University of Crete
Department of Biology

Population genomic analysis of the European anchovy (*Engraulis encrasicolus*) in the Western and Central Mediterranean

Master thesis

Damianos Alexandridis

Scientific committee:
Tsigenopoulos C., Manousaki T., Somarakis S.

Heraklion, 2024

Πανεπιστήμιο Κρήτης
Τμήμα Βιολογίας

Γονιδιωματική πληθυσμών του Ευρωπαϊκού γαύρου (*Engraulis encrasicolus*) της Δυτικής και Κεντρικής Μεσογείου

Μεταπτυχιακή διατριβή

Δαμιανός Αλεξανδρίδης

Επιστημονική επιτροπή:
Τσιγγενόπουλος Κ., Μανουσάκη Τ., Σωμαράκης Σ.

Ηράκλειο, 2024

Contents

Abstract

Genomic studies have recently become a powerful tool to gain compelling insights into species recent evolutionary history, patterns of spatial genetic structure and connectivity. The European anchovy (*Engraulis encrasicolus*), one of the most important pelagic fish in the Mediterranean Sea, constitutes a keystone species within the marine ecosystem due to its role in regulating trophic interactions, while also contributing significantly to the region's fishing industry. The aim of the study was to assess the European anchovy's population genetic structure in the Western and Central Mediterranean using a genomic approach. To identify a great number of polymorphic loci, we first assembled a draft reference genome of 1.69 Gb and 79.8% BUSCO completeness serving as a powerful mean for our downstream analysis. Then, we mapped ddRAD-sequencing data, generated from 398 anchovy individuals across twelve Western and Central Mediterranean sites, which led to the successful genotyping and further study based on 9,497 single nucleotide polymorphisms (SNPs). Results showed that our samples were divided into two clusters: one grouping the individuals from the Atlantic and the Alboran Sea, and the other including individuals from the Northwestern and the Central Mediterranean Seas (F_{ST}=0.09). This differentiation highlights the presence of two distinct genetic pools, which is in alignment with prior research on European anchovy, indicating once again the Almeria–Oran front as a potential dispersal barrier for gene flow in the species. Consequently, our research sets the ground for delineating distinct stocks and implementing targeted conservation strategies to safeguard the species from over-exploitation and habitat degradation.

Περίληψη

Οι γονιδιωματικές μελέτες έχουν γίνει πρόσφατα ένα ισχυρό εργαλείο για την απόκτηση γνώσεων σχετικά με την πρόσφατη εξελικτική ιστορία των ειδών, τα πρότυπα της χωρικής γενετικής δομής και συνδεσιμότητας. Ο ευρωπαϊκός γαύρος (*Engraulis encrasicolus*), ένα από τα σημαντικότερα πελαγικά ψάρια στη Μεσόγειο Θάλασσα, αποτελεί βασικό είδος στο θαλάσσιο οικοσύστημα λόγω του ρόλου του στη ρύθμιση των τροφικών αλληλεπιδράσεων, ενώ παράλληλα συμβάλλει σημαντικά στην αλιευτική βιομηχανία της περιοχής. Σκοπός της παρούσας έρευνας ήταν η αξιολόγηση της γενετικής δομής του πληθυσμού του ευρωπαϊκού γαύρου στη Δυτική και Κεντρική Μεσόγειο με τη χρήση γονιδιωματικής προσέγγισης. Για τον εντοπισμό μεγάλου αριθμού πολυμορφικών τόπων, συναρμολογήσαμε αρχικά ένα πρόχειρο γονιδίωμα αναφοράς 1.69 Gb και 79.8% γονιδιακής περιεκτικότητας -BUSCO- που χρησιμεύει ως ισχυρό μέσο για την περαιτέρω ανάλυσή μας. Στη συνέχεια, στοιχίσαμε τα δεδομένα αλληλούχησης ddRAD, από 398 άτομα γαύρου σε δώδεκα περιοχές της Δυτικής και Κεντρικής Μεσογείου, τα οποία οδήγησαν στην επιτυχή γονοτύπηση και περαιτέρω μελέτη με βάση 9.497 μονονουκλεοτιδικούς πολυμορφισμούς (SNPs). Τα αποτελέσματα έδειξαν ότι τα δείγματά μας χωρίστηκαν σε δύο ομάδες: η μία περιλαμβάνει τον Ατλαντικό και την Θάλασσα Alboran και η άλλη τη Βορειοδυτική και Κεντρική Μεσόγειο Θάλασσα (F_{ST}=0,09). Αυτή η διαφοροποίηση αναδεικνύει την παρουσία δύο διαφορετικών γενετικών δεξαμενών, η οποία ευθυγραμμίζεται με προηγούμενες έρευνες για τον ευρωπαϊκό γαύρο, υποδεικνύοντας για άλλη μια φορά το μέτωπο Αλμερίας-Οράν ως πιθανό φραγμό γενετικής ροής στο είδος. Κατά συνέπεια, η έρευνά μας θέτει τις βάσεις για την οριοθέτηση διακριτών ιχθυαποθεμάτων και την εφαρμογή στοχευμένων στρατηγικών διατήρησης για τη διαφύλαξη του είδους από την υπεραλίευση και την υποβάθμιση των ενδιαιτημάτων.

INTRODUCTION

The European anchovy: biology and distribution

The European anchovy (*Engraulis encrasicolus* L. 1758) belongs to the Engraulidae family, a group of small, schooling fish found in marine and brackish waters (Checkley et al., 2017). The *Engraulis* genus consists of nine species occurring in temperate and subtropical coastal waters worldwide (FAO-FIGIS, 2001). The European anchovy is widely distributed along the north eastern and central Atlantic, throughout the entire Mediterranean, Black and Azov seas and along the coast of West to South Africa (**Figure 1**,Whitehead et al., 1984; Whitehead et al., 1988).

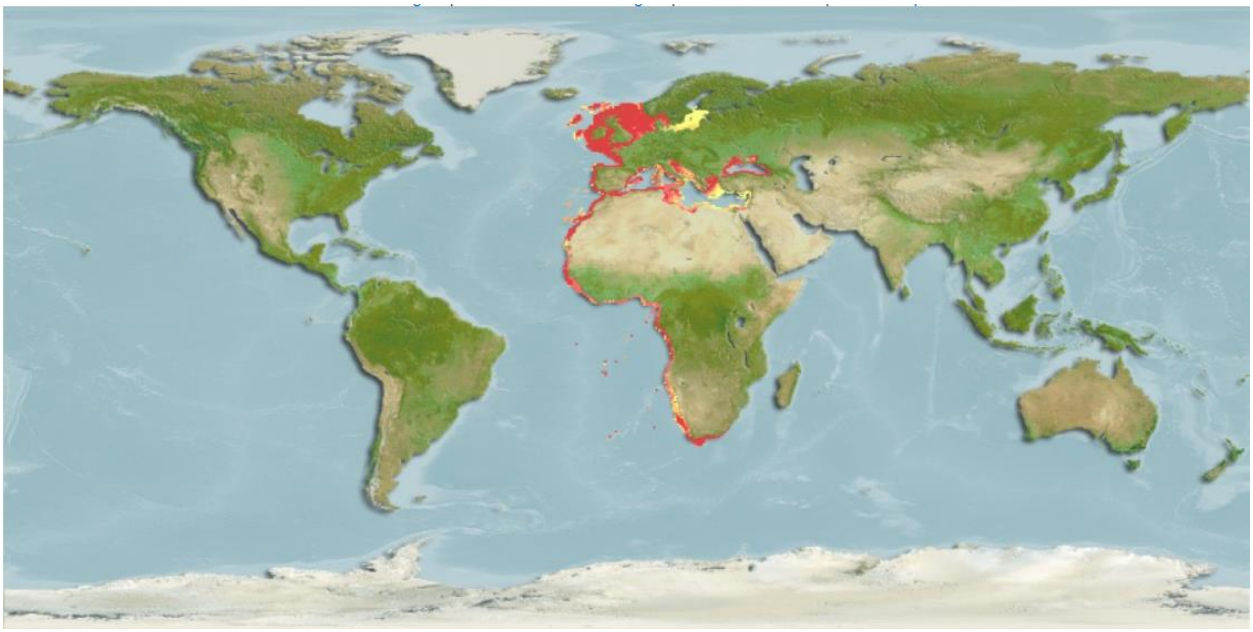

Figure 1: Distribution of *Engraulis encrasicolus* (European anchovy) (see geographical distribution in the FAO—Fishbase)

It is a coastal pelagic species with the capacity to live in a wide range of salinity and temperature conditions. This short-lived planktivorous organism, spawns multiple times from spring to autumn (in Mediterranean) with peaks in June-July (Palomera et al., 2007; Somarakis et al., 2004). The onset of its spawning is highly temperature-dependent but other environmental variables might also be the cause of the difference in peak-spawning (Millân, 1999; Motos et al., 1996; Palomera, 1992).

Anchovy revealed: ecosystem, socio-economic and evolutionary importance

Anchovy as part of the marine ecosystem

Small pelagic fish are known to play a key ecological role in coastal ecosystems, ensuring energy transfer between lower and higher trophic levels in the food web, while forming narrow "wasp-waists" characterized by low species diversity but high abundance (Bakun, 2006; Jordán et al., 2005). Canonical wasp-waist ecosystems, often found in upwelling regions, consist of high species diversity at the lower (e.g. planktonic) and higher (e.g. predatory fishes, seabirds, marine mammals) trophic levels, and a single species, or at most few species, of small planktivorous fishes like anchovy (forage fish), entirely dominating the middle trophic level (inter alia: Bakun, 2006). These forage species are responsible for the transfer of energy from plankton to predators.

Changes in the relative abundance of forage fish are expected to have both bottom-up and top-down effects (Schwartzlose et al., 1999). Bottom-up regulation is observed when there is a reduction of forage fish populations e.g. by overfishing (Hunt & McKinnell, 2006), whereas increases in prey populations upon the removal of their predators (e.g., by fisheries) have been taken as evidence of top-down effects (Furness, 2002).

European anchovy: a high-value currency

Despite the anchovy's crucial role in marine ecosystems, its economic importance in supporting fisheries cannot be overlooked. In the Western Mediterranean basin, local pelagic fisheries highlight not only its contribution in the local economy but its cultural significance in the region (Pertierra & Lleonart, 1996). In 2020 and 2021, the European anchovy was one of the main landed species in the Western and Central Mediterranean accounting for 26,623 tonnes (13%) and 8,263 tonnes (4.7%), respectively (FAO, GFCM, 2023). Anchovy is a key target in Northwestern Mediterranean fisheries, caught mainly by purse seine and trawl (Pertierra & Lleonart, 1996). Moreover, the European anchovy is also known for its highly nutritional profile, since essential fatty acids like Ω-3 (DHA, EPA) and Ω-6 (ARA) are found in its flesh (Zlatanos & Laskaridis, 2007). Finally, due to anthropogenic pressures in the Mediterranean Sea (climate change, fishing pressure, pollutants - microplastics), more research and conservation efforts on the anchovies are essential for both commercial sustainability and overall health of the Mediterranean marine ecosystems.

European anchovy shaped by evolutionary events

Genetic studies on the European anchovy became popular for several reasons. Initially, morphological studies were focused on its taxonomy (Fage, 1911; Aleksandrov, 1927). Subsequent research using morphological (Spanakis et al., 1989; Tudela, 1999), acoustic (Chashchin, 1996) and molecular data (Bembo et al., 1996; Borsa, 2002;Magoulas et al., 1996; Spanakis et al., 1989) provided more information regarding the species' populations, stocks and evolutionary history. Lastly, due to these primary insights and its high economic value, a high number of genetic studies followed (e.g. Catanese et al., 2020; Grant, 2005; Magoulas et al., 2006; Turan et al., 2004; Viñas et al., 2013).

Phylogeographic studies on the European anchovy have revealed interesting genetic patterns that are due to ocean-climate changes that date back to Pliocene and Pleistocene epochs (Magoulas et al., 1996 & 2006). These patterns can be traced to particular geological events, such as the Messinian salinity crisis (around 5.5 million years ago), the connection between Mediterranean and Black Sea at the end of Pliocene, around 2.58 mya and the following glaciation events (2.58 mya - 11.700 y), all of which have led to colonization events, population bottlenecks, long term isolation and subsequent mixing of populations. Thus, such dynamic processes mostly explain the deep divergence in populations of Mediterranean marine species (Magoulas et al., 1996). Mitochondrial DNA analysis conducted by Magoulas et al. (1996, 2006), revealed that, due to the previous events, the mitochondrial genetic component of anchovy populations in the Mediterranean and nearby regions, is segregated into two groups. Finally, the mitochondrial genome of recent populations seems to contain part of both mito-groups, with varying proportions observed across different geographic areas (references above).

Population genetic studies, using nuclear and mitochondrial data, on the European anchovy, revealed remarkable population structure throughout its distribution area (e.g. Grant, 2005). In particular, Western Mediterranean populations (except the Alboran Sea), were found to be genetically closer to those of North Sea, English Channel and Bay of Biscay than to those of East Atlantic coast (Gulf of Cadiz, Alboran Sea, Portugal, Canary islands) (Zarraonaindia et al., 2012). The observed genetic pattern has also been reported by other studies (Magoulas et al., (1996, 2006); Sanz et al., 2008; Zarraonaindia et al., 2009).

In the Mediterranean basin, there is a pronounced population structure, in offshore anchovies (see below), which results from recent and historic factors (Grant, 2005; Magoulas et al., 2006). In the basin, the population structure varies between the eastern and western part which may be due to sharp water salinity changes (Magoulas et al., 1996) in conjunction with physical barriers limiting the gene flow within the basin. In particular, studies using morphometric and genetic data, in the Adriatic, Ionian and Aegean Sea, have shown population substructure in the Eastern Mediterranean (references above). On the other hand, no genetic differentiation is observed in the Western Mediterranean anchovy populations (Sanz et al., 2008; Tudela, 1999; Viñas et al., 2013).

In particular, only the Alboran population (GSA01) is found genetically distinct from those in Balearic Sea, Gulf of Lion, Lingurian and Tyrrhenian Seas, whereas they do not differentiate genetically from each other (e.g. Zarraonaindia et al., 2012).

Furthermore, other studies indicate the existence of two partially isolated ecotypes within the *E. encrasicolus* species (Catanese et al., 2020 and literature therein). One lives in coastal waters (i.e. nearshore, lagoons, estuaries) and the other in open sea (i.e. marine, offshore waters). They probably correspond to two different groups of spawners that breed and overwinter in coastal and marine areas, respectively (Zarraonaindia et al., 2012 and cited references). Such local ecotypic differentiation refers to populations of the same species which have evolved heritable physiological, morphological and behavioral differences that are closely associated with environmental variation (Johannesson, 2001). For example, Le Moan et al., (2016) found that inshore anchovy populations from the Bay of Biscay and the Gulf of Lion were genetically closer than with their corresponding offshore populations. Accordingly, genetic distinctiveness of marine and coastal anchovy is also observed in Tyrrhenian Sea, North Adriatic Sea and Mersin Bay in Eastern Mediterranean (Borsa, 2002; Catanese et al., 2017; Karahan et al., 2014; Ruggeri et al., 2016).

Aim

In this study, we aim at providing a better understanding of the current population structure of the European anchovy in the Western and Central Mediterranean by analyzing genome-wide single nucleotide polymorphisms (SNPs). To better achieve this, we first attempted to generate the draft whole genome assembly of the species. This decision was motivated by the need to have a high number of loci resulting in high resolution in our dataset. This research is crucial for fishery and conservation strategies, as our large genomic dataset allowed us to conduct an in-depth population analysis, establishing the genetic relationships of *Engraulis encrasicolus* stocks across various geographical sub-areas in the Western and Central Mediterranean, as well as adjacent Atlantic regions.

MATERIALS AND METHODS

Sampling

Anchovy individuals were collected from twelve different locations (**Figure 2**) from Western and Central Mediterranean and the adjacent Atlantic waters, during scientific surveys (Mediterranean International bottom Trawl Survey, MEDITS; and MEDiterranean International Acoustic Survey, MEDIAS) or commercial hauls in the period November 2017 – June 2018. A total of 408 anchovies (197 males, 176 females, 35 immatures) were collected. These samples were used for the double-digest RAD library construction.

Additionally, one anchovy (sexed as female) was collected from the Northern Euboean Gulf in May 2023 with the purse-seine method at a depth of 50m and its genomic DNA was used as template for the construction of the European anchovy draft genome.

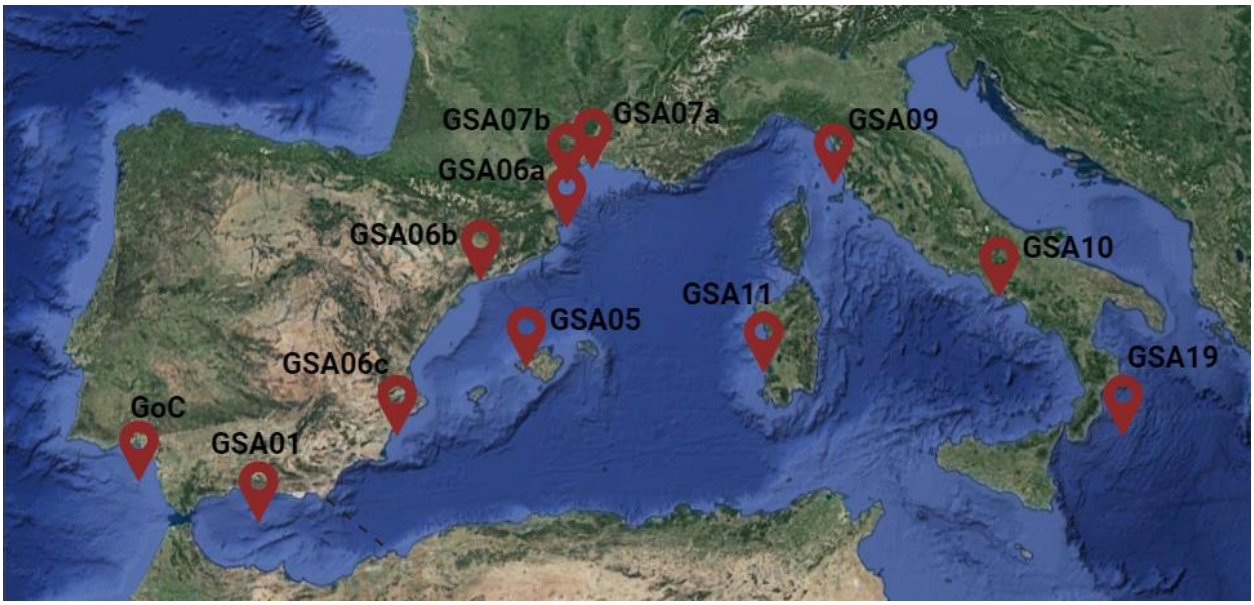

Figure 2: Geographical distribution of sampling sites. Almeria-Oran Front (AOF) is indicated by a dashed line.

Building the European anchovy draft genome: DNA extraction, library preparation, Nanopore sequencing, Genome assembly and evaluation

DNA extraction was performed with a standard salt-based protocol by Miller et al. (1988) which is a common, reliable, less expensive, and non-toxic method (**Figure 3**). The extracted DNA was checked qualitatively and quantitatively with gel electrophoresis, Nanodrop (ND-1000) and Qubit. For the genome assembly, since the extracted DNA was moderately fragmented, low molecular weight DNA was depleted with a size selection protocol using PEG 8000 (Polyethylene Glycol) and NaCl which are known to be compatible with downstream processes (Tyson, 2020). Library construction and sequencing was done in two rounds. In the first round, a library was constructed and sequenced using a SQK-LSK110 sequencing kit and a R9.4.1 flow-cell (Oxford Nanopore). This kit and flow-cell type have a raw median accuracy of around 96%. During sequencing, it became obvious that the number of sequences produced was low. A subsequent library was constructed with a more recent and with higher accuracy Ligation Sequencing Kit V14 (SQK-LSK114) protocol which is recommended for achieving median raw reads accuracy of around 99%. The library was loaded onto two MinION R10.4.1 flow-cells and Nanopore sequencing was conducted. Each flow-cell run for approximately 100 hours. The library exhibited a high degree of so-called pore blocking; therefore, an unblocking procedure, which consisted of digesting the library on the flow cell followed by loading a fresh library aliquot, was performed approximately every 12 hours to increase the amount of sequences produced. During the sequencing run, a fast and low-accuracy live basecalling was performed to get preliminary statistics (**Figure 3**). Basecalling is the process of translating the electronic raw signal of the sequencer into nucleotide bases, i.e., A, C, T, or G.

Nanopore basecalling

After completion of the runs, a slower but higher accuracy basecalling was performed. Two state-of-the-art Oxford Nanopore's basecallers, Guppy and Dorado, were used (**Figure 3**). Guppy version 6.5.7 is a neural-network-based basecaller that was used for basecalling of the first round of data (SQK-LSK110). Dorado version 0.4.1 is a high-performance, open source basecaller for nanopore reads. Dorado was used together with Samtools version 1.9 for basecalling the second round of data (SQK-LSK114). Dorado provides the potential for both simplex and duplex basecalling. In duplex basecalling, the complement DNA strand is read immediately after the template strand and the consensus basecalling leads to a further increase in accuracy up to 99.9% (Q30). Only a small percentage (typically ~10%) of reads are duplex reads. Basecalling was done with a compatible GPU (NVIDIA GTX1080). The generated basecalled data is in *fastq* format, which is compatible for downstream analyses.

Assembly and polishing of the reference genome

The generated Oxford Nanopore single-molecule sequencing reads (*.fastq*) are assembled into contigs and scaffolds using the flye *de novo* assembler (Kolmogorov et al., 2019) (**Figure 3**). Flye, version 2.9, is designed for a wide range of datasets, from small bacterial projects to large mammalian-scale assemblies. The assembler uses a repeat graph, as the core data structure, which is built using approximate sequence matches. The edges of the repeat graph represent the genomic sequences and the nodes define the junctions. Repeat graphs reveal the repeat structure of the genome, which helps to reconstruct an optimal assembly. In particular, the following parameters were used: *--genome-size 1.5g* to provide the anchovy's estimated genome size of 1.5 Gb (representing the length of the diploid genome) as GoaT (Genomes on a Tree) estimated and *--scaffold* to enable scaffolding (Challis et al., 2023).

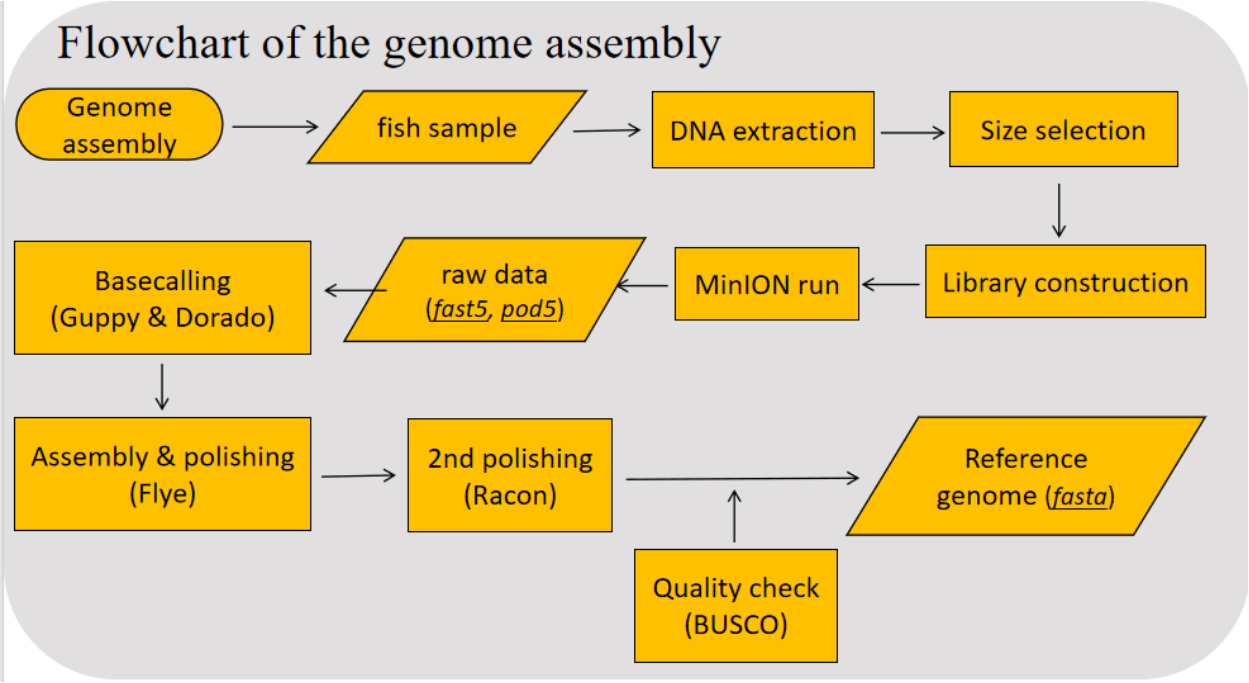

Figure 3: Flowchart of the genome assembly. File formats are specified with underlined italics.

One of the most critical processes to achieve high-quality genome for downstream analysis is the polishing process (Lee et al., 2021). Flye can also function as a stand-alone polishing tool (**Figure 3**). The polishing process aligns the reads to the assembled contigs and makes corrections to the contigs where warranted, i.e., correct the remaining errors in the initial genome assembly, improving the local and overall accuracy. Racon, version 1.4.20, a standard nanopore-based polishing tool, was used for a second polishing process (Lee et al., 2021; Vaser et al., 2017) (**Figure 3**). Racon takes as input a set of backbone sequences (contigs), a set of reads (with which the assembly was generated) and a set of mappings between the reads and the backbone sequences.

Minimap2, version 2.17, is a sequence alignment program that aligns DNA sequences against a reference database (in some cases, like here, can be a genome). Minimap2 was used as the mapper since it is fast and provides reasonable results (Li, 2018).

Genome assembly evaluation

The generated assembly was evaluated in terms of expected genes content using BUSCO (Benchmarking Universal Single-Copy Orthologs, Manni et al., 2021) (**Figure 3**). BUSCO is an assessment tool that, through the orthoDB database (http://www.orthodb.org/; a database with ortholog genes for each set of major lineages e.g. acidobacteria, vibrionales, actinopterygii) uses a set of genes (BUSCOs) for each lineage and searches for them in the genome assembly under assessment. The genes that make up each set of each lineage are selected from orthologous groups with genes present as single-copy orthologs in at least 90% of the species that make up each lineage. While allowing for rare gene duplications or losses, this establishes an evolutionarily-informed expectation that these genes should be found as single-copy orthologs in any newly-sequenced genome. The evolutionary expectation means that if the expected genes-BUSCOs cannot be identified in a genome assembly, it is possible that the sequencing and/or assembly approaches have failed to capture the expected "complete" gene content. Lastly, the *actinopterygii_odb10* lineage dataset was used with 3,640 total BUSCOs (single-copy orthologs) to be searched.

Bioinformatic analysis of ddRAD data

ddRAD library preparation and sequencing

Genomic DNA was extracted from tissue samples and the libraries were prepared using the double-digest restriction site associated DNA (ddRAD) sequencing approach developed by Kess et al., (2016).

Raw reads quality control

The Illumina raw data consists of two *fastq* files per individual: R1.fastq.gz and R2.fastq.gz, representing the forward and reverse reads (paired-end), respectively, with a read length of 150bp. Before performing any downstream analyses, the raw data passed through a quality control using fastQC version 0.11.9 (Andrews, 2010). Raw reads were analyzed in Stacks version 2.64 pipeline (Catchen et al., 2011), and the main idea is to track a set of loci and the alleles present at those loci, in a population of organisms. In this case, Stacks searches for SNPs among the individuals.

Raw data recovery and cleaning were performed in *process_radtags* program in Stacks. The parameters that were used here were 1) *--renz-1 PstI --renz-2 BgIII*, specifying the restriction enzymes that cleaved the genomic DNA, 2) *-c* parameter which cleaned the data and removed

reads with uncalled bases, 3) *-q* parameter which discarded reads with low quality scores (Phred score < 10) and 4) *-r* that corrected sequencing errors in RAD cut sites (**Figure 4**).

De novo analysis of ddRAD sequencing data

A *de novo* analysis, i.e., without using a reference genome, in Stacks proceeds in six major stages. After the reads go through the *process_radtags* program (mentioned above), the next three stages comprise the main Stacks pipeline: building loci (*ustacks*), creating the catalog of loci (*cstacks*), and matching against the catalog (*sstacks*). In the fifth stage, the *gstacks* program is executed to assemble and merge paired-end contigs, call variant sites in the population and genotypes in each sample. In the final stage, the *populations* program is executed, which can filter data, calculate population genetics statistics, and export a variety of data formats.

Ustacks takes as input a set of reads and aligns them into "exactly"-matching *stacks.* Here, the *M* parameter specifies the maximum distance -in nucleotides- allowed between stacks and the *m* parameter is referred to the minimum coverage required to create a stack. Moreover, cstacks will create a catalog of loci and the *n* parameter specifies the number of mismatches allowed between sample loci when building the catalog.

The *M*,*m* and *n* parameters must be estimated since they are dataset-dependent. Here, the protocols of Rochette & Catchen, (2017) were followed to optimize the three parameters. Firstly, a sub-dataset was created for the parameters optimization, consisting of five individuals per population (based on the number of retained reads) and the "80% rule" (Paris et al., 2017) was used to check the maximum number of polymorphic loci. To estimate the *m* parameter first, the pipeline was run with the default values for *M* and *n i.e.,* 2 and 1, respectively. Then, the m parameter was set to the "best" value and the pipeline was run again to optimize the others. The parameters were set to *m*=13, *M*=5 and *n*=7. The *populations*-specific parameters are the same in *de novo* and reference-based approaches (see below).

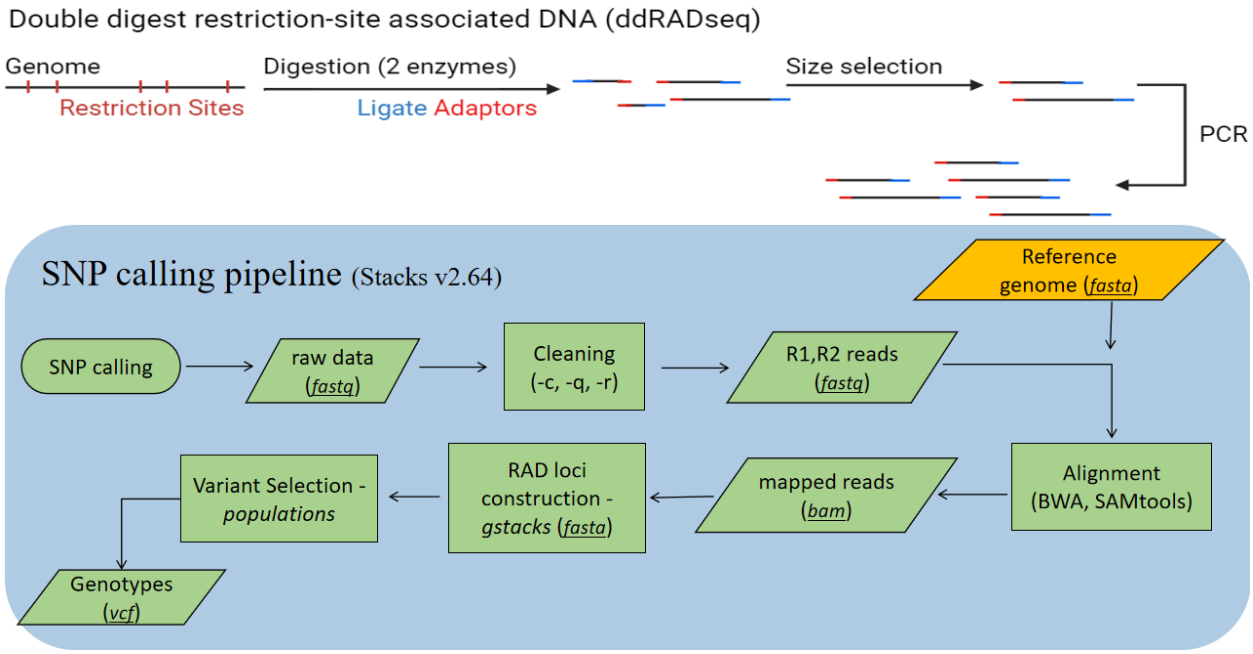

Figure 4: ddRADseq main process and Stacks pipeline. The reference genome was generated following the genome assembly pipeline. Different file formats are specified with underlined italics.

Reads mapping against the genome

The recovered and cleaned reads were aligned against the generated reference genome. To that purpose, the Burrows-Wheeler Alignment (bwa), version 0.7.15, software package was employed (Li & Durbin, 2009, **Figure 4**). BWA first indexed, with the *index* command, the reference genome and then mapped the paired-end reads of each individual on it by applying the bwa *mem* command. The *–M* parameter was added to flag additional hits of a read on the reference genome as secondary. The generated *sam* files were converted to *bam* files with *sort* command in SAMtools version 1.9 (Li et al., 2009) and the mapping rate was assessed with the *flagstat* command (**Figure 4**).

Genotyping

The merging of the forward and reverse reads leads to the paired-end contigs which are further analyzed in Stacks. In detail, the bam files, one for each individual, are given as input to the Stacks component *gstacks*, which assembles the paired-end contigs, calls variable loci (loci with SNPs) in the dataset and genotypes individuals at those sites. In addition, *gstacks* creates a catalog with the consensus sequence for each assembled locus in the dataset (**Figure 4**).

Finally, the *populations* component of the Stacks pipeline performs the final loci selection based on certain criteria (**Figure 4**). The catalog from the *gstacks* is passed to *populations* which outputs the genotyped polymorphic loci, using the following variables: *-R 0.8* which indicates that a locus needs to be genotyped in at least 80% of individuals within the 12 geographical sites, *--min-maf 0.05* which indicates the minimum minor allele frequency, *--max-obs-het 0.5* which is the maximum observed heterozygosity required to process a nucleotide site at a locus and *--write-random-snp* to specify that for each ddRAD locus, only a single, randomly selected, SNP to be kept. Finally, after excluding loci and individuals with >20% missing data, with PLINK version 1.9 (Purcell et al., 2007), the *populations* program was run again on the updated data set and the output *vcf* file was used for downstream analysis.

Linkage disequilibrium-based filtering of single nucleotide polymorphisms

Linkage disequilibrium (LD), the non-random association of alleles at two or more loci, is an important concept in various areas of genetics (Lewontin & Kojima, 1960). Often expressed with the squared correlation (r^2) between alleles at two loci (Qu et al., 2020). Here, we opted to exclude highly associated loci since downstream analyses assume their independence. Highly linked loci are likely co-inherited, share common evolutionary history and thus enhance the same phylogenetic information. This could introduce redundancy and potentially lead to the overestimation of our results. Thus, to reduce the influence of highly associated loci on the assessment of population stratification, an r^2 threshold of 0.5 for the SNP pairs in a sliding window of 50 SNPs was implemented. The command was *--indep-pairwise 50 5 0.5,* specifying the sliding window (50), the variant count to shift the window at the end of each step (5) and the LD threshold (0.5).

Population genomics

Outliers detection

Outlier loci are genomic locations that show behavior or patterns of variation that are extremely divergent from the rest of the genome (locus-specific effects), as revealed by simulations or statistical tests (Luikart et al., 2003). Outlier loci were detected with the following approaches.

PCA-based approach

The principal component analysis (PCA-based) approach is implemented on individual genotype data using the PCAdapt R package (Luu et al., 2017). Then, the Mahalanobis distance is computed for each SNP, which quantifies the association of each SNP with the K PCs that mostly support the population structure. The “outliers” were selected with an adjusted p-value = 0,01 using the Benjamini-Hochberg correction.

Bayesian approach

A Bayesian statistical method was utilized with BayeScan version 2.1 estimating the relative posterior probabilities for each locus being under either diversifying or balancing selection (Foll & Gaggiotti, 2008). BayeScan, firstly, calculates F_{ST} coefficients for each SNP and then partitions these into a population-specific component (β), common to all loci, and a locus-specific component (α) shared by all the populations using a logistic regression. Allele frequencies are assumed to follow a Dirichlet distribution. Selection is detected when α is significantly different from 0. When $\alpha > 0$, it is assumed that directional selection acting on the locus under analysis, while $\alpha < 0$, suggests balancing or purifying selection. Here, BayeScan analyses used 20 pilot runs of 5,000 iterations (to estimate posterior odds), a burn-in of 50,000 iterations, a thinning interval of 10, and a prior odds ratio of 100, indicating that selection model has to be 100 times more likely, for a SNP, than the balancing model in order to infer "outliers".

Arlequin

We implemented an F_{ST}-based method proposed by Excoffier et al. (2009) with the Arlequin program. Here, a distribution of F_{ST} values across loci is generated and "outliers" are identified as those being present in the tails of this distribution. Here, 40,000 simulations under the hierarchical island model took place, for 100 simulated demes in 10 simulated groups, with the allowed level of missing data/SNP set to 1 (Slatkin & Voelm, 1991). The genetic structure was defined a priori for K=2, and loci with a p-value≤0,001 were considered outliers.

Clustering analyses

Bayesian approach

STRUCTURE software (Pritchard et al., 2000) analyzes differences in the distribution of genetic variants among populations with a Bayesian iterative algorithm by assigning samples into groups whose members share similar patterns of variation (Porras-Hurtado et al., 2013). STRUCTURE starts with a "burn-in" stage where the simulation runs for a specified number of steps till the convergence of summary statistics (such as α, F, divergence distances among populations, likelihood, etc.). Then, MCMC (Markov Chain Monte Carlo) simulations take place to best estimate parameters, like admixture coefficients. STRUCTURE applies a model to the data of K assumed populations or genetic groups, each characterized by a subset of allele frequencies identified in the data. The uppermost hierarchical level of structure or, simply, "best K" is selected by calculating the change in the log probability of data between K values (Evanno et al., 2005). During each analysis, membership coefficients summing to one are assigned to individuals for each population/group generating the "individual Q-matrix" table (Porras-Hurtado et al., 2013).

In present data, the admixture model with correlated allele frequencies between populations was used with a burn-in length of 250,000 and 500,000 MCMC iterations. The number of predetermined clusters tested were K=7 with R=10 replicates, indicating that ten replicate runs will be performed testing the posterior probability for K=1-7 genetic groups.

Structure threader was used to parallelize and automate the runs of STRUCTURE (Pina-Martins et al., 2017). Structure harvester was utilized for the implementation of the Evanno method and the provision of Q-matrices for DISTRUCT and CLUMPP to graphical display the population structure (Earl & vonHoldt, 2012; Jakobsson & Rosenberg, 2007; Rosenberg, 2004).

Analysis of Molecular Variance (AMOVA)

AMOVA is a statistical method used in population genetics to understand how genetic variation is partitioned among different hierarchical levels within a species. F_{CT} index describes the variance explained by groups of populations, F_{SC} describes the variance among populations (within groups) and F_{ST} index describes the variance within groups among all samples.

A locus-by-locus AMOVA, was implemented on our data using Arlequin setting to 50,000 permutations and the level of missing data/SNP to 1. Here, four different clustering scenarios were tested, based mainly on geographical criteria and genetic data (Table 1).

Table 1: Population clustering scenarios tested by a locus-by-locus AMOVA based on geographical and genetic data.

Scenario	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	GoC,GSA01	GSA06c,GSA05, GSA06a,GSA06b, GSA09,GSA10, GSA11,GSA07a, GSA07b, GSA19		
2	GoC,GSA01	GSA06c,GSA05, GSA06a,GSA06b, GSA09,GSA10, GSA11,GSA07a, GSA07b	GSA19	
3	GoC	GSA01	GSA06c,GSA05,GSA06a,GSA06b,GSA09,GSA10,GSA11,GSA07a,GSA07b,GSA19	
4	GoC	GSA01	GSA06c,GSA05,GSA06a,GSA06b,GSA09,GSA10,GSA11,GSA07a,GSA07b,	GSA19

Multivariate Analysis

Discriminant Analysis of Principal Components (DAPC) is a multivariate statistical approach, in which the genetic variance in the data set is divided into a between-group and within-group component, aiming to maximize discrimination between groups and minimize the second. DAPC uses principal component analysis (PCA), at first, for data transformation reducing the variables (PCs) to ones that best describe our data. Then, a k-means clustering algorithm is applied on the data with increasing values of k, and different clustering solutions are compared using Bayesian Information Criterion (BIC). Cross-validation was implemented with the *xvalDapc* command, in order to retain the appropriate number of PCs. A Discriminant Analysis is performed on the retained PCs and a number of Discriminant Analysis eigenvalues (DAs) can be selected. The final result of the DAPC is a scatter plot. Finally, DAPC was implemented with *dapc* command using the *adegenet 2.0.0* (Jombart, 2015) and *ade4* (Dray & Dufour, 2007) R packages.

F-statistics

The metrics F_{ST} (fixation index) and F_{IS} (inbreeding coefficient) are a set of functions of allelic-state correlations introduced by Wright (1951) and are referred to in population genetics as F-statistics. These metrics are important measures in population genetics and help to understand genetic variation within, among and in the total sample/population. F_{ST} measures the extent of genetic differentiation among populations. F_{ST} values range from 0, indicating no differentiation (populations are genetically identical) to 1, indicating populations being completely differentiated. On the other hand, the F_{IS} index, measures the level of inbreeding within a population relative to the expectations under random mating. F_{IS} ranges from -1 to 1, where 0 indicates random mating (as expected), negative values suggest an excess of heterozygotes, and positive values indicate an excess of homozygotes due to inbreeding. Lastly, the Arlequin version 3.5 and hierfstat R package (Goudet, 2005) were used for the F_{ST} and F_{IS} calculations, respectively, after 10,000 permutation and 0.05 significance level.

RESULTS

Genome assembly

Nanopore sequencing yielded around 14.7 Gb data in total. Nearly 13.45 Gb data was successfully basecalled, of which 94.4% was simplex and 5.6% was duplex. The base called data accounts for 3,668,883 reads with N50 of 5,923 bp.

Flye pipeline assembled a genome of total length 1.69 Gb, which contains 76,551 contigs (contig N50: 57,970) and 48 scaffolds with 9x coverage in total. BUSCO assessment found 2,905 (79.8%) completed, 257 fragmented (7.1%) and 478 (13.1%) missing BUSCOs.

SNP calling, individual genotyping and filtering (reference-based approach)

Illumina sequencing yielded 983,601,116 raw reads, 2.5% of which were discarded from the *process_radtags* quality control; 1,233,193 due to low quality and 23,084,689 for absent RAD cut-site. After alignment against the genome, more than 98% of the retained reads per individual were mapped to the genome. The *gstacks* program generated a catalog of 410,701 loci with a mean insert length of 192.2 bp (sd: 63.5) and effective coverage per-individual of 28.7x (stdev=7.8x). *Populations* program removed 396,901 out of 410,701 loci that did not pass sample or population constraints. From the remaining 13,800 loci, 9,645 variant sites occurred with mean genotyped sites per locus of 229.25 bp (stderr 0.59).

In the *populations*' output file (*vcf* format), additional filtering was applied in terms of data missingness and association between sites/SNPs. After applying the first filter, 398 out of 408 individuals, and 9,618 SNPs were retained with less than 20% of missing data. Lastly, 9,497 polymorphic SNPs (the *de novo* approach yielded 952 polymorphic loci) remained with less than 50% (r^2=0.5) genetic association (**Figure 5**).

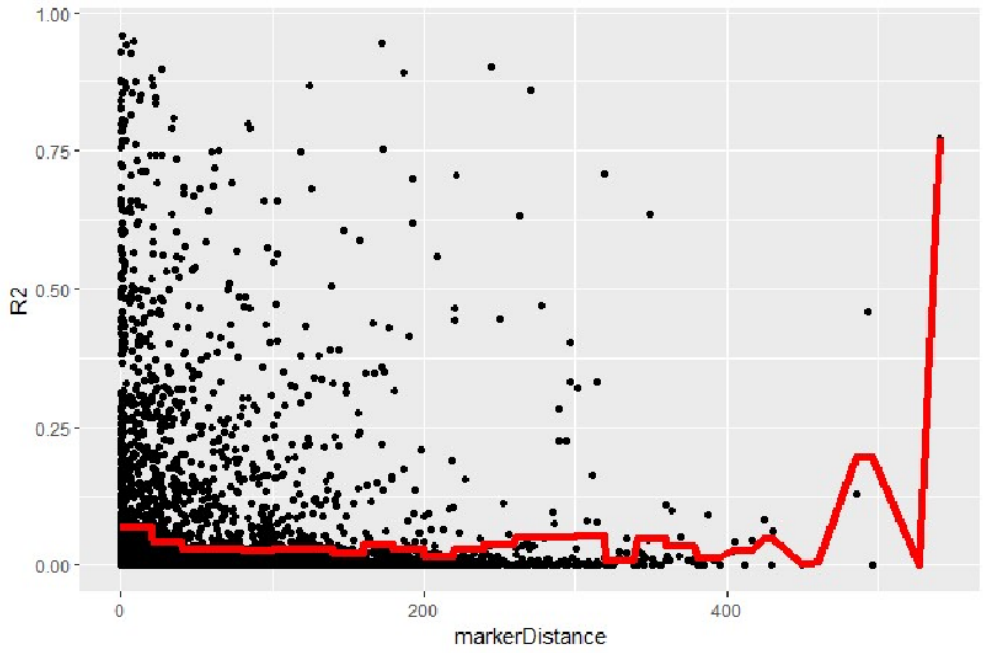

Figure 5: Genetic association measured with r-squared (y-axis) and intermarker distance measured in thousands of bp (Kb, x-axis). Red line is the mean r^2 per inter-marker distance.

Outlier loci detection

The number of potential outlier loci identified from PCAdapt, Arlequin and BayeScan were 1,266, 96 and 7, respectively. PCAdapt shares 93 outliers with Arlequin and 4 with BayeScan, while BayeScan shares 3 outliers with Arlequin (**Figure 6**). Three outlier loci were found from all approaches. Finally, 100 outlier loci were identified from at least two approaches and were used for downstream analysis.

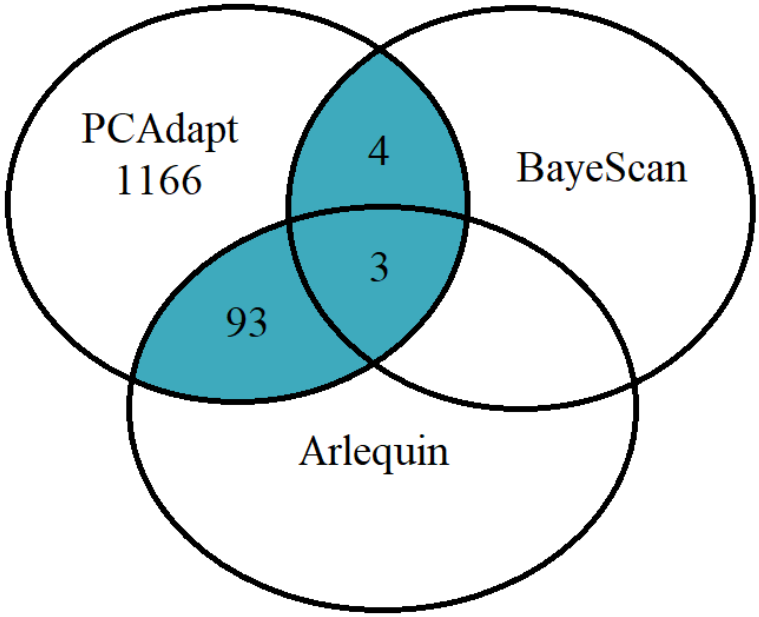

Figure 6: Venn diagram representing the potential outliers that occurred by three approaches and the overlapping sites. 1,266 loci were identified by pcadapt, 96 by Arlequin and 7 by BayeScan. 100 loci in total were identified by at least two methods.

Cluster Analyses

Bayesian Analysis

Typically, Bayesian clustering analysis is the first step in examining population structure or to infer the origins of individuals with unknown population characteristics, especially when population admixture has occurred (Porras-Hurtado et al., 2013). Here, Bayesian clustering analysis with STRUCTURE was conducted on three datasets consisting of: the total number of loci (9,497), candidate outlier loci (100) and neutral loci (9,397) respectively. The Evanno method determined that the optimal number of clusters that best fit our data is achieved by having two clusters (K=2) for all data sets (**Figure 7**). Individuals with membership coefficients (q values) ≥ 0.9 were assigned to either cluster with high confidence while samples with intermediate values were considered as admixed, that is having a mixed ancestry from the two clusters. In all datasets, the first cluster (ATL from now on) consists of two stocks, GoC (Gulf of Cadiz) and GSA01 (Alboran Sea), while the rest ones belong to the second cluster (MED from now on). Moreover, all stocks seemed to have admixed individuals, to some extent (**Figure 7**). The GSA01, the GoC and the GSA06c stocks have the most admixed individuals with 100%, 80% and 54%, respectively, while the GSA06b and GSA06a have the less admixed individuals with just 11%. Outliers' STRUCTURE bar plot showed the same pattern as the other datasets but managed to clearly sort out some previously-admixed samples to either cluster. For example, the GSA01, the GoC and

GSA06c stocks have now less admixed individuals i.e., 82%, 48% and 40%, respectively (**Figure 7**).

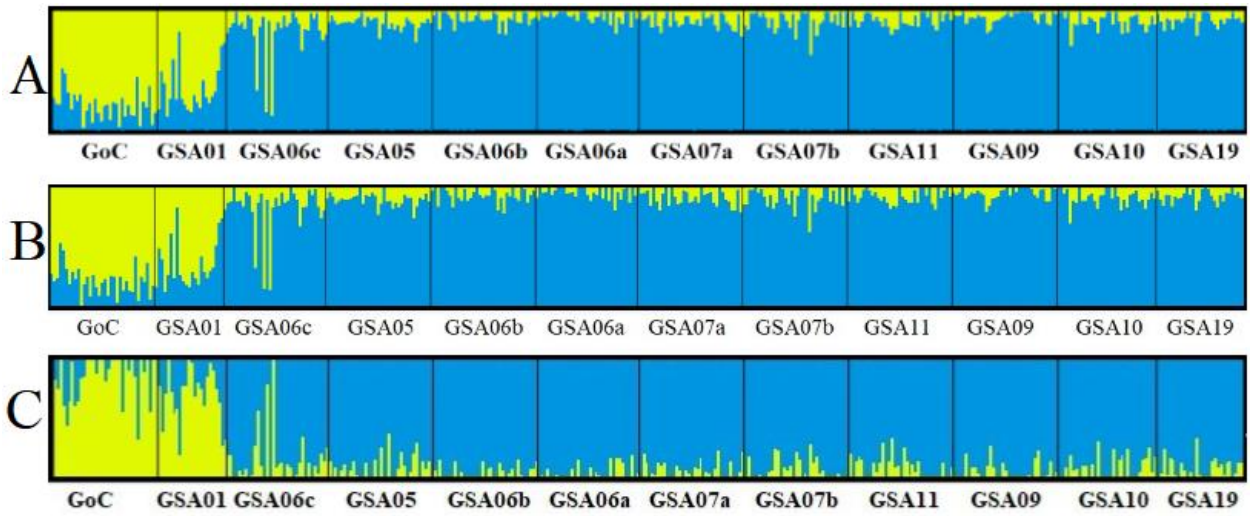

Figure 7: Population structure of European anchovy in the Eastern Atlantic and Western and Central Mediterranean Sea as inferred by STRUCTURE for K = 2. Bar plots of individual Q-values after CLUMPP analysis (plot created with CLUMPP and DISTRUCT). GoC: Gulf of Cadiz, GSA01: Northern Alboran Sea, GSA06: Northern Spain, GSA05: Balearic Islands, GSA07: Gulf of Lions, GSA11: Sardinia, GSA09: Ligurian Sea, GSA10: Tyrrhenian Sea, GSA19: Western Ionian Sea.

Analysis of Molecular Variance

Four clustering scenarios were tested with locus-by-locus AMOVA. These scenarios were based on geographical and genetic data. The source of variation (groups, populations or individuals) from each AMOVA analysis is presented in Table 2. The genetic variation among groups (F_{CT}) is higher in scenarios 1 and 3 which share almost identical values. This means that among the tested scenarios, scenario 1 (two clusters; ATL and MED) and scenario 3 (three clusters; GoC, GSA01 and MED) explain most of the variability of our data.

Table 2: European anchovy's analysis of molecular variance (AMOVA) with percentage of total variance, F-statistics (F_{CT}, F_{SC}, F_{ST}) and its significance (p) based on 10,000 permutations.

Structure tested	%total variance	F statistics	p
Scenario 1			
Among groups	9.63152	F_{CT}=0.09632	0.00000
Among populations	0.52459	F_{SC}=0.00580	0.00000
Within populations	89.84390	F_{ST}=0.10156	0.00000
Scenario 2			
Among groups	6.76806	F_{CT}=0.06768	0.00000
Among populations	0.51135	F_{SC}=0.00548	0.00000
Within populations	92.72059	F_{ST}=0.07279	0.00000
Scenario 3			
Among groups	9.66343	F_{CT}=0.09663	0.00000
Among populations	0.42989	F_{SC}=0.00476	0.00000
Within populations	89.90669	F_{ST}=0.10093	0.00000
Scenario 4			
Among groups	6.96	F_{CT}=0.06957	0.00458
Among populations	-0.18	F_{SC}=-0.00196	1
Within populations	93.22	F_{ST}=0.26845	0.00000

Multivariate Analysis

DAPC tested the potential structure on the whole dataset (all populations) using either all (**Figure 8**) or outlier or neutral loci. Also, DAPC was implemented to test potential substructuring inside the MED cluster using all loci (**Figure 9**). The whole data set was divided into two clusters using any type of loci dataset: ATL and MED clusters (**Figure 8**). DAPC using neutral and outlier loci only, showed exactly the same pattern (not shown here). Lastly, no clear substructuring occurred in the MED cluster, but some individuals from the GSA19 population (West Ionian Sea) seemed to be a little differentiated from the remaining populations (**Figure 9**).

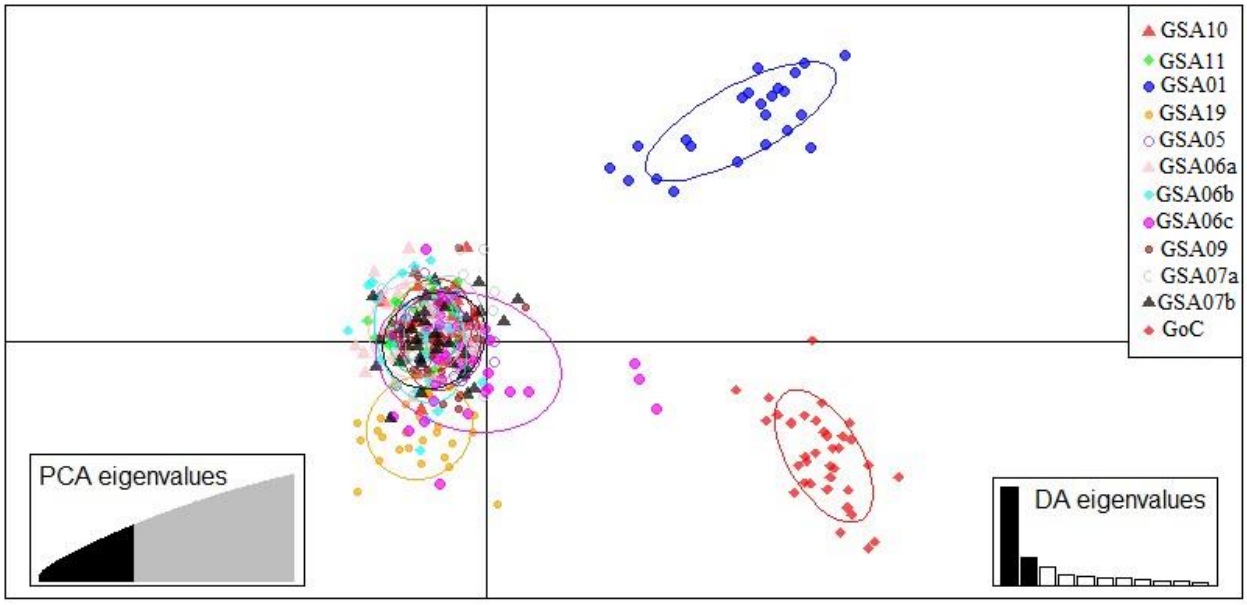

Figure 8: DAPC in the twelve anchovy's stocks using 9,497 loci (all loci). 150 Principal components were used with 2 discriminant factors.

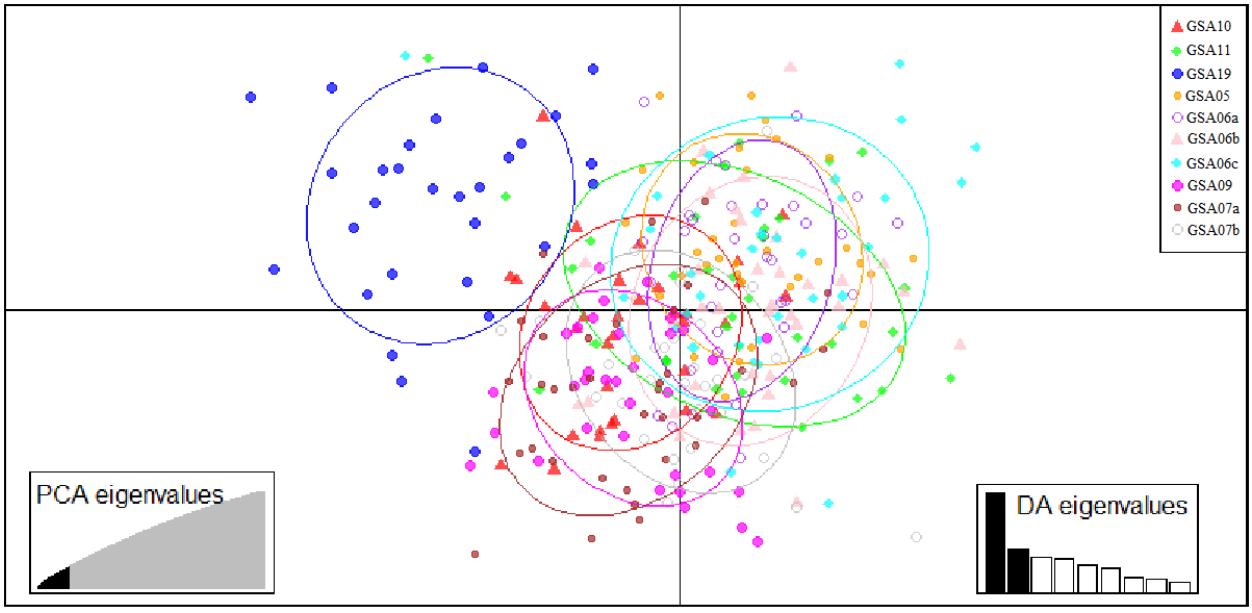

Figure 9: DAPC in the MED cluster using 9,486 loci (all loci). 50 Principal components were used with 2 discriminant factors.

Genetic differentiation

Pairwise F_{ST}

The GoC and GSA01 anchovy stocks share quite high F_{ST} values with all the other Mediterranean populations and at the same time they have low F_{ST} (0.0043) between them (Table 3). Moreover, all the Mediterranean populations, except the GSA01, are not genetically distinct from one another. Finally, all populations show an excess of homozygotes (F_{IS} >0) but with no statistical support for most of them.

Table 3: Pairwise F_{ST} values with significance level of 0.05 and 10,000 permutations and inbreeding coefficient (F_{IS}) among the studied areas. Statistically significant values are in bold.

Area	**GoC**	**GSA01**	**GSA06c**	**GSA05**	**GSA06b**	**GSA06a**	**GSA07a**	**GSA07b**	**GSA11**	**GSA09**	**GSA10**	***F_{IS}***
GoC												0.212
GSA01	0.0043											0.165
GSA06c	**0.0851**	**0.0504**										0.173
GSA05	**0.1077**	**0.0689**	0.0009									0.147
GSA06b	**0.1190**	**0.0774**	0.0028	0.0019								0.149
GSA06a	**0.1277**	**0.0867**	**0.0069**	**0.0026**	0.0009							0.136
GSA07a	**0.1090**	**0.0725**	-0.0014	-0.0017	-0.0028	-0.0017						0.15
GSA07b	**0.11363**	**0.0752**	0.0006	-0.0021	-0.027	-0.0021	-0.023					0.133
GSA11	**0.1166**	**0.0751**	**0.0041**	0.0020	0.0018	0.0014	-0.0127	-0.0016				0.161
GSA09	**0.1232**	**0.0807**	0.0009	-0.0016	-0.0035	-0.0026	0.0018	-0.0014	-0.0025			0.143
GSA10	**0.1081**	**0.0732**	-0.0024	-0.0041	-0.0042	-0.009	0.0011	0.0012	-0.0035	0.0021		**0.142**
GSA19	**0.1143**	**0.0769**	**0.0057**	**0.0033**	**0.0039**	**0.0040**	0.0023	**0.0029**	0.0039	0.0024	-0.0001	0.157

DISCUSSION

Our genomic dataset revealed two distinct genetic clusters in the Western and Central Mediterranean and Eastern Atlantic. The Almeria-Oran Front is suggested to be a potential mechanism contributing to the observed clustering. Furthermore, the genetic homogeneity observed in anchovy stocks within the Western Mediterranean basin is likely due to specific water current systems in the region.

The *Engraulis encrasicolus* genome

Constructing the reference genome of a species is one of the most important tasks in genomics (Chen et al., 2021). In the last decade, many small pelagic fish genomes were published but not the European anchovy's (Kongsstovu et al., 2019; Lee et al., 2023; Louro et al., 2019; Machado et al., 2022; Sukumaran et al., 2023). Population genomics analyses based on a reference genome usually provide a much larger number of loci compared to *de novo* ones (e.g. Ulaszewski et al., 2021). Therefore, we opted to assemble a draft genome and use it as template to analyze our RADseq data. As a result, the reference-based approach increased ten times the number of available loci.

A few months ago, researchers from the University of Istanbul submitted to NCBI a near chromosome-level assembly for the European anchovy (GCA_034702125.1). Their genome was reported to have a size of 1.4Gb, distributed across 24 chromosomes. Here, the generated genome has a size of 1.69Gb, close to the above assembly. Finally, our BUSCO results, which are comparable with other assemblies (Lee et al., 2023; Sukumaran et al., 2023 and references above), and the very high mapping rate of the Illumina reads (ddRAD-seq data) indicate the relatively good quality of our genome assembly.

Population structure of anchovy

Contrary to the general view of limited barriers in marine waters and high levels of migration and gene flow in fish (Waples, 1998), the European anchovy is known to be one of the exceptions with a pronounced population structure throughout its distribution area (Ouazzani et al., 2017 and references therein). In this study, all clustering analyses (STRUCTURE, AMOVA, DAPC) and the genetic differentiation indices, reveal an apparent population structure of the European anchovy in the Eastern Atlantic and Western and Central Mediterranean Sea. In particular, the clustering analyses show that the two-cluster scenario (scenario 1) best fits our genomic data. The first cluster (ATL) consists of the GoC (Gulf of Cadiz) and GSA01 (Alboran Sea) populations whereas the rest of the populations belong to the second one (MED). In fact, the observed pattern

of genetic differentiation has also been reported in other studies (Bouchenak-Khelladi et al., 2008; Zarraonaindia et al., 2012).

Furthermore, another statistic for measuring the genetic differentiation between populations is the F_{ST}. In marine fish species, relatively low F_{ST} values are observed (Magoulas et al., 2006; Ward et al., 1994). The European anchovy is characterized by relative high F_{ST} values: Keskin & Atar (2012), using the COI mitochondrial gene, found overall F_{ST} =0.131 in Black Sea and Aegean anchovy populations; Magoulas et al. (2006), using mitochondrial RFLPs, reported average F_{ST}=0.15 for the European anchovy populations; Catanese et al. (2020) used a 96 SNP-panel for the Western Italy anchovy populations and found F_{ST} values relative to our study, and lastly; Zarraonaindia et al. (2012) using 90 SNPs (nuclear and mitochondrial) found a range of F_{ST} values (0.04 to 0.318) from Denmark to S. Africa (average F_{ST} =0.078) (see also Huret et al., 2020). In this study, the suggested clustering scenario (scenario 1) is also observed in the pairwise F_{ST} matrix (table 3). For instance, when comparing the Gulf of Cadiz population with ones from the MED cluster, a substantial F_{ST} value exceeding 0.085 is observed. Our pairwise F_{ST} measurements, focusing only on statistically significant results, ranged from 0.026 to 0.1277. Low F_{ST} values were found in the MED cluster and high ones when comparing populations from the two clusters. Notably, the two clusters display a significant pairwise F_{ST} of 0.097 (statistically significant). Finally, our F_{ST} measurements, based on a genome-wide SNP dataset, not only offer comparable results but also highlight the observed genetic differentiation among specific anchovy populations.

Atlantic vs Mediterranean isolation: the case of Almeria–Oran front

The Alboran Sea constitutes a transition zone between the Atlantic Ocean and the Mediterranean Sea. Its waters are much closer, in terms of chemical components, to those of the North East Atlantic than those of Western Mediterranean, shaping the Almeria-Oran Front (AOF; Tintore J. et al., 1998). It is characterized by a 200-300m thick, one-way surface current of Atlantic water into the Mediterranean and the presence of gyres that form a well-defined hydro-geographical boundary of surface waters between Almeria and Oran. Its oceanographic conditions act as a barrier for dispersal for many marine organisms as well as for the European anchovy e.g. Bouchenak-Khelladi et al. (2008). Our genetic data account for two genetic clusters, ATL and MED, namely, which seem to be separated at the point where the AOF is located (**Figure 1**). In fact, multiple studies have reported the genetic distinctiveness of the Alboran population from the rest of Mediterranean and at the same time its genetic similarity with the Cadiz population (Bembo et al., 1996; Bouchenak-Khelladi et al., 2008; Magoulas et al., 2006; Sanz et al., 2008; Zarraonaindia et al., 2012). However, Viñas et al. (2013), using the mtDNA control region, first suggested the genetic dissimilarity of the Alboran Sea anchovy from the Gulf of Cadiz. This might result either from the endogenous recovery of the population after the fishery collapse in the 1990s (Pertierra & Lleonart, 1996) or anchovies from other regions of the Eastern Atlantic might have migrated there and influenced that genetic pool.

The biogeographic boundary that is created by the AOF might not be strong enough to maintain absolute isolation of anchovy populations. The AOF is bounded from the GSA01 (Alboran sea) and the GSA06c (Southeastern Spain) populations, in which we found many individuals with mixed ancestries. All individuals from the GSA01 and 19 (54%) from GSA06c were admixed. Zarraonaindia et al. (2012) also found admixed individuals in the Alboran population. Therefore, the identification of admixed individuals in Cadiz and other Mediterranean populations, is expected, as they are connected with the Alboran and GSA06c (southeastern Spain) populations, respectively.

Population substructure in the Mediterranean basin

While the European anchovy exhibits an interesting population structure in its overall distribution area, no structure is observed within the Western Mediterranean populations (e.g. Sanz et al., 2008). The genetic homogeneity of the Western Mediterranean anchovy's stocks (except Alboran population) might be mostly explained by present-day factors. Marine currents play the most important role in egg and larvae mixing because they displace them by passive transport from the spawning area (Catanese et al., 2020). Western Mediterranean is characterized by specific water circulation patterns and hence the mixing of individuals from remote regions (Figure 10). In particular, on the west coasts of the Italian peninsula, Tyrrhenian and Ligurian Seas communicate via the Corsica Channel (West Corsica and Tyrrhenian currents, Figure 10, Astraldi et al. (1995)). This is the beginning of a huge current, the "Ligurian Current" or "Liguro–Provenço–Catalan Current" or simply "Northern Current", that is actually an entity flowing along the continental slope, at least as far as the Channel of Ibiza (Millot, 1999). Astraldi et al. (1994) and Viettia et al. 2010) have reported westward movements of European anchovy's eggs and larvae by coastal currents in the Tyrrhenian and Ligurian Seas. Also, Rubín (1996) reported a mechanism of egg and larval transport from the Gulf of Cadiz to the Alboran Sea via the inflow of the Atlantic jet. Therefore, anchovy eggs and larvae could passively migrate, taking advantage of such currents, then mixing and reproducing in remote spawning grounds, and thus generating the specific genetic patterns.

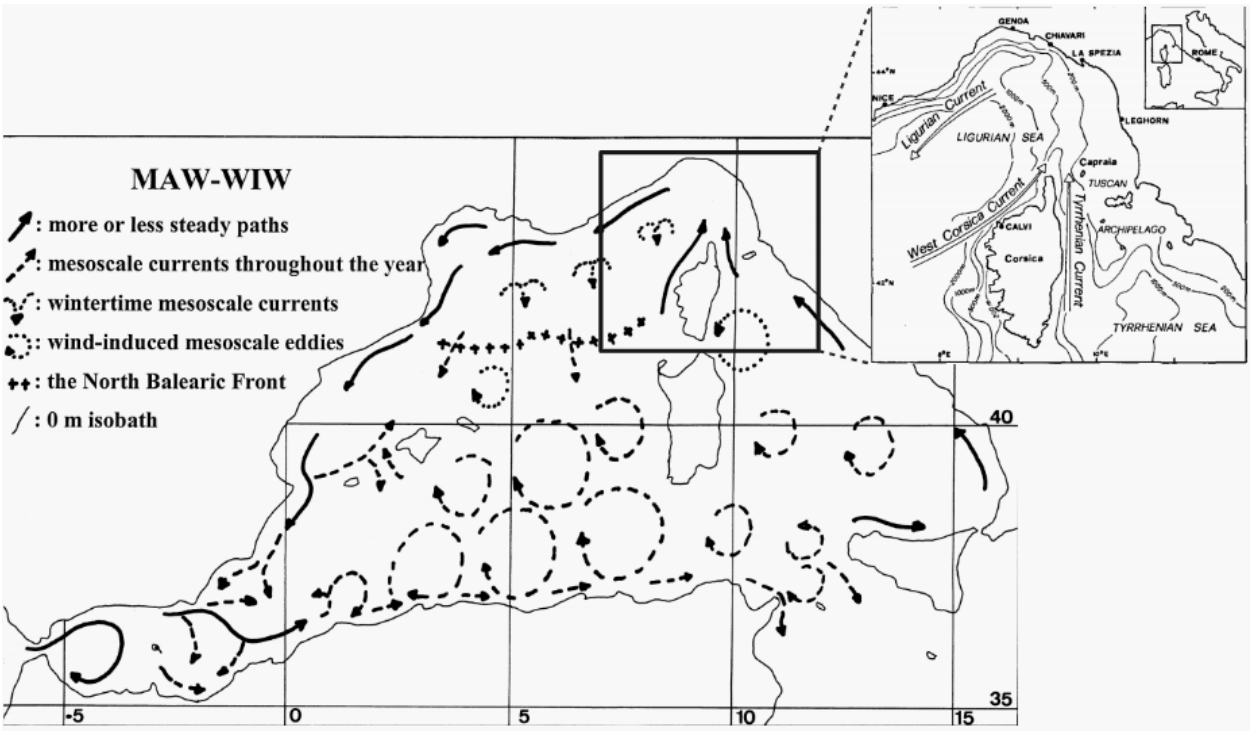

Figure 10: Circulation of the Modified Atlantic Water and the Winter Intermediate Water (Millot, 1999). Figure in top right: area in the north-western corner of the Mediterranean, showing the boundary between the Ligurian and Tyrrhenian Seas (Tuscan Archipelago). Arrows illustrate the major current systems (graph from Astraldi et al. (1995)).

Another biogeographic boundary within the Mediterranean sea is the Siculo-Tunisian Strait (Bahri-Sfar et al., 2000). The bathymetry of the Siculo-Tunisian Strait (STS) is very complex and exerts a strong influence on the water masses flowing through it (eastward: Atlantic-Ionian Stream (AIS, Robinson et al. (1999)) and westward: Ionian current), causing local currents which create a turbulent mixing of the two water bodies. Such hydrographic patterns bound the natural dispersal across the STS for particular marine fish species (Mejri et al., 2009). Such species are: *Pomatoschistus tortonesei* (Mejri et al., 2009), *Pomatoschistus minutus* (Stefanni & Thorley, 2003) and *Scomber scombrus* (Zardoya et al., 2004), whereas for other marine fish, e.g. *Psetta maxima* (Suzuki et al., 2004) and *Dicentrarchus labrax* (Bahri-Sfar et al., 2000), the STS might contribute to their genetic differentiation between western and eastern Mediterranean basins. There are a few genetic studies (Bembo et al., 1996; Borsa, 2002; Cuttitta et al., 2015; Ruggeri et al., 2016) on the stocks' connectivity of the *Engraulis encrasicolus* around the STS and thus little information is provided about the genetic relationship between the GSA19, GSA10, GSA15 and GSA16 anchovy stocks. Ruggeri et al. (2016) used eleven microsatellite markers and found high genetic similarity between the southern Adriatic i.e. GSA17-18 (Bari and Pescara regions), the Northwestern Sicily i.e. GSA10 (Castellammaere del Golfo region) and the Western Italy i.e. GSA10 (Sperloga region). Here, the GSA19 anchovy stock (Western Ionian Sea) found to be

genetically homogeneous with the GSA10 (Tyrrhenian Sea) and GSA09 stocks (Ligurian Sea) (STRUCTURE bar plot, AMOVA, pairwise-F_{ST} matrix). Thus, the genetic similarity found in GSA 17-18-19-10-09 anchovy stocks, implies potential contact.

As described above, the genetic homogeneity between the western Mediterranean anchovy stocks (GSA06-05-07-11) was explained with the particular oceanographic features in the region such as currents, waterfronts and features related to anchovy's life history, such as egg and larval dispersal (see also Cuttitta et al., 2015). Several studies, e.g. Bonanno et al. (2013); Falcini et al. (2015); García Lafuente et al. (2002), have described the eastward anchovy's egg and larval transport through the Sicily Channel (STS). Falcini et al. (2015) showed the role of the wind-induced coastal currents in influencing the fate and distribution of European anchovy larvae within the Sicily Channel. Moreover, the hydrographic circulation in the Southern Sicily plays a key role in dispersion and retention of the anchovy eggs and larvae. The Atlantic-Ionian Stream (AIS) controls the surface circulation of the Sicilian Channel and has a key role in transferring the anchovy's eggs and larvae eastward (**Error! Reference source not found.**; García Lafuente et al. (2002)). Therefore, the particular oceanographic features of the STS might be a mechanism of genetic connectivity between either side of Sicily.

However, wind-induced northern currents in conjunction with minor branches of AIS could lead to offshore dispersion of the anchovy larvae, leading to potential higher mortalities due to starvation (García Lafuente et al., 2002). For example, such an advection mechanism is located in the Andenture Bank -crucial spawning area in the Southern Sicily- where the pelagic early life stages could be transport offshore (Figure 11). Such factors that partially prevent long-distance transportation of anchovy eggs and larvae in conjunction with the less favored conditions, for spawning and larval-juvenile growth, in the Northeastern Sicily (Giannoulaki et al., 2013), could possibly explain the reason that some anchovy individuals from the GSA19 stock seemed to be a little differentiated from the MED cluster (Figure 9). In any case, Bembo et al. (1996) stated "genetic analyses are generally over-sensitive to gene flow. Relatively low levels of exchange between stocks, negligible from a management perspective, may be sufficient to ensure genetic homogeneity".

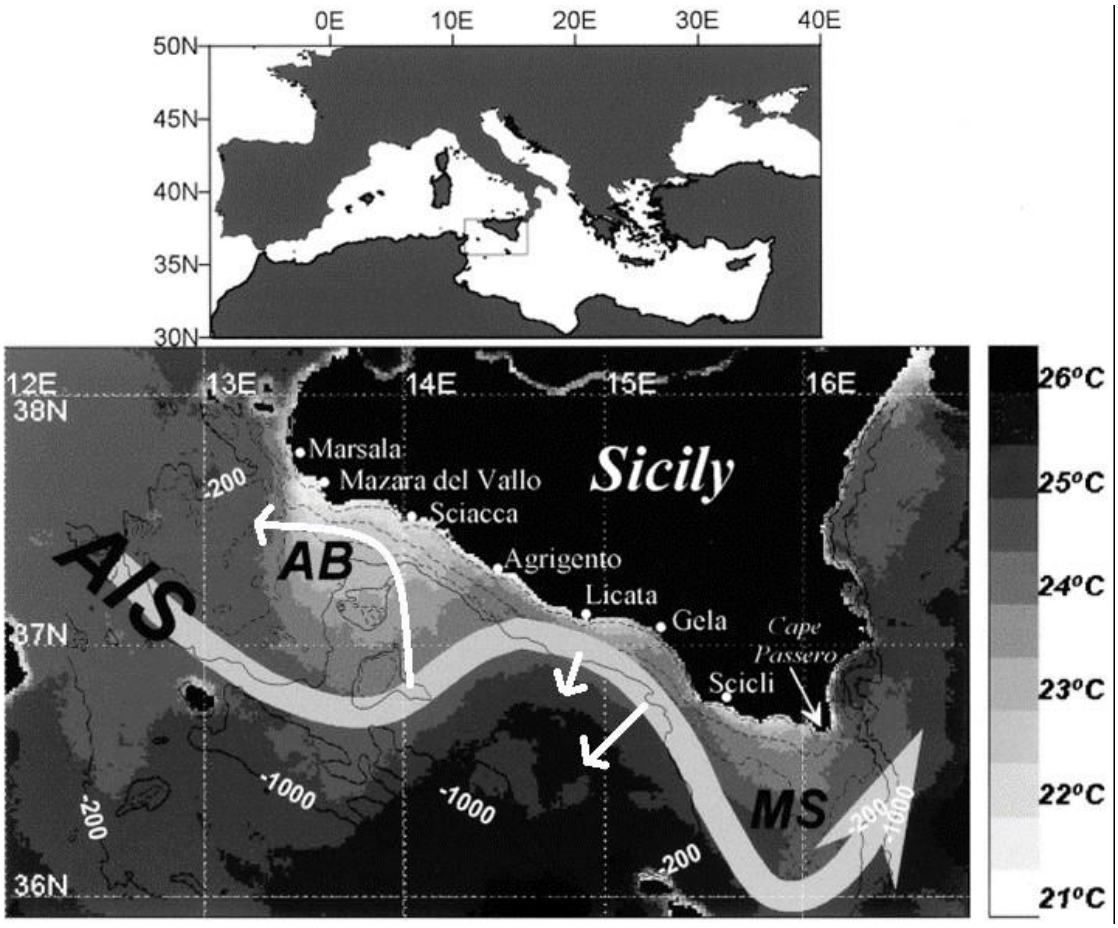

Figure 11: Figure from Garcia Lafuente et al. (2002). The mean path of the Atlantic-Ionian Stream (AIS) is sketched. Adventure Bank (AB) and Maltese Shelf (MS) are shown, which have high egg and larval concentrations, respectively. White arrows indicate the offshore advection.

Management implications

Management policies for anchovy fishery stocks should incorporate genetic studies that shed light on the population structure of the species (Viñas et al., 2013). The worst-case scenario is not detecting population structure when it actually exists (Viñas et al., 2011; Waples et al., 2008). Genetically distinct populations within the same management area should be treated as independent fishery stocks; for instance, the Alboran population should continue to be managed separately from the northwestern Mediterranean ones due to its genetic distinctiveness. On the other hand, the northwestern Mediterranean, having multiple anchovy stocks, is characterized by a genetically homogeneous anchovy population since high gene flow between the geographic sub-areas is observed. Similarly, the Alboran and Cadiz anchovy populations are genetically homogeneous. Finally, commercial fishery collapse, and even species extinction, might be inevitable if continuous overfishing takes place (Hutchings, 2001). Thus, the implementation of genetic monitoring for anchovy stocks is crucial to safeguard *Engraulis encrasicolus* both as a biological species and for its high commercial value (Sanz et al., 2008).

REFERENCES

Aleksandrov A.I. (1927). Anchovies of the Azov-Black Sea Basin, their origin and taxonomic designations. Proceedings of the Kerch Scientifc Fisheries Station 1(2–3):37–99 (in Russian)

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Astraldi, M., Bianchi, C. N., Gasparini, G. P., Morri, C., & Lerici, D. (1995). Oceanologica acta- vol. 18-N°2 ~ Climatic fluctuations, current variability and marine species distribution: Variabilité du climat Espèces thermophiles Dispersion Epibenthos Biogéographie marine abstract résumé (vol. 18).

Astraldi, M., Gasparini, G. Pietro, & Sparnocchia, S. (1994). The seasonal and interannual variability in the Ligurian-Provençal Basin. *46*, 93–113. https://doi.org/10.1029/ce046p0093

Bahri-Sfar, L., Lemaire, C., Kalthoum, O., Hassine, B., & Ois Bonhomme, F. (2000). Fragmentation of sea bass populations in the western and eastern Mediterranean as revealed by microsatellite polymorphism. https://royalsocietypublishing.org/

Bakun, A. (1996). Patterns in the Ocean. Ocean processes and marine population dynamics. California Sea Grant College System/NOAA/Centro de Investigaciones Biologicas del Noroeste, La Paz, Mexico. 323 pp., ISBN 1-888691-01-8.

Bakun, A. (2006). Wasp-waist populations and marine ecosystem dynamics: Navigating the "predator pit" topographies. Progress in Oceanography, 68(2–4), 271–288. https://doi.org/10.1016/j.pocean.2006.02.004

Bembo, D. G., Carvalho, G. R., Cingolani, N., Bembo, T. J. P., Carvalho, D. G., & Cingolani, G. R. (1996). Electrophoretic analysis of stock structure in Northern Mediterranean anchovies, *Engraulis encrasicolus*. In ICES J. mar. Sci (Vol. 53). https://academic.oup.com/icesjms/article/53/1/115/731545

Bonanno, A., Zgozi, S., Cuttitta, A., El Turki, A., Di Nieri, A., Ghmati, H., Basilone, G., Aronica, S., Hamza, M., Barra, M., Genovese, S., Falco, F., Knittweis, L., Mifsud, R., Patti, B., Bahri, T., Giacalone, G., Fontana, I., Tranchida, G., & Mazzola, S. (2013). Influence of environmental variability on anchovy early life stages (*Engraulis encrasicolus*) in two different areas of the Central Mediterranean Sea. Hydrobiologia, 701(1), 273–287. https://doi.org/10.1007/s10750-012-1285-8

Borsa P. (2002). Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). Biological Journal of the Linnean Society, 75, 261–269.

Bouchenak-Khelladi, Y., Durand, J. D., Magoulas, A., & Borsa, P. (2008). Geographic structure of European anchovy: A nuclear-DNA study. Journal of Sea Research, 59(4), 269–278. https://doi.org/10.1016/j.seares.2008.03.001

Catanese, G., Di Capua, I., Iriondo, M., Bonanno, A., Estonba, A., & Procaccini, G. (2020). Application of high-throughput single nucleotide polymorphism genotyping for assessing the origin of *Engraulis encrasicolus* eggs. Aquatic Conservation: Marine and Freshwater Ecosystems, 30(7), 1313–1324. https://doi.org/10.1002/aqc.3321

Catanese, G., Watteaux, R., Montes, I., Barra, M., Rumolo, P., Borme, D., Nardelli, B. B., Botte, V., Mazzocchi, M. G., Genovese, S., Di Capua, I., Iriondo, M., Estonba, A., Ruggeri, P.,

Tirelli, V., Caputo-Barucchi, V., Basilone, G., Bonanno, A., Iudicone, D., & Procaccini, G. (2017). Insights on the drivers of genetic divergence in the European anchovy. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-03926-z

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci *de novo* from short-read sequences. G3: Genes, Genomes, Genetics, 1(3), 171–182. https://doi.org/10.1534/g3.111.000240

Chahdi Ouazzani, K., Benazzou, T., Charouki, N., Bonhomme, F., & Chlaida, M. (2017). Genetic differentiation of European anchovy (*Engraulis encrasicolus*) along the Moroccan coast reveals a phylogeographic break around the 25th parallel North. Marine Biology Research, 13(3), 342–350. https://doi.org/10.1080/17451000.2016.1257811

Challis, R., Kumar, S., Sotero-Caio, C., Brown, M., & Blaxter, M. (2023). Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. Wellcome Open Research, 8. https://doi.org/10.12688/wellcomeopenres.18658.1

Checkley, D. M., Asch, R. G., & Rykaczewski, R. R. (2017). Climate, Anchovy, and Sardine. Annual Review of Marine Science, 9(1), 469–493. https://doi.org/10.1146/annurev-marine-122414-033819

Chen, Y., Zhang, Y., Wang, A. Y., Gao, M., & Chong, Z. (2021). Accurate long-read *de novo* assembly evaluation with Inspector. Genome Biology, 22(1). https://doi.org/10.1186/s13059-021-02527-4

Cuttitta, A., Patti, B., Maggio, T., Quinci, E. M., Pappalardo, A. M., Ferrito, V., De Pinto, V., Torri, M., Falco, F., Nicosia, A., Musco, M., Armeri, G. M., Placenti, F., Tranchida, G., Mifsud, R., Bonanno, A., & Mazzola, S. (2015). Larval population structure of *Engraulis encrasicolus* in the Strait of Sicily as revealed by morphometric and genetic analysis. Fisheries Oceanography, 24(2), 135–149. https://doi.org/10.1111/fog.12098

Dray, S., & Dufour, A.-B. (2007). Journal of Statistical Software The ade4 Package: Implementing the Duality Diagram for Ecologists. http://www.jstatsoft.org/

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

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

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. Heredity, 103(4), 285–298. https://doi.org/10.1038/hdy.2009.74

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources, 10(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fage, L. 1911. Récherches sur la biologie de l'anchois (*Engraulis encrasicolus* Linné): races-agenigrations. Ann. Ins. Oceanog. 2 (4)

Fage, L. 1920. Engraulidae-Clupeidae. Report on the Danish Oceanographic Expeditions 1908-1910 to Mediterranean and Adjacent Seas. 2. Biology. 137 pp

Falcini, F., Palatella, L., Cuttitta, A., Nardelli, B. B., Lacorata, G., Lanotte, A. S., Patti, B., & Santoleri, R. (2015). The role of hydrodynamic processes on anchovy eggs and larvae distribution in the sicily channel (Mediterranean Sea): A case study for the 2004 data set. PLoS ONE, 10(4). https://doi.org/10.1371/journal.pone.0123213

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

Furness, R. W. (2002). Management implications of interactions between fisheries and sandeel-dependent seabirds and seals in the North Sea. ICES Journal of Marine Science, 59(2), 261–269. https://doi.org/10.1006/jmsc.2001.1155

García Lafuente, J., García, A., Mazzola, L., Quintanilla, J., Delgado, J., Cuttita, A., & Patti, B. (2002). Hydrographic phenomena influencing early life stages of the Sicilian Channel Anchovy. Fisheries Oceanography, 11(1), 31–44. https://doi.org/10.1046/j.1365-2419.2002.00186.x

Giannoulaki, M., Iglesias, M., Tugores, M. P., Bonanno, A., Patti, B., De Felice, A., Leonori, I., Bigot, J. L., Tičina, V., Pyrounaki, M. M., Tsagarakis, K., Machias, A., Somarakis, S., Schismenou, E., Quinci, E., Basilone, G., Cuttitta, A., Campanella, F., Miquel, J., … Valavanis, V. (2013). Characterizing the potential habitat of European anchovy *Engraulis encrasicolus* in the Mediterranean Sea, at different life stages. Fisheries Oceanography, 22(2), 69–89. https://doi.org/10.1111/fog.12005

Goudet, J. (2005). HIERFSTAT , a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes, 5, 184–186. https://doi.org/10.1111/j.1471-8278

Grant, W. S. (2005). A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): Assessing models of population structure and the Black Sea isolation hypothesis. Genetica, 125(2–3), 293–309. https://doi.org/10.1007/s10709-005-0717-z

Hunt, G. L., & McKinnell, S. (2006). Interplay between top-down, bottom-up, and wasp-waist control in marine ecosystems. Progress in Oceanography, 68(2–4), 115–124. https://doi.org/10.1016/j.pocean.2006.02.008

Huret, M., Lebigre, C., Iriondo, M., Montes, I., & Estonba, A. (2020). Genetic population structure of anchovy (*Engraulis encrasicolus*) in North-western Europe and variability in the seasonal distribution of the stocks. Fisheries Research, 229, 105619.

Hutchings, J. A. (2001). Influence of population decline, fishing, and spawner variability on the recovery of marine fishes. Journal of Fish Biology, 59(SUPPL. A), 306–322. https://doi.org/10.1006/jfbi.2001.1756

Kongsstovu, S., Mikalsen, S. O., Homrum, E., Jacobsen, J. A., Flicek, P., & Dahl, H. A. (2019). Using long and linked reads to improve an Atlantic herring (Clupea harengus) genome assembly. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-54151-9

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Johannesson, K. (2001). Review Review Parallel speciation: a key to sympatric divergence. In TRENDS in Ecology & Evolution (Vol. 16, Issue 3). http://tree.trends.com0169

Jombart, T. (2015). A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 1.3-4. 1–43.

Jordán, F., Liu, W. C., & Wyatt, T. (2005). Topological constraints on the dynamics of wasp-waist ecosystems. *Journal of Marine Systems*, *57*(3–4), 250–263. https://doi.org/10.1016/j.jmarsys.2005.05.002

Pertierra J.P., & Lleonart J. (1996). NW Mediterranean anchovy fisheries. Scientia marina, 60,

257-267.
Karahan, A., Borsa, P., Gucu, A. C., Kandemir, I., Ozkan, E., Orek, Y. A., Acan, S. C., Koban, E., & Togan, I. (2014). Geometric morphometrics, Fourier analysis of otolith shape, and nuclear-DNA markers distinguish two anchovy species (*Engraulis* spp.) in the Eastern Mediterranean Sea. Fisheries Research, 159, 45–55. https://doi.org/10.1016/j.fishres.2014.05.009
Keskin, E., & Atar, H. H. (2012). Genetic structuring of European anchovy (*Engraulis encrasicolus*) populations through mitochondrial DNA sequences. Mitochondrial DNA, 23(2), 62–69. https://doi.org/10.3109/19401736.2011.653798
Kess, T., Gross, J., Harper, F., & Boulding, E. G. (2016). Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. Journal of Molluscan Studies, 82(1), 104–109. https://doi.org/10.1093/mollus/eyv042
Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology, 37(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8
Le Moan, A., Gagnaire, P. A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Molecular Ecology, 25(13), 3187–3202. https://doi.org/10.1111/mec.13627
Lee, J. Y., Kong, M., Oh, J., Lim, J. S., Chung, S. H., Kim, J. M., Kim, J. S., Kim, K. H., Yoo, J. C., & Kwak, W. (2021). Comparative evaluation of Nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-00178-w
Lee, Y. H., Abueg, L., Kim, J. K., Kim, Y. W., Fedrigo, O., Balacco, J., Formenti, G., Howe, K., Tracey, A., Wood, J., Thibaud-Nissen, F., Nam, B. H., No, E. S., Kim, H. R., Lee, C., Jarvis, E. D., & Kim, H. (2023). Chromosome-level genome assembly of chub mackerel (*Scomber japonicus*) from the Indo-Pacific Ocean. Scientific Data, 10(1). https://doi.org/10.1038/s41597-023-02782-z
Lewontin, R. C., & Kojima, K. (1960). The evolutionary dynamics of complex polymorphisms,. Evolution, 14(4), 458–472. https://doi.org/10.1111/j.1558-5646.1960.tb03113.x
Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Louro, B., De Moro, G., Garcia, C., Cox, C. J., Veríssimo, A., Sabatino, S. J., Santos, A. M., & Canário, A. V. M. (2019). A haplotype-resolved draft genome of the European sardine (*Sardina pilchardus*). GigaScience, 8(5). https://doi.org/10.1093/gigascience/giz059
Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. In Nature Reviews Genetics (Vol. 4, Issue 12, pp. 981–994). https://doi.org/10.1038/nrg1226
Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources, 17(1),

67–77. https://doi.org/10.1111/1755-0998.12592
Machado, A. M., Gomes-Dos-Santos, A., Fonseca, M. M., Da Fonseca, R. R., Veríssimo, A., Felício, M., Capela, R., Alves, N., Santos, M., Salvador-Caramelo, F., Domingues, M., Ruivo, R., Froufe, E., Filipe, L., & Castro, C. (2022). A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource. GIGAbYte, 1–21. https://doi.org/10.1101/2021.11.19.468211
Magoulas, A., Castilho, R., Caetano, S., Marcato, S., & Patarnello, T. (2006). Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). Molecular Phylogenetics and Evolution, 39(3), 734–746. https://doi.org/10.1016/j.ympev.2006.01.016
Magoulas, Antonios, Tsimenides, N., & Zouros, E. (1996). Mitochondrial DNA Phylogeny and the Reconstruction of the Population History of a Species: The Case of the European Anchovy (*Engraulis encrasicolus*). https://academic.oup.com/mbe/article/13/1/178/1055491
Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. Current Protocols, 1(12). https://doi.org/10.1002/cpz1.323
Mejri, R., Lo Brutto, S., Hassine, O. K. Ben, & Arculeo, M. (2009). A study on *Pomatoschistus tortonesei* Miller 1968 (Perciformes, Gobiidae) reveals the Siculo-Tunisian Strait (STS) as a breakpoint to gene flow in the Mediterranean basin. Molecular Phylogenetics and Evolution, 53(2), 596–601. https://doi.org/10.1016/j.ympev.2009.04.018
Millân M. (1999). Reproductive characteristics and condition status of anchovy *Engraulis encrasicolus* L. from the Bay of Cadiz (SW Spain).
Miller, S., Dykes, D., & Polesky, H. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic acids res, 16, 1215. doi: 10.1093/nar/16.3.1215
Millot, C. (1999). Circulation in the Western Mediterranean Sea. In Journal of Marine Systems (Vol. 20).
Motos, L., Uriarte, A., & Valencia, V. (1996). The spawning environment of the Bay of Biscay anchovy (*Engraulis encrasicolus* L.)*. In the european anchovy and its environment (Vol. 60, Issue 2).
Palomera, I., Olivar, M. P., Salat, J., Sabatés, A., Coll, M., García, A., & Morales-Nin, B. (2007). Small pelagic fish in the NW Mediterranean Sea: An ecological review. Progress in Oceanography, 74(2–3), 377–396. https://doi.org/10.1016/j.pocean.2007.04.012
Palomera, Isabel. (1992). Spawning of anchovy *Engraulis encrasicolus* in the Northwestern Mediterranean relative to hydrographic features in the region.
Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution, 8(10), 1360–1373. https://doi.org/10.1111/2041-210X.12775
Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Molecular Ecology Resources, 17(6), e268–e274. https://doi.org/10.1111/1755-0998.12702
Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. Frontiers in Genetics, 4(MAY). https://doi.org/10.3389/fgene.2013.00098
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using

Multilocus Genotype Data. http://www.stats.ox.ac.uk/pritch/home.html.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795
Qu, J., Kachman, S. D., Garrick, D., Fernando, R. L., & Cheng, H. (2020). Exact distribution of linkage disequilibrium in the presence of mutation, selection, or minor allele frequency filtering. Frontiers in Genetics, 11. https://doi.org/10.3389/fgene.2020.00362
Robinson, A. R., Sellschopp, J., Warn-Varnas, A., Leslie, W. G., Lozano, C. J., Haley, P. J., Anderson, L. A., & Lermusiaux, P. F. J. (1999). The Atlantic Ionian Stream. In Journal of Marine Systems (Vol. 20).
Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols, 12(12), 2640–2659. https://doi.org/10.1038/nprot.2017.123
Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. Molecular Ecology Notes, 4(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x
Rubín, J.P. – 1996. El ictioplancton estival en el Mar de Alborán: Composición cualitativa-cuantitativa y distribución espacial (Junio de 1993). Publicaciones Especiales (IEO), 23:33-54
Ruggeri, P., Splendiani, A., Occhipinti, G., Fioravanti, T., Santojanni, A., Leonori, I., De Felice, A., Arneri, E., Procaccini, G., Catanese, G., Ticina, V., Bonanno, A., Cerioni, P. N., Giovannotti, M., Grant, W. S., & Barucchi, V. C. (2016). Biocomplexity in populations of european anchovy in the adriatic sea. PLoS ONE, 11(4). https://doi.org/10.1371/journal.pone.0153061
Sanz, N., García-Marín, J. L., Viñas, J., Roldán, M., & Pla, C. (2008). Spawning groups of European anchovy: Population structure and management implications. ICES Journal of Marine Science, 65(9), 1635–1644. https://doi.org/10.1093/icesjms/fsn128
Schwartzlose, R. A., Alheit, J., Bakun, A., Baumgartner, T. R., Cloete, R., Crawford, R. J. M., Fletcher, W. J., Green-Ruiz, Y., Hagen, E., Kawasaki, T., Lluch-Belda, D., Lluch-Cota, S. E., Maccall, A. D., Matsuura, Y., Nevárez-Martínez, M. O., Parrish, R. H., Roy, C., Serra, R., Shust, K. V, … Zuzunaga, J. Z. (1999). Worldwide large-scale fluctuations of sardine and anchovy populations. In Afr. J. mar. Sci (Vol. 21).
Slatkin, M., & Voelm, L. (1991). F_{ST} in a Hierarchical Island Model.
Somarakis, S., Palomera, I., Garcia, A., Quintanilla, L., Koutsikopoulos, C., Uriarte, A., & Motos, L. (2004). Daily egg production of anchovy in European waters. ICES Journal of Marine Science, 61(6), 944–958. https://doi.org/10.1016/j.icesjms.2004.07.018
Spanakis, E., Tsimenides, N., & Zouros, E. (1989). Genetic differences between populations of sardine, *Sardina pilchardus*, and anchovy, *Engraulis encrasicolus*, in the Aegean and Ionian seas. Journal of Fish Biology, 35(3), 417–437. https://doi.org/10.1111/j.1095-8649.1989.tb02993.x
Stefanni, S., & Thorley, J. L. (2003). Mitochondrial DNA phylogeography reveals the existence of an Evolutionarily Significant Unit of the sand goby *Pomatoschistus minutus* in the Adriatic (Eastern Mediterranean). Molecular Phylogenetics and Evolution, 28(3), 601–609. https://doi.org/10.1016/S1055-7903(03)00054-X
Sukumaran, S., Sebastian, W., Gopalakrishnan, A., Mathew, O. K., Vysakh, V. G., Rohit, P., & Jena, J. K. (2023). The sequence and *de novo* assembly of the genome of the Indian oil

sardine, *Sardinella longiceps*. Scientific Data, 10(1). https://doi.org/10.1038/s41597-023-02481-9

Suzuki, N., Nishida, M., ÜstÜndağ K. Y., Şahin T. & Amaoka, D. K. (2004). Phylogeographic relationships within the Mediterranean turbot inferred by mitochondrial DNA haplotype variation. Journal of Fish Biology, 65, 580–585. https://doi.org/10.1111/j.1095-8649.2004.00433.x

Tintore J., La Violette P.E, Blade I., & Cruzado A. (1998). Tintore et al, 1988. Journal of Physical Oceanography.

Tudela, S. (1999). Morphological variability in a Mediterranean, genetically homogeneous population of the European anchovy, *Engraulis encrasicolus*.

Turan, C., Ergüden, D., Gürlek, M., Başusta, N., & Turan, F. (2004). Morphometric Structuring of the Anchovy (*Engraulis encrasicolus* L.) in the Black, Aegean and Northeastern Mediterranean Seas.

Tyson, J. (2020). Rocky Mountain adventures in Genomic DNA sample preparation, ligation protocol optimisation/simplification and ultra long read generation. protocols.io https://dx.doi.org/10.17504/protocols.io.7euhjew

Ulaszewski, B., Meger, J., & Burczyk, J. (2021). Comparative analysis of SNP discovery and genotyping in Fagus sylvatica L. and Quercus robur L. using RADseq, GBS, and ddRAD methods. Forests, 12(2), 1–17. https://doi.org/10.3390/f12020222

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Research, 27(5), 737–746. https://doi.org/10.1101/gr.214270.116

Viettia, R. C., Albertellia, G., Alianib, S., Bava, S., Bavestrello, G., Cecchi, L. B., Bianchi, C. N., Bozzo, E., Capello, M., Castellano, M., Cerrano, C., Chiantore, M., Corradia, N., Cocito, S., Cutroneo, L., Diviacco, G., Fabiano, M., Faimali, M., Ferrari, M., … Wurtz, M. (2010). The Ligurian Sea: Present status, problems and perspectives. In Chemistry and Ecology (Vol. 26, Issue SUPPL. 1, pp. 319–340). https://doi.org/10.1080/02757541003689845

Viñas, J., Gordoa, A., Fernández-Cebrián, R., Pla, C., Vahdet, Ü., & Araguas, R. M. (2011). Facts and uncertainties about the genetic population structure of Atlantic bluefin tuna (*Thunnus thynnus*) in the Mediterranean. Implications for fishery management. Reviews in Fish Biology and Fisheries, 21(3), 527–541. https://doi.org/10.1007/s11160-010-9174-6

Viñas, J., Sanz, N., Peñarrubia, L., Araguas, R. M., García-Marín, J. L., Roldán, M. I., & Pla, C. (2013). Genetic population structure of European anchovy in the Mediterranean Sea and the Northeast Atlantic Ocean using sequence analysis of the mitochondrial DNA control region. ICES Journal of Marine Science, 71(2), 391–397. https://doi.org/10.1093/icesjms/fst132

Waples, R S. (1998). Separating the Wheat From the Chaff: Patterns of Genetic Differentiation in High Gene Flow Species. In The American Genetic Association (Vol. 438).

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

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

Whitehead, P. J. P. 1984. Engraulidae. [in:] Whitehead. P. J.P., M.-L. Bauchot, J. -C. Hureau, J. Nielsen and E. Tortonese. Fishes of the North-eastern Atlantic and the Mediterranean. Vol.

I. Unesco: 282-283
Whitehead, P. J. P., G. J. Nelson and Th. Wongratana 1988. FAO Species Catalogue vol. 7 Clupeoid fishes of the World (Suborder Clupeoidei) Part 2 Engraulidae. FAO Fish. Synop. no. 123, vol. 7, part 2: 305-579.
Wright, S. (1951). The genetical structure of populations. Annals of Eugenics, 15(4), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x
Zardoya, R., Castilho, R., Grande, C., Favre-Krey, L., Caetano, S., Marcato, S., Krey, G., & Patarnello, T. (2004). Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus*) and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. Molecular Ecology, 13(7), 1785–1798. https://doi.org/10.1111/j.1365-294X.2004.02198.x
Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M. A., Manzano, C., Grant, W. S., Irigoien, X., & Estonba, A. (2012). Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in european anchovy (*Engraulis encrasicolus* L.). PLoS ONE, 7(7). https://doi.org/10.1371/journal.pone.0042201
Zarraonaindia, I., Pardo, M. A., Iriondo, M., Manzano, C., & Estonba, A. (2009). Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. https://academic.oup.com/icesjms/article/66/10/2176/679546
Zlatanos, S., & Laskaridis, K. (2007). Seasonal variation in the fatty acid composition of three Mediterranean fish - sardine (*Sardina pilchardus*), anchovy (*Engraulis encrasicholus*) and picarel (*Spicara smaris*). Food Chemistry, 103(3), 725–728. https://doi.org/10.1016/j.foodchem.2006.09.013