on the right and low on the left) on close-kin abundance estimate and its coefficient of variation (*CV*). Small fish on the top of the figure correspond to juveniles whereas adults are the bigger fish below and only the dark colored ones from both groups are sampled. The same number of matches (indicated by the orange lines between dark colored adults and juveniles) is found in both situations and hence they produce the same number of POPs. Nonetheless, for species with high reproductive variability, the matches involve fewer adults and thus, a low precision of the estimate $\hat{N}_{adult}$ (i.e. larger *CV*), as it is disproportionately affected by the number of sampled "super-parents". In the figure, 6 out of 10 and 6 out of 12 juveniles and adults, respectively, are sampled in both situations, and 6 POPs are found, resulting in the same $\hat{N}_{adult}$ =6*6*(2/6)= 12. However, if bigger specimens are more fecund and also more likely to be caught as in the right panel, and only the large individuals are sampled, $\hat{N}_{adult}$ will be underestimated, unless selectivity is accounted for. Figure redrawn from (Bravington et al., 2014).

implementation in terms of sampling design, data requirements and analysis, requires tailored refinements.

Particularities related to the biology of species require attention as they can potentially lead to biased CKMR estimates, as introduced in the previous section. For example, the study of sequentially hermaphroditic species (e.g. Asian seabass or groupers (Bayona-Vásquez et al., 2019)), species showing unbalanced sex ratios due to geographical segregation of the sexes (e.g. shortfin mako (Mucientes et al., 2009)), or due to ontogenetic shifts (e.g. common octopus (Alonso-Fernández et al., 2017)), introduces important considerations in terms of experimental design and might even prevent the application of the method. In fact, the life-history of some aquatic organisms make them ill-suited to CKMR. Facultative parthenogenesis (i.e. the occurrence of asexual reproduction in otherwise sexually producing species) prevents the application of the method. This type of reproduction has been described in at least 20 species of fish, all freshwater and brackish water species not targeted by fisheries (e.g. spined loaches *Cobitis* spp., Amazon molly *Poecilia formosa*), but also in elasmobranchs. Semelparous species (breed-once-then-die) that cannot be sampled as adults, including octopus, squid and eels, prevent the analysis of CKMR as they only allow the sampling of a single cohort. Moreover, species with very long lifespans that cannot be sampled young, such as the orange roughy that lives over 100 years (Andrews et al., 2009) are unlikely candidates for CKMR because estimates are back-dated to juvenile birth and would have a high uncertainty.

Biological knowledge about population or stock structure, habitat use and patterns of social structure (random or nonrandom association of individuals) is also important for sampling design. This is because the estimation of demographic parameters using CKMR requires the assumption that sampling is random with respect to kin, and this may not be satisfied if these aspect are ignored (Carroll et al., 2018). Thus, sampling should consider the existence of spatio-temporal differences in the distribution due to ontogeny, seasonal patterns, or regime shifts among other causes.

Collection of both, adults and juveniles, is highly recommended as the combination of both types of close-kin analysis is more powerful than either alone (Bravington, 2014). POPs and HSPs provide independent estimates of adult abundance, more close-kin pairs per sample and also allow the estimation of structural parameters than cannot be calculated from either directly, such as age/size-specific fecundity, selectivity, survival and mortality (see section 3.3. for further details) (Bravington, 2014; Rodriguez-Ezpeleta et al., 2020). Nonetheless, both components are not always essential, as sampling of just one might be adequate for species with fecundity that does not change with age (e.g. sharks, Hillary et al., 2018).

In principle, the whole spatial range of the stock of interest should be sampled, to ensure enough coverage to uncover potential spatial heterogeneity (Maunder et al., 2021). Nonetheless, it has been shown that low to moderate bias in spatial sampling does not greatly affect CKMR estimates, provided that sample sizes are large enough (Conn et al., 2020; Trenkel et al., 2022). However, the demographic composition of the sample can affect the precision of the CKMR estimates and needs to include a range of birth years (cohorts), especially for medium and long-lived species (Trenkel et al., 2022). For parents, the age at sampling informs on their age at maturation while for juveniles, the age at sampling informs on their year of birth. As specimens grow older, both reference years are further in the past and the associated uncertainty in the estimates of maturity rate and age increases significantly (Trenkel et al., 2022). Since CKMR models are built around the conception of adult population size at time of juvenile birth, obtaining reliable information on the age of sampled individuals is vital for the success of CKMR studies. Likewise, the selectivity of sampling gear is an important factor that may bias the age composition of the sample and has to be carefully considered.

Summarizing, studies that collect sufficient auxiliary data of sampled individuals and target species for which a deep biological knowledge is available (including life-history traits, reproductive biology, structure and habitat use), lead to more robust CKMR models.

Nonetheless, in most study cases, at least some of the biological parameters are unknown, requiring inevitably some assumptions. Natural mortality, for instance, is difficult to estimate and often not known in fish, but, for many species, it may be a valid assumption that it has little variation after maturity. Constant mortality rates over the period of juvenile birthdates can be assumed in these species, when several cohorts are included in a CKMR study, if collection of auxiliary data is not possible. Predicting other parameters such as the relationship between age, fecundity and catchability, for instance, is more complicated in the absence of direct evidence. In such cases, it is essential to make biologically reasonable assumptions according to the species under study. Although any assumption necessarily entails some level of undefined uncertainty, it can be reduced, or at least assessed, through an adequate sampling design.

Besides the importance of the particularities of the species and the amount biological information available, to achieve a representative and suitable CKMR sampling design, another crucial consideration to assess the feasibility of any CKMR project, is the sample size needed to produce precise and accurate parameter estimations. CKMR estimates rely on finding reasonable rates of "recaptures" or kin pairs. Assuming the collection of an equal number of juveniles and adults $m_J=m_A$, is easily derived from Eq. 1 that the sample size requirements scale with $\sqrt{N_{adult}}$. While bigger populations demand bigger sample sizes in absolute terms, the sample size relative to abundance is actually lower (Bravington et al., 2016a). Nonetheless, most fishes are characterized by large populations of mobile and dispersed specimens, and, therefore, still require extensive efforts to obtain sufficient recaptures. This is one of the greatest challenges for the application of CKMR to exploited teleost populations. The use of early developmental stage individuals (larvae) might have the potential to alleviate this problem to a certain extent. However, caution is needed since high levels of sibship among larval samples, which could be originated by the formation of kin-aggregations or by reproductive hyperallometry (i.e., when fecundity increases disproportionately to mass and only a few large individuals capitalize reproduction), could bias CKMR estimations (Selwyn et al., 2016; Marshall et al., 2021; Barry et al., 2022). McDowell et al. (2022) tested this hypothesis in Atlantic bluefin tuna and found that the levels of sibship detected were not enough to cause severe problems for CKMR based on POPs. Although this might not be true for other fish species, it might be worth exploring this alternative source to increment sample collection. Moreover, the use of larvae provides up-to-date estimates of CKMR abundance as these are intrinsically back-dated to the birth-years of the juvenile samples.

Given some *a priori* notion of N_{adult}, the formula $10^* \sqrt{N_{adult}}$ provides a crude approximation of the sample size needed in POP-based studies to obtain an accurate parameter estimation (≈15% coefficient of variation), assuming a single sampling event and a balanced number of juveniles and adults (Bravington et al., 2016b). Nonetheless, the authors highlight that this multiplier 10 is not universal and remark that a serious approximation should consider, at least, the expected number of comparisons and the likely probabilities of kinship to determine the number of samples required to obtain a reasonably precise estimation of abundance (see (ICES, 2016; ICES, 2017) for an example). This level of precision is usually expressed via the coefficient of variation (*CV*) of the abundance estimate as follows, where m is the combined sample size of adults and juveniles (distributed equally 50:50 for optimality) (Bravington and Grewe, 2007):

$$CV\left(\hat{N}_{adult}\right) \approx \frac{\sqrt{2}}{m}\sqrt{N_{adult}}$$

For commercially targeted species, the output of stock assessments based on CPUE data can be used as a starting point for the calculation and should aim at a small coefficient of variation ($CV \approx 0.15$) (Bravington et al., 2016a).

The R package CKMRpop offers the possibility of simulating pedigrees within age-structured populations under different life-history scenarios to assess the feasibility and potential accuracy of a CKMR approach (Anderson, 2022). Simulations can be performed *in silico* for a variety of sampling schemes targeting the collection of a distinct fraction of individuals in the population, and the software reports the expected number of kinship pairs and their distribution. Previous knowledge of basic life-history parameters of the species of interest are useful to narrow down the possible scenarios of the simulations.

3.2 Genetic identification of close-kin pairs

If the feasibility assessment reveals the adequacy of the method to the case study of interest, the next decision involves the genetic identification of close-kin pairs. Two key aspects that concern the type of both, marker and kinship to be applied in the analysis, need to be decided.

3.2.1 Selecting the markers for kin pair finding

Two types of markers have been used in CKMR studies to date, microsatellites and SNPs, each with its own characteristics that are out of the scope of this revision (but see (Flanagan and Jones, 2019) for an overview). Considering the resolving power of both markers in a kinship context, together with the large number of techniques that can be used to scan across genomes in search of polymorphisms at an affordable price, SNPs would be the natural choice for any project initiating today. Nonetheless, this has to be evaluated on a case-by-case basis depending on the prior knowledge of markers available for the organism of interest, the researcher´s skills, the laboratory infrastructure available and the project´s budget.

Independently of the marker of choice, it is essential to select a marker set that allows an accurate estimation of kinship pairs, avoiding false genetic associations. False-positives arise when an unrelated pair shares alleles by chance. They are especially problematic in CKMR studies since these generally involve a very large number of comparisons to detect a small number of kin pairs. The inclusion of spurious kin (false-positives) can, therefore, have

large effects on the estimates (Bravington et al., 2016a; Bravington et al., 2017; Rodriguez-Ezpeleta et al., 2020). The probability of false-positives can be assessed in advance from the allele frequencies, and this step is essential in determining whether enough loci are being used. Accurate kinship analyses require large numbers of markers and is infrequent to score successfully all loci across all samples. In comparisons involving less loci, the false-positive probability might increase substantially. False-negatives, on the contrary, are the result of true kinship pairs *appearing* not to share one or more alleles that are actually common. This could rarely arise through mutation or more often through scoring errors, whereby the true alleles are incorrectly recorded. Scoring errors can be kept at minimum controlling the quality of the DNA, a careful selection of loci and robust protocols.

The expected number of false-positives generated by both type of errors will be negligible compared to the number of true positives if enough number of loci are used, the loci meet the quality requirements and fulfill regular population genetic assumptions (no linkage, etc.).

3.2.2 Selecting the type of kinship

Similarly, studies of CKMR published so far have analyzed either first or second-degree relationships, or a combination of both. The simplest form of CKMR is based on the identification of first-order kin relations (POPs). Every diploid animal has two alleles at each locus, one inherited from each parent and thus, a POP must share, at least, one allele at every locus (Städele and Vigilant, 2016). This allele sharing pattern can occur, by chance, between non-POP individuals but this probability can be decreased by increasing the number and variability of loci examined (Bravington et al., 2012). There are several analytical approaches that can be used to assign parentage and they can be grouped into three broad categories: exclusion-based, relatedness-based and likelihood-based methods (Huisman, 2017). Exclusion methods follow the principle of treating a pair as a POP if, and only if, the two animals share, at least, one allele at every loci, although normally allow some genotyping errors to avoid the exclusion of "true" POPs. Methods in the second category, estimate pairwise relatedness or kinship coefficients while the last group of methods assess likelihood ratios, being the most powerful of all but computationally more demanding (Städele and Vigilant, 2016).

The genetic identification of Full Sibling Pairs (FSPs, share both parents) and Half Sibling Pairs (HSPs, share only one parent) is more complicated. Detecting genetically more distant relationships is more demanding than detecting POPs (Maunder et al., 2021). FSPs share, on average, the same amount of DNA as a POP (50%), but with a different inheritance pattern, while HSPs, share, on average, only 25% of their DNA, although due to the random recombination process, this percentage varies among half-sibs (Figure 4) (Städele and Vigilant, 2016).

Finding half-siblings requires many more SNPs and as the number of locus increases so does the possibility of linkage among them (coinheritance). However, in theory, having a genome assembly or a linkage map could provide information about the physical linkages between genetic markers (i.e. whether they are on the same chromosome and how close together they are, allowing assessment of likely linkage disequilibrium). Sets of markers that are close together on the genome are more likely to be shared by siblings than by pairs with more distant antecedents. Thus, information on linkage disequilibrium may allow for better discrimination between HSPs and more distant relatives but also for differentiation from other second-degree relatives (Bradford et al.,

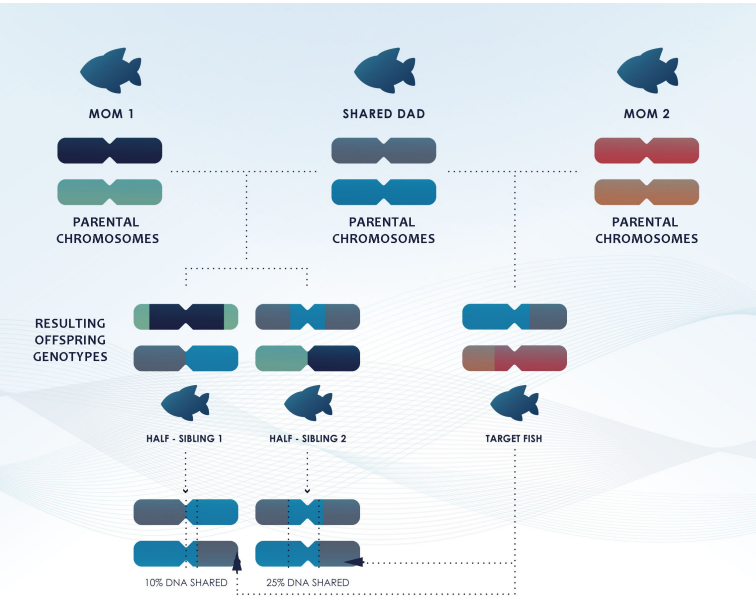


FIGURE 4
Simplified diagram illustrating that while half-siblings share, on average, 25% of their DNA, the recombination process shuffles de DNA differently for each offspring and can result in higher or lower proportions of DNA shared among siblings and half-siblings.

2018). Half-siblings share an identical percentage of the genome, 25%, on average, than avuncular (e.g., aunt-niece) and grandparent-grandchild relationships, but their patterns of inheritance of genome segments overlap little (Qiao et al., 2021). Thus, it should be evaluated whether considering HSPs might introduce kinship assignment errors, depending on the samples collected, the auxiliary data available for them (mainly age), and the biology of the species.

Teleosts exhibit a wide array of reproductive strategies but most of the exploited species are largely characterized by large numbers of tiny offspring with a very variable survival (Murua and Saborido-Rey, 2003; Anderson and Gillooly, 2021). Strong variations in larval survival from offspring-to-offspring or from "litter-to-litter", to follow Bravington et al., 2016b notation, cause a systematic over-representation of within-cohort half siblings due to the "lucky litter effect" (large survival of individuals from the same spawning/birthing event) (Bravington et al., 2016a; Bravington, 2019; Rodriguez-Ezpeleta et al., 2020; Maunder et al., 2021). This can affect independence of samples and requires filtering out within-cohort pairs as it could lead to a negative bias in CKMR estimates (Hillary et al., 2018; Maunder et al., 2021). Elasmobranch species display a lower fecundity and might, accordingly, be less affected by these within-cohort effects. However, CKMR-based estimations in these animals also requires the exclusion of same-cohort half-siblings (Hillary et al., 2018, for further information see appendix there).

Moreover, as most teleost species are promiscuous, it is extremely rare to find full-sibs across cohorts in random-mating populations with millions of adults and thus, FSP are normally not considered (Davies et al., 2020). This might not be true for other species of potential interest in fisheries, such as rays and sharks, as females of certain species have the ability to store sperm and thus, commonly produce cross-cohort FSPs (Trenkel et al., 2022). Still, this type of sibship should not be considered for the analysis as only cross-cohort half-sibling comparisons are suitable for estimating abundance (Hillary et al., 2018; Waples and Feutry, 2022). However, they might be useful to recognize sweepstake reproductive events which occur in highly fecund populations when only a small number of parents contribute to the next cohort as most adults do not successfully reproduce (Christie, 2010; Vendrami et al., 2021). If the sibship incidence is very high, it might be an indication of an infeasible close-kin project (Bravington et al., 2014). In studies where cohorts are easily distinguishable (i.e. their size distribution barely overlaps), FSPs can also be used to compare known-same-cohort-individuals to study variability in length-at-age in the absence of age data (Bravington et al., 2019).

Analyses using HSPs are based on an offspring-centric view of relatedness that calculates the probability of two randomly chosen juveniles in a sample having the same parent. Although the HSPs themselves only involve juvenile fish, HSP-based CKMR still provides information only about adults (Bravington, 2014). HSPs can be the result of sharing either the mother (maternal HSP) or the father (paternal HSP) and discrimination among both can be achieved by analyzing the mitochondrial DNA (mtDNA), always inherited from the mother, of identified HSPs. The comparison of maternal and paternal HSP provides insights into differences in how fecundity varies with age between males and females, and on the sex ratio of adults (Davies et al., 2018; Thompson et al., 2020).

For some species such as sharks, characterized by a significantly different reproductive biology compared to teleosts with much lower fecundity, and often little lifetime change in fecundity after maturity, it is feasible to base CKMR analysis in HSPs alone. However, for teleosts is almost always necessary to include POPs as they enable disentangling the effects of increasing fecundity in adult life (Davies et al., 2015).

Including both types of kinship relationships, POPs and HSPs, is advisable whenever possible, as is more powerful than either one alone (Bravington et al., 2015). Additionally, provides more information on the population as it may allow estimating additional (adult) parameters (see next section). On the other hand, considering two kinship categories instead of one increases the probability of finding kin pairs. This results in an improvement of the statistical power, addressing one of the main limitations on *CV* (coefficient of variation) of CKMR estimates, leads to more robust CKMR modelling and lowers sample size requirements (Bravington et al., 2015).

The use of more distant kin relationships (half-first cousins or great uncles, for instance) would decrease further the demanding requirements of CKMR, in terms of number of samples. In the near future, it might become possible to assign reliably these more distant kin pairs, allowing their use in a CKMR framework, as long as their frequency of occurrence can be also predicted (Anderson, 2022).

3.3 Population parameters that can be estimated via CKMR

CKMR provides a fishery-independent tool for monitoring of adult or spawning stock biomass and key biological parameters of adult population dynamics relevant for the management of exploited populations. The experimental design, biology of the species and the ancillary data collected for the specimens (length, age, etc) determines the population parameters that can be estimated using this tool.

3.3.1 Abundance

Reliable estimations of population abundance are paramount for assessing the status and trends of exploited fish populations. The CKMR methodology provides estimates of recent absolute adult abundances (N_{adult}) based on kinship relationships (POPs and/or HSPs). POPs allow estimation of total adult abundance irrespective of an individual's contribution to reproductive output while HSPs permit inference of the number of breeding adults (Bradford et al., 2018). Nonetheless, the incidence of siblings is also a widely-used method to estimate the unrelated genetic concept of effective population size (N_e) (Wang, 2009; Waples et al., 2018). Although both are underpinned by genetics, CKMR is based on mark-recapture principles rather than population genetics theory (Bravington and Grewe, 2007). Therefore, it is essential to identify conditions under which CKMR methods based on siblings estimate N_{adult}, when they estimate N_e, effective number

of breeders per year (N_b) and when they estimate something else altogether. A key factor is whether the siblings are from the same or different cohorts. A good overview is provided by (Waples and Feutry, 2022).

The CKMR approach is also capable of detecting temporal trends in adult abundance when the time span of samples covers a sufficiently long period, which duration depends on the lifespan of the species (Bruce et al., 2018) and/or sufficient number of offspring cohorts are sampled.

3.3.2 Adult mortality rates

Mortality rates of fishes are also crucial inputs for stock assessments as they are directly related to the sustainable yield of a fishery. The estimation of adult mortality rates cannot rely alone on annual age and length compositions because of statistical confounding with fishery selectivity (Bravington et al., 2016a). CKMR studies based on POPs can provide additional relevant data about mortality as the average interval between juvenile birth and adult 'recapture' is negatively correlated with the adult mortality. Nonetheless, estimating adult mortality rate from POPs alone is not possible without auxiliary data on female fecundity, if fecundity varies through adult´s life, because the same reproductive output can be produced by a small number of high fecundity or a larger number of low-fecundity adults (Bravington et al., 2016a; Trenkel et al., 2022). CKMR studies based on HSPs over multiple cohorts permit a direct estimate of adult survival in species for which fecundity does not change across the lifespan, and in combination with catch data, can be used to separate natural from fishing mortality (Bravington et al., 2017; Rodriguez-Ezpeleta et al., 2020; Maunder et al., 2021). The presence of HSPs from different cohorts in a sample "marks" the shared parent and, at the same time, informs that the parent was alive at the time of birth of each sibling. The larger the age gap between the sibling cohorts, the longer the parent had to survive. As the difference in age increases, the rate of HSPs decreases and this rate of decline is related to adult survival. Nonetheless, data from HSPs alone is insufficient in teleosts as it does not provide any information of fecundity-size relationship (Figure 5). Variable fecundity-at-size (or age) implies that large specimens produce disproportionately more progeny than newly matured individuals and therefore have a higher probability of producing HSPs. In the absence of further data sources is not possible to disentangle mortality from the extent of such variation in fecundity from HSPs alone (Trenkel et al., 2022). Information on maturity at age might be sufficient for some teleost, for which is reasonable to assume a reproductive output proportional to age/size/weight (Maunder et al., 2021). However, in most fishes, an accurate estimation of adult mortality requires having both, POPs and HSPs together with length/age-compositions, to estimate how reproductive success changes with age and separate it from adult mortality (Maunder et al., 2021). In general, it is better to collect both, juveniles and adults, and estimate POPs and HSPs whenever possible.

3.3.3 Size-specific fecundity (when length data are available)

Knowledge of fish fecundity is required for estimating the stock reproductive potential, for understanding the relationship between adult or spawning stock biomass and recruitment and hence for building suitable statistical models for assessing sustainable catch levels (Pérez-Rodríguez et al., 2010).

Assuming that ancillary length and/or age data has been recorded and that fecundity depends on size/age, CKMR based on POPs can provide information on this parameter through the analysis of differences in numbers of POPs found among parents of different sizes/ages (Bravington et al., 2016a). Fecundity-at-size/age can be inferred by comparing the length/age distributions of adult females vs. identified mothers. Both distributions are similarly influenced by (i) selectivity; and (ii) total mortality but the latter is also weighted by fecundity-at-size/age (Bravington et al., 2016a). This is the main determinant of the reproductive output, which is higher in larger/older females that are tagged more often by the offspring, affecting the number of POPs (Bravington et al., 2016a).

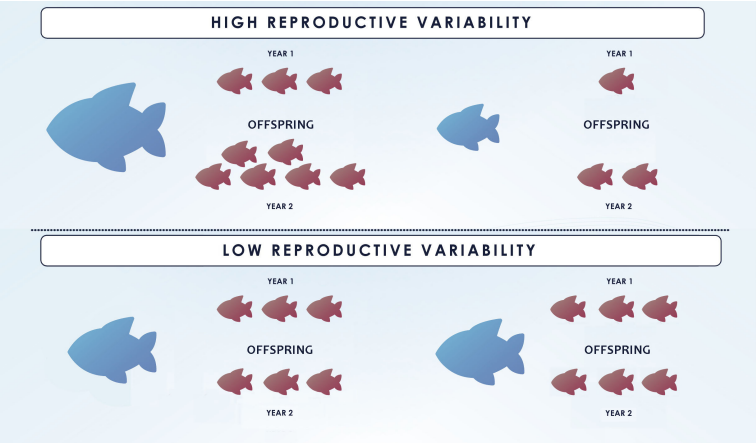


FIGURE 5
Impact of reproductive variability on HSP-based CKMR. Parents are represented in blue while offspring is coloured in red. The figure displays a variable fecundity scenario on top and a constant fecundity scenario below. The upper and lower sets of offspring within each scenario represent different cohorts. The upper scenario is by far more common in teleosts and entails a higher probability of producing HSPs for large individuals as they have disproportionately more progeny than newly matured fish. Figure redrawn from (Bravington et al., 2014).

So, by "dividing" the two distributions, is possible to estimate relative fecundity-at-age (Davies et al., 2015).

Studies based on HSPs, on the contrary, are not useful to obtain any information on size-specific fecundity, because HSPs give no information on which adults were responsible for them.

3.3.4 Selectivity

Selectivity determines the probability of extraction of fish individuals and results in a specific distribution of fishing mortality across components of a population. The interpretation of kinship (the number of kinship pairs found, and their patterns with age and time) in a CKMR study is affected not just by adult abundance, mortality and fecundity, but also by the complication of adult selectivity.

On the one hand, as animals age, the chances of finding kinship diminishes at the mortality rate because the parent might die, and also diminishes at the population rate of increase because the proportion of non-Parents and non-Half-Siblings increases with time. On the other hand, when a juvenile is born, assuming that the parent does survive and is still around to be captured and/or to reproduce, for POP-based studies of teleosts, the parent becomes more catchable each year because of increasing selectivity with age. For HSPs-based studies, more half-sibs are available for capture each year, because the parent becomes more fecund over time. These trends in probability determine the mean time gap for POPs and for HSPs. The difference between the mean gaps for both types of kinship can be used to disentangle age-specific selectivity and fecundity (Bravington et al., 2015).

3.3.5 Other parameters

Throughout the CKMR literature, kin pair data has been also useful to provide qualitative information on several other parameters, which are normally not included in assessment models but are essential to assess population dynamics. Although CKMR has been mostly applied in single, well-mixed populations, it can also be useful within fragmented or different populations. Akita (2022) developed an extension of CKMR, based on POPs and HSPs, to estimate migration number or rate of adults between two populations in iteroparous species. The author formulates the probabilities of kinship pairs and provides two estimators that only require genetic data, as they are (approximately) independent of the reproductive potential. Information on animal movements and demographic connectivity can be alternatively obtained from the spatial distribution of close relatives (Delaval et al., 2022; Patterson et al., 2022). In addition, CKMR studies have the potential to provide insight into the breeding dynamics and the structure of a population (Feutry et al., 2017; Feutry et al., 2020; Patterson et al., 2022).

4 Case studies - from teleosts to elasmobranchs

The method of CKMR offers great promise for the estimation of population parameters often difficult to infer by other methodologies. Nonetheless, few practical applications of CKMR across aquatic animals have been published to date. Here, we provide an overview of the most important aspects highlighted by studies involving teleosts and elasmobranchs (skates, rays and sharks) (Table 1) and analyze their relevance in a framework of fisheries assessments.

The first full application of the CKMR method was a study on the southern bluefin tuna (SBT), a heavily exploited marine species (FAO, 2022). This study, published in 2016, sparked great interests in the methodology (Bravington et al., 2016a). SBT fulfills a series of characteristics that facilitates the experimental design and sampling, making it an ideal target for close-kin studies. These, include an extensive biological knowledge of the species, absence of population structure, a single spawning site and availability of a large number of samples collected across several years. The stock is severely depleted after an intense exploitation for decades but a large uncertainty remains about its absolute abundance and the level of depletion of its reproductive component (Kolody et al., 2008; Kurota et al., 2010). Using CKMR, the authors assessed abundance of SBT, with 25 microsatellite markers that were used to genotype ~14,000 tissue samples of juvenile and adult tuna individuals. Kinship inference revealed 45 POPs that were subsequently used to infer absolute adult abundance and adult survival. Resulting estimates revealed a less depleted and more productive stock than previously suggested by traditional stock assessments. Interestingly, the methodology provided evidence of a strong nonlinear relationship between fecundity and weight. Big females have a much larger reproductive contribution than expected from their bodyweight, a highly relevant information for the productivity and resilience of the species, largely reported in the literature from life-history analyses and maternal effects (Green, 2008; Marshall, 2009). Thus, CKMR was able to provide an independent check of the assessment model in SBT, reducing the uncertainty and has been incorporated into current management procedures.

This CKMR study remains the more relevant in the context of fisheries assessments among those involving teleosts as most other applications to date have targeted much smaller populations. Ruzzante et al. (2019) analyzed seven populations of brook trout inhabiting coastal streams along the shore of Nova Scotia. These authors used in parallel CKMR and standard mark-recapture (MR) to compare their respective estimates of population abundance. Using 33 microsatellite markers, a number of POPs ranging from ≈8 to ≥ 50, depending on the population, were detected among 2,400 individuals. These kinship pairs were used to infer CKMR-derived abundances that proved to be statistically indistinguishable from those obtained by standard MR. This is the first study that validates CKMR estimates, but is important to emphasize that population structure and life history of this species is unusually simple. Moreover, confidence intervals of the estimations were wide in most populations, indicating the need of a significant increase of the sampling effort to achieve precision levels similar to those obtained for SBT (i.e., $CV \approx 0.15$) (Bravington et al., 2016a). Based on their results, the authors conclude that the method is unlikely to be applicable to stocks with tens of millions of individuals or larger due to the dependence of the sample size requirements on population size (Ruzzante et al., 2019). Another study of brook trout from the Honnedaga lake (New York) used 44 microsatellites to genotype 304 individuals, identifying 72 POPs that allowed the estimation of population abundance (Marcy-Quay et al., 2020). The authors compared the values obtained with abundance estimates

TABLE 1 Summary table of CKMR studies involving fish species published to date in peer-reviewed journals, detailing the number of juvenile and adult individuals, type and number of markers, type and number of kinship used in the study, the estimation of abundance reported and the corresponding coefficient of variation.

Study	Bravington et al. (2016a)	Ruzzante et al. (2019)	Marcy-Quay et al. (2020)	Prystupa et al. (2021)	Wacker et al. (2021)	Hillary et al. (2018)	Trenkel et al. (2022)	Delaval et al. (2022)	Patterson et al. (2022)
Species	**Southern bluefin tuna**	**Brook trout***	**Brook trout**	**Artic grayling**	**Atlantic salmon**	**White shark**	**Thornback ray**	**Blue skate**	**Speartooth shark**
Number of adults	5,755	110	127	507	113	0	6,555	662	0
Number of juveniles	7,448	116	257	597	278	75	0	0	226
Type of marker	M	M	M	M	SNP	SNP	SNP	SNP	SNP
Number of markers	25	33	44	38	164	2,186	3,668	6,291	1400
Type of kinship	POP	POP	POP	POP	POP	HSP	POP	HSP	HSP
Number kinship pairs	45	16	72	37	60	20	73	15	41
Abundance	under 2,000,00	1146 (591-1701)	1525**	1858 (1259–2457)	621 (472–769)	2,500–6,750	135,000	25,582 (10,484–52,664)	2020
CV	~0.17	~0.25	unreported	~0.16	~0.12	unreported	~0.19	unreported	0.28

CV indicates coefficient of variation; M denotes microsatellite, SNP denotes single nucleotide polymorphism, POP refers to Parent-Offspring-Pair and HSP to Half-Sibling-Pair.
*Only one population (VWU) is reflected in the table while seven are included in the publication.
**Abundance is calculated for 5 years separately in the original publication, here we provide the average.
Abundance estimates (with their corresponding confidence intervals, when reported) are provided together with the coefficient of variation.

from classical MR, finding a good agreement between both. The CKMR approach was also used to infer the abundance of the Arctic Grayling at the Lubbock system river in Yukon (Canada). A total of 967 specimens (split approximately equally between adults and juveniles), were collected and genotyped at 38 microsatellites, revealing 37 POPs that served to estimate adult population abundance with reasonable precision (*CV* ≈0.16) (Prystupa et al., 2021). In addition, Wacker et al. (2021) compared the estimates of spawner abundance of adult Atlantic salmon in a Norwegian river using three methods; conventional surveys, CKMR and CKMR combined with tagging. A total of 278 juveniles and 113 adults were genotyped at 164 SNP loci, revealing 80 POPs that allowed the estimation of CKMR abundances. CKMR estimates were considerably higher than those obtained in conventional surveys, and these, in turn, were smaller than those obtained by the combination of CKMR with tagging.

The results of these teleosts studies provide essential information for their management, however, they also highlight that most applications to date have targeted small populations inhabiting rivers, as this feature facilitates reaching the demanding requirements, in terms of sample sizes, needed to find relevant numbers of kinship pairs to produce estimates with reasonable precision. Also, most of these populations can be considered "closed" (low migration, high self-recruitment), making assumptions more straightforward. Furthermore, although few studies compare abundance estimates obtained from classical methods with those produced by CKMR, not all find a good agreement, highlighting the need for further studies to understand the biases that affect each of them.

The generally smaller population sizes of elasmobranchs compared to teleosts together with a poor conservation status and a lack of any abundance estimates affecting many species has made them also a frequent choice in CKMR studies. The first study that applied a CKMR framework to estimate the population of an elasmobranch targeted white sharks in eastern Australia and New Zealand (Hillary et al., 2018). This analysis was based solely on HSPs due to the unfeasibility of sampling adults in useful numbers. Genetic sampling of juveniles is easier as there are known aggregation sites. In this study, a total of 75 juveniles were genotyped using 2,186 SNPs, revealing 20 HSPs used to estimate population size. Kinship also served to produce estimates of adult survival rates and to inform of adult sex ratios. Despite the low precision of the results, acknowledged by the authors, mainly due to the small sample size and the limited range of sampled cohorts, the estimates obtained are highly valuable, given the lack of data for this population, as they provide crucial data for assessing its status. This data deficiency is common in elasmobranchs, with many populations considered severely depleted and endangered worldwide (Juan-Jordá et al., 2022), urging the application of methodologies like CKMR to assess their conservation status and protect them.

The abundance of the thornback ray in the Bay of Biscay was assessed by genotyping over 6,500 specimens using 3,668 SNPs. The analysis revealed 73 "usable" POPs and 431 FSPs that could not be considered for abundance calculations due to the lack of biological data (particularly, age) and the limited knowledge on population structure available for the species. The large number of specimens analyzed still permitted the estimation of adult abundance with

reasonable precision (*CV*=0.19). The authors provide a good overview of practical lessons learned from the application of CKMR to their case study, worth to be considered by anyone intending to apply the CKMR method for abundance estimation (Trenkel et al., 2022).

A CKMR study of the blue skate in the Celtic Sea assessed its abundance using over 6,000 SNPs to genotype 662 specimens collected across four years (Delaval et al., 2022). The results revealed 15 cross-cohort half-sibling pairs while no POPs were detected, possibly due to a limitation in the number of cohorts sampled as samples mostly comprised young adults. The authors inferred abundance and annual adult survival rate but due to the small number of kinship pairs detected the estimates involved a large uncertainty. Interestingly, the authors compared the CKMR estimated abundance with CPUE-based estimates. Although the latter reflects the abundance of all the individuals in a population, not just the breeders, a good general agreement was found. Both approaches indicated a stable, possibly expanding population of blue skate in the Celtic Sea during the analyzed period.

Finally, Patterson et al. (2022) used CKMR to assess abundance and connectivity in a critically endangered euryhaline elasmobranch, the speartooth shark (*Glyphis glyphis*). The analysis involved only juveniles as finding mature individuals is extremely difficult, similarly to other elasmobranchs. With only three encounters of adults in over seven years in Australian waters, the abundance was unknown. A total of 226 juveniles collected in two Australian river systems across four years were genotyped using 1,400 SNPs revealing the presence of 21 and 41 full- and half-siblings, respectively. The authors also used information from the mitogenomes to determine whether the inferred HSPs share the mother or the father. These relationships served to calculate the total adult abundance of the whole population and at each river system but also to estimate other parameters, including sex-specific connectivity and annual survival as well as to infer several aspects of the reproductive dynamics of this critically endangered species. This study demonstrates that the method is capable of providing essential parameters in a short period of time for a long-lived species, even when the collection of auxiliary data (age) involved a substantial uncertainty (Patterson et al., 2022). Nonetheless, it also highlights a remarkable complication that derives from the difficulties in assigning age accurately in elasmobranchs, which often requires lethal sampling to obtain vertebral growth readings. As this is not possible for many protected species, the potential of sample collection is extremely reduced but the alternative non-lethal sampling entails a high uncertainty in cohort assignment that can significantly lower the number of usable kinship pairs. Thus, even when using a large number of markers for genotyping that allows the inclusion of second-degree relationships (HSP), lack of accurate biological ancillary data can undermine the power of a CKMR study. Nonetheless, these elasmobranch studies, also reveal the potential of the methodology for inferring abundance in species spanning a large variety of life-histories, even when sampling and background biological knowledge are scarce, conditional on the model being tailored to each species-specific characteristics (Rodriguez-Ezpeleta et al., 2020).

5 Discussion

Unbiased and precise estimates of demographic parameters are essential to ensure the sustainable exploitation of fish stocks. Reliable estimates of abundance and life history traits (e.g., fecundity, mortality) are central components of population models, but often extremely difficult to obtain, as they must rely on data that is often scarce and suffer from recognized biases (Bradshaw et al., 2007; Schmidt et al., 2015; Hoenig et al., 2016).

Close-kin mark–recapture (CKMR) has recently emerged as a promising technique for estimating some of these parameters in animal populations from the frequency, and the distribution in space and time of kinship relationships observed in genetic samples (Davies et al., 2015). This alternative has potentially several advantages over traditional methods, such as the use of data from CPUE, scientific surveys or tagging experiments, as it does not suffer from many of the potential sources of bias affecting these (Rodriguez-Ezpeleta et al., 2020). CKMR can be applied to tissue samples from live or dead animals and although it could be theoretically used with any set of individuals that can be divided into "parents" and "juveniles, most applications have shown the importance of specimen´s ancillary data and sampling design considerations (e.g. Trenkel et al., 2022). Life histories of most species determines the coexistence of several cohorts for most fish stocks, requiring a precise assignment of the time of birth to take full advantage of the method. In some cases, accurate length estimates can be used for this purpose but most case studies highly benefit from accurate age information of the specimens (Coggins et al., 2013; Patterson et al., 2022). If age data is available, it becomes possible to extend the basic method (POPs) to more distant kinship (HSP) and use a short-term, cross-sectional sample of a population to produce reliable estimations, opposite to traditional methods that require decades-long datasets (Bravington et al., 2016b).

Therefore, CKMR might have the potential to become more informative, robust and cost-efficient than current methods for estimating demographic parameters. Nonetheless, the use of this new genetic tool in fisheries science is still in its infancy. The theory and promise of close-kin analysis beyond the target species of CKMR studies published so far - mostly characterized by relatively small population sizes and high uncertainty in abundance estimation by traditional means - still needs to be demonstrated and validated by further studies (Friedman et al., 2022). In this regard, it is crucial to validate the method in stocks with good quality estimations of abundance based in well-established stock assessment models.

The method in its current form has several limitations. First, it is not suited to all fish populations as it relies on being able to differentiating parental and offspring generations and also requires the independent sampling of the different cohorts. To fulfill these requirements, knowledge of the biology of the species, especially life-history strategies and reproductive biology as well as of the patterns of social structure and habitat-use, is needed, but these parameters are unknown for a vast number of species.

Even when an in-deep biological knowledge is accessible, there are some aquatic organisms for which CKMR is unlikely to produce reliable estimations. These, include species with very long lifespans and those displaying semelparity or facultative parthenogenesis

(Bravington et al., 2016b). The latest has been described in, at the minimum, six species of sharks and rays, which have shown the ability to reproduce asexually in the absence of males (hammerhead shark (Chapman et al., 2007), blacktip shark (Chapman et al., 2008), zebra shark (Robinson et al., 2011), bamboo shark (Feldheim et al., 2010), white tip reef shark (Portnoy et al., 2014), small tooth sawfish (Fields et al., 2015) and spotted eagle ray (Harmon et al., 2016)). Although it is unknown how frequently parthenogenesis occurs in elasmobranchs in the wild, it has been proposed to be facultative in situations where females have difficulty encountering suitable mates. As elasmobranchs numbers decline worldwide the difficulty of finding mates increases, possibly promoting parthenogenesis (Bernal et al., 2015). This should be taken into account when applying CKMR to this group of animals, preferentially targeted in the studies published so far due to their low population densities that are, in many cases, exacerbated by overexploitation (e.g. Hillary et al., 2018; Patterson et al., 2022). It is essential to pay special attention to low genetic diversity and significant homozygosity all across the genome in order to avoid wrong estimations of abundance with CKMR.

Nonetheless, the major drawback of the method affects generally all species and is linked to the number of samples required for the analysis, which is proportional to the square root of the true population abundance (Bravington et al., 2013; Bravington et al., 2014). In large populations, entails the collection of a substantial number of samples to ensure the detection of sufficient numbers of kinship pairs, implying high effort and high costs (Casey et al., 2016). A simulation exercise (ICES, 2017) estimated that, for a species with an estimated adult population of ~1.5 million individuals, a sample size of ~17 000 individuals and 70 POPs would be required to obtain reliable estimates of abundance (CV =10%). Many commercially relevant fish species exceed these abundances by several orders of magnitude and thus, adhering to these standards would be economically daunting even without considering the sampling costs (ICES, 2017). Super-abundant species, such as krill, would possibly never be a candidate for CKMR studies as they require colossal sample sizes, even taking into account that this number is proportional to $\sqrt{N_{adult}}$ rather than to N_{adult} itself (Bravington et al., 2016b). On the other hand, applying CKMR to very small populations would imply little cost but requires the collection of an impractically high fraction of the population to yield a useful number of kin pairs, particularly in marine ecosystems that are characterized by few physical barriers to dispersal (Bravington et al., 2016b).

The method has other important drawbacks for its application in the context of fisheries. CKMR is not able to provide precise estimates of abundance-at-age for the whole population, a required parameter to be used in age-structured stock assessment models, widely used to provide scientific advice in fisheries management. Another important limitation is that it provides information only for the adult component of the stock, as is not able to produce any estimates on recruitment or juvenile abundance. This prevents to forecast stock abundance at mid- and long-term. More importantly, the large economic and sampling requirements of CKMR are too demanding to produce annual estimates of stock abundance to be used in stock assessment models (irrespective of the models being age-structured or not) for most exploited species. In its current state, CKMR cannot replace the estimation of abundance based in research surveys and analytical stock assessment models. Another technical challenge is that close-kin mark-recapture analysis requires specialized knowledge that is usually outside of the experience of population geneticists who have the skills to generate and analyze the genetic data. Besides the population genetics knowledge, expertise in statistics and mathematics to develop statistical mark-recapture and population-dynamic models is considered essential to perform this type of studies (Ovenden et al., 2015; Bravington et al., 2016a). It is imperative to build bridges across these disciplines to promote advances in the methodology.

Despite the challenges, we believe that the method can bring enormous advancements to the field of fisheries management. We foresee several topics where CKMR may contribute importantly in scientific advice.

First, creating or modifying Biological Reference Points (BRP). These are standardized stock indicators used to compare stock status and inform fisheries managers about stock's status relative to various management objectives. Therefore, they are key components in scientific advice and management as they establish limits - mostly based in mortality and/or biomass - above or below which stock sustainability is jeopardized. The estimation of BRPs is subject to important uncertainties very much depending on the model used to estimate them, but often leading to a loss of fishing yield to avoid the risk of overexploitation (Mildenberger et al., 2022). CKMR outputs, particularly abundance and mortality, should help in understanding how accurate are the defined BRPs, even if CKMR is not applied annually, and contribute to adjust current BRPs or to build new ones during stock benchmarking. For example, the first CKMR study on bluefin tuna (Bravington et al., 2016a) revealed a significantly higher abundance of adult abundance than that estimated by the Operating Model (OM) used to assess this stock. The method was subsequently implemented and has been used to observe trends in abundance in combination with other indexes based on surveys or fisheries CPUE. The abundance estimates from CKMR should allow to define a new BRP or modify the target reference points used in many stock assessments. This is a field that should be further explored and it is very much linked with the need of establishing dynamic reference points that cope with ecosystem changes (Berger, 2019).

Second, stock-recruitment relationship is a cornerstone in fisheries management. It is the basis to define some of the BRPs used in fisheries management and is also used to predict future recruitment and forecast stock abundance for different management options (catch projections). However, stock-recruitment relationships are normally poor, impeding the ability to predict recruitment. One of the reasons for this is the use of spawning stock biomass (SSB), which does not consider variations in the stock reproductive potential (Saborido-Rey and Trippel, 2013) and specifically patterns of variation in individual reproductive success, i.e. the probability that an adult will produce offspring (recruitment) differs within and across-years. CKMR based on HSPs can inform about the effective number of breeders per year, N_b (Waples and Feutry, 2022), and has the potential to improve stock-recruitment relationships, especially when combined with CKMR adult abundance estimations. Although the importance of N_b in fisheries management has been diminished (Hillary et al., 2018), it is actually a relevant parameter.

Finally, there is a range of situations where CKMR estimations of abundance can be very valuable. For many populations, bottom trawl research surveys are not able to provide realistic estimates of abundance due to low catchability, either because of low abundance, or to low/null accessibility (e.g. coastal and littoral species, species inhabiting rocky areas, river streams or vulnerable ecosystems). The vast majority of these are data-limited stocks that lack complex stock assessments or forecasts (Rosenberg et al., 2014). The stock size of many pelagic species is assessed based on acoustic surveys, which still contain uncertainty in spite of the huge advances in technology. CKMR may contribute to reduce uncertainty, although the size of some stocks (especially small pelagic stocks like sardines, anchovies, herrings, mackerels) would require extremely large sample sizes. However, for widely distributed stocks (redfishes, i.e. Atlantic *Sebastes* spp.), or large pelagic fish stocks (sharks, tunas, billfishes) the application of CKMR seems to be more feasible and might provide valuable information as estimations of abundance by traditional research surveys are limited. The method has the potential to offer as well a better understanding of stocks productivity, the relative importance of different age classes to total stock reproductive potential of stocks and more generally, a better understanding of the biology of exploited species (e.g. Bravington et al., 2016a, Patterson et al., 2022). Moreover, CKMR can also inform about parameters often ignored in stock assessments, like connectivity and migration in iteroparous species (Akita, 2022; Patterson et al., 2022). This information is normally not included due to the lack of data but can lead to biased estimates of population parameters (Van Beveren et al., 2019) and is essential to understand population dynamics as well as to identify appropriate management units (Goethel et al., 2012).

The major impediment on the application of CKMR outputs as stock indicators is still the quantification of the uncertainty and its impact in scientific advice and the subsequent management options that can easily lead to stock overexploitation. However, in stocks where there is lack of data to build proper stock indicators (as BRPs), CKMR outputs may provide valuable information for stock benchmarking if used with caution.

Integrating the collection of tissues (finclips) for genetic analysis into the regular scientific fisheries surveys has little cost, and would be a framework that provides, at the same time, the specimen´s biological information (e.g.: age from the otoliths, sex or maturity stage) needed for the application of CKMR. Pilot studies could start in parallel with regular assessments to maintain the long time-series datasets and facilitate validation of the CKMR methodology. The implementation of CKMR into assessment programs will clearly depend on its capacity to reduce uncertainty and contribute to establishing or modifying BRPs but its potential uptake would also be promoted if fisheries scientist become familiar with the methodology. Finally, CKMR can be crucial to determine the status of data-limited stocks since a single CKMR study can be adequate to estimate abundance and other population parameters. If an appropriate sampling scheme can be maintained over time, the time series of abundance could be used to establish harvest rates in species that lack a full stock assessment (Maunder et al., 2021).

Author's note

The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods EMFF/2018/015" signed with the European Climate, Infrastructure and Environment Executive Agency CINEA and funded by the European Union. The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of CINEA or of the European Commission. Neither CINEA nor the European Commission can guarantee the accuracy of the scientific data/information collected under the above Specific Contract or the data/information included in this publication. Neither CINEA nor the European Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Author contributions

All authors contributed to the article and approved the submitted version.

Funding

The information and views set out in this manuscript are based on scientific data and information collected under Service Contract "Improving cost-efficiency of fisheries research surveys and fish stocks assessments using next-generation genetic sequencing methods [EMFF/2018/015]" signed with the European Climate, Infrastructure and Environment Executive Agency (CINEA) and funded by the European Union.

Acknowledgments

The authors would like to express their gratitude to Gary Carvalho, Laszlo Orban and Allan Tucker for their advice and expert revision of the State of the Art review in which this manuscript is based on. We would also wish to thank the reviewers for their critical reading of the manuscript, as their inputs have led to significant improvements.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Akita, T. (2022). Estimating contemporary migration numbers of adults based on kinship relationships in iteroparous species. *Mol. Ecol. Resour.* 22, 3006–3017. doi: 10.1111/1755-0998.13682

Alonso-Fernández, A., Otero, J., Bañón, R., Campelos, J. M., Santos, J., and Mucientes, G. (2017). Sex ratio variation in an exploited population of common octopus: ontogenic shifts and spatio-temporal dynamics. *Hydrobiologia* 794, 1–16. doi: 10.1007/s10750-016-3065-3

Anderson, E. C. (2022). CKMRpop: forward-in-time simulation and tabulation of pairwise kin relationships in age-structured populations. *Mol. Ecol. Resour.* 22, 1190–1199. doi: 10.1111/1755-0998.13513

Anderson, D. M., and Gillooly, J. F. (2021). Evaluating the tradeoff between offspring number and survivorship across fishes, amphibians, reptiles and mammals. *Oikos* 130, 798–807. doi: 10.1111/oik.07569

Andrews, A. H., Tracey, D. M., and Dunn, M. R. (2009). Lead–radium dating of orange roughy (Hoplostethus atlanticus): validation of a centenarian life span. *Can. J. Fish. Aquat. Sci.* 66, 1130–1140. doi: 10.1139/F09-059

Barry, P., Broquet, T., and Gagnaire, P. (2022). Age-specific survivorship and fecundity shape genetic diversity in marine fishes. *Evol. Lett.* 6, 46–62. doi: 10.1002/evl3.265

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., et al. (2019). Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ* 7, e7724. doi: 10.7717/peerj.7724

Berger, A. M. (2019). Character of temporal variability in stock productivity influences the utility of dynamic reference points. *Fish. Res.* 217, 185–197. doi: 10.1016/j.fishres.2018.11.028

Bernal, M. A., Sinai, N. L., Rocha, C., Gaither, M. R., Dunker, F., and Rocha, L. A. (2015). Long-term sperm storage in the brownbanded bamboo shark chiloscyllium punctatum. *J. Fish Biol.* 86, 1171–1176. doi: 10.1111/jfb.12606

Bradford, R. W., Thomson, R., Bravington, M., Foote, D., Gunasekera, R., Bruce, B. D., et al. (2018). A close-kin mark-recapture estimate the population size and trend of east coast grey nurse shark. *Rep. to Natl. Environ. Sci. Program*, Mar. Biodivers. Hub. CSIRO Ocean. Atmos, Hobart, Australia. 44pp. doi: 10.13140/RG.2.2.26338.58560

Bradshaw, C. J. A., Mollet, H. F., and Meekan, M. G. (2007). Inferring population trends for the world's largest fish from mark-recapture estimates of survival. *J. Anim. Ecol.* 76, 480–489. doi: 10.1111/j.1365-2656.2006.01201.x

Bravington, M. V. (2014). Close-kin mark-recapture for SBT: options for the longer term. *Consult. Rep. CCSBT*. CSIRO Ocean. Atmos, Hobart, Australia. 28pp. Available at: https://www.ccsbt.org/system/files/ESC22_BGD12_CCSBT_CkgSBT_LongTerm_Rev1.pdf.

Bravington, M. (2019). Shortfin mako sharks in the Atlantic: outline design for close-kin mark-recapture.

Bravington, M. V., Eveson, J. P., Grewe, P. M., and Davies, C. R. (2015). SBT close-kin mark-recapture: options for the medium term. *Contract Rep. CCSBT*. CSIRO Ocean. Atmos, Hobart, Australia. 38pp. Available at: https://www.ccsbt.org/system/files/resource/en/55c99097e8e7b/ESC20_19_CCSBT_SBTCKMR_LongTerm_2015.pdf.

Bravington, M. V., Eveson, J. P., Grewe, P. M., and 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/* 29pp. Available at: https://www.ccsbt.org/system/files/ESC22_12_AU_CKMR_POP_HSP.pdf

Bravington, M., Feutry, P., Pillans, R. D., Hillary, R., Johnson, G., Saunders, T., et al. (2019). Close-kin mark-recapture population size estimate of *Glyphis garricki* in the Northern Territory. *Rep. to Natl. Environ. Sci. Program Mar. Biodiversity Hub*. CSIRO Oceans & Atmosphere, Hobart. 31 pp. Available at: https://www.nespmarine.edu.au/system/files/Bravington etal_Glyphis garricki CKMR population estimate_FINAL_4Dec19.pdf.

Bravington, M. V., and Grewe, J. P. (2007). A method for estimating the absolute spawning stock size of SBT, using close-kin genetics. *CCSBT-ESC/0709/18*. CSIRO, Hobart, Australia. 25pp. Available at: https://www.ccsbt.org/ja/system/files/resource/ja/4d93fe45b53f2/18.pdf.

Bravington, M., Grewe., P., and Davies, C. (2012). Report of the close-kin project: estimating the absolute spawning stock size of SBT using genetics. *CCSBT-ESC/1208/19*. CSIRO, Hobart, Australia. 39pp. Available at: https://www.ccsbt.org/system/files/resource/ja/502c7824304e8/ESC17_19-_AU_Close_kin.pdf.

Bravington, M., Grewe, P., and Davies, C. (2013). Available at: https://www.ccsbt.org/system/files/resource/ja/521ebb122ae21/ESC18_BGD03_Australia_Close_kin.pdf.

Bravington, M. V., Grewe, P. G., and Davies, C. R. (2014). "Fishery-independent estimate of spawning biomass of southern bluefin tuna through identification of close-kin using genetic markers," in *FRDC report 2007/034* (CSIRO, Australia), 147pp. Available at: https://www.researchgate.net/publication/261835492_Fishery-independent_estimate_of_spawning_biomass_of_Southern_Bluefin_Tuna_through_identification_of_close-kin_using_genetic_markers

Bravington, M. V., Grewe, P. M., and Davies, C. R. (2016a). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat. Commun.* 7, 1–8. doi: 10.1038/ncomms13162

Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016b). Close-kin mark-recapture. *Stat. Sci.* 31, 259–274. doi: 10.1214/16-STS552

Bruce, B., Bradford, R., Bravington, M., Feutry, P., Grewe, P., Gunasekera, R., et al. (2018). "A national assessment of the status of white sharks," in *National environmental science programme* (Marine Biodiversity Hub, CSIRO, Australia), 64pp. Available at: https://www.nespmarine.edu.au/system/files/NESP-A3-A_national_assessment_of_the_status_of_white_sharks_Feb_2018.pdf.

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., et al. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evol. Appl.* 11, 1094–1119. doi: 10.1111/eva.12600

Casey, J., Jardim, E., and Martinsohn, J. T. (2016). The role of genetics in fisheries management under the E.U. common fisheries policy. *J. Fish Biol.* 89, 2755–2767. doi: 10.1111/jfb.13151

Chapman, D. D., Firchau, B., and Shivji, M. S. (2008). Parthenogenesis in a large-bodied requiem shark, the blacktip carcharhinus limbatus. *J. Fish Biol.* 73, 1473–1477. doi: 10.1111/j.1095-8649.2008.02018.x

Chapman, D. D., Shivji, M. S., Louis, E., Sommer, J., Fletcher, H., and Prodöhl, P. A. (2007). Virgin birth in a hammerhead shark. *Biol. Lett.* 3, 425–427. doi: 10.1098/rsbl.2007.0189

Cheminée, A., Le Direach, L., Rouanet, E., Astruch, P., Goujard, A., Blanfuné, A., et al. (2021). All shallow coastal habitats matter as nurseries for Mediterranean juvenile fish. *Sci. Rep.* 11, 14631. doi: 10.1038/s41598-021-93557-2

Christie, M. R. (2010). Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Mol. Ecol. Resour.* 10, 115–128. doi: 10.1111/j.1755-0998.2009.02687.x

Coggins, L. G., Gwinn, D. C., and Allen, M. S. (2013). Evaluation of age–length key sample sizes required to estimate fish total mortality and growth. *Trans. Am. Fish. Soc* 142, 832–840. doi: 10.1080/00028487.2013.768550

Conn, P. B., Bravington, M. V., Baylis, S., and Ver Hoef, J. M. (2020). Robustness of close-kin mark–recapture estimators to dispersal limitation and spatially varying sampling probabilities. *Ecol. Evol.* 10, 5558–5569. doi: 10.1002/ece3.6296

Davies, C. R., Bravington, M. V., Eveson, J. P., Lansdell, M., Aulich, J., and Grewe, P. M. (2020). "Next-generation close-kin mark recapture: using SNPs to identify half-sibling pairs in southern bluefin tuna and estimate abundance, mortality and selectivity," in *FRDC project no 2016-044* (CSIRO Australia, Australia), 63pp. Available at: https://www.frdc.com.au/sites/default/files/products/2016-044-DLD.pdf.

Davies, C. R., Bravington, M. V., Grewe, P. D., Eveson, P. J., Lansdell, M., Hill, P., et al. (2018) *Close-kin project report*. Available at: https://www.ccsbt.org/ja/system/files/ESC23_14_AU_CloseKinProjectReport_0.pdf.

Davies, C., Bravington, M. V., and Thomson, R. (2015). Advice on close-kin mark-recapture for estimating abundance of eastern Atlantic blue n tuna: a scoping study. *GBYP 07c/201*. CSIRO, Hobart, Australia. 34pp. Available at: https://www.iccat.int/GBYP/Docs/Close_Kin_Mark_Recapture_Phase_6_Scoping_Study.pdf

Delaval, A., Bendall, V., Hetherington, S. J., Skaug, H. J., Frost, M., Jones, C. S., et al. (2022). Evaluating the suitability of close-kin mark-recapture as a demographic modelling tool for a critically endangered elasmobranch population. *Evol. Appl.* 16 (2):461–473. doi: 10.1111/eva.13474

de Pontual, H., Groison, A. L., Piñeiro, C., and Bertignac, M. (2006). Evidence of underestimation of European hake growth in the bay of Biscay, and its relationship with bias in the agreed method of age estimation. *ICES J. Mar. Sci.* 63, 1674–1681. doi: 10.1016/j.icesjms.2006.07.007

FAO. (2022). The State of World Fisheries and Aquaculture 2022. *Towards Blue Transformation*. Rome, FAO. doi: 10.4060/cc0461en

Feldheim, K. A., Chapman, D. D., Sweet, D., Fitzpatrick, S., Prodohl, P. A., Shivji, M. S., et al. (2010). Shark virgin birth produces multiple, viable offspring. *J. Hered.* 101, 374–377. doi: 10.1093/jhered/esp129

Feutry, P., Berry, O., Kyne, P. M., Pillans, R. D., Hillary, R. M., Grewe, P. M., et al. (2017). Inferring contemporary and historical genetic connectivity from juveniles. *Mol. Ecol.* 26, 444–456. doi: 10.1111/mec.13929

Feutry, P., Devloo-Delva, F., Tran Lu Y, A., Mona, S., Gunasekera, R. M., Johnson, G., et al. (2020). One panel to rule them all: DArTcap genotyping for population structure, historical demography, and kinship analyses, and its application to a threatened shark. *Mol. Ecol. Resour.* 20, 1470–1485. doi: 10.1111/1755-0998.13204

Fields, A. T., Feldheim, K. A., Poulakis, G. R., and Chapman, D. D. (2015). Facultative parthenogenesis in a critically endangered wild vertebrate. *Curr. Biol.* 25, R446–R447. doi: 10.1016/j.cub.2015.04.018

Flanagan, S. P., and Jones, A. G. (2019). The future of parentage analysis: from microsatellites to SNPs and beyond. *Mol. Ecol.* 28, 544–567. doi: 10.1111/mec.14988

Friedman, K. J., Bartley, D. M., Rodríguez-Ezpeleta, N., Mair, G. C., Ban, N., Beveridge, M., et al. (2022). *Current and future genetic technologies for fisheries and aquaculture: implications for the work of FAO* (Rome: FAO). doi: 10.4060/cc1236en

Goethel, D. R., Kerr, L. A., and Cadrin, S. X. (2012). Incorporating spatial population structure in stock assessment models of marine species. *ICES CM 2012/N:03, Bergen, Norway*, 14pp. Available at: http://ices.dk/sites/pub/CM Doccuments/CM-2012/N/N0312.pdf.

Green, B. S. (2008). Maternal effects in fish populations. *Adv. Mar. Biol.* 54, 1–105. doi: 10.1016/S0065-2881(08)00001-1

Harmon, T. S., Kamerman, T. Y., Corwin, A. L., and Sellas, A. B. (2016). Consecutive parthenogenetic births in a spotted eagle ray aetobatus narinari. *J. Fish Biol.* 88, 741–745. doi: 10.1111/jfb.12819

Hillary, R. M., Bravington, M. V., Patterson, T. A., Grewe, P., Bradford, R., Feutry, P., et al. (2018). Genetic relatedness reveals total population size of white sharks in eastern Australia and new Zealand. *Sci. Rep.* 8, 2661. doi: 10.1038/s41598-018-20593-w

Hoenig, J. M., Then, A. Y.-H., Babcock, E. A., Hall, N. G., Hewitt, D. A., and Hesp, S. A. (2016). The logic of comparative life history studies for estimating key parameters, with a focus on natural mortality rate. *ICES J. Mar. Sci. J. du Cons.* 73, 2453–2467. doi: 10.1093/icesjms/fsw089

Hoffmayer, E. R., McKinney, J. A., Franks, J. S., Hendon, J. M., Driggers, W. B., Falterman, B. J., et al. (2021). Seasonal occurrence, horizontal movements, and habitat use patterns of whale sharks (Rhincodon typus) in the gulf of Mexico. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.598515

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour.* 17, 1009–1024. doi: 10.1111/1755-0998.12665

ICES (2016). *Interim report of the working group on application of genetics in fisheries and mariculture (WGAGFM), 11–13 May 2016*, Belfast, UK. ICES CM2016/SSGEPI:15. 41 pp. Available at: https://ices-library.figshare.com/ndownloader/files/33392261.

ICES (2017). *Report of the working group on the application of genetics in fisheries and mariculture (WGAGFM), 2–5 May 2017, Olhão, Portugal. ICES CM 2017/SSGEPI:13.* 75 pp. Available at: https://ices-library.figshare.com/ndownloader/files/33393416

Juan-Jordá, M. J., Murua, H., Arrizabalaga, H., Merino, G., Pacoureau, N., and Dulvy, N. K. (2022). Seventy years of tunas, billfishes, and sharks as sentinels of global ocean health. *Science*, 378(6620):eabj0211. doi: 10.1126/science.abj0211

Kindsvater, H. K., Mangel, M., Reynolds, J. D., and Dulvy, N. K. (2016). Ten principles from evolutionary ecology essential for effective marine conservation. *Ecol. Evol.* 6, 2125–2138. doi: 10.1002/ece3.2012

Kolody, D., Polacheck, T., Basson, M., and Davies, C. (2008). Salvaged pearls: lessons learned from a floundering attempt to develop a management procedure for southern bluefin tuna. *Fish. Res.* 94, 339–350. doi: 10.1016/j.fishres.2008.08.016

Kurota, H., Hiramatsu, K., Takahashi, N., Shono, H., Itoh, T., and Tsuji, S. (2010). Developing a management procedure robust to uncertainty for southern bluefin tuna: a somewhat frustrating struggle to bridge the gap between ideals and reality. *Popul. Ecol.* 52, 359–372. doi: 10.1007/s10144-010-0201-1

Lowerre-Barbieri, S. K. (2009). "Reproduction in relation to conservation and exploitation of marine fishes," in *Reproductive biology and phylogeny of fishes (Agnathans and bony fishes) Science Publishers, Enfield, CT*. Ed. B. G. M. Jamieson (CRC Press), 371–394. doi: 10.13140/2.1.2048.0643

Marcy-Quay, B., Sethi, S. A., Therkildsen, N. O., and Kraft, C. E. (2020). Expanding the feasibility of fish and wildlife assessments with close-kin mark–recapture. *Ecosphere* 11(10):e03259. doi: 10.1002/ecs2.3259

Marshall, C. T. (2009). "Mplementing information on stock reproductive potential in fisheries management: the motivation, challenges and opportunities," in *Fish reproductive biology: implications for assessment and management*. Eds. T. Jakobsen, M. J. Fogarty, B. A. Megrey and E. Moksness (Oxford, UK: Wiley-Blackwell), 395–420. doi: 10.1002/9781444312133.ch11

Marshall, D. J., Bode, M., Mangel, M., Arlinghaus, R., and Dick, E. J. (2021). Reproductive hyperallometry and managing the world's fisheries. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2100695118. doi: 10.1073/PNAS.2100695118

Maunder, M. N., Lennert-Cody, C. E., Aires-da-Silva, A. M., and Xu, H. (2021) *Considerations for conducting close kin mark recapture of stocks managed by the IATCC.* Available at: https://www.iattc.org/GetAttachment/ddc8410e-4de7-401c-87a9-73fb33090f23/SAC-12-14_Considerations-for-conducting-Close-Kin-Mark-Recapture-of-stocks-managed-by-IATTC.pdf.

Maunder, M. N., and Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. *ICES J. Mar. Sci.* 72, 7–18. doi: 10.1093/icesjms/fsu015

McDowell, J. R., Bravington, M., Grewe, P. M., Lauretta, M., Walter, J. F., Baylis, S. M., et al. (2022). Low levels of sibship encourage use of larvae in western Atlantic bluefin tuna abundance estimation by close-kin mark-recapture. *Sci. Rep.* 12, 18606. doi: 10.1038/s41598-022-20862-9

Mildenberger, T. K., Berg, C. W., Kokkalis, A., Hordyk, A. R., Wetzel, C., Jacobsen, N. S., et al. (2022). Implementing the precautionary approach into fisheries management: biomass reference points and uncertainty buffers. *Fish Fish.* 23, 73–92. doi: 10.1111/faf.12599

Mucientes, G. R., Queiroz, N., Sousa, L. L., Tarroso, P., and Sims, D. W. (2009). Sexual segregation of pelagic sharks and the potential threat from fisheries. *Biol. Lett.* 5, 156–159. doi: 10.1098/rsbl.2008.0761

Murua, H., and Saborido-Rey, F. (2003). Female reproductive strategies of marine fish species of the north Atlantic. *J. Northwest Atl. Fish. Sci.* 33, 23–31. doi: 10.2960/J.v33.a2

Nielsen, R., Mattila, D. K., Clapham, P. J., and Palsbøll, P. J. (2001). Statistical approaches to paternity analysis in natural populations and applications to the north Atlantic humpback whale. *Genetics* 157, 1673–1682. doi: 10.1093/genetics/157.4.1673

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., and 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 Fish.* 16, 125–159. doi: 10.1111/faf.12052

Patterson, T. A., Hillary, R. M., Kyne, P. M., Pillans, R. D., Gunasekera, R. M., Marthick, J. R., et al. (2022). Rapid assessment of adult abundance and demographic connectivity from juvenile kin pairs in a critically endangered species. *Sci. Adv.* 8(51): eadd1679. doi: 10.1126/sciadv.add1679

Pérez-Rodríguez, A., Morgan, M., Rideout, R., Domínguez-Petit, R., and Saborido-Rey, F. (2010). Study of the relationship between total egg production, female spawning stock biomass, and recruitment of Flemish cap cod (Gadus morhua). *Cienc. Mar.* 37, 675–687. doi: 10.7773/cm.v37i4B.1785

Pennino, M. G., Conesa, D., López-Quílez, A., Muñoz, F., Fernández, A., and Bellido, J. M. (2016). Fishery-dependent and -independent data lead to consistent estimations of essential habitats. *ICES J Mar Sci* 73, 2302–2310. doi: doi 10.1093/icesjms/fsw062

Portnoy, D. S., Hollenbeck, C. M., Johnston, J. S., Casman, H. M., and Gold, J. R. (2014). Parthenogenesis in a whitetip reef shark triaenodon obesus involves a reduction in ploidy. *J. Fish Biol.* 85, 502–508. doi: 10.1111/jfb.12415

Prystupa, S., McCracken, G. R., Perry, R., and Ruzzante, D. E. (2021). Population abundance in arctic grayling using genetics and close-kin mark-recapture. *Ecol. Evol.* 11, 4763–4773. doi: 10.1002/ece3.7378

Qiao, Y., Sannerud, J. G., Basu-Roy, S., Hayward, C., and Williams, A. L. (2021). Distinguishing pedigree relationships via multi-way identity by descent sharing and sex-specific genetic maps. *Am. J. Hum. Genet.* 108, 68–83. doi: 10.1016/j.ajhg.2020.12.004

Ramírez-Macías, D., Queiroz, N., Pierce, S. J., Humphries, N. E., Sims, D. W., and Brunnschweiler, J. M. (2017). Oceanic adults, coastal juveniles: tracking the habitat use of whale sharks off the pacific coast of Mexico. *PeerJ* 5, e3271. doi: 10.7717/peerj.3271

Robinson, D. P., Baverstock, W., Al-Jaru, A., Hyland, K., and Khazanehdari, K. A. (2011). Annually recurring parthenogenesis in a zebra shark stegostoma fasciatum. *J. Fish Biol.* 79, 1376–1382. doi: 10.1111/j.1095-8649.2011.03110.x

Rodriguez-Ezpeleta, N., Patterson, T. A., Pereda, I., Grande, M., Davies, C. R., Lezama-Ochoa, N., et al. (2020). Feasibility study on applying close-kin mark recapture abundance estimates to Indian ocean tuna commission shark species. *Final Report to IOTC.* San Sebastian, Spain: AZTI; 2020. csiro:EP2022-2718. Available at: http://hdl.handle.net/102.100.100/442298?index=1

Rosenberg, A. A., Fogarty, M. J., Cooper, A. B., Dickey-Collas, M., Fulton, E. A., Gutiérrez, N. L., et al. (2014). "Developing new approaches to global stock status assessment and fishery production potential of the seas," in *FAO Fisheries and Aquaculture Circular No. 1086*(Rome: FAO). Available at: http://www.fao.org/fi/oldsite/eims_search/1_dett.asp?calling=simple_s_result&lang=en&pub_id=314932.

Ruzzante, D. E., McCracken, G. R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., et al. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. *Methods Ecol. Evol.* 10, 1445–1453. doi: 10.1111/2041-210X.13243

Saborido-Rey, F., and Kjesbu, O. S. (2005). *Growth and maturation dynamics* (CSIC), 26pp. Available at: http://hdl.handle.net/10261/47150.

Saborido-Rey, F., and Trippel, E. A. (2013). Fish reproduction and fisheries. *Fish. Res.* 138, 1–4. doi: 10.1016/j.fishres.2012.11.003

Schmidt, J. H., Johnson, D. S., Lindberg, M. S., and Adams, L. G. (2015). Estimating demographic parameters using a combination of known-fate and open n -mixture models. *Ecology* 96, 2583–2589. doi: 10.1890/15-0385.1

Seber, G. A. F., and Schofield, M. R. (2019). "Tagging methods and tag loss," in Capture-recapture: parameter estimation for open animal populations. *Statistics for Biology and Health* (Cham: Springer), 13–37. doi: 10.1007/978-3-030-18187-1_2

Selwyn, J. D., Hogan, J. D., Downey-Wall, A. M., Gurski, L. M., Portnoy, D. S., and Heath, D. D. (2016). Kin-aggregations explain chaotic genetic patchiness, a commonly observed genetic pattern, in a marine fish. *PloS One* 11, e0153381. doi: 10.1371/journal.pone.0153381

Skaug, H. J. (2001). Allele-sharing methods for estimation of population size. *Biometrics* 57, 750–756. doi: 10.1111/j.0006-341X.2001.00750.x

Städele, V., and Vigilant, L. (2016). Strategies for determining kinship in wild populations using genetic data. *Ecol. Evol.* 6, 6107–6120. doi: 10.1002/ece3.2346

Stamatopoulos, C. (2002). "Sample-based fishery surveys," in *A technical handbook* (FAO Fisheries Technical Paper) 425. Available at: http://www.fao.org/documents/card/en/c/b8a043cb-4537-5ce6-bc9f-14fe84d689df/.

Swain, D. P. (2011). Life-history evolution and elevated natural mortality in a population of Atlantic cod (gadus morhua). *Evol. Appl.* 4, 18–29. doi: 10.1111/j.1752-4571.2010.00128.x

Thompson, R. B., Bravington, M. V., Feutry, P., Gunasekera, R., and Grewe, P. (2020) *Close kin mark recapture for school shark in the SESSF*. Available at: https://www.frdc.com.au/sites/default/files/products/2014-024-DLD.pdf.

Trenkel, V. M., Charrier, G., Lorance, P., and Bravington, M. V. (2022). Close-kin mark–recapture abundance estimation: practical insights and lessons learned. *ICES J. Mar. Sci.* 79, 413–422. doi: 10.1093/icesjms/fsac002

Van Beveren, E., Duplisea, D. E., Brosset, P., and Castonguay, M. (2019). Assessment modelling approaches for stocks with spawning components, seasonal and spatial dynamics, and limited resources for data collection. *PloS One* 14, e0222472. doi: 10.1371/journal.pone.0222472

Vendrami, D. L. J., Peck, L. S., Clark, M. S., Eldon, B., Meredith, M., and Hoffman, J. I. (2021). Sweepstake reproductive success and collective dispersal produce chaotic genetic patchiness in a broadcast spawner. *Sci. Adv.* 7(37):eabj4713. doi: 10.1126/sciadv.abj4713

Wacker, S., Skaug, H. J., Forseth, T., Solem, Ø., Ulvan, E. M., Fiske, P., et al. (2021). Considering sampling bias in close-kin mark–recapture abundance estimates of Atlantic salmon. *Ecol. Evol.* 11, 3917–3932. doi: 10.1002/ece3.7279

Wang, J. (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Mol. Ecol.* 18, 2148–2164. doi: 10.1111/j.1365-294X.2009.04175.x

Wang, J. (2014). Estimation of migration rates from marker-based parentage analysis. *Mol. Ecol.* 23, 3191–3213. doi: 10.1111/mec.12806

Waples, R. S., and Feutry, P. (2022). Close-kin methods to estimate census size and effective population size. *Fish Fish.* 23, 273–293. doi: 10.1111/faf.12615

Waples, R. S., Grewe, P. M., Bravington, M. W., Hillary, R., and Feutry, P. (2018). Robust estimates of a high n e / n ratio in a top marine predator, southern bluefin tuna. *Sci. Adv.* 4, eaar7759. doi: 10.1126/sciadv.aar7759

Wright, P. J., and Trippel, E. A. (2009). Fishery-induced demographic changes in the timing of spawning: consequences for reproductive success. *Fish Fish.* 10, 283–304. doi: 10.1111/j.1467-2979.2008.00322.x

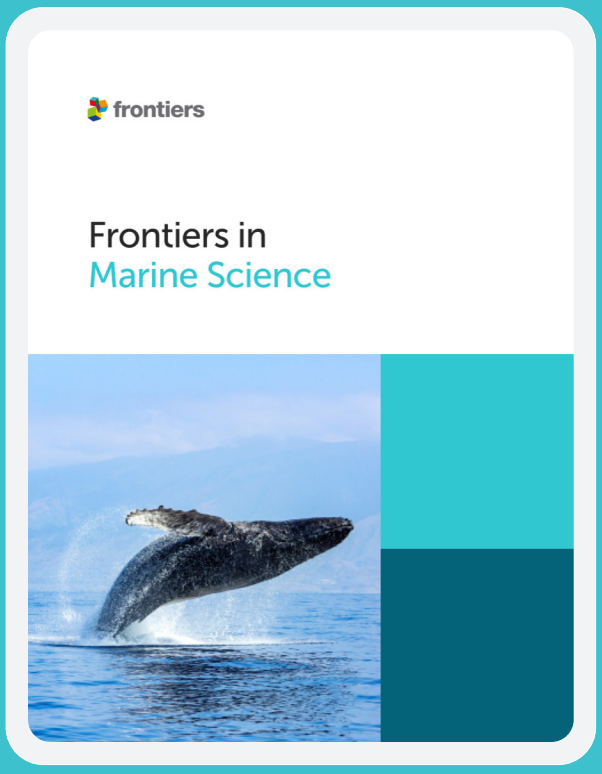
Frontiers in Marine Science

Explores ocean-based solutions for emerging global challenges

The third most-cited marine and freshwater biology journal, advancing our understanding of marine systems and addressing global challenges including overfishing, pollution, and climate change.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact