BMC Ecology and Evolution

RESEARCH Open Access

Genomic insights into the population structure and adaptive variation of *Mullus barbatus* in the Mediterranean Sea

Piergiorgio Massa[1*], Henrique G. Leitão[2], Tereza Manousaki[3], Alessia Cariani[1,4], Hannes Svardal[2,5*], Dimitris Tsaparis[3], Rita Cannas[4,6], Lorenzo Zane[4,7], Tim De Pooter[8,9], Maria Teresa Spedicato[10], Mojca Strazisar[8,9], Claudio Ciofi[11], Alice Ferrari[1,4], Genevieve Diedericks[2], Maria Angela Diroma[11], Giulio Formenti[12], Petya Ivanova[13], Ann M. Mc Cartney[14], Alice Mouton[15], Chiara Natali[11], Ilaria AM Marino[7,16] and Costas S. Tsigenopoulos[3*]

Abstract

Background Red mullet (*Mullus barbatus*) is a key species in Mediterranean fisheries, yet its stock structure and population dynamics remain poorly understood due to a lack of comprehensive genomic resources. This study provides the first high-quality reference genome for *M. barbatus* and a comprehensive set of SNP markers to investigate its population structure and adaptive potential across the Mediterranean.

Results Using the newly generated chromosome-level reference genome, we re-analyzed a Mediterranean-wide reduced-representation genomic dataset. Our analysis reveals a panmictic population structure with strong genetic connectivity across the species' range, likely driven by extensive larval dispersal and multigenerational gene flow. Despite minimal genome-wide differentiation, outlier analysis identified candidate loci under directional selection, linked to key biological processes such as ontogeny and environmental adaptation.

Conclusions This study presents the first genomic resource for *M. barbatus*, providing valuable insights into its genetic structure and adaptive mechanisms. While the identification of loci under selection offers promising leads, these findings are preliminary due to the limited genomic coverage of the dataset. Nonetheless, they pave the way for future genomic studies to explore how *M. barbatus* adapts to environmental and anthropogenic pressures. These results hold significant implications for the sustainable management of Mediterranean fisheries, especially in the context of climate change and conservation.

Keywords Reference genome, Genome assembly, Genome annotation, Population structure, Genetic connectivity, Genomic adaptation, Fisheries management, Mediterranean fisheries, Red mullet, Demersal species

*Correspondence:
Piergiorgio Massa
piergiorgio.massa2@unibo.it
Hannes Svardal
hannes.svardal@uantwerpen.be
Costas S. Tsigenopoulos
tsigeno@hcmr.gr

Full list of author information is available at the end of the article

Background

Red mullet as a key commercial fish species

The red mullet (*Mullus barbatus* L., 1758) is a demersal fish species with a broad geographic distribution spanning the Mediterranean and Black Seas, as well as the eastern Atlantic Ocean, from the North Sea to Senegal [1]. It shares its range with the congeneric striped red mullet (*Mullus surmuletus* L., 1758). These two species co-occur on continental shelves at depths reaching 200 m, though striped red mullet can be also present in deeper waters [2, 3]. Despite their overlapping habitats, they exhibit notable differences in substrate preference with *M. barbatus* favoring muddy seabeds, while *M. surmuletus* is typically associated with rougher bottom types [1, 2].

Both species are key targets for Mediterranean demersal fisheries [3–5]. The red mullet, in particular, is a target for coastal bottom-trawl fisheries [3]. The stocks of this species were considered to be under high fishing pressure in the past [6, 7]. A reduction in effective population sizes (*Ne*) for both species was observed, attributed to factors such as excessive fishing pressure, habitat fragmentation, and natural population fluctuations [8]. Nonetheless, recent assessments indicate that stocks of red mullet are currently sustainably exploited in most of the Mediterranean Sea [9], with biomass increasing since 2008 in several GFCM Geographical SubAreas (GSAs) [3].

Genetic insights into the Mediterranean stock structure

Larval dispersal plays a major role in determining the spatial scales over which marine populations are connected. Marine environments, with their putative continuity, often foster high connectivity and unrestricted dispersal among populations, which can lead to low genetic differentiation [10–13]. In species with high dispersal capabilities during their larval or adult stages, significant gene flow frequently results in weak population structure [14–18]. Red mullet exhibits low adult mobility, but larvae are able to remain in the water column for 25–35 days before settling in a suitable habitat [19]. Gargano et al. (2017) [20] reported that larvae and pre-recruits from the Strait of Sicily can reach suitable coastal areas to recruit even far from their natal area, supporting the existence of a meta-population structure. This complex dispersal dynamic challenges traditional stock assessments, making it difficult to define discrete population units. Misclassifications in stock structure can undermine management efforts and jeopardize fishery sustainability [21]. Furthermore, aligning biological processes with management strategies is crucial for sustainable fisheries management [22].

Studies on the population genetics of *M. barbatus* within the Mediterranean Sea have relied largely on nuclear microsatellite markers [8, 23–26]. However, methodological differences have often led to conflicting results. For instance, while Matic-Skoko et al. (2018) [8] reported genetic homogeneity between Western and Eastern Mediterranean populations, Galarza et al. (2009) [24] identified distinct genetic clusters. No local structure was evident along the Spanish Mediterranean coast [26], while the Adriatic Sea appeared genetically distinct in some instances [23, 25]. Interestingly, Matic-Skoko et al. (2018) [8] identified three overlapping genetic clusters within the Adriatic region.

First genomic efforts

The population structure of the congeneric *M. surmuletus* has been explored at high resolution thanks to numerous population genomic studies [16, 27–30]. In contrast, genomic investigation of *M. barbatus* has remained limited. The MED_UNITs project [31], funded by the European Commission, has been, to date, the only initiative to apply genomic techniques to *M. barbatus*, with the goal of aligning biological and management stock units across the Mediterranean for this and several other key commercial species.

This project employed a multidisciplinary approach, combining otolith shape analysis and microchemistry, with genomic methods, including Reduced Representation DNA Sequencing (RRS). RRS is a common genomic method employed in studies aimed at non-model species [32] as it provides genome-wide data from large sample sizes at a relatively low cost [33]. Although RRS captures only a fraction of the genome, it provides enough resolution to assess population structure in dispersive marine fish [34]. Indeed, a dataset of 853 SNPs, called in at least 70% of the 771 retained individuals, was used to infer genetic structure. The analysis revealed very low genetic differentiation values, suggesting a lack of distinct population structure in the Mediterranean. This finding was further supported by otolith shape analysis and microchemistry, which provided only weak and inconclusive evidence for the presence of two to three sub-populations in the region.

Reference genome improvements

High-quality reference genomes have significantly advanced genomic studies, providing unparalleled insights into species biology at increasingly affordable costs [35]. Initiatives like the Earth BioGenome Project (EBP) and its European node, the European Reference Genome Atlas (ERGA), are spearheading efforts to create comprehensive reference resources for eukaryotic life [36, 37].

When MED_UNITs conducted its genomic study, no reference genome was available for *M. barbatus*. As a result, genomic loci were constructed de novo [38], leading to a low number of high-quality loci. Aligning

genomic sequences to the species' reference genome enhances the accuracy of inferences obtained from RRS data [39, 40]. Furthermore, reference genomes allow in silico optimization of RAD-seq protocols, enhancing genotyping accuracy and SNP marker quality [41].

This study addresses a major gap in Mediterranean fisheries genomics by producing a high-quality reference genome for *M. barbatus* as part of the ERGA Pilot Project (I). The reference is then used to refine the first and only available genomic dataset for this species (II), improving the resolution of population structure analyses and informing sustainable fishery management strategies (III). Finally, we explore signatures of potential adaptive selection across the Mediterranean and report a preliminary set of candidate loci (IV).

Methods

Reference genome

Sampling and specimen information

An adult male *Mullus barbatus* (fMulBarb1, ERGA_PI_BG_01) was collected from the Black Sea, Bulgarian coast (Varna Bay) on July 20, 2021, using a trawl. This individual was used for reference genome assembly. For transcriptome sequencing, fin tissue was taken from the same reference individual, while gonad and kidney were sampled from additional specimen fMulBarb6 (ERGA_PI_BG_06) and muscle from fMulBarb2 (ERGA_PI_BG_02).

DNA extraction, library preparation, and long-read sequencing

High-molecular-weight (HMW) DNA was extracted from brain tissue of fMulBarb1 using the Nanobind Tissue Big DNA kit (PacBio). For Oxford Nanopore Technologies (ONT) long-read sequencing, two libraries were prepared: one fragmented to approximately 20 kb (Megaruptor, Diagenode) to optimize sequencing yield, and one unfragmented to maximize read length, with size selection performed using Short Read Eliminator (PacBio), only for the latter. Quality of the extracted and treated samples was monitored using UV/VIS for purity (Little Lunatic, Unchained Labs), fluorescence for concentration (Qubit, ThermoFisherScientific) and capillary electrophoresis for sizing and integrity (5200 Fragment Analyzer, Agilent). Library was prepared following an updated protocol 1D Genomic DNA by ligation protocol (SQK-LSK109, version: GDE_9063_v109_revQ_14Aug2019). Libraries were sequenced on one FLO-PRO002 flow cell using PromethION P24 for 72 h with 4 subsequent DNase flushes and library reloads for increased yield and fragment sizing.

Hi-C library preparation

Hi-C library preparation was performed using the Dovetail Genomics Omni-C kit, following the manufacturer's protocol. Liver tissue from the reference individual was manually homogenized prior to library preparation. Library fragment size distribution and concentration were evaluated using High Sensitivity D5000 ScreenTapes on a TapeStation 4150 (Agilent Technologies) and a Qubit Fluorometer (Thermo Fisher Scientific).

RNA extraction and library preparation

RNA was extracted from fin, gonad, kidney, and muscle tissues using the Quick-RNA Miniprep Plus Kit (Zymo Research). RNA integrity and concentration were assessed using RNA ScreenTapes (TapeStation 4150) and a Qubit Fluorometer. RNA library preparation followed a polyA capture protocol with the Illumina Stranded mRNA Prep kit, using MagBio HighPrep PCR magnetic beads for clean-up. Library fragment size and concentration were assessed with High Sensitivity D5000 ScreenTapes (TapeStation 4150) and a Qubit Fluorometer.

Hi-C and RNA library sequencing

The Omni-C Library was denatured and loaded for paired-end sequencing on an Illumina NovaSeq 6000 system and run in XP mode using two separate NovaSeq 6000 SP Reagent Kits v1.5 (300 cycles). For the Omni-C Library we set a single index running mode to 6:151:151:0 cycles. The RNA Library was denatured and run in dual-indexed mode with 10:151:151:10 bp cycles.

Genome assembly

High-accuracy basecalling of ONT reads was performed using Guppy (v5.0.13 [42]), followed by filtering with Filtlong (v0.2.1 [43]), to retain high-quality reads with a minimum read length of 1,000 bp and a targeted total base count of 45 Gbp. The expected genome size was calculated via genome profiling with the ONT data, using a k-mer spectra analysis (k = 21) with meryl (v1.3 [44]), and GenomeScope (v2.0 [45]). The genome was assembled with Flye (v2.9 [46]), using default parameters and an estimated genome size of 0.56 Gb based on the Genomes on a Tree (GoaT) database [47]. Polishing of the initial assembly was performed in two rounds with Racon (v1.4.3 [48]), using the ONT reads aligned to the draft assembly with Minimap2 (v2.17 [49]), followed by a consensus polishing step using Medaka (v1.4.4 [50]). Haplotypic duplicates were removed using purge_dups (v1.2.5 [51]). Scaffolding was performed with yahs (v1.2a [52]), using Omni-C data, which was aligned to the polished assembly with bwa-mem (v0.7.17 [53]). Valid ligation events were recorded, and PCR duplicates were removed using pairtools (v0.3.0 [54]), generating a.bam file for scaffolding. Assembly contiguity metrics were obtained with Gfastats (v1.3.1 [55]), and Merqury (v1.3.1 [44]), was used to assess assembly quality. Assembly completeness was evaluated using BUSCO (v5.2.2 [56]), with the

default BUSCO_MetaEuk workflow option and the actinopterygii_odb10 data set. The final assembly was manually curated by the Genome Reference Informatics Team at the Wellcome Sanger Institute. A Hi-C contact map was created and visualised with PretextMap (v0.1.9 [57]), PretextView (v0.2.5 [58]), and PretextSnapshot (v0.0.4 [59]). The mitochondrial genome was assembled using MitoHiFi (v2.2 [60]), with the mitochondrial genome of *Mullus surmuletus* (NC_052759) as the most closely related reference available.

Genome annotation

To identify the geneset, we mapped the produced RNA-seq reads against the genome reference using HISAT2 (v.2.2 [61]). The resulting bam files were merged and sorted using samtools (v1.7 [62]). Then to predict the gene models we used BRAKER3 (v3.0 [63]), providing as evidence the RNA alignments and similarity with known proteins as well. To produce the protein similarity hint file, we downloaded known proteins from UniProt and aligned them against the *M. barbatus* reference genome using miniprot (v0.13 [64]). The predicted geneset was reduced to one isoform per gene with AGAT (v1.4 [65]), and was then evaluated using BUSCO (v5.3.1 [56]). To functionally annotate the predicted geneset, we conducted a BLASTp [66] search against UniProt Database [67] and ran InterProScan (v5 [68]).

Population genomics

Sampling design and protocol

The genomic data analyzed in this study were sourced from the MED_UNITs project. This project aimed to integrate biological and genomic approaches to align stock management units for key commercial species across the Mediterranean, with *M. barbatus* as one of its focal species.*M. barbatus* biological samples were obtained through the activities of the EU Data Collection Framework (DCF, EU Reg. 1004/2017), which included biological sampling from commercial fisheries and experimental trawl surveys conducted as part of the Mediterranean International bottom Trawl Survey program (MEDITS [69]). The sampling strategy aimed to achieve comprehensive spatial coverage, encompassing 38 geographical sites across the Mediterranean Sea, from the Alboran to the Levantine Sea. A total of 2,133 specimens were collected between June 2019 and July 2020. A standardized sampling and shipping protocol was distributed to sampling partners emphasizing the collection of skeletal muscle tissues from freshly caught specimens. Alternatively, frozen tissues (−20 °C) were accepted if immediately frozen post-capture. Samples were preserved in 96% ethanol and stored at −20 °C until DNA extraction. Despite these guidelines, logistical challenges onboard commercial vessels sometimes precluded optimal preservation conditions. Consequently, tissues were collected from freshly caught specimens (35.2%) and frozen specimens (32.1%), while no preservation details were provided for 32.7% of samples.

DNA extraction

DNA was extracted following a modified salt extraction protocol [70], with an added RNase treatment step to prevent DNA content inflation. DNA integrity was assessed through electrophoresis gel imaging, categorizing extracts into three quality tiers (good, medium, poor). DNA concentrations were assessed through Qubit Fluorometer and Qubit dsDNA BR Assay. High variability in DNA quality was observed across and within GSAs. In general, DNA of good quality was easily obtained from fresh tissues, while frozen ones generated larger fraction of medium or even unusable DNA. These differences underscored the importance of adhering to optimal sampling protocols. Further details on DNA extraction and quality control protocols are provided in Supplementary Material 1, Additional File 1.

Library preparation and sequencing

A subset of 1,373 specimens from 32 geographical sites were obtained by including the majority of the medium quality DNA extracts (Supplementary Fig. 2.1, Additional File 1) (Supplementary Table 0.1, Additional File 2). Genomic libraries were prepared following a typical ddRAD protocol [71] using the *Sbf*I-*Sph*I enzyme pair, commonly used for fish species [72–75]. Libraries were PCR-amplified with 19 cycles and a gel cut-window of 400–700 bp was used. Five multiplexed libraries (288 specimens each, Supplementary Fig. 2.2, Additional File 1) were finally sequenced on an Illumina HiSeq 4000 platform (150 bp, paired-end) at the Norwegian Sequencing Center.

Reads processing and mapping

Demultiplexing, trimming, and quality-filtering of the raw sequence reads were performed using the process_radtags pipeline in Stacks (v2.62 [40]). The filtered reads were then mapped to the chromosome-level reference genome described in this study using the bwa-mem algorithm (v0.7.17 [53]). Quality of the resulting alignment files was assessed using FastQC (v0.12.1 [76]), MultiQC (v1.19 [77]), and samtools (v1.7 [62]), ensuring high standards for Downstream genomic analyses. Further details on raw data processing and genome mapping are provided in Supplementary Material 3, Additional File 1.

Genotype calling

The RAD loci catalog was constructed using the gstacks module in the Stacks reference-based pipeline, and biallelic single nucleotide polymorphisms (SNPs) were called

from the catalog. SNPs were defined using the populations module of Stacks based on the following criteria: a minor allele frequency of 0.01 (--min-maf 0.01), a maximum observed heterozygosity of 0.70 (--max-obs-het 0.70), and a minimum call rate threshold of 80% within each geographical site (-r 80 -p $N_POP). The resulting set of SNPs was examined for outlier detection. For population structure analyses, one random SNP per RAD locus was selected from the Stacks populations module (--write-random-snp) to minimize Linkage bias. The stringent settings applied ensured that high-quality SNPs were consistently called across geographical sites, thereby mitigating biases introduced by poorly sequenced samples and low Library complexities - a recurrent issue in this study due to variable sample collection conditions. Further details on call rate thresholds and missing data patterns are provided in Supplementary Material 4, Additional File 1.

Dataset filtering

The initial dataset of 1,373 specimens exhibited low genome-wide SNP call rates, likely due to the presence of poorly sequenced samples, as indicated by inconsistencies in sequencing quality across geographical sites (Supplementary Material 3–4, Additional File 1). These issues were attributed to challenges in tissue collection and preservation during fieldwork. To address these challenges, an iterative filtering approach was implemented to remove specimens with high levels of missing data. Iterative filtering strategies have been widely demonstrated to improve dataset quality by optimizing the balance between specimen retention and SNP call rate [78, 79]. An in-depth description of the iterative filtering applied is described in Supplementary Material 5, Additional File 1. Following filtering, the final number of retained RAD loci was evaluated against predictions made using Radinitio (v1.2.3 [80]), a pipeline for the assessment of RADseq experiments where reference genome is in silico digested to obtain a series of reference RAD loci. The analysis used the SbfI-SphI restriction enzyme pair and an insert size range of 400–700 bp, consistent with the parameters used during library preparation for most sequencing libraries.

Population structure

Prior to the population structure analysis, the statistical power of the markers employed was assessed with Powsim (v4.1 [81]), using Powsim_b that can accommodate up to 5000 loci. We tested a range of expected divergences to estimate the smallest detectable F_{ST}. Given the large amount of loci, the power was tested by running only 10 replicates. At the end of each simulation, genotypic samples with number and sizes corresponding to those of the empirical data set were collected and statistical testing of homogeneity performed using chi-square and Fisher testing approaches. The proportion of significant outcomes ($p < 0.05$) among replicates was used to estimate statistical power. Contemporary effective population size (Ne) was initially estimated from the final 1,592 SNP dataset using the linkage disequilibrium (LD) method implemented in NeEstimator (v2.0 [82]). However, this method is known to suffer from several limitations in large marine populations, particularly due to high sensitivity to sample size and allele frequency thresholds [83–85]. To address these concerns, we also employed CurrentNe (v1.0 [86]), a complementary LD-based approach specifically designed to provide robust Ne estimates under challenging conditions such as large Ne or limited sample sizes. Finally, to validate and contextualize contemporary Ne estimates, we applied the Multiple Sequentially Markovian Coalescent method (MSMC2, v2.1.4 [87, 88]). Given the method's reliance on a large and dense set of markers [89, 90], MSMC2 was applied to the long-read sequencing data from the fMulBarb1 reference individual. Full methodological details and results for these analyses are provided in Supplementary Material 8, Additional File 1. Principal Component Analysis (PCA) was conducted using Adegenet (v2.1.10 [91]), employing mean imputation for missing data. Given recent findings on PCA-related biases in genetic analyses [92, 93], we interpreted these results cautiously. Bayesian clustering was performed using STRUCTURE (v2.3.4 [94]), to infer the most likely number of populations (K). Following the Evanno method [95], we conducted 10 independent runs for each K value (1-10) under the admixture model with correlated allele frequencies, comprising 100,000 burn-in steps and 200,000 MCMC replicates. The ad hoc ΔK statistic was calculated using STRUCTURE HARVESTER (v0.7 [96]), and the results were visualized and merged using CLUMPAK [97]. Pairwise genetic differentiation (F_{ST}) between geographical sites was calculated using the Weir and Cockerham (1984) [98] estimator implemented in StAMPP (v1.6.3 [99]). Statistical significance was assessed using 100,000 permutations following Benjamini-Hochberg correction for multiple comparisons ($\alpha < 0.05$) [100]. To further evaluate the differentiation within or among populations, Analysis of MOlecular VAriance (AMOVA) was performed in pegas (v1.3 [101]). Statistical significance ($p < 0.05$) was assessed based on 100,000 permutations and geographical sites were hierarchically clustered into five macro-areas based on the subregional technical units defined by the General Fisheries Commission for the Mediterranean (GFCM [102]), (Supplementary Fig. 6.1, Additional File 1). This clustering ensured a balanced sample size at higher hierarchical levels. The relationship between pairwise genetic distance (F_{ST}) and geographic distance was evaluated by the Mantel test implemented in vegan (v2.6 [103]), to verify

whether species fit the Isolation-By-Distance (IBD) patterns among geographical sites. The least-cost geographical distance was used to account for land barriers [104] and 100,000 permutations were performed to assess statistical significance ($p < 0.05$).

Outliers detection

Filtered SNPs were analyzed to detect loci under directional selection using three F_{ST} outlier detection methods: Bayescan (v2.1 [105]), FstHet (v1.0.1 [106]), and OutFLANK (v0.2 [107]). Each method's parameters and results are thoroughly described in Supplementary Material 9, Additional File 1. Given the panmictic population structure identified in this study, geographical sites were grouped into the five macro-areas already identified based on the subregional technical units defined by the GFCM to ensure a balanced sample size. SNPs identified as outliers by at least two pipelines were retained as the strongest candidates for directional selection. Genomic regions containing candidate SNPs were identified using IGV (v2.18.4 [108]), and gffread (v0.12.8 [109]), and protein sequences for candidate genes were extracted. Further insights into the candidate genes function were obtained by similarity searches (BLASTp) against the NCBI Reference Protein Database (RefSeq [110]), best hits were selected.

Results

Reference genome

Genome assembly

ONT long-read sequencing generated 7.37 million reads, yielding a total 90.61 Gb with N50 18.48Kb Gbp. GenomeScope analysis estimated a haploid genome size of approximately 554 Mbp (Fig. 1b). A chromosome-scale genome assembly for a male *Mullus barbatus* was produced by combining ONT sequencing and paired-end chromosomal conformation capture data (Omni-C; 65.6 million reads) (Fig. 1a). The final genome assembly, after polishing, haplotypic duplication removal, scaffolding, and manual curation, had a total length of 538.9 Mbp. The assembly achieved a contig NG50 of 12.54 Mbp and a scaffold NG50 of 23.7 Mbp. Chromosomal scaffolds ($n = 22$) accounted for 535.65 Mbp, covering 99.39% of the assembly length. Genome completeness was evaluated using 3,640 conserved orthologous genes, identifying 96.7% as complete with 95.9% in single copy and 0.8% as duplicated. Additionally, 1.1% of genes were fragmented, and 2.2% were missing. K-mer-based quality assessment using Merqury resulted in a quality value (QV) of 32.1 and k-mer completeness of 84.96% (Fig. 1c), indicating that most k-mers from the sequencing data were successfully incorporated into the final genome assembly.

Genome annotation

RNA sequencing produced 524.34 million reads that were mapped against the genome with an average of 85% mapping rate. Gene prediction yielded 35,727 genes including 43,217 transcripts. From those 39,625 had a hit against UniProt and 33,003 had a protein Domain assigned through InterProScan search. Finally, BUSCO completeness assessment showed that 92.7% of the searched genes are present in the predicted geneset, 91.5% in single copy and 1.2% as duplicated. Additionally, 3.1% of genes were fragmented, and 4.2% were missing.

Population genomics

Final dataset

Approximately 3 billion reads were sequenced across the dataset of 1,373 individuals, with per-individual read counts ranging from 568 to 59,036,450, highlighting substantial variation (Supplementary Fig. 3.1, Additional File 1). A similar trend was observed in mapping performance, which ranged from 11.31 to 99.88% (Supplementary Fig. 3.2, Additional File 1). An initial catalog of 45,751 loci was constructed from the mapped reads, but this number dropped to zero after applying call rate filtering (Supplementary Fig. 4.1, Additional File 1). The iterative filtering strategy significantly reduced the initial dataset retaining 350 specimens, with approximately 20 individuals representing each of the 18 Mediterranean geographical localities ultimately included (Fig. 2a). Despite the reduction, the obtained dataset retained most of the starting Mediterranean geographic coverage (Supplementary Fig. 5.4, Additional File 1). Nonetheless, we additionally constructed a supplementary dataset by re-integrating a subset of individuals (58) that had been excluded in earlier filtering steps but exhibited high call rates at the final 1,592 SNP loci. This complementary dataset was used to qualitatively explore whether the inclusion of individuals from previously underrepresented locations might affect the observed patterns of structure (see Supplementary Figs. 5.5 and 7.5, Additional File 1). The filtering strategy adopted effectively standardized sequencing yield and mapping performance, now ranging from 526,068 to 7,108,806 reads per individual and 96.48–99.86% mapped reads, respectively (Supplementary Fig. 5.6, Additional File 1). A refined catalog of 30,551 loci was built from the mapped reads, which decreased to 1,947 after call rate filtering. In silico estimates using radinitio predicted 2,660 RAD loci suggesting Limited potential for further improvement through additional filtering given the 1,947 RAD loci ultimately retained. These loci provided uniform genome coverage, representing 0.106% of the assembled genome, with an average inter-locus distance of 333,040 bp. Accordingly, the number of high-quality SNPs increased from none in the initial dataset to 7,569, with a final set of 1,592

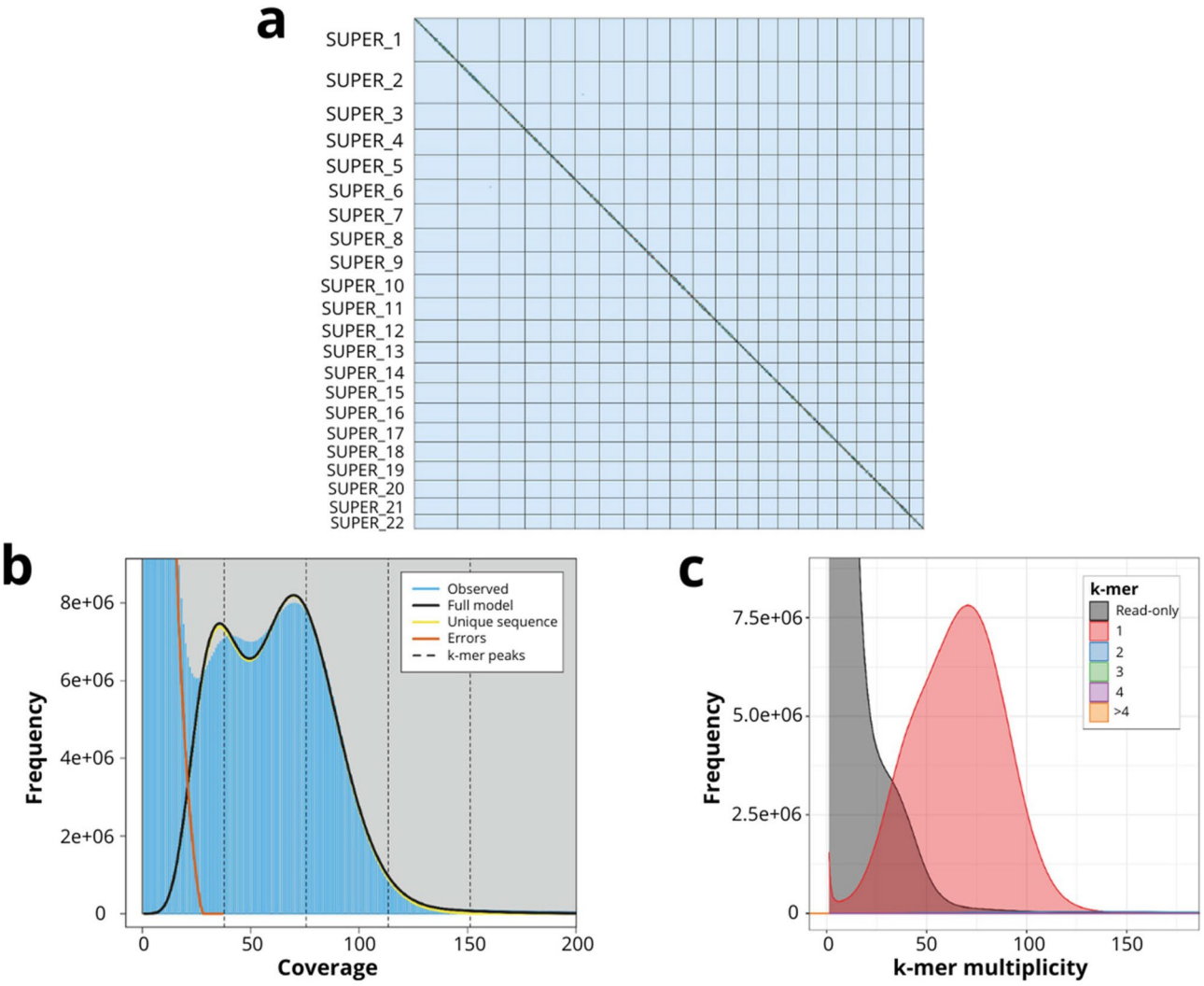

Fig. 1 Reference genome assembly overview. (**a**) Hi-C contact map of the fMulbar1.1 assembly visualized in PretextMap. Chromosomes are ordered by size, from top left to bottom right. (**b**) k-mer profile of ONT reads generated using GenomeScope (len:554,396,671 bp, uniq:81.2%, aa:98.7%, ab:1.3%, kcov:37.8, err:3.31%, dup:3.18, k:21, p:2), showing k-mer distribution. (**c**) k-mer spectra comparison between the assembly and raw sequencing data, generated using Merqury

retained when randomly selecting one SNP per locus, as applied in the population structure analysis (Supplementary Fig. 5.7, Additional File 1). This final dataset al.so exhibited low and uniform levels of missing data across geographical localities (Supplementary Fig. 5.9, Additional File 1).

Population structure

NeEstimator yielded an effective population size of 7,228 individuals (Supplementary Table 7.2, Additional File 1). Comparable estimates were obtained using CurrentNe (Supplementary Table 7.3, Additional File 1), and the trend was further supported by MSMC2 analysis on the whole-genome sequence of the reference individual fMulBarb1 (Supplementary Fig. 8.1, Additional File 1). Simulations conducted with Powsim demonstrated that our SNP dataset has sufficient statistical power to detect F_{ST} values as low as 0.0014, given the empirical sample sizes (Supplementary Table 7.1, Additional File 1). Such a minimal level of genetic differentiation could arise in as few as 20 generations of complete isolation, assuming the estimated effective population size. Principal Component Analysis (PCA) revealed no distinct genetic structure within the empirical dataset (Fig. 2b). The first few principal components represented only a minimal percentage of the total genetic variance, with PC1 and PC2 accounting for 0.894% and 0.885%, respectively (Supplementary FIgure 7.1, Additional File 1). A low variance retention by the primary components further suggests limited differentiation among the sampled geographical localities, indicating a lack of pronounced genetic structure. This pattern was consistent when using the complementary

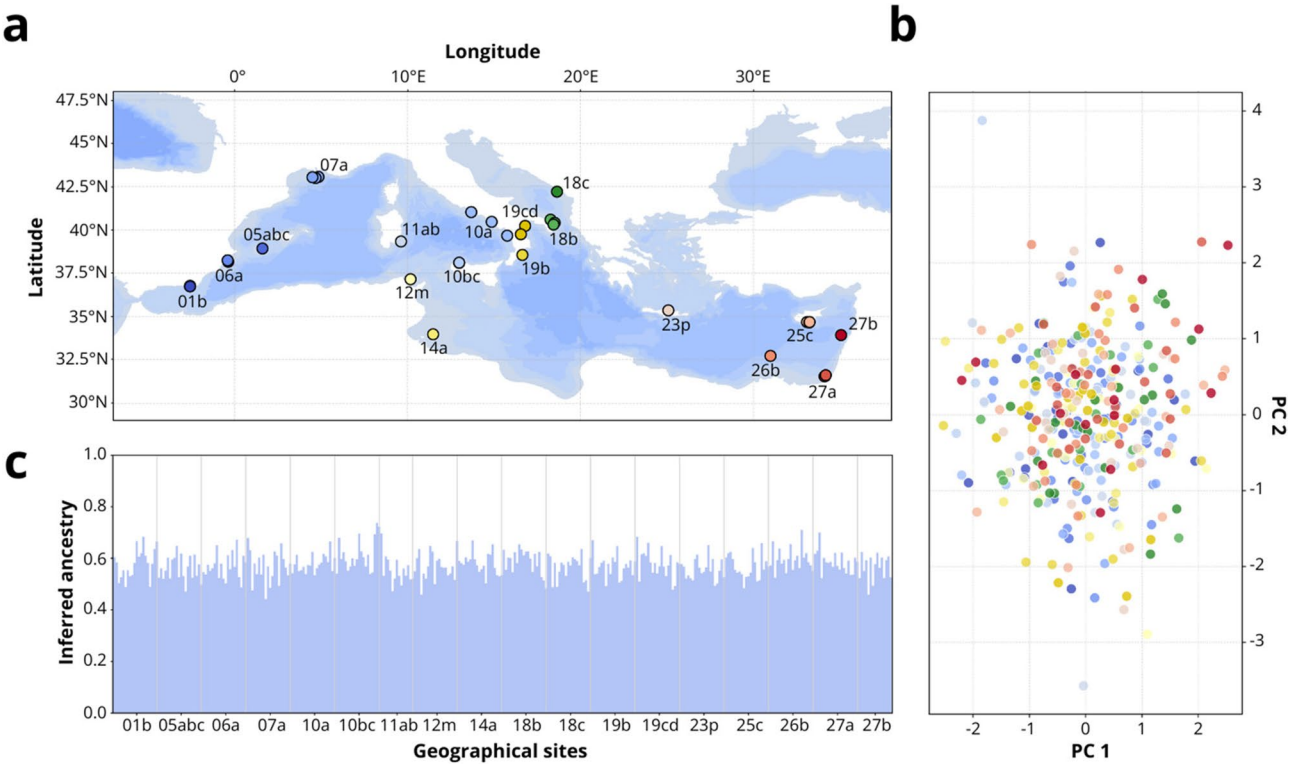

Fig. 2 Genetic structure of red mullet across the Mediterranean Sea. (**a**) Map showing the locations of 350 *Mullus barbatus* specimens from 18 distinct geographical sites across the Mediterranean Sea labelled according to the corresponding Geographical SubAreas (GSAs) : 01b: Northern Alboran Sea; 05abc: Balearic Islands; 06a: Northern Spain; 07a: Gulf of Lion; 10a: Southern and Central Tyrrhenian Sea; 10bc: Northern Sicily; 11ab: Eastern Sardinia; 12 m: Northern Tunisia: 14a: Gulf of Gabès; 18b: Southern Adriatic Sea (Western); 18c: Southern Adriatic Sea (Eastern); 19b: Western Ionian Sea (Sothern); 19 cd: Western Ionian Sea (Northern); 23p: Northern Crete; 25c: Southern Cyprus; 26b: Southern Levant Sea; 27a: Eastern Levant Sea (Southern); 27b: Eastern Levant Sea (Northern). Each site is color-coded according to subregional technical units defined by GFCM (blue for Western Mediterranean, yellow for Central Mediterranean, red for Eastern Mediterranean and green for Adriatic Sea). Within each sub regional technical unit, sites located in individual Geographical SubAreas (GSAs) are further differentiated using various shades of the main color associated with the region. Seabed bathymetry is depicted in blue. (**b**) Principal Component Analysis (PCA) performed on 1,592 high-quality SNPs reveals no significant clustering, as indicated by the minimal variance explained by the first two principal components (0.894% and 0.885%, respectively) (Supplementary Fig. 7.1, Additional File 1). The color scheme matches that of part (a), maintaining visual consistency across sub-regions. (**c**) Results from a STRUCTURE analysis with K = 2 show the ancestral proportions (blue for K1 and white for K2) for each specimen, grouped by geographical sites. This visual representation supports the conclusion of limited genetic differentiation among the sampled localities, underscoring a panmictic population structure across the Mediterranean

dataset, which included additional low-missingness individuals from previously excluded geographical regions. The reintegrated samples did not introduce any new clustering signal, confirming that the lack of structure is maintained even when geographic coverage is extended (see Supplementary Figs. 5.5 and 7.5, Additional File 1). The Evanno method suggested the most likely number of genetic clusters (K) to be 2 (Supplementary Fig. 7.2a, Additional File 1). Although the ad hoc ΔK statistic is effective for detecting population structure in complex hierarchical migration patterns, it cannot identify K = 1 as the optimal solution [95]. The mean posterior probability of the data for each K, as the model choice criterion implemented in STRUCTURE to detect the true K [94], suggested K = 1 to be more likely than K = 2 (Supplementary Fig. 7.2b, Additional File 1). Likewise, the symmetrical distribution of inferred ancestry components between clusters (Fig. 2c, Supplementary Fig. 7.3, Additional File 1) further indicated a lack of clear population structure [111]. Pairwise F_{ST} values ranged narrowly from −0.004 to 0.005, with none showing statistical significance following Benjamini-Hochberg correction for multiple comparisons ($\alpha < 0.05$) (Supplementary Fig. 7.4, Additional File 1). AMOVA results revealed that most of the variance (~68.6%) is attributed to the individual level rather than among macro-areas (i.e. GFCM subregional technical units) or geographical sites. The latter indeed exhibited small σ^2 (0.036124 and 0.044339, respectively) and non-significant p-values (0.2465 and 0.3148, respectively) (Supplementary Table 7.3, Additional File 1). The Mantel test revealed no significant positive correlation between geographic and genetic distances ($r = 0.039$, $p = 0.354$). These findings are consistent with a panmictic population structure across the Mediterranean range.

Outliers detection

The Bayescan, FstHet, and OutFLANK pipelines identified 10, 203, and 6 Fst outlier SNPs, respectively (Supplementary Figs. 9.1-9.2-9.3, Additional File 1). Among these, 9 SNPs were consistently detected by at least two of the methods employed and were subsequently classified as the strongest candidates for directional selection (Supplementary Figs. 9.4–9.5, Additional File 1). Nearly half of these candidate SNPs (4) were clustered within a 2 Mb region on chromosome SUPER_8, suggesting a potential genomic hotspot under directional selection across Mediterranean macro areas. Three candidate SNPs were distributed across a broader 15 Mb region on chromosome SUPER_9, while the remaining two were scattered on chromosomes SUPER_2 and SUPER_20 (Fig. 3).

Most of the candidate SNPs were positioned within intronic regions of genes, with only two (SUPER_8:21764708 and SUPER_9:154553) being located outside the genes but still in close proximity, approximately 100 bp and 1,000 bp upstream of neighboring genes, respectively. Sequence similarity searches revealed strong homology to genes from other teleost species (Table 1). These findings suggest potential relevance of the identified candidate genomic regions for directional selection.

Detailed summary and associated gene homology for SNPs identified as the strongest candidates for directional selection within the *M. barbatus* genome. The chromosomal position of each SNP, the outlier detection pipelines (flagged by at least two out of three methods), and associated gene names and elements from the genome structural annotation are reported. Additionally, each entry includes the best hit derived from BLASTp searches against the NCBI Reference Protein Database. Protein accession number, description, and species name are provided, along with query coverage and E-values.

Discussion

The first chromosome-scale reference genome for *Mullus barbatus*

The fMulBar1.1 genome assembly achieved chromosome-scale contiguity and high completeness, surpassing key Earth BioGenome Project (EBP) benchmarks for reference genomes. With a scaffold NG50 of 23.7 Mbp and contig NG50 of 12.54 Mbp, the assembly exceeds the EBP minimum threshold for contiguity (NG50 > 1 Mbp for contigs) and meets the criterion for chromosomal-scale scaffolding [113]. Despite its high contiguity and completeness, the genome's base-level accuracy (QV 32.1, ~ 99.93%) falls below the EBP-recommended threshold (QV > 40) for reference-grade genomes [113]. This is consistent with other ONT-based genome assemblies from the ERGA Pilot Project, which tend to have lower QV scores than those generated with PacBio HiFi sequencing [35]. The k-mer completeness of 84.96% also reflects common challenges in ONT-based haploid genome assemblies, where some sequence loss may occur due to haplotypic purging, error correction, and the higher error rate of ONT long-read sequencing (Fig. 1b and c). Overall, the fMulBar1,1 genome assembly represents a highly contiguous and complete chromosome-scale reference genome, making it a valuable resource for population genomics and functional analyses in *Mullus barbatus* and related species.

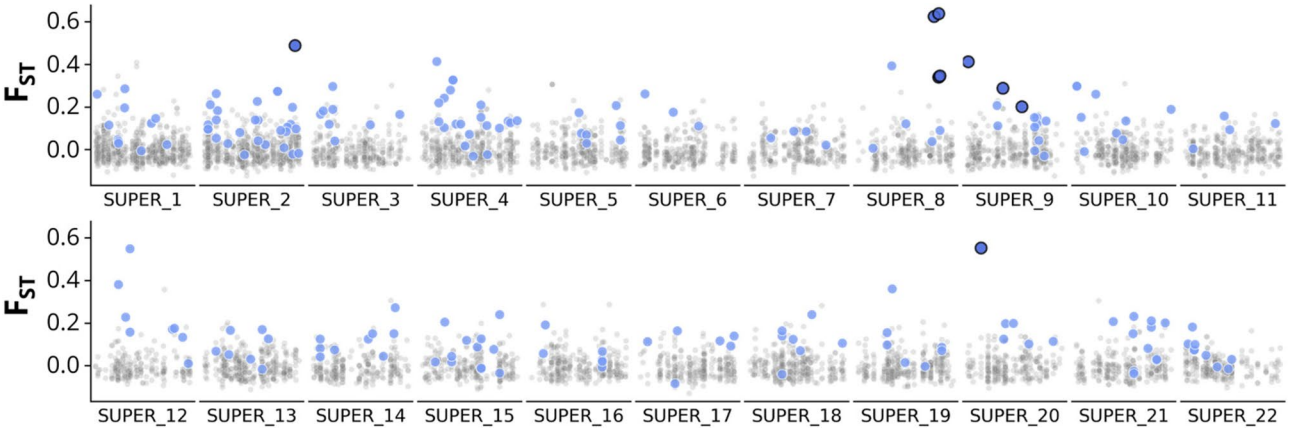

Fig. 3 Genomic localization of candidate loci for directional selection. Genomic positions of candidate SNPs identified as potentially under directional selection within the *Mullus barbatus* genome using the Bayescan, FstHet, and OutFLANK pipelines. The strongest candidates for directional selection are marked with large, dark blue dots, predominantly clustered on chromosome SUPER_8, suggesting a genomic hotspot for selection pressures. Smaller, Light blue Dots represent outliers detected by a single pipeline, indicating additional areas of interest with less robust support. The background of 7,569 SNPs from the dataset is visualized with faint dots to emphasize their overall genomic distribution. F_{ST} values (hierfstat v0.5.11 [112]), are reported for each SNP

Table 1 Summary of candidate loci under directional selection and associated gene homology

Chromosome	Position (bp)	Outliers detection pipeline	Gene name	Gene element	blastp protein accession	blastp protein description	blastp species name	blastp query cover	blastp E-value
SUPER_2	41,814,511	Bayescan, FstHet	g19994	Intron 24 (/64)	XP_044044271.1	cilia- and flagella-associated protein 47-like isoform X6	*Siniperca chuatsi*	94%	0.0
SUPER_8	20,231,385	Bayescan, FstHet, OutFLANK	g33565.t1	Intron 22 (/35)	XP_053733186.1	low-density Lipoprotein receptor-related protein 1B isoform X4	*Synchiropus splendidus*	99%	0.0
SUPER_8	21,439,221	Bayescan, FstHet	g33622	Intron 8 (/12)	XP_033472921.1	polycystic kidney disease protein 1-like 2	*Epinephelus lanceolatus*	92%	0.0
SUPER_8	21,439,227	Bayescan, FstHet, OutFLANK	g33622	Intron 8 (/12)	XP_033472921.1	polycystic kidney disease protein 1-like 2	*Epinephelus lanceolatus*	92%	0.0
SUPER_8	21,764,708	Bayescan, FstHet, OutFLANK	g33637.t1	88 bp upstream	XP_051937411.1	neuropilin-2-like isoform X1	*Hippocampus zosterae*	73%	1,00E-62
SUPER_9	154,553	Bayescan, FstHet, OutFLANK	g34000	1115 bp upstream	XP_033478532.1	myomegalin isoform X7	*Epinephelus lanceolatus*	81%	0.0
SUPER_9	8,799,729	Bayescan, FstHet	g34639.t1	Intron 20 (/24)	XP_044055393.1	adhesion G-protein coupled receptor D2 isoform X1	*Siniperca chuatsi*	99%	0.0
SUPER_9	13,510,067	Bayescan, FstHet	g34862.t1	Intron 7 (/8)	XP_053178793.1	transmembrane protein 59-like	*Scomber japonicus*	79%	0.0
SUPER_20	2,648,926	Bayescan, FstHet, OutFLANK	g20512	Intron 27 (/33)	XP_039983191.1	ribonuclease 3	*Xiphias gladius*	98%	0.0

Challenges in large-scale population genomics of marine species

In this study, we experienced a significant reduction in the number of samples and data ultimately available for the population genomics analysis. This can arise from two main sources: inappropriate selection or application of molecular protocols, and issues encountered during library preparation and sequencing [80]. Additional challenges may stem from insufficient filtering of sequencing data [78].

The availability of a reference genome enables the optimization of molecular protocols through in silico simulations [80]. However, in this study, the lack of a reference genome during dataset production led to the use of ddRAD protocols based on prior studies of other marine fishes [72–75] in which less frequent cutting restriction enzymes (*Sbf*I-*Sph*I) were used aiming to identify SNPs supported by high read coverage.

Issues during library preparation and sequencing represent another significant obstacle. High library complexity - defined by the diversity of unique molecular species present prior to amplification - is crucial for genomic studies, especially for non-model species with limited sample availability [41, 80]. In RAD protocols, low library complexity exacerbates issues as the extensive PCR amplification required amplifies duplicates, reducing data quality and introducing analytical biases. This limitation is particularly problematic in ddRAD protocols, where PCR duplicates cannot be efficiently removed [114–118]. Nonetheless, studies suggest that reliable results can still be obtained with high duplicate rates, provided data are carefully filtered [119, 120].

Sampling protocols adopted in this study emphasized proper tissue preservation methods; however, the operational realities of commercial vessels often prevented optimal preservation conditions. Scientific surveys at sea provide a good opportunity to implement a proper sampling design for genomic analyses. However, sampling for genomic analyses cannot be a collateral task to the many activities carried out on board by scientific staff. Adequate resources should therefore be allocated to this specific activity to ensure proper performance.

Non-optimal preservation conditions are likely to have resulted in reduced DNA quality and increased levels of missing data in our study (Supplementary Materials 3–4, Additional File 1). Missing loci in RADseq data are often attributed to polymorphisms at restriction sites, although this typically has a minor effect unless genetic diversity is exceedingly high [80, 121, 122]. Accordingly, poor library

complexity is likely the primary cause of missing data in this study [41, 80].

High missingness can obscure biological signals in population structure analyses [92]. While imputation can alleviate this issue, it is unsuitable when missingness levels are uneven or excessive [93]. In such cases, filtering loci with high missingness is preferable [39, 78, 123]. However, when poorly sequenced specimens are widespread, this approach can drastically reduce the SNP set. This study indeed faced an inflation of poorly sequenced specimens, which initially resulted in the exclusion of the entire SNP dataset. After an iterative filtering process, the dataset was reduced to 350 specimens, yielding 7,569 SNPs from 1,947 RAD loci widespread across the genome. The final set of high-quality SNPs retained for population structure analyses (1,592) was sufficient to detect subtle genetic differentiation, as supported by in silico simulations. Nonetheless, increased genomic coverage might further enhance the resolution of fine-scale population structure.

This strategy, while ensuring robust data quality, also led to underrepresentation of certain Mediterranean regions. Such sampling imbalance may reduce the ability to detect localized or fine-scale population structure. To explore this further, a complementary dataset was assembled by re-integrating low-missingness individuals from previously excluded localities. Principal Component Analysis confirmed the absence of genetic structuring even within this expanded dataset.

Nonetheless, the unequal sample sizes among locations, along with the complete absence of several Mediterranean regions, still limit the interpretability of fine-scale patterns. Additionally, temporal replication could help assess the stability of the observed panmixia over time. In our case, sampling was explicitly designed to minimize cohort effects by avoiding the inclusion of individuals from different age classes within the same locality. However, implementing a temporally replicated sampling strategy would require substantial logistical effort and was beyond the scope of the present study.

Previous population genomics analyses of the same dataset were conducted at the time the RRS was performed. Due to the absence of a reference genome, RAD loci were built *de novo* [38], which resulted in a Limited number of high-quality SNPs. However, the alignment of sequenced reads to a reference genome, as Done in the present study, significantly increased the number of retained SNPs available for population genomics inferences. Specifically, the number of SNPs for population genomics analyses rose from 853 to 1,592, thereby enhancing the reliability of the inferences [39, 40].

Filtering and parameter optimization protocols may not be fully effective if sequencing data capture little of the true biological signal [80]. Low library complexity can lead to preferential amplification of one allele during PCR, distorting allelic ratios at heterozygous loci and reducing genotype call accuracy [41, 80]. When low-complexity libraries are sequenced at high depth, PCR duplicates accumulate, further masking allelic imbalance and increasing confidence of false homozygote calls for true heterozygotes [41]. Since ddRAD protocols do not allow direct estimation of PCR duplicate rates, quantifying these biases is challenging [41]. In this study, the combination of suboptimal DNA quality and high sequencing depth (Supplementary Materials 3–4, Additional File 1) likely contributed to allelic dropout. Given these limitations, careful consideration is required when interpreting population genomics inferences from the final dataset. To mitigate these limitations in future studies, we recommend prioritizing optimal tissue preservation and employing whole-genome resequencing to enhance data quality and the detection of subtle genetic differentiation by sampling a greater number of linkage groups across the genome. The ability to identify genomic regions under directional selection would be improved as well.

Pan-Mediterranean genetic connectivity in the red mullet

This study is the first to use genomic data to investigate the population structure of *M. barbatus*, suggesting a panmictic population across the Mediterranean. While the previously discussed limitations of the dataset should be considered, the findings still indicate strong genetic flow across the species' range. This conclusion is further supported by the results from otolith shape and microchemistry analyses performed within the MED_UNITs framework, which provided only weak and inconclusive evidence for the presence of two to three sub-populations in the studied region.

Minimal genetic differentiation was already reported from genetic studies [8], but this contradicts other earlier findings suggesting localized genetic structure [8, 23, 25]. Previous genetic analyses were largely based on microsatellite markers, which remain widely used in phylogeographic and conservation genetics studies due to their cost-effectiveness [124]. However, microsatellites have known limitations in resolving fine-scale genetic diversity and population structure [125]. Discrepancies in previous findings may also stem from variations in the number of markers used, methodologies applied, and sampling efforts.

Weak genetic differentiation is a characteristic trait of many marine species, primarily due to extensive dispersal during their larval or adult stages [14, 15, 17, 18]. Red mullet exhibits low adult mobility, but larvae are reported to reach suitable coastal areas to recruit even far from their natal area [20]. Comparable dispersal patterns have been observed in the congeneric species *M. surmuletus*, where genomic analyses have revealed low

genetic differentiation across Mediterranean locations [16, 28]. In this species, larval dispersal appears to be a critical mechanism for maintaining gene flow at small to intermediate spatial scales [28]. On larger spatial scales, connectivity in *M. surmuletus* has been attributed to a multigenerational stepping-stone dispersal model, underscoring the significance of intermediate habitats in preserving genetic connectivity [29, 126–128].

A similar mechanism could explain the observed patterns in *M. barbatus*, as the species' larvae settle on widely distributed muddy seabeds, facilitating effective long-distance dispersal. This phenomenon might also extend to the Adriatic Sea, where genetic differentiation was reported but not consistently observed [8, 23, 25].

Although genetic structure was not detected in this study, the possibility of stock structure cannot be entirely excluded. Self-sustaining stocks have been identified through traditional population dynamics studies based on biomass estimates and Lagrangian models of larval dispersal [3, 129]. This likely reflects the differing temporal scales captured by these approaches as genetic studies provide insights into evolutionary dynamics, whereas classical population dynamics examine ecological-scale variations, which occur over shorter and more recent timescales [3]. Moreover, the inherent limitations of most population genetic approaches pose challenges to accurately inferring demographic connectivity for stock management [13]. Apparent genetic panmixia can arise under various scenarios, ranging from the existence of a single, truly panmictic population to nearly complete demographic independence among large populations [13]. In large marine populations, even minimal migration rates can prevent genetic differentiation [10, 13, 130]. This is likely the case for *M. barbatus*, whose census size is expected to be several orders of magnitude larger than the estimated effective population size of approximately 10^4 individuals reported in this study [83, 131]. Despite the low Ne, this estimate is consistent across multiple methods.

However, estimates of contemporary Ne in marine fishes are known to be downwardly biased, particularly when using linkage disequilibrium. This bias has been attributed mainly to methodological constraints—chiefly the requirement for extremely large sample sizes to produce precise estimates in populations with large census sizes—rather than to biological factors such as variance in reproductive success or overlapping generations [83–85, 131, 132]. Programs such as CurrentNe were specifically developed to address some of these limitations, particularly in cases of high Ne or limited sample sizes [86]. Still, even under ideal conditions, simulations suggest that up to 1% of the total census size may need to be sampled to obtain reliable Ne estimates in large marine populations [83–85, 131, 132].

Interestingly, similar Ne values were reported for *M. surmuletus* using analogous methodologies [16]. Furthermore, historical Ne estimates obtained with MSMC2 showed recent values in the same order of magnitude, providing additional support for the low contemporary estimates observed. Nevertheless, these results must be interpreted with caution, due to the limited sample size available for both LD-based and coalescent-based methods, and the reliance on a single individual for MSMC2.

In such cases, where neutral markers may fail to capture the current level of demographic connectivity, selection can serve as a more effective counterforce to genetic homogenization than drift, particularly as its efficiency increases with population size [13, 133]. Recent conservation genetic studies have proposed delineating locally adapted units using signals from outlier loci, though this approach comes with specific challenges [13]. The small number of outlier loci detected in this study was too small to draw population structure inferences, likely due to the limited portion of the genome sampled with the sequencing method used. Future studies employing whole-genome resequencing could improve stock structure resolution, even in such complex scenarios, by capturing a broader range of linkage groups and, in turn, a greater number of outlier loci [18, 134]. Broader genomic coverage may also uncover fine-scale population structure that could remain undetected with reduced-representation sequencing approaches.

Preliminary evidence for a genomic hotspot under directional selection in *Mullus barbatus*

The identification of loci under directional selection remains a fundamental aspect of population genetics, especially for understanding local adaptation processes. Early outlier detection methods focused on identifying loci with elevated F_{ST} values relative to neutral expectations [135, 136]. Recent advancements in genomic tools have improved the accuracy of these methods, enabling more robust analyses of adaptive divergence [97–99, 137, 138]. Nevertheless, F_{ST}-outlier methods are susceptible to false positives under non-equilibrium scenarios [13, 139], emphasizing the importance of integrating multiple approaches to improve reliability [140–144].

Previous studies have successfully applied such approaches, leading to the identification of genes under putative directional selection in related species. For instance, four genes linked to salinity adaptation were identified in *M. surmuletus*, associated with metabolic functions [27]. More recently, outlier SNP analyses uncovered candidate genes tied to temperature adaptation in the same species, suggesting local adaptation to thermal conditions [16].

In the present study, nine candidate loci for directional selection in *M. barbatus* were identified by intersecting

the results from multiple F_{ST}-outlier detection methods (Table 1). These loci were linked, through homology searches, to genes involved in biological processes relevant to development and environmental response in other teleost species. However, we fully acknowledge that the presence of a well-supported homolog alone is not sufficient to define a gene as a strong candidate for adaptation. Given the relatively limited genomic coverage of the dataset, the genes reported here should be regarded as provisional targets for future investigation. Broader genomic coverage and functional validation will be essential to assess whether these and other loci are truly involved in adaptive processes in *M. barbatus*.

Four candidate loci located within a 2 Mb region on chromosome SUPER_8, suggesting a potential genomic hotspot under directional selection associated with critical biological functions. Notably, one locus was Linked to the low-density Lipoprotein receptor-related protein 1B (LRP1B), a gene involved in oocyte development in oviparous vertebrates [145, 146]. Two additional loci were mapped to the polycystic kidney disease protein 1-like 2 (PKD1L2) gene, which plays a key role in vertebrate development [147]. Another locus was located near the neuropilin-2-like (NRP2L) gene, a highly conserved receptor implicated in neurogenesis and cardiovascular development [148]. Gene similarity analyses of loci in the SUPER_8 region suggests this genomic region may be associated with various aspects of teleost ontogenesis. Natural selection acting on this region could regulate the expression of these genes, potentially influencing fish development in response to environmental conditions or fishing pressures.

Similarly, other candidate loci for directional selection potentially suggest additional genomic hotspots. However, the resolution of these findings is Limited by the relatively small number of loci examined and the preliminary nature of functional annotation derived from sequence similarity. These genes might be involved in essential cellular functions, reflecting the complex evolutionary pressures driving the adaptation of this species. For instance, mutations in the cilia- and flagella-associated protein 47 (CFAP47), located on chromosome SUPER_2, have been linked to abnormalities in sperm morphogenesis and motility [149]. Similarly, the adhesion G protein-coupled receptor D2 (ADGRD2/GPR64), found on chromosome SUPER_9, plays a critical role in cell signaling and adhesion, acting as a key component in the male reproductive system [150]. Another gene on SUPER_9, myomegalin (MMG), particularly its isoform X7, is essential for the organization of the Golgi apparatus and microtubules [151, 152]. The transmembrane protein 59-like (TMEM59L), also on SUPER_9, is involved in intracellular protein trafficking and processing and has been associated with behaviors in mice, including anxiety, depression, and memory [153]. Finally, ribonuclease 3 (RNase3), located on chromosome SUPER_20, plays a critical role in immune defense [154, 155].

Associating these candidate loci with environmental heterogeneity provides valuable insights into the mechanisms underlying local adaptation [16, 27, 140, 141, 156]. However, the limitations of reduced-representation sequencing methods, such as lower genomic coverage, likely exclude many loci that contribute to local adaptation, limiting the scope of our conclusions. As a result, the few candidate loci identified should be considered as preliminary evidence of directional selection. Moreover, the functional interpretation of these loci relies on homology-based annotations from other teleost species, which may not fully capture lineage-specific gene functions in M. barbatus. These annotations offer useful first approximations but should be validated through species-specific functional studies.

Furthermore, the data quality issues extensively discussed in previous sections should be carefully considered when interpreting the final results. In this way, whole-genome resequencing offers a more comprehensive alternative for future studies seeking to uncover additional adaptive loci [122, 157].

Conclusions

This study presents the first high-quality reference genome for *Mullus barbatus*, significantly enhancing our understanding of its population structure, adaptive potential, and evolutionary processes. By integrating this genome with a Mediterranean-wide genomic dataset, we revealed a panmictic population structure with strong genetic connectivity, likely driven by extensive larval dispersal and multigenerational gene flow. Despite minimal genome-wide differentiation, outlier analysis identified candidate loci potentially under directional selection, associated with key aspects of teleost ontogenesis. Nonetheless, given the constrained genomic coverage, these candidate loci should be regarded as a preliminary step toward identifying the full spectrum of adaptive variation in this species. These findings have important implications for the sustainable management of *M. barbatus* fisheries in the Mediterranean, providing evidence for strong genetic connectivity across distant populations. Additionally, the newly generated genome assembly offers the foundation for developing genomic tools that can guide conservation strategies. As climate change continues to reshape marine ecosystems, understanding the genetic basis of adaptation will be essential for managing fish stocks and ensuring their long-term resilience. The preliminary identification of candidate adaptive loci highlights the potential of genomic approaches and paves the way for future studies using whole-genome

sequencing to more comprehensively investigate the adaptive potential of *M. barbatus*, ultimately contributing to the conservation of marine biodiversity and the sustainable use of marine resources.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-025-02443-2.

Supplementary Material 1. DNA extraction and quality control; Supplementary Material 2 - Library preparation and sequencing; Supplementary Material 3 - Raw data processing and genome mapping; Supplementary Material 4 - Call rate thresholds and missing data patterns; Supplementary Material 5 - Iterative filtering; Supplementary Material 6 - Macro-areas subdivision for AMOVA and outliers detection; Supplementary Material 7 - Population structure; Supplementary Material 8 - Population demography with MSMC2; Supplementary Material 9 - Parameters and results of outliers detection pipelines

Supplementary Material 2. Sequenced sample metadata

Acknowledgements
The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Commission and EASME. Neither EASME nor the European Commission can guarantee the accuracy of the data included in this publication. Neither EASME nor the European Commission or any person acting on the EASME's or on the European Commission's behalf may be held responsible for the use which may be made of the information contained therein. We acknowledge the whole MED_UNITs Consortium for the provision of data and previous analyses results, and all MED_UNITs contributors for their efforts in the collection and provision of biological samples. Sequencing of the M. barbatus genome was part of the pilot project of the European Reference Genome Atlas initiative. We acknowledge the contributions of Jo Wood, Dominic Absolon, and James Torrance from the Genome Reference Informatics Team at the Wellcome Sanger Institute for manual curation of the genome assembly. We are grateful to the CoNISMa staff, and particularly to Maddalena Laggini and Chiara Innocenti, for their constant and precious administrative support. The research leading to these results has been carried out under PM's PhD Program in "Earth, Life and Environmental Sciences".

Authors' contributions
Language used to describe roles below uses the CRediT Taxonomy [162]. Conceptualization: P.M., T.M., A.C., H.S., D.T., R.C., L.Z., G.F., A.M.C., A.M., C.S.T. Data curation: P.M., T.M., H.G.L., M.A.D. Formal Analysis: P.M., T.M., H.G.L., D.T., L.Z., C.S.T. Funding acquisition: A.C., H.S., R.C., L.Z., M.T.S., C.S.T. Investigation: P.M., T.M., H.G.L., D.T., T.D.P., M.S., G.D., C.N. Methodology: P.M., T.M., H.G.L., D.T., T.D.P., M.S. Project administration: A.C., H.S., R.C., M.T.S., C.S.T. Resources: A.C., H.S., R.C., L.Z., T.D.P., M.T.S., M.S., C.C., P.I., C.S.T. Software: P.M. Supervision: T.M., A.C., H.S., D.T., A.F., C.S.T. Validation: P.M., H.G.L. Visualization: P.M., H.G.L., A.F., C.C. Writing original draft: P.M., H.G.L., C.S.T. All authors commented on the article before submission. All authors read and approved the final manuscript.

Funding
MED_UNITs project - Study on Advancing fisheries assessment and management advice in the Mediterranean and Black Sea by aligning biological and management units of priority species - was carried out within the European Research Contract (EASME/EMFF/2017/1.3.2.3/01/SI2.793201, implementing the Framework Contract EASME/EMFF/2016/032) funded by the Executive Agency for Small-and Medium-sized Enterprises (EASME) of the European Commission (DG MARE). H.S. acknowledges support from the Flemish University Research Fund (BOF) and from the University of Antwerp through a startup grant. H.G.L. was supported by the Research Foundation—Flanders (FWO PhD fellowship 1156622 N). Sequencing of Omni-C and RNA libraries was supported by the University of Florence through internal funding to CC. NovaSeq 6000 Reagent Kits for Omni-C and RNA libraries sequencing were provided by Illumina EMEA support to the European Reference Genome Atlas pilot project. We also would like to thank Dovetail Genomics, Part of Cantata Bio LLC; Integrated DNA Technologies (IDT); MagBio Genomics Europe GmbH; Zymo Research; Agilent Technologies that have kindly donated kits, reagents to the ERGA pilot Library preparation Hubs. A.C., P.M. and A.F. acknowledge support from the University of Bologna (Canziani legacy and RFO grants).

Data availability
The genome assembly and associated sequencing data have been deposited in the NCBI databases under BioProject PRJNA1244255 [158]. The genome annotation file is available on Zenodo at DOI: 10.5281/zenodo.15209764. [159]. Raw ddRAD sequencing data used for population genomics analyses are deposited in the NCBI database under BioProject PRJNA1244614 [160]. Codes and genetic variant files used for population genomics can be accessed in the PiergiorgioMassa/mbarb_medunits_popgen GitHub repository [161].

Declarations

Ethics approval and consent to participate
All samples used in this study were obtained from either commercial fishing operations or scientific research surveys, conducted in accordance with the European Union's Common Fisheries Policy (Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013). To facilitate the collection of samples from southern and eastern Mediterranean countries, a liaison was established between the Directorate-General for Maritime Affairs and Fisheries (DG MARE), the General Fisheries Commission for the Mediterranean (GFCM), and the Food and Agriculture Organization (FAO) Regional Projects, within the framework of the Data Collection Regulation Framework (DCRF). The fishing operations involved the harvest of fish through standard practices, which inherently result in mortality. No experimental procedures were conducted on live animals, and as such, ethical approval was not required.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Laboratory of Genetics & Genomics of Marine Resources and Environment (GenoDREAM), Department of Biological, Geological and Environmental Sciences, University of Bologna, Via Sant'Alberto 163, Ravenna 48123, Italy
[2]Evolutionary Ecology, Department of Biology, University of Antwerp, Antwerp, Belgium
[3]Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (H.C.M.R.), Heraklion 715 00, Crete, Greece
[4]National Inter-University Consortium for Marine Sciences (CoNISMa), Piazzale Flaminio 9, Roma 00196, Italy
[5]Naturalis Biodiversity Center, Leiden, The Netherlands
[6]Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy
[7]Department of Biology (DiBio), University of Padova, Via Ugo Bassi 58/b, Padova 35131, Italy
[8]Neuromics Support Facility, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium
[9]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[10]Fondazione COISPA ETS, Bari, Italy
[11]Department of Biology, University of Florence, Sesto Fiorentino (FI), Florence, Italy
[12]The Vertebrate Genome Laboratory, The Rockefeller University, 1230 York Ave, New York 10021, USA
[13]Institute of Oceanology – Bulgarian Academy of Sciences, Parvi may Street 40, PO 152 9000, Varna, Bulgaria
[14]Genomics Institute, University of California, Santa Cruz, California, USA
[15]Department of Socio-économie, Environnement et Développement (SEED), University of Liège, Liège, Belgium
[16]National Biodiversity Future Center (NBFC), Piazza Marina 61, Palermo 90133, Italy

Received: 3 April 2025 / Accepted: 27 August 2025
Published online: 07 October 2025

References

1. Fischer W, Schneider M, Bauchot M-L, Fisheries and Aquaculture Management Division. Fiches FAO d'identification des espèces pour les besoins de La Pêche. (Révision 1). Méditerranée et Mer noire. Zone de Pêche 37. II: vertébrés. Publication préparée par La FAO, résultant d'un accord entre La FAO et La commission des communautés Européennes (Projet GCP/INT/422/EEC), financé conjointement par Ces Deux organisations. Rome FAO. 1987;2:761–1530. https://openknowledge.fao.org/handle/20.500.14283/x0170f. Accessed 3 April 2025.
2. Lombarte A, Recasens L, González M, de Sola LG. Spatial segregation of two species of mullidae (*Mullus surmuletus* and *M. barbatus*) in relation to habitat. Mar Ecol Prog Ser. 2000;206:239–49.
3. Tserpes G, Massutí E, Fiorentino F, Facchini MT, Viva C, Jadaud A, et al. Distribution and spatio-temporal biomass trends of red mullets across the Mediterranean. Sci Mar. 2019;83(S1):43–55.
4. Papaconstantinou C, Farrugio H. Fisheries in the Mediterranean. Mediterr Mar Sci. 2000;1(1):5–18.
5. FAO. The State of Mediterranean and Black Sea Fisheries 2023 – Special edition. General Fisheries Commission for the Mediterranean. Rome. 2023. https://doi.org/10.4060/cc8888en. Accessed 3 April 2025.
6. Caddy JF, Pere O. Some future perspectives for assessment and management of mediterranean fisheries for demersal and shellfish resources, and small pelagic fish. Resource Environ Issues Relevant Mediterranean Fisheries Manage. 1996;Studies and Reviews GFCM:66–19. https://www.fao.org/docrep/pdf/006/w0509e/w0509e03.pdf. Accessed 3 April 2025.
7. Farrugio H, Oliver P, Biagi F. An overview of the history, knowledge, recent and future research trends in mediterranean fisheries. Sci Mar. 1993;57(2–3):105–19.
8. Matić-Skoko S, Šegvić-Bubić T, Mandić I, Izquierdo-Gomez D, Arneri E, Carbonara P, et al. Evidence of subtle genetic structure in the sympatric species *mullus barbatus* and *mullus surmuletus* (Linnaeus, 1758) in the Mediterranean Sea. Sci Rep. 2018;8(1):676.
9. GFCM. Report of the twenty-fifth session of the Scientific Advisory Committee on Fisheries. Marseille, France, 24–27 June 202/Rapport de la vingt-cinquième session du Comité scientifique consultatif des pêches, Marseille, France, 24–27 juin 2024. https://www.fao.org/gfcm/statutory-meetings/detail/en/c/1712734. Accessed 3 April 2025.
10. Waples R. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. J Hered. 1998;89(5):438–50.
11. Hellberg ME. Gene flow and isolation among populations of marine animals. Annu Rev Ecol Evol Syst. 2009;40(40, 2009):291–310.
12. Selkoe KA, Toonen RJ. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Mar Ecol Prog Ser. 2011;436:291–305.
13. Gagnaire PA, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F, et al. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evol Appl. 2015;8(8):769–86.
14. Moore GI, Chaplin JA. Population genetic structures of three congeneric species of coastal pelagic fishes (*Arripis*: Arripidae) with extensive larval, postsettlement and adult movements. Environ Biol Fishes. 2013;96(9):1087–99.
15. Gkagkavouzis K, Karaiskou N, Katopodi T, Leonardos I, Abatzopoulos TJ, Triantafyllidis A. The genetic population structure and temporal genetic stability of gilthead sea bream *Sparus aurata* populations in the Aegean and Ionian seas, using microsatellite DNA markers. J Fish Biol. 2019;94(4):606–13.
16. Boulanger E, Benestan L, Guerin PE, Dalongeville A, Mouillot D, Manel S. Climate differently influences the genomic patterns of two sympatric marine fish species. J Anim Ecol. 2022;91(6):1180–95.
17. Caballero-Huertas M, Frigola-Tepe X, Coll M, Muñoz M, Viñas J. The current knowledge status of the genetic population structure of the European sardine (*Sardina pilchardus*): uncertainties to be solved for an appropriate fishery management. Rev Fish Biol Fish. 2022;32(3):745–63.
18. Andrews AJ, Eriksen EF, Star B, Præbel K, Natale AD, Malca E, et al. Ancient DNA reveals historical demographic decline and genetic erosion in the Atlantic Bluefin tuna. BioRxiv. 2024. https://doi.org/10.1101/2024.09.14.613028. Accessed 3 April 2025.
19. Macpherson E, Raventos N. Relationship between pelagic larval duration and geographic distribution of Mediterranean littoral fishes. Mar Ecol Prog Ser. 2006;327:257–65.
20. Gargano F, Garofalo G, Fiorentino F. Exploring connectivity between spawning and nursery areas of *mullus barbatus* (L., 1758) in the mediterranean through a dispersal model. Fish Oceanogr. 2017;26(4):476–97.
21. Cadrin SX. Defining spatial structure for fishery stock assessment. Fish Res. 2020;221:105397.
22. Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish Fish. 2009;10(4):361–95.
23. Garoia F, Guarniero I, Piccinetti C, Tinti F. First microsatellite loci of red mullet (*Mullus barbatus*) and their application to genetic structure analysis of Adriatic shared stock. Mar Biotechnol. 2004;6(5):446–52.
24. Galarza JA, Turner GF, Macpherson E, Rico C. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). Can J Fish Aquat Sci. 2009;66(9):1478–90.
25. Maggio T, Lo Brutto S, Garoia F, Tinti F, Arculeo M. Microsatellite analysis of red mullet *Mullus barbatus* (Perciformes, Mullidae) reveals the isolation of the Adriatic basin in the Mediterranean Sea. ICES J Mar Sci. 2009;66(9):1883–91.
26. Félix-Hackradt FC, Hackradt CW, Pérez-Ruzafa Á, García-Charton JA. Discordant patterns of genetic connectivity between two sympatric species, *Mullus barbatus* (Linnaeus, 1758) and *Mullus surmuletus* (Linnaeus, 1758), in southwestern Mediterranean Sea. Mar Environ Res. 2013;92:23–34.
27. Dalongeville A, Benestan L, Mouillot D, Lobreaux S, Manel S. Combining six genome scan methods to detect candidate genes to salinity in the mediterranean striped red mullet (*Mullus surmuletus*). BMC Genomics. 2018;19(1):1–13.
28. Dalongeville A, Andrello M, Mouillot D, Lobreaux S, Fortin MJ, Lasram F, et al. Geographic isolation and larval dispersal shape seascape genetic patterns differently according to spatial scale. Evol Appl. 2018;11(8):1437–47.
29. Boulanger E, Dalongeville A, Andrello M, Mouillot D, Manel S. Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. Ecography. 2020;43(8):1167–79.
30. Fietz K, Trofimenko E, Guerin PE, Arnal V, Torres-Oliva M, Lobréaux S, et al. New genomic resources for three exploited mediterranean fishes. Genomics. 2020;112(6):4297–303.
31. European Commission: Agencia Estatal Consejo superior de Investigaciones Científicas CSIC, COISPA Tecnologia &, Ricerca C, Nazionale delle Ricerche IRBIM, Consorzio Nazionale Interuniversitario per le Scienze del Mare CoNISMa, Consorzio per il Centro Interuniversitario di Biologia Marina Ecologia Applicata G. Bacci di Livorno (CIBM), European Climate, Infrastructure and Environment Executive Agency, Hellenic Centre of Marine Research HCMR, IFREMER Département Ressources Biologiques et Environnement - Boulogne sur mer, Instituto Español de Oceanografía IEO, Istituto Nazionale di Oceanografica e di Geofisica, Sperimentale OGS, Spedicato M, Cannas R, Mahé K, Morales B, Tsigenopoulos C, Zane L et al. Study on advancing fisheries assessment and management advice in the Mediterranean by aligning biological and management units of priority species MED_UNITs: final report. Publications Office of the European Union. 2022. https://doi.org/10.2926/909535. Accessed 3 April 2025.
32. Hohenlohe PA, Funk WC, Rajora OP. Population genomics for wildlife conservation and management. Mol Ecol. 2021;30(1):62–82.
33. Theissinger K, Fernandes C, Formenti G, Bista I, Berg PR, Bleidorn C, et al. How genomics can help biodiversity conservation. Trends Genet. 2023;39(7):545–59.
34. D'Aloia CC, Andrés JA, Bogdanowicz SM, McCune AR, Harrison RG, Buston PM. Unraveling hierarchical genetic structure in a marine metapopulation: a comparison of three high-throughput genotyping approaches. Mol Ecol. 2020;29(12):2189–203.
35. Mc Cartney AM, Formenti G, Mouton A, De Panis D, Marins LS, Leitão HG, et al. The European reference genome atlas: piloting a decentralised approach to equitable biodiversity genomics. Npj Biodivers. 2024;3(1):1–17.
36. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth biogenome project: sequencing life for the future of life. Proc Natl Acad Sci U S A. 2018;115(17):4325–33.
37. Mazzoni CJ, Ciofi C, Waterhouse RM. Biodiversity: an atlas of European reference genomes. Nature. 2023;619(7969):252–252.
38. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22(11):3124–40.
39. Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for stacks. Methods Ecol Evol. 2017;8(10):1360–73.

40. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol. 2019;28(21):4737–54.
41. Rochette NC, Rivera-Colón AG, Walsh J, Sanger TJ, Campbell-Staton SC, Catchen JM. On the causes, consequences, and avoidance of PCR duplicates: towards a theory of library complexity. Mol Ecol Resour. 2023;23(6):1299–318.
42. Oxford Nanopore Technologies. Guppy protocol. 2018. https://nanoporetech.com/document/Guppy-protocol. Accessed 3 April 2025.
43. Wick R, rrwick/Filtlong GH. 2025. https://github.com/rrwick/Filtlong. Accessed 3 April 2025.
44. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):245.
45. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11(1):1432.
46. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6.
47. Challis R, Kumar S, Sotero-Caio C, et al. Genomes on a tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. Wellcome Open Res. 2023;8(24). https://doi.org/10.12688/18658.1version 1; peer review: 2 approved. Accessed 3 April 2025.
48. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.
49. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
50. Oxford Nanopore Technologies, nanoporetech/medaka GH. 2025. https://github.com/nanoporetech/medaka. Accessed 3 April 2025.
51. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36(9):2896–8.
52. Zhou C, McCarthy SA, Durbin R. Yahs: yet another Hi-C scaffolding tool. Bioinformatics. 2023;39(1):btac808.
53. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. https://doi.org/10.48550/arXiv.1303.3997. Accessed 3 April 2025.
54. Open C 2, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, et al. Pairtools: from sequencing data to chromosome contacts. PLoS Comput Biol. 2024;20(5):e1012164.
55. Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. Bioinformatics. 2022;38(17):4214–6.
56. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54.
57. Tree of Life programme. sanger-tol/PretextMap, GitHub. 2025. https://github.com/sanger-tol/PretextMap. Accessed 3 April 2025.
58. Tree of Life programme. sanger-tol/PretextView: OpenGL Powered Pretext Contact Map Viewer, GitHub. https://github.com/sanger-tol/PretextView. Accessed 3 April 2025.
59. Tree of Life programme. sanger-tol/PretextSnapshot, GitHub. https://github.com/sanger-tol/PretextSnapshot. Accessed 3 April 2025.
60. Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Blaxter M, Mieszkowska N, Hall N, et al. Mitohifi: a python pipeline for mitochondrial genome assembly from Pacbio high fidelity reads. BMC Bioinformatics. 2023;24(1):288.
61. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25(16):2078–9.
63. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024;34(5):769–77.
64. Li H. Protein-to-genome alignment with miniprot. Bioinformatics. 2023;39(1):btad014.
65. NBIS - National Bioinformatics Infrastructure Sweden. NBISweden/AGAT, GitHub. https://github.com/NBISweden/AGAT. Accessed 3 April 2025.
66. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
67. The UniProt Consortium. The universal protein resource (UniProt). Nucleic Acids Res. 2008;36(suppl1):D190-5.
68. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. Interproscan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
69. Spedicato MT, Massutí E, Mérigot B, Tserpes G, Jadaud A, Relini G. The MEDITS trawl survey specifications in an ecosystem approach to fishery management. Sci Mar. 2019;83(S1):9–20.
70. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16(3):1215–1215.
71. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest radseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE. 2012;7(5):e37135.
72. Manousaki T, Tsakogiannis A, Taggart JB, Palaiokostas C, Tsaparis D, Lagnel J, et al. Exploring a nonmodel teleost genome through RAD sequencing—linkage mapping in common pandora, *Pagellus erythrinus* and comparative genomic analysis. G3: Genes, genomes, genetics. 2016;6(3):509–19.
73. Kyriakis D, Kanterakis A, Manousaki T, Tsakogiannis A, Tsagris M, Tsamardinos I, et al. Scanning of genetic variants and genetic mapping of phenotypic traits in Gilthead sea bream through DdRAD sequencing. Front Genet. 2019;10:675.
74. Oikonomou S, Tsakogiannis A, Kriaridou C, Danis T, Manousaki T, Chatziplis D, et al. First linkage maps and a pilot QTL analysis for early growth performance in common dentex (*Dentex dentex*) and sharpsnout seabream (*Diplodus puntazzo*). Aquac Rep. 2021;21:100855.
75. Katirtzoglou A, Tsaparis D, Kolios E, Magoulas A, Mylonas CC, Fakriadis I, et al. Population genomic analysis of the greater Amberjack (*Seriola dumerili*) in the Mediterranean and the Northeast Atlantic, based on SNPs, microsatellites, and mitochondrial DNA sequences. Front Fish Sci. 2024;2:1356313.
76. Andrews S, s-andrews/FastQC GH. 2025. https://github.com/s-andrews/FastQC. Accessed 3 April 2025.
77. Ewels P, Magnusson M, Lundin S, Käller M. Multiqc: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–8.
78. Hemstrom W, Grummer JA, Luikart G, Christie MR. Next-generation data filtering in the genomics era. Nat Rev Genet. 2024;25(11):750–67.
79. O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. These aren't the loci you'e looking for: principles of effective SNP filtering for molecular ecologists. Mol Ecol. 2018;27(16):3193–206.
80. Rivera-Colón AG, Rochette NC, Catchen JM. Simulation with radinitio improves RADseq experimental design and sheds light on sources of missing data. Mol Ecol Resour. 2021;21(2):363–78.
81. Ryman N, Palm S. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Mol Ecol Notes. 2006;6(3):600–2.
82. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. Neestimator v2: re-implementation of software for the estimation of contemporary effective population size (N) from genetic data. Mol Ecol Resour. 2014;14(1):209–14.
83. Marandel F, Lorance P, Berthelé O, Trenkel VM, Waples RS, Lamy JB. Estimating effective population size of large marine populations, is it feasible? Fish Fish. 2019;20(1):189–98.
84. Waples RS. Practical application of the linkage disequilibrium method for estimating contemporary effective population size: a review. Mol Ecol Resour. 2024;24(1):e13879.
85. Mamoozadeh NR, Wade MJ, Reid BN, Bardwell E, Collins EE, Hugentobler SA, et al. A practical introduction to effective population size for fisheries management. Trans Am Fish Soc. 2025;154(4):352–71.
86. Santiago E, Caballero A, Köpke C, Novo I. Estimation of the contemporary effective population size from SNP data while accounting for mating structure. Mol Ecol Resour. 2024;24(1):e13890.
87. Schiffels S, Wang K. MSMC and MSMC2: the multiple sequentially Markovian coalescent. In: Dutheil JY, editor. Statistical population genomics. New York (NY): Humana; 2020. pp. 147–65.
88. Wang K, Mathieson I, O'Connell J, Schiffels S. Tracking human population structure through time from whole genome sequences. PLoS Genet. 2020;16(3):e1008552.
89. Liu S, Hansen MM. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. Mol Ecol Resour. 2017;17(4):631–41.
90. Beichman AC, Huerta-Sanchez E, Lohmueller KE. Using genomic data to infer historic population dynamics of nonmodel organisms. Annu Rev Ecol Evol Syst. 2018;49(49, 2018):433–56.

91. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403–5.
92. Elhaik E. Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. Sci Rep. 2022;12(1):14683.
93. Yi X, Latch EK. Nonrandom missing data can bias principal component analysis inference of population genetic structure. Mol Ecol Resour. 2022;22(2):602–11.
94. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59.
95. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005;14(8):2611–20.
96. Earl DA, von Holdt BM. Structure harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61.
97. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 2015;15(5):1179–91.
98. Weir BS, Cockerham CC. Estimating F-Statistics for the analysis of population structure. Evolution. 1984;38(6):1358–70.
99. Pembleton LW, Cogan NOI, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol Ecol Resour. 2013;13(5):946–52.
100. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.
101. Paradis E. Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 2010;26(3):419–20.
102. Area of application | General Fisheries Commission for the Mediterranean - GFCM | Food and Agriculture Organization of the United Nations. https://www.fao.org/gfcm/about/area-of-application/en/. Accessed 3 April 2025.
103. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14(6):927–30.
104. Jeffery N, Stanley R. rystanley/CartDist: CartDist: Re-projection tool for complex marine systems (1.0.1). Zenodo. 2017. https://doi.org/10.5281/zenodo.802875. Accessed 3 April 2025.
105. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics. 2008;180(2):977–93.
106. Flanagan SP, Jones AG. Constraints on the FST–heterozygosity outlier approach. J Hered. 2017;108(5):561–73.
107. Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. Am Nat. 2015;186(S1):S24-36.
108. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.
109. Pertea G, Pertea MGFF, Utilities. GffRead and GffCompare [version 2; peer review: 3 approved]. F1000Research. 2020,9:304. https://doi.org/10.12688/f1000research.23297.2. Accessed 3 April 2025.
110. RefSeq:. April NCBI Reference Sequence Database. https://www.ncbi.nlm.nih.gov/refseq/. Accessed 3 2025.
111. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV. An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet. 2013;4:98.
112. Goudet J. Hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes. 2005;5(1):184–6.
113. Lawniczak MKN, Durbin R, Flicek P, Lindblad-Toh K, Wei X, Archibald JM, et al. Standards recommendations for the Earth biogenome project. Proc Natl Acad Sci U S A. 2022;119(4):e2115639118.
114. DaCosta JM, Sorenson MD. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS ONE. 2014;9(9):e106713.
115. Schweyen H, Rozenberg A, Leese F. Detection and removal of PCR duplicates in population genomic DdRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. Biol Bull. 2014;227(2):146–60.
116. Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, et al. Radcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Resour. 2016;16(5):1264–78.
117. Franchini P, Monné Parera D, Kautt AF, Meyer A. QuaddRAD: a new high-multiplexing and PCR duplicate removal DdRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. Mol Ecol. 2017;26(10):2783–95.
118. Magbanua ZV, Hsu CY, Pechanova O, Arick M, Grover CE, Peterson DG. Innovations in double digest restriction-site associated DNA sequencing (ddRAD-Seq) method for more efficient SNP identification. Anal Biochem. 2023;662:115001.
119. Díaz-Arce N, Rodríguez-Ezpeleta N. Selecting RAD-seq data analysis parameters for population genetics: the more the better?? Front Genet. 2019;10:533.
120. Euclide PT, McKinney GJ, Bootsma M, Tarsa C, Meek MH, Larson WA. Attack of the PCR clones: rates of clonality have little effect on RAD-seq genotype calls. Mol Ecol Resour. 2020;20(1):66–78.
121. Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol. 2013;22(11):3165–78.
122. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17(2):81–92.
123. Cerca J, Maurstad MF, Rochette NC, Rivera-Colón AG, Rayamajhi N, Catchen JM, et al. Removing the bad apples: a simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. Methods Ecol Evol. 2021;12(5):805–17.
124. Abdul-Muneer PM. Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. Genet Res Int. 2014;2014(1):691759.
125. Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol Evol. 2014;4(22):4399–428.
126. Poortvliet M, Longo GC, Selkoe K, Barber PH, White C, Caselle JE, et al. Phylogeography of the California sheephead, *Emicossyphus pulcher*: the role of deep reefs as stepping stones and pathways to antitropicality. Ecol Evol. 2013;3(13):4558–71.
127. D'Aloia CC, Bogdanowicz SM, Francis RK, Majoris JE, Harrison RG, Buston PM. Patterns, causes, and consequences of marine larval dispersal. Proc Natl Acad Sci U S A. 2015;112(45):13940–5.
128. Crandall ED, Toonen RJ, Laboratory T, Selkoe KA. A coalescent sampler successfully detects biologically meaningful population structure overlooked by F-statistics. Evol Appl. 2019;12(2):255–65.
129. Quattrocchi F, Fiorentino F, Gargano F, Garofalo G. The role of larval transport on recruitment dynamics of red mullet (*Mullus barbatus*) in the central Mediterranean Sea. Mar Environ Res. 2024;202:106814.
130. Lowe WH, Allendorf FW. What can genetics tell us about population connectivity? Mol Ecol. 2010;19(15):3038–51.
131. Waples RS. Tiny estimates of the ne/n ratio in marine fishes: are they real? J Fish Biol. 2016;89(6):2479–504.
132. Delord C, Arnaud-Haond S, Leone A, Noskova E, Tournebize R, Jacques P, et al. Effective population size estimation in large marine populations: considering current challenges and opportunities when simulating large data sets with high-density genomic information. Evol Appl. 2025;18(8):e70121.
133. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11(10):697–709.
134. Weist P, Kusche H, Tørresen O, Hermida M, Lopes E, Jentoft S, et al. Genomic differentiation and interoceanic population structure of two large pelagic scombrid species. Glob Ecol Conserv. 2024;54:e03117.
135. Lewontin RC, Krakauer J, DISTRIBUTION OF GENE FREQUENCY AS A TEST OF, THE THEORY OF THE SELECTIVE NEUTRALITY OF POLYMORPHISMS. Genetics. 1973;74(1):175–95.
136. Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B Biol Sci. 1997;263(1377):1619–26.
137. Guillot G, Vitalis R, Rouzic A, le, Gautier M. Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. Spat Stat. 2014;8:145–55.
138. Luu K, Bazin E, Blum MGB. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 2017;17(1):67–77.
139. Bierne N, Roze D, Welch JJ. Pervasive selection or is it… why are FST outliers sometimes so frequent? Mol Ecol. 2013;22(8):2061–4.
140. Antoniou A, Manousaki T, Ramírez F, Cariani A, Cannas R, Kasapidis P, et al. Sardines at a junction: seascape genomics reveals ecological and oceanographic drivers of variation in the NW mediterranean sea. Mol Ecol. 2023;32(7):1608–28.

141. Corti R, Piazza E, Armelloni EN, Ferrari A, Geffen AJ, Maes GE, et al. A multidisciplinary approach to describe population structure of *Solea solea* in the Mediterranean Sea. Front Mar Sci. 2024;11:1372743.
142. Cunha RL, Robalo JI, Francisco SM, Farias I, Castilho R, Figueiredo I. Genomics goes deeper in fisheries science: the case of the blackspot seabream (*Pagellus bogaraveo*) in the Northeast Atlantic. Fish Res. 2024;270:106891.
143. Leone A, Arnaud-Haond S, Babbucci M, Bargelloni L, Coscia I, Damalas D, et al. Population genomics of the blue shark, *Prionace glauca*, reveals different populations in the Mediterranean Sea and the Northeast Atlantic. Evol Appl. 2024;17(9):e70005.
144. Ollé-Vilanova J, Hajjej G, Macias D, Saber S, Lino PG, Muñoz-Lechuga R, et al. Atlantic bonito (*Sarda sarda*) genomic analysis reveals population differentiation across Northeast Atlantic and Mediterranean locations: implications for fishery management. Mar Environ Res. 2024;196:106408.
145. Reading BJ, Hiramatsu N, Schilling J, Molloy KT, Glassbrook N, Mizuta H, et al. Lrp13 is a novel vertebrate lipoprotein receptor that binds vitellogenins in teleost fishes. J Lipid Res. 2014;55(11):2287–95.
146. Babio L, Damsteegt EL, Lokman PM. Lipoprotein receptors in ovary of eel, *Anguilla australis*: molecular characterisation of putative vitellogenin receptors. Fish Physiol Biochem. 2023;49(1):117–37.
147. England SJ, Campbell PC, Banerjee S, Swanson AJ, Lewis KE. Identification and expression analysis of the complete family of zebrafish Pkd genes. Front Cell Dev Biol. 2017;5:5.
148. Schellenburg S, Schulz A, Poitz DM, Muders MH. Role of neuropilin-2 in the immune system. Mol Immunol. 2017;90:239–44.
149. Liu C, Tu C, Wang L, Wu H, Houston BJ, Mastrorosa FK, et al. Deleterious variants in X-linked CFAP47 induce asthenoteratozoospermia and primary male infertility. Am J Hum Genet. 2021;108(2):309–23.
150. Demberg LM, Rothemund S, Schöneberg T, Liebscher I. Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor GPR64/ADGRG2. Biochem Biophys Res Commun. 2015;464(3):743–7.
151. Roubin R, Acquaviva C, Chevrier V, Sedjaï F, Zyss D, Birnbaum D, et al. Myomegalin is necessary for the formation of centrosomal and Golgi-derived microtubules. Biol Open. 2012;2(2):238–50.
152. Wang Z, Zhang C, Qi RZ. A newly identified myomegalin isoform functions in golgi microtubule organization and ER–Golgi transport. J Cell Sci. 2014;127(22):4904–17.
153. Zheng Q, Zheng X, Zhang L, Luo H, Qian L, Fu X, et al. The neuron-specific protein TMEM59L mediates oxidative stress-induced cell death. Mol Neurobiol. 2017;54(6):4189–200.
154. Domachowske JB, Dyer KD, Adams AG, Leto TL, Rosenberg HF. Eosinophil cationic protein/rnase 3 is another RNase A-family ribonuclease with direct antiviral activity. Nucleic Acids Res. 1998;26(14):3358–63.
155. Pulido D, Torrent M, Andreu D, Nogués MV, Boix E. Two human host defense ribonucleases against mycobacteria, the eosinophil cationic protein (RNase 3) and RNase 7. Antimicrob Agents Chemother. 2013;57(8):3797–805.
156. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24(17):4348–70.
157. Martchenko D, Shafer ABA. Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. Heredity. 2023;131(4):273–81.
158. Bioproject NCBI. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1244255. Accessed 14/04/2025.
159. Repository. - Zenodo. https://doi.org/10.5281/zenodo.15209763. Accessed 14/04/2025.
160. BioProject. April – NCBI. https://www.ncbi.nlm.nih.gov/bioproject/1244614. Accessed 3 2025.
161. PiergiorgioMassa, PiergiorgioMassa/mbarb_medunits_popgen GH. 2025. https://github.com/PiergiorgioMassa/mbarb_medunits_popgen. Accessed 3 April 2025.
162. CRediT. https://credit.niso.org/. Accessed 3 April 2025.

Publisher's Note