scientific reports

Check for updates

OPEN

Mitochondrial genomes of the European sardine (*Sardina pilchardus*) reveal Pliocene diversification, extensive gene flow and pervasive purifying selection

Ana Rita Vieira[1,2,5]✉, Filipe de Sousa[2,3,5], João Bilro[3], Mariana Bray Viegas[3], Richard Svanbäck[4], Leonel S. Gordo[1,2] & Octávio S. Paulo[2,3]

The development of management strategies for the promotion of sustainable fisheries relies on a deep knowledge of ecological and evolutionary processes driving the diversification and genetic variation of marine organisms. Sustainability strategies are especially relevant for marine species such as the European sardine (*Sardina pilchardus*), a small pelagic fish with high ecological and socioeconomic importance, especially in Southern Europe, whose stock has declined since 2006, possibly due to environmental factors. Here, we generated sequences for 139 mitochondrial genomes from individuals from 19 different geographical locations across most of the species distribution range, which was used to assess genetic diversity, diversification history and genomic signatures of selection. Our data supported an extensive gene flow in European sardine. However, phylogenetic analyses of mitogenomes revealed diversification patterns related to climate shifts in the late Miocene and Pliocene that may indicate past divergence related to rapid demographic expansion. Tests of selection showed a significant signature of purifying selection, but positive selection was also detected in different sites and specific mitochondrial lineages. Our results showed that European sardine diversification has been strongly driven by climate shifts, and rapid changes in marine environmental conditions are likely to strongly affect the distribution and stock size of this species.

Keywords Small pelagic fish, Mitogenome, Geographic structure, Positive selection, Late Miocene cooling, Mid-Pliocene warm period

Understanding the ecological and evolutionary mechanisms driving the diversification and dispersal of marine organisms is essential for elucidating their genetic variation patterns. This knowledge underpins the development of conservation and management strategies for species of economic and ecological importance[1,2]. Advances in marine genomics provide new insights into the evolutionary history and population structure of marine organisms, as well as into the evolutionary consequences of selective harvest, local adaptation, and response to climate change[2–5]. Genomic data have also become increasingly relevant for the assessment and promotion of sustainable fisheries[6,7], for example, by enabling demographic analyses for stock identification and management and the assessment of connectivity among geographically delimited stocks.

Population structure and adaptive divergence in marine fishes have been presumed to be minimal, since marine environments have fewer barriers to gene flow compared to terrestrial ecosystems, resulting in high levels of connectivity among populations and large population sizes in marine species[1,8]. However, large population sizes may increase the probability of retention of advantageous alleles, a phenomenon facilitated by local selective

[1]MARE – Marine and Environmental Sciences Centre & ARNET – Aquatic Research Network, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal. [2]Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal. [3]cE3c – Centre for Ecology, Evolution and Environmental Changes & CHANGE – Global Change and Sustainability Institute, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal. [4]Department of Ecology and Genetics, Section of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden. [5]Ana Rita Vieira and Filipe de Sousa have contributed equally to this work. ✉email: arivieira@ciencias.ulisboa.pt

pressures[9], and adaptive processes have been shown to shape genetic patterns in oceanic fish populations[8,10–16]. Adaptive processes in oceanic fish populations may be correlated with sea surface temperatures[8], which are known to have fluctuated significantly during geological periods such as the Late-Miocene Cooling[17] and the Mid-Pliocene Warm Period[18]. Local adaptation is the driving force behind divergent selection[19], when different alleles are selected in different subpopulations, in contrast with global adaptation, when the same allele is selected across all species populations. Local adaptation results from two antagonistic forces, natural selection, which promotes intraspecific differentiation, and gene flow, which promotes homogenization. Identifying genomic signatures of natural selection is pivotal for unravelling the molecular mechanisms underlying adaptation[20]. Genomic signatures can result from different types of natural selection, manifesting in two main forms: positive and negative (or purifying) selection. Positive selection can be divided into balancing selection, preserving genetic polymorphisms, and directional selection, driving advantageous alleles to fixation, in contrast to purifying selection that works to eliminate deleterious mutations within a population[21].

The European sardine, *Sardina pilchardus* (Walbaum, 1792), is a small pelagic fish from the Alosidae family, inhabiting the Northeast Atlantic Ocean, from the North Sea to Mauritania and Senegal and with populations in the Azores, Madeira and Canaries, and the Mediterranean Sea[22]. The European sardine is a migratory and schooling species that forms schools potentially comprising millions of individuals and is known to prefer colder water for living and spawning[23–26]. The European sardine plays an important role in marine ecosystems, as both a consumer of plankton and a prey for larger predators[27,28]. Moreover, it is one of the most important marine fish resources in Southern Europe and Morocco[29], especially in the Iberian Peninsula[30], where its landings represent ~ 40% of the total capture[31]. It constitutes the main target species for the purse-seine fleets operating in Portugal and Spain, thereby serving as a critical revenue stream for the respective local economies[32]. The biomass of the Ibero-Atlantic stock has been declining since 2006, as its recruitment is strongly related to environmental conditions[33]. This decline has led sardine abundance to fall to its historical minimum[34], triggering profound socio-economic impacts on fishing communities. Consequently, this species has been subject to numerous studies, namely on its biology and ecology[23,24,35–38], phenotypic variation[39–44], population genetics, besides a complete genome sequencing[45,46].

The mitochondrial genome (mitogenome), a maternally inherited, circular DNA molecule, has been a focal point in the study of evolutionary biology and population genetics due to its relatively high mutation rate, lack of recombination, and haploid nature. The positive and negative features of mitochondrial DNA (mtDNA) for population genetics, phylogeographic and phylogenetic studies have been extensively discussed[47–49]. Nevertheless, mitogenomes can provide valuable information at a relatively low cost as a byproduct of whole genome resequencing. Positive selection in mtDNA can be detected due to direct selection on the mitogenome or indirect selection in the nuclear genes that compose the mito-nuclear complex. The mitogenome contains 13 protein-coding genes that contribute to four Electron Transport System (ETS) complexes, whose function requires over 500 proteins encoded in the nuclear genome[50]. This interaction between the mitogenome and the nuclear genome could generate diverse response patterns with compensatory mechanisms and coevolution between the two genomic compartments. A study on 70 mitogenomes of Clupeoid fishes[51], including the European sardine, concluded for the prevalence of purifying selection, but also for the observed shift in codon preference patterns between marine and euryhaline/freshwater Clupeoids, indicating possible selection for improved translational efficiency while adapting to low-salinity habitats. This mitogenomic plasticity and enhanced efficiency of the metabolic machinery may have contributed to the evolutionary success and abundance of Clupeoid fish[51]. Mitogenomes can also harbour rare mutations that provide a selective advantage through the interaction with environmental factors such as temperature[47], raising the hypothesis that the wide temperature range of the European sardine increases its potential for local adaptation due to divergent selection.

Previous studies on European sardine[52] have largely focused on mtDNA fragments to infer population structure, historical demography, and signatures of molecular adaptation. However, the resolution provided by partial mitochondrial data is limited, often obscuring finer-scale evolutionary processes and the detection of adaptive genetic variation. In this study, complete mitochondrial genomes from individuals collected across the entire distribution range of the European sardine were sequenced and analysed with three main objectives: 1) to assess the population genetic structure inside the species; 2) to reconstruct the phylogenetic relationships among lineages and explore the timing of diversification events within the species; and 3) to evaluate the role of natural selection (both positive and negative) in shaping mitochondrial genome variation, and its potential for local adaptation or association with intraspecific lineage.

Materials and methods

Sampling, gDNA extraction, and sequencing

A total of 380 European sardine muscle samples were collected from 19 different geographical locations covering, as much as possible, the species' entire distribution area (Fig. 1, Table 1). Samples were collected from fresh fish landed in fishing ports (15 locations) and from scientific research campaigns (4 locations) and preserved in absolute ethanol (Merck KGaA, Darmstadt, Germany) at 4 °C.

Total genomic DNA (gDNA) was extracted using Qiagen DNeasy Blood & Tissue Kits (Hilden, Germany) according to the manufacturer's instructions with the following modifications: 1) RNase A, added to the digestion product, was incubated for 30 min at room temperature; 2) -20 °C absolute ethanol was immediately added after the buffer AL; and 3) 65 µl of buffer AE at 37 °C were pipetted directly onto the DNeasy membranes to increase gDNA yield. DNA integrity was confirmed in 1% TBE agarose gel electrophoresis (9 V cm^{-1}) and DNA purity and concentration were verified by spectrophotometry using a Nanodrop (Thermo Fisher Scientific, Massachusetts, U.S.A.), and fluorometry with a Qubit (Thermo Fisher Scientific, Massachusetts, U.S.A.).

The SNP&SEQ Technology Platform (Uppsala, Sweden) constructed the TruSeq Nano DNA libraries and sequenced paired-end reads (150 base pairs, bp) in 2 lanes of a S4 flowcell using one NovaSeq 6000 and v1.5

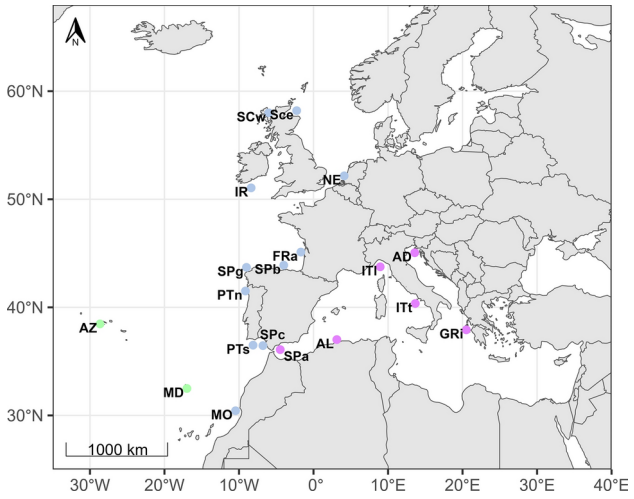

Fig. 1. Sampling map. Sampling locations of the European sardine (*Sardina pilchardus*) across the species' distribution geographic area. Produced with R Statistical Software (version 4.3.0 – https://www.R-project.org/). Sampling points coloured following the population clusters with maximum differentiation in AMOVA analysis (Fig. 2). Correspondence of each sampling point given in Table 1.

Location code	Country	Region	Origin	Sampling date	Number of collected samples	Number of sequenced samples
AD	Croatia	Adriatic Sea (Split)	Fishing port	Aug-2020	15	7
AL	Algerie	Algiers	Fishing port	Dec-2020	15	7
AZ	Portugal	Horta, Azores	Fishing port	Feb-2020	30	9
FRa	France	Arcachon-Royan Bay	Fishing port	Sep-2019	15	6
GRi	Greece	Ionian Sea (Messolonghi)	Fishing port	Jul-2019	15	6
IR	Irland	ICES division 7.g	Scientific research campaign	Oct-2019	15	10
ITl	Italy	Lingurian Sea (Genoa)	Fishing port	Sep-2019	15	7
ITt	Italy	Tyrrhenian Sea (Naples)	Fishing port	Sep-2019	15	8
MD	Portugal	Funchal, Madeira	Fishing port	Dec-2019	30	10
MO	Morocco	Agadir	Fishing port	Nov-2019	15	9
NE	Netherlands	ICES division 4.c	Fishing port	Jun-2020	15	1
PTn	Portugal	Viana do Castelo	Fishing port	Aug-2019	30	8
PTs	Portugal	Quarteira	Fishing port	Aug-2019	30	9
Sce	Scotland	ICES division 4.a	Scientific research campaign	Jan-2020	15	8
SCw	Scotland	ICES division 6.a	Scientific research campaign	Mar-2020	15	8
SPa	Spain	Alboran Sea (Fuengirola)	Fishing port	Sep-2019	15	8
SPb	Spain	Suances	Fishing port	Aug-2020	20	8
SPc	Spain	Cadiz	Scientific research campaign	Feb-2020	30	3
SPg	Spain	La Coruña	Fishing port	Jul-2020	30	7
				Total	380	139

Table 1. Overview of sample collection and sequencing by location. Summary of geographic location, samples origin, sampling date and number of collected and sequenced samples for the European sardine used in this study.

sequencing chemistry (Illumina Inc). In total, 139 European sardine specimens were sequenced (Table 1) with a sequencing depth of around 10x.

Quality control at different phases of the assembly process was carried out with FastQC v. 0.11.9[53] and MultiQC v. 1.11[54]. Fastp v. 0.23.2[55] was used to remove adaptors, poly G tails (-g), overrepresented sequences (-p), duplicate reads (-D), reads shorter than 80 bp (-l) and reads with a qualified quality phred score below 20 (-q). After this cleaning step, mitogenomes were assembled using GetOrganelle v. 1.7.7.0[56], with maximum extension rounds of 10 (-R), SPAdes k-mer settings (-k) at 21,45,65,85, 105 with the seed file for the mitochondrial DNA of sardine MH329246[46].

Alignment and annotation

Sequences of the 139 assembled circular genomes were manually edited to have matching start and end positions. The length of assembled genomes varied from 17,571 bp (AZ8) to 17,531 bp (AZ30). Circular genomes were annotated in GenBank format using the automated online pipeline MitoAnnotator, available at the Mitochondrial Genome Database of Fish[57–59]. Individual features (genes, tRNA, rRNA, D-loop/control region) were extracted from each GenBank-format file, using a custom script. Whole genome sequences were aligned using MAFFT v. 7.3[60] and the "auto" setting. The script catfasta2phyml.pl[61] was used to concatenate all 13 protein-coding genes. For a calibration analysis, an additional alignment of 12 concatenated genes was generated, that excluded the gene ND6 and included 3 outgroup sequences from two species in genus *Sardinops*: *S. melanostictus* (Schlegel, 1846) and *S. sagax* (Jenyns, 1842). The whole genome alignment had 17,581 bp and the concatenated alignments of 13 and 12 genes had 11,508 bp and 10,907 bp, respectively. The list of genes included in each dataset is presented in Table 2. The concatenated alignments and input files for Iqtree and BEAST are available in Zenodo (https://doi.org/https://doi.org/10.5281/zenodo.13247537).

Diversity and phylogenetic analyses

To test which grouping of populations showed the greatest amount of differentiation, a standard AMOVA was run on the whole genome alignment using Arlequin v. 3.5.2[62], with pairwise differences and 1000 permutations. Seven different arrangements of two to five groups were tested. A median-joining network (MJN) was calculated from the 13 gene alignment using PopArt v1.7 with default settings[63]. The number of haplotypes, haplotype diversity, nucleotide diversity (π), Tajima's D, Watterson's estimator (θ), and R_2[64] were calculated from the whole genome data set, for each sampling location and each cluster identified in the MJN, using the packages pegas v. 1.2[65], adegenet v. 2.1.10[66,67] and ape v. 5.7.1[68] as implemented in R v. 4.2.1[69] and the RStudio environment[70]. Pairwise F_{ST} was calculated using the package hierfstat v. 0.5–11[71]. The package pegas v. 1.2[65], implemented in R v. 4.2.1[69], was used for mismatch distribution analyses of the whole genome dataset. Mismatch distribution was calculated for the entire dataset (139 samples) and for each cluster observed in the MJN. Expansion times for each cluster were calculated using the equation $t = tau/2u$, where tau is the mode of the mismatch distribution, and $u = \mu k$ (μ = substitution rate per generation; k = number nucleotides in the sequence). A principal components analysis (PCA) based on the whole genome dataset with 139 samples was run on R using the package ADEGENET v 2.1.10 and visualised using the package ggplot2 v. 3.5.1[72]. The best-fit partitioning scheme and nucleotide substitution models for the concatenated datasets were estimated in PartitionFinder v. 2.1.1[73] using the greedy algorithm, linked branch lengths and the corrected Akaike Information Criterion (AICc) for choice of models[73–75]. For the maximum-likelihood analysis, data blocks included different codon positions for the 13 protein-coding genes, and all models were considered. A maximum-likelihood analysis of the 13 gene concatenated data set was run using IQTree v. 2.2.2.3[76], with a 16-partition scheme as determined by PartitionFinder. Support for the best tree was obtained with 300 bootstrap replicates. For the calibration analysis, datablocks consisted of the 12 individual genes and the models considered were JC, HKY and GTR. A time-calibrated phylogenetic tree was inferred using BEAST v. 2.7.4[77], based on the 12 gene concatenated alignment, with a 6-partition scheme as determined by PartitionFinder. To obtain a calibration point, three individuals of genus *Sardinops* Hubbs, 1929 were used as outgroups. Mitochondrial genome sequences of the three *Sardinops* individuals were retrieved from Genbank (Sard 16: *S. melanostictus* NC_002616; Sard 17: *S. sagax* NC_057117; Sard 41: *S. sagax* OR482441). Since using all 139 European sardine samples would likely result in poor MCMC convergence, the 12-gene dataset was reduced to 63 samples, 20 samples randomly chosen from each apparent cluster observed on the MJN (G1, G2 and G3), and 3 outgroup taxa. Sardine samples were chosen by generating a random number list. The dataset for the time-calibration analysis was partitioned following the best scheme indicated by PartitionFinder, with the 12 genes arranged into six subsets. The models employed were GTR + Gamma for four of the subsets, and HKY + Gamma for two of the subsets. Clock and tree

Gene name	13-gene alignment	12-gene alignment
ATPase6	X	X
ATPase8	X	X
COXI	X	X
COXII	X	X
COXIII	X	X
CYTB	X	X
ND1	X	X
ND2	X	X
ND3	X	X
ND4	X	X
ND4L	X	X
ND5	X	X
ND6	X	–

Table 2. Protein-coding genes. List of protein-coding genes included in the concatenated alignments used in Maximum-likelihood (13 genes) and calibration (12 genes) analyses.

models were linked for all partitions. A Yule tree model and an optimised relaxed clock (ORC) were defined, and a Gamma prior was used for "birthrate" (alpha = 0.001, beta = 1000). No consensus exists on molecular rates for fish mitogenomes, although much work has been done on the subject[78–80]. Bagley et al.[79] report a "fish rate" with a range of 0.017×10^{-8}-0.14×10^{-8}. To approximate this range and its mean, the ORC ucld.mean was set with a lognormal prior: mean (in real space) = 0.1×10^{-8}, sigma = 0.695 (median: 7.85×10^{-10}). The divergence of genus *Sardina* and *Sardinops* was estimated at c. 23 Ma by Bloom and Lovejoy[81]. Considering the reported age, the time to the most recent common ancestor (TMRCA) for *Sardina*/*Sardinops* was defined with a lognormal prior: mean (real space) = 23×10^6, sigma = 0.35 (95% interval: 1.09×10^7–4.3×10^7). The analysis ran for 200×10^6 MCMC generations on the CIPRES Science Gateway[82]. Convergence was validated using Tracer v. 1.7.2., after a 10% burnin. A maximum clade credibility tree with common ancestor heights was obtained with TreeAnnotator v. 2.7.4 using a 10% burnin. In addition to the calibrated tree analysis, a polymorphisms-aware phylogenetic model (PoMo) was used to analyse the 12 gene dataset[83,84], excluding the outgroup as no root age was incorporated in the analysis. The 12 gene sequence alignment was converted from FASTA to counts format using the script "FastaToCounts.py" (available at https://github.com/pomo-dev/PoMo). A PoMo analysis was run on IQTree with a GTR + G substitution model and default settings (weighted binomial sampling (+ WB) and virtual population size = 9 (+ N9)). Branch support was obtained with 1000 ultrafast bootstrap replicates.

Tests of selection

The signal of positive selection in the 139 mitochondrial genomes was investigated using computational methods that rely on maximum likelihood (ML) and Bayesian inference (BI) estimates to calculate the ratio of nonsynonymous to synonymous substitution rates (ω = dN/dS) across codons, based on a given phylogeny. Analyses were run on the 13 individual gene alignments using the ML phylogeny inferred with IQTree. The methods used were FUBAR (Fast Unconstrained Bayesian AppRoximation method)[85], MEME (Mixed Effects Model of Evolution)[86], and FEL (Fixed effects likelihood)[87], which were implemented in the program HyPhy[88]. FUBAR detects pervasive selection (positive and purifying) across codons, whereas MEME detects positive selection, both pervasive and episodic (affecting a subset of the phylogeny). FEL calculates ω assuming that selective pressures are constant across the phylogeny. The default values were used to test the significance for each model (p-value < 0.1 for MEME, posterior probability > 0.9 for FUBAR, p-value < 0.1 for FEL). Selective pressures were further investigated using site, branch and branch-site models in the program CodeML, which is integrated in the package PAML v. 4.9j[89]. Branch and branch-site models were run for each of the haplogroups that were established by the AMOVA analysis by defining them as foreground branches in different runs. The necessary tree annotations were performed on the ML tree using the web server-distributed version of phylotree.js[90].For sites under selection, "ancestral" codons were inferred by comparison with the mitochondrial genomes of three Alosidae species: *Alosa alosa* (Linnaeus, 1758), *Brevoortia tyrannus* (Latrobe, 1802) and *Sardinops melanostictus* (Temminck & Schlegel, 1846) (current accepted name *Sardinops sagax* (Jenyns, 1842)).

Results

Alignment and annotation

A total of 139 European sardine samples were sequenced producing a total of 12.5×10^9 raw reads, with an average of 89 ± 15 million raw reads per individual. After the quality control phase and removal of adapters, contaminants and duplicates, a total of 10.9×10^9 reads remained with an average of 77 ± 13 million reads per individual sample. Assembled mtDNA for all samples, from GetOrganelles, were obtained with a number of reads between 13 and 20×10^6 and an average kmer-coverage of around 157 and a base coverage of around 500.

Annotation using the automated online pipeline MitoAnnotator resulted in annotated mitogenomes containing 13 protein-coding genes, 22 tRNAs, 2 rRNAs and the D-loop. One gene (ND6) and eight tRNA genes were annotated as complement sequences, as expected[91]. The 139 total sardine mtDNA sequences are available on GenBank with the accession numbers PQ010125-PQ010263.

Geographic structure

The AMOVA analysis on different population clusters, based on whole genome data, revealed an overall low variation among populations and clusters (Supplementary Table S1). The largest differentiation among the tested clusters ($F_{CT} = 0.05$) was obtained when considering three geographic clusters, one comprising all Mediterranean populations, a second comprising Madeira and Azores populations, and the third containing all other Atlantic populations. The MJN inferred from the 13 gene concatenated dataset recovers two large and distinct groups (G1 and G2) connected to a central group (G3) of smaller lineages through several unsampled or extinct haplotypes (Fig. 2). Samples were coloured following the population clusters with maximum differentiation as per AMOVA analysis. No clear geographical correspondence is evident in the three haplogroups, as samples from the same site appear scattered through the network, although one of the groups (G2) is predominantly Mediterranean, whereas the others are predominantly Atlantic (G1 and G3). The PCA analysis (Fig. 3) also shows the three distinct groups G1, G2 and G3, the latter appearing much more scattered than the other groups, especially on axes 1 and 3. The number of haplotypes found was equal to the number of samples for every sampling site, and haplotype diversity was thus maximum (1) for all locations. Nucleotide diversity varied from 0.0035 (ITl) to 0.0064 (AZ), with an average of 0.0049. Tajima's D was negative in all sampling sites but values were non-significant, whereas R_2 values were low and significant. Theta varied from 0.0052 (ITl) to 0.0092 (AZ).p

Diversity analyses on the three clusters (Table 3) show that G3 has the highest nucleotide diversity and theta. Tajima's D and R_2 were negative with a significance of 0.01 for G1 and G2. The plot for the mismatch distribution analysis, based on whole genome data and 139 samples, showed a deviation from the expected distribution with a multimodal, nearly bimodal profile (Supplementary Fig. S1a). Plots for clusters G1 and G2 have unimodal profile (Supplementary Fig. S1b, c), and the plot for cluster G3 has a multimodal profile (Supplementary Fig.

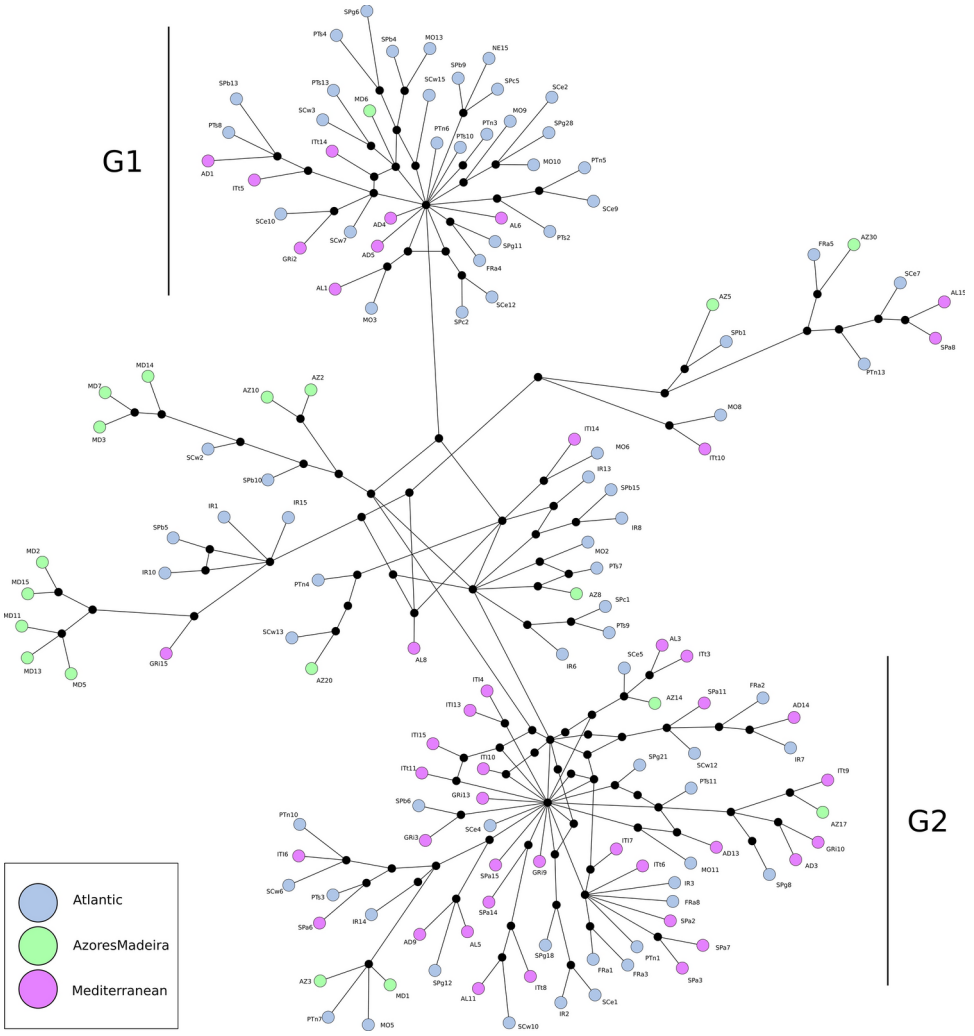

Fig. 2. Haplotype median-joining network. Haplotype median-joining network calculated from the 13 gene alignment using PopArt. Samples coloured according to the population clusters with maximum differentiation in AMOVA analysis.

S1d). Time since expansion, assuming a mean generation time of two years[37,92], was estimated at 3.27 Ma for G1, 2.96 Ma for G2 and 6.2 Ma for G3.

Phylogenetic analysis

The best-fit partitioning scheme for the 13 gene concatenated dataset and 39 data blocks (different codon positions for the 13 protein-coding genes) comprised 16 partitions and 8 different substitution models. For the 12 gene concatenated dataset that included 3 outgroup sequences, the best-fit partitioning scheme comprised 6 partitions and 2 substitution models. The tree obtained from the maximum-likelihood partitioned analysis of the 13-gene dataset confirms the presence of two large fully supported (BS = 100) haplogroups and several smaller, supported lineages (Supplementary Fig. S2).

The MCMC of the calibrated phylogenetic analysis converged within 200E6 MCMC generations, with all parameters reaching ESS values above 200. The tree obtained shows two distinct, diverse haplogroups with full support (PP = 1.0) and several lineages of which five are supported (PP > 0.95). The root of the tree was estimated at 20.1 Ma. The crown age for European sardine is estimated at 5.6 Ma (95% HPD = 1.96–10.07), whereas the crown age of the two larger, supported haplogroups is estimated at 2.94 Ma (95% HPD G1 = 0.99–5.44, G2 = 0.96–5.36). Relationships among the main lineages are unsupported (Fig. 4). The best scoring tree of the PoMo analysis (Supplementary Fig. S3) shows a topology compatible with the calibrated tree, with no conflicting supported branches and most supported branches in the calibrated tree also supported in the PoMo tree (BS ≥ 80%).

Tests of selection

Positive selection was detected, by at least two methods, for five candidate sites in four different genes (Table 4). Two sites under positive selection were detected in gene ND2, and one site in genes ND4, ND5 and ND6.

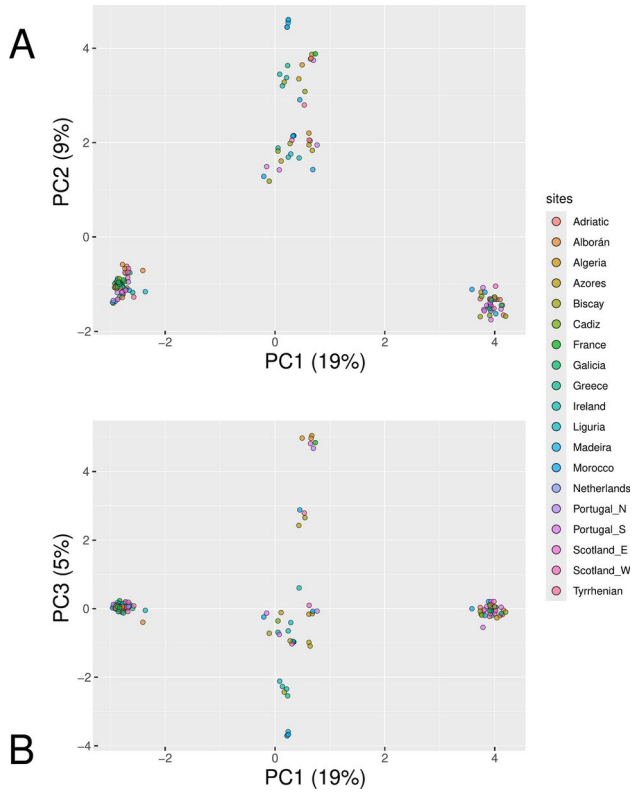

Fig. 3. Principal components analysis. PCA of the whole mitochondrial genome dataset with 139 samples; (**A**) PC1 + PC2; (**B**) PC1 + PC3. Principal components 1, 2 and 3 explain 19%, 9% and 5% of the variance, respectively.

All sites were detected with the program codeml (Table 4; Supplementary Table S2) except for site ND4/95. FUBAR detected sites ND2/218, ND5/419 and ND6/195; MEME detected sites ND2/319, ND4/95 and ND6/195; FEL detected ND2/319, ND4/95 and ND6/195. Codon replacement in European sardine affected five unrelated samples in site ND2/218 and four unrelated samples ND5/419 (Table 4; Supplementary Fig. S2). In site ND6/195, 14 samples had codon replacement, of which six form a clade. Site ND6/195 is located in a 63 bp insertion at the end of the gene that is present in all samples of *S. pilchardus*. In ND4/95, 138 samples had codon replacement, one of them representing a secondary replacement from the "derived" Threonine to Alanine, and one representing a reversal to Methionine. In sites ND2/218 and ND2/319, 5 and 126 samples were affected (Table 4; Supplementary Fig. S2). Purifying selection was detected by FUBAR and FEL for 18.09% and 15.97% of the total number of codon sites, respectively. The number of codon sites detected and shared by both methods was 12.91% and the cumulative number of sites was 21.14% (Supplementary Table S3).

Discussion

Diversity and gene flow

Analyses of the 139 European sardine mitogenomes show extensive gene flow, in which the maternally inherited mtDNA is a poor predictor of geographic origin. Under the best AMOVA scenario, maximizing the value of F_{CT}, the obtained three clusters, Atlantic, Mediterranean and Azores/Madeira, explained only 5% of the variation while variation within populations explained 94% of the total variation.

Strikingly, in the MJN, samples collected on the same site appear scattered in different haplogroups and associated with haplotypes from geographically distant sites (Fig. 2). For example, Gri2 and Gri10, two samples from the easternmost Mediterranean population (Ionian Sea, Greece), have mitogenomes nested in the two larger, distinct haplogroups, and closest to haplotypes from Scotland and the Adriatic Sea, respectively (Fig. 2). A separation between individuals from Atlantic and Mediterranean sites is not clear on the network, although G2 seems to predominantly contain Mediterranean samples (Fig. 2).

Evidence for a lack of genetic structure in the European sardine mitogenomes had also been previously reported for analyses on 96 samples from five locations around the Iberian Peninsula[52] and 108 samples from 15 locations across most of the species distribution area[93]. This lack of geographic structure and extensive gene flow for the mtDNA contrasts with the results for the nuclear genomes that supported the existence of 3 genetic clusters[93], coinciding approximately with the three AMOVA clusters from our study. However, those nuclear data show evidence of variable admixture for most of the populations analysed, except for Madeira and Azores[93].

Genetic diversity estimates (Table 3) for the two distinct groups (G1 and G2) recovered from the MJN show signs of population expansion, with negative and significant values of Tajima's D and R_2 for all three groups (Table

Sampling site/Group	N	N hap	hap. div	nuc. div	D	R_2
Adriatic (AD)	7	7	1	0.0044	-1.0291	0.0759*
Algeria (AL)	7	7	1	0.0055	-1.3092	0.0653*
Azores (AZ)	9	9	1	0.0064	-1.5856	0.0705*
France (FRa)	6	6	1	0.0047	-1.3224	0.0887*
Greece (GRi)	6	6	1	0.0045	-1.4404	0.0884*
Ireland (IR)	10	10	1	0.0051	-1.5065	0.0509*
Liguria (ITI)	7	7	1	0.0035	-1.8732	0.0889*
Tyrrhenian (ITt)	8	8	1	0.0048	-1.3764	0.0596*
Madeira (MD)	10	10	1	0.0049	-0.7335	0.1149*
Morocco (MO)	9	9	1	0.0056	-1.5648	0.0479*
Portugal North (PTn)	8	8	1	0.0056	-1.2691	0.0635*
Portugal South (PTs)	9	9	1	0.0048	-1.5443	0.0584*
Scotland East (SCe)	8	8	1	0.0051	-1.2922	0.0607*
Scotland West (SCw)	8	8	1	0.0050	-1.2789	0.0640*
Alborán (SPa)	8	8	1	0.0040	-1.8445	0.0859*
Biscay (SPb)	8	8	1	0.0056	-1.6816	0.0447*
Cadiz (SPc)	3	3	1	0.0049	NA	0.1758*
Galicia (SPg)	7	7	1	0.0046	-1.3339	0.0666*
G1	38	38	1	0.0032	-2.6679*	0.0147*
G2	59	59	1	0.003	-2.6863*	0.0148*
G3	42	42	1	0.0056	-2.1076*	0.0356*

Table 3. Genetic diversity. The number of individuals (N), number of haplotypes (N hap.), haplotype diversity (hap. div.), nucleotide diversity (nuc. div.), Tajima's D (D), Watterson's theta (θ) and Ramos-Onsins & Rozas R_2 are presented for each sampling site and the three groups. The Netherlands (NE) sampling site was excluded as it only contains one sample. Significant values of D and R_2 (p-value ≤ 0.05) are marked with *

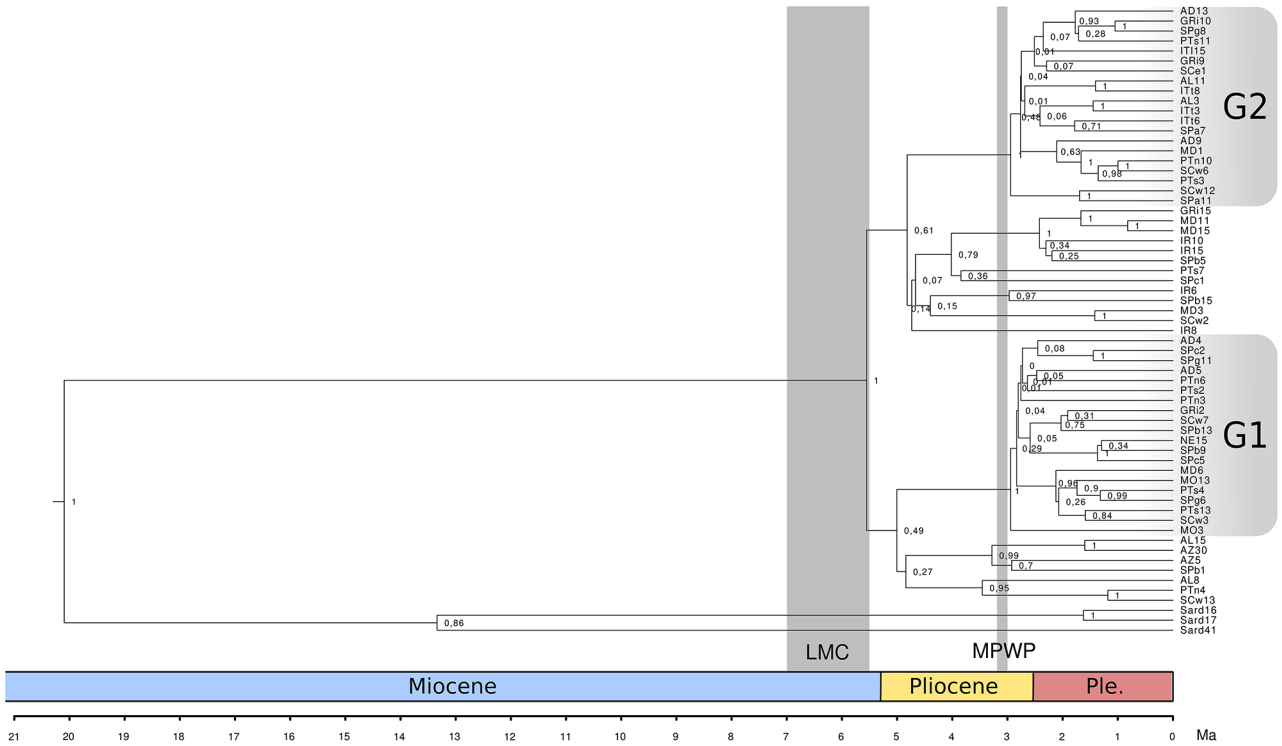

Fig. 4. Time-calibrated phylogenetic reconstruction of Sardina pilchardus haplotypes. Tree showing the phylogenetic reconstruction obtained with a time-calibrated analysis of 63 haplotypes in BEAST2. The inferred tree is scaled to geological time in units of million years (Ma). Node support values represent posterior probabilities. Ple. – Pleistocene.

Gene	Site (amino acid)	FUBAR	FUBAR	MEME	FEL	CodeML							
		Prob[alpha > beta]	Prob[alpha < beta]	p-value	p-value		ancestral codon	ancestral aa	derived codon	derived aa	N ind	ancestral aa polarity	derived aa polarity
ND2	218	0.026	0.935	-	-	yes	ACC	*Thr*	GCC	*Ala*	5	polar	non-polar
ND2	319	-	-	0.04	0.0884	yes	ACA	*Thr*	GCA	*Ala*	126	polar	non-polar
ND4	95	-	-	0.06	0.0980	no	ATA	*Met*	GCA, ACA	*Ala, Thr*	1, 137	non-polar	non-polar, polar
ND5	419	0.063	0.901	-	-	yes	ATT	*Ile*	GTT	*Val*	4	non-polar	non-polar
ND6	195	0.015	0.974	0.05	0.033	yes	CTG	*Leu*	CCG	*Pro*	14	non-polar	non-polar

Table 4. Positive selection. List of candidate genes and sites under positive selection, detected by at least two methods. Probabilities and p-values indicated by the Hyphy methods FUBAR, MEME and FEL are shown for each site. The "ancestral" and "derived" codons and the respective amino acid (aa), as well as the number of individuals with the derived aa (N ind.) are shown for each site. Amino acid abbreviation meaning: Ala – Alanine; Ile – Isoleucine; Leu – Leucine; Met – Methionine; Pro – Proline; Thr – Threonine; Val – Valine.

3). These results are supported by the mismatch distribution analyses on groups G1 and G2, which do suggest sudden population growth (Supplementary Fig. S1b, c), albeit less pronounced in G3 (Supplementary Fig. S1d). In contrast, the mismatch distribution analysis of the entire dataset (Supplementary Fig. S1a) shows a nearly bimodal profile, likely due to the presence of separate lineages in the data set. At the level of sampling locations, diversity estimates do not show significant differences between sites, which all have maximum haplotype diversity and comparable nucleotide diversity and values for Tajima's D and R_2. Values of D are negative but non-significant, whereas R_2 values are significant, but both statistics have higher values than those obtained for the three groups. Mitochondrial data thus seems to lack the power to discriminate between sampling sites. However, it is possible that, in future studies, more informative diversity measures at the level of sampling locations can be estimated from nuclear sequence data.

Estimation of divergence times

Running a calibrated phylogenetic analysis of European sardine mitogenomes proved challenging because long branches unite *Sardina*/*Sardinops* and *Alosa*/*Brevoortia*, resulting in poor convergence of the clock model and failure to recover the correct tree root when samples from all four genera were included. Nevertheless, a calibrated phylogeny was approximated using an outgroup of three *Sardinops* samples. The use of a gene-tree based approach for divergence time estimation, as adopted here, may result in overestimated node ages, as genetic divergence occurring before diversification events is not accounted for. Gene trees are contained within the species tree, so gene divergence times precede the true timing of speciation[94]. Furthermore, standard phylogenetic models may not be fully adequate for inferring within-species divergence times, as not all observed polymorphism are bound to become fixed in the population, although the presence of multiple sequences from the same species has been considered to be determinant for the correct inference of speciation times, as it allows for the distinction between within-species and between-species variation[95]. To test whether the calibrated tree topology is robust under alternative model assumptions, the PoMo species-tree method, which models the different contribution to substitution patterns of relative mutation rates and fixation biases[83,] was applied to the 12 gene dataset. The PoMo method allows for multiple samples from the same taxon and for sites to be polymorphic, and accounts for ancestral shared polymorphisms and incomplete lineage sorting[84], although the assumption of unlinked sites is violated with our data, which corresponds to a single linked locus, the mitogenome. The PoMo tree (Supplementary Fig. S3) recovers a topology compatible with that of the calibrated tree (Fig. 4), showing both G1 and G2 on long, supported branches, although the two trees are not directly comparable in terms of branch lengths. Nevertheless, a comparison of branch lengths for G1 and G2 between the PoMo and the ML analyses (Supplementary Fig. S2) shows a difference of c. 0.3 in both analyses, i.e. the branch length proportions in the two groups were not significantly altered under the PoMo analysis. These results demonstrate that the dataset is robust to alternative model assumptions, and that the effect of polymorphism in the sample does not have a significant impact on tree topology. Nevertheless, it was observed that, in calibration analyses of the present data set, both root placement and age estimates are highly sensitive to changes in outgroup composition and clock-rate priors, and the inferred ages should be cautiously considered as a mere approximation.

Lineage diversification patterns

The pattern seen on the MJN (Fig. 2) is suggestive of early radiation in the European sardine. This pattern is also observed in the phylogenetic trees, which do not recover supported relationships between the major mitochondrial lineages (Figs. 3, S2). A second, more recent period of rapid diversification is patent in Atlantic (G1) and Mediterranean (G2) groups (Fig. 4). . The calibrated phylogeny (Fig. 4) shows two distinguishable diversification events. The first event occurred c. 5.5 Ma, which corresponds to the lower Messinian, the last stage of the Miocene epoch[17] , and the second apparent event of rapid diversification occurred in the two larger clades c. 2.9 Ma, in the Pliocene. Age estimates on the calibrated phylogeny have a significant margin of error, but assuming that the estimated crown age of *S. pilchardus* is within realistic boundaries, and corresponds broadly to the Miocene/Pliocene transition, an explanation for the observed diversification pattern can be hypothesized. The two diversification events correspond to two documented climate transition periods, the Late Miocene Cooling (LMC)[17] and the mid-Pliocene warm period (mPWP)[18].

The first diversification of the European sardine mitogenomes, towards the end of the Miocene, could be related to the steep decrease of sea surface temperature that occurred in the LMC, between 7 and 5.5 Ma[17,96,97]. For example, the Benguela system of cold-water upwelling, which sustains an abundance of phytoplankton, is thought to have originated during this period[98]. The LMC is associated with the late Miocene biogenic boom (LMBB), a marine event which was defined based on an increase in biogenic deposits in the Pacific Ocean[99], and which may have promoted the diversification in pelagic fish such as *Sardina*. The LMBB is thought to have been caused by increased nutrient supply from land into the oceans or by a reorganization of oceanic circulation that promoted the redistribution of nutrients[100]. The onset of the LMBB is estimated at 8–6.7 Ma[99,101] and its end at 4.4–4.5 Ma[99,101,102]. In the North Atlantic, a maximum accumulation of deposits is estimated at 5 Ma[103]. Herbert et al.[104] report a change in sea surface temperature (SST) in the North Atlantic from above 25 °C to 12–21 °C during the late LMC. The temperature preference range of *S. pilchardus* lies between 13.5 and 15 °C, whereas tolerance ranges between 12 and 17°C[23]. Spawning occurs mainly at 14 to 15 °C, decreasing at water temperatures below 12 °C and above 16°C[24–26], and recruitment is also decreased outside this range[105]. The temperature shift at the end of the LMC may have led to a temperature optimum for the European sardine, while increasing both nutrient availability and levels of dissolved oxygen in the North Atlantic, thus promoting population growth and dispersal in this species that resulted in the observed rapid diversification pattern.

The presence of several distinct, supported groups with long stems on the calibrated tree suggests that the fixation of mitochondrial lineages followed this initial diversification burst. Fixation of distinct mitolineages may have been caused by geographical or ecological isolation between populations and decreased population sizes or even high extinction rates, possibly related to the mPWP, which occurred c. 3.2–3.0 Ma[18,106]. During this period, SST in the Northeastern Atlantic may have risen above the optimum for *S. pilchardus*. Middle to high-latitude SST in the Atlantic is estimated to have been up to 6 °C higher than present day[18,107] and surface water productivity was low, as the North Atlantic Current transported warm, nutrient-poor waters northwards[107] even into the Arctic Ocean[108]. Sea levels were also up to 22 m higher during this period[109,110]. The second diversification burst, observed in the two large clades, which accounts for most of present-day diversity, is likely to have resulted from the cooling phase that followed the mPWP, c. 3 Ma[107]. As a result of northern hemisphere glaciations during this phase[111], surface water productivity increased, and SST decreased as low as 12 °C c. 2.5 Ma[107]. This temperature shift may have led to another temperature optimum for the European sardine which likely promoted population growth, and the burst of diversification observed in Atlantic (G1) and Mediterranean (G2) clades, which accounts for most of its present-day mitogenome diversity.

During the lower Messinian, the Mediterranean experienced consecutive desiccation events, likely resulting from tectonic activity[17], leading to the so-called Messinian Salinity Crisis (MSC), which ended in the early Pliocene[112]. It is not necessary to invoke the MSC to explain the diversification pattern seen here in the European sardine mitogenomes. However, if an extreme scenario is speculated upon, fixation of mitochondrial lineages, after the first diversification event, could in theory have resulted from the persistence of isolated sardine populations in the Mediterranean during the MSC. This would imply that at least some areas of the Mediterranean maintained ecological conditions for the persistence of marine ichthyofauna. In a survey of Messinian fossils from the Mediterranean basin, Landini and Menesini[113] found evidence of the presence of several marine and freshwater fish during the Lower Messinian, including *S. pilchardus*, although the authors also consider that current fish communities in the Mediterranean derive essentially from faunal renewal from the Atlantic in the lower Pliocene. The presence of Messinian ichthyofauna has also been attributed to flooding events from the Atlantic before the final opening of the Mediterranean in the Pliocene[114,115]. *S. pilchardus* is an euryhaline species[116], therefore sardine populations may have tolerated salinity fluctuations during the Messinian. If European sardine populations persisted in isolation in the Mediterranean for a period of time, or if sporadic colonisation events from the Atlantic occurred throughout the Messinian, the fixation of mitochondrial lineages in the Mediterranean may have been possible, and these lineages could eventually have been retained after secondary contact between Mediterranean and Atlantic populations in the early Pliocene. However, high SST in the Mediterranean during the Messinian[104,117] was likely not ideal for the persistence of significant sardine populations.

The observed pattern of mitochondrial haplotype relationships indicates that there is extensive gene flow between the European sardine populations, even between individuals from opposite sides of its geographical distribution and does not indicate a structured population. Nevertheless, information from nuclear genomes may provide fine-scale insights into contemporary gene flow in the European sardine. Importantly, an extensive scan of sardine nuclear genomes may reveal the presence of environment-associated polymorphisms, which may in turn allow for modelling of populations resilience under future climate.

Positive and purifying selection in sardine mitogenomes

Analyses of 13 protein-coding genes of the European sardine mitogenome dataset detected between 12.9% and 21.1% of the total number of codon sites to be under purifying selection. Strong purifying selection, which eliminates deleterious mutations, has been reported in freshwater fish mitogenomes, particularly in populations from cool environments[118], and appears to be a major evolutionary force affecting mitogenome evolution in the European sardine.

Positive selection was detected, by at least two methods, in five candidate sites, but it is possible that relaxed purifying selection, instead of positive selection, explains at least part of these results. For example, sites ND2/218 and ND5/419 have codon changes in a few unrelated individuals, which may indicate that these are simply instances of fixation, by drift, of possibly neutral or mildly deleterious mutations that have not yet been eliminated by purifying selection[21]. In sites ND4/95 and ND2/319, the "derived" amino acid state is apomorphic to the European sardine, and changes observed represent reversals to the "ancestral" amino acid, except for one lineage, which has a secondary "derived" amino acid, in site ND4/95. These reversals seem to affect mostly

unrelated individuals and may correspond to instances of relaxed purifying selection, except in the case of eight individuals with the reversal in site ND2/319, for which it is less clear whether true positive selection is present, as they form an unsupported clade. However, in the case of site ND6/195, positive selection may be acting over at least one lineage. Site ND6/195 is located in an insertion that is present in the European sardine but not in its closely related species. Individuals displaying the "derived" amino acid (proline) are all part of G3 and are distributed over few lineages. One of these lineages is fully supported and only two out of 10 individuals do not display the "derived" character. Site ND6/195 was detected to be under positive selection by site and branch-site models in codeml, whereas branch models indicate significant variation in dN/dS for ND6 in G2 and G3 when each of these was considered as the foreground branches. It may not be possible to ascertain, at this point, whether the evolutionary force driving the fixation of this character state in these groups is positive selection or, instead, relaxed purifying selection combined with a demographic effect such as a bottleneck, which allowed for the persistence of the mutation in a diversified clade. Nevertheless, if present, positive selection in sardine mitogenomes is likely episodic and not a major driver of diversification in the mitogenome of this species.

Conclusions

This study, which presents the most extensive mitochondrial genome sampling of *S. pilchardus* so far, with 139 samples across 19 locations spanning from the North Sea to the Moroccan Atlantic coast and from the Azores to the Ionian Sea, provides evidence supporting the lack of geographic structure and extensive gene flow.

The extensive geographical and mitogenomic data provided by this study allowed us to contextualize its findings within the broader scope of historical climate change, especially given the evidence that past global warming events, similar to those predicted for the near future, have had significant impacts in marine life. Climate conditions in the mPWP represent the most recent event of global warming comparable to what is predicted for the second half of the twenty-first century[115]. The global warmth of 2–3 °C above current temperatures estimated for the mPWP lies within projections for global conditions towards the end of the twenty-first century[119]. If indeed the apparent diversity loss seen in the European sardine mitogenomes in the period up to 3 Ma was due to increased global temperatures in the mPWP causing population contraction or extinction, then global conditions expected towards the end of the century may also have a negative impact on the European sardine genetic diversity and stocks, which may warrant new or additional conservation and management actions for this species.

Data availability

Aligned sequences used in data analyses are available on Zenodo (https://doi.org/10.5281/zenodo.13247537).

Received: 3 May 2024; Accepted: 2 December 2024
Published online: 28 December 2024

References

1. Conover, D. O., Clarke, L. M., Munch, S. B. & Wagner, G. N. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *J. Fish Biol.* **69**, 21–47. https://doi.org/10.1111/j.1095-8649.2006.01274.x (2006).
2. Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F. & Bekkevold, D. Population genomics of marine fishes: Identifying adaptive variation in space and time. *Mol. Ecol.* **18**, 3128–3150 (2009).
3. Czech, L. & Exposito-Alonso, M. grenepipe: A flexible, scalable and reproducible pipeline to automate variant calling from sequence reads. *Bioinformatics* **38**, 4809–4811. https://doi.org/10.1093/bioinformatics/btac600 (2022).
4. da Fonseca, R. R. et al. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar. Genomics* **30**, 3–13. https://doi.org/10.1016/j.margen.2016.04.012 (2016).
5. Lou, R. N., Jacobs, A., Wilder, A. & Therkildsen, N. O. A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol. Ecol.* **30**, 5966–5993. https://doi.org/10.1111/mec.16077 (2021).
6. Andersson, L. *et al.* How fish population genomics can promote sustainable fisheries: A road map. *Annu. Rev. Anim. Biosci.* https://doi.org/10.1146/annurev-animal-021122-102933 (2024).
7. Baltazar-Soares, M., Lima, A. R. A., Silva, G. & Gaget, E. Towards a unified eco-evolutionary framework for fisheries management: Coupling advances in next-generation sequencing with species distribution modelling. *Front. Mar. Sci.* https://doi.org/10.3389/fmars.2022.1014361 (2023).
8. Zhang, B.-D., Xue, D.-X., Li, Y.-L. & Liu, J.-X. RAD genotyping reveals fine-scale population structure and provides evidence for adaptive divergence in a commercially important fish from the northwestern Pacific Ocean. *PeerJ* **7**, e7242. https://doi.org/10.7717/peerj.7242 (2019).
9. Nielsen, E. E. et al. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Ecol. Evol.* **9**, 276. https://doi.org/10.1186/1471-2148-9-276 (2009).
10. Barth, J. M. I. et al. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Mol. Ecol.* **26**, 4452–4466. https://doi.org/10.1111/mec.14207 (2017).
11. Clucas, G. V., Lou, R. N., Therkildsen, N. O. & Kovach, A. I. Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolut. Appl.* **12**, 1971–1987. https://doi.org/10.1111/eva.12861 (2019).
12. Hotaling, S., Desvignes, T., Sproul, J. S., Lins, L. S. F. & Kelley, J. L. Pathways to polar adaptation in fishes revealed by long-read sequencing. *Mol. Ecol.* **32**, 1381–1397. https://doi.org/10.1111/mec.16501 (2023).
13. Kess, T. et al. A putative structural variant and environmental variation associated with genomic divergence across the Northwest Atlantic in Atlantic Halibut. *ICES J. Mar. Sci.* **78**, 2371–2384. https://doi.org/10.1093/icesjms/fsab061 (2021).
14. Knutsen, H. et al. Combining population genomics with demographic analyses highlights habitat patchiness and larval dispersal as determinants of connectivity in coastal fish species. *Mol. Ecol.* **31**, 2562–2577. https://doi.org/10.1111/mec.16415 (2022).
15. Layton, K. K. S. et al. Genomic evidence of past and future climate-linked loss in a migratory Arctic fish. *Nat. Clim. Change* **11**, 158–165. https://doi.org/10.1038/s41558-020-00959-7 (2021).
16. Pujolar, J. M., Jacobsen, M. W. & Bertolini, F. Comparative genomics and signatures of selection in North Atlantic eels. *Mar. Genomics* **62**, 100933. https://doi.org/10.1016/j.margen.2022.100933 (2022).
17. Hodell, D. A., Curtis, J. H., Sierro, F. J. & Raymo, M. E. Correlation of Late Miocene to Early Pliocene sequences between the Mediterranean and North Atlantic. *Paleoceanogr. Paleoclimatol.* **16**, 164–178. https://doi.org/10.1029/1999pa000487 (2001).

18. Dowsett, H. J., Chandler, M. A., Cronin, T. M. & Dwyer, G. S. Middle Pliocene sea surface temperature variability. *Paleoceanography.* https://doi.org/10.1029/2005pa001133 (2005).
19. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
20. Modesto, I. S. et al. Identifying signatures of natural selection in cork oak (*Quercus suber* L.) genes through SNP analysis. *Tree Genet. Genomes* **10**, 1645–1660. https://doi.org/10.1007/s11295-014-0786-1 (2014).
21. Hughes, A. L. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* **99**, 364–373. https://doi.org/10.1038/sj.hdy.6801031 (2007).
22. Parrish, R. H., Serra, R. & Grant, W. S. The monotypic sardines, *Sardina* and *Sardinops*: Their taxonomy, distribution, stock structure, and zoogeography. *Can. J. Fish. Aquat. Sci.* **46**, 2019–2036. https://doi.org/10.1139/f89-251 (1989).
23. Bernal, M. et al. Sardine spawning off the European Atlantic coast: Characterization of and spatio-temporal variability in spawning habitat. *Progress in Oceanography* **74**, 210–227. https://doi.org/10.1016/j.pocean.2007.04.018 (2007).
24. Garrido, S. et al. Effect of temperature on the growth, survival, development and foraging behaviour of *Sardina pilchardus* larvae. *Mar. Ecol. Prog. Ser.* **559**, 131–145 (2016).
25. Peck, M. A., Reglero, P., Takahashi, M. & Catalán, I. A. Life cycle ecophysiology of small pelagic fish and climate-driven changes in populations. *Prog. Oceanogr.* **116**, 220–245. https://doi.org/10.1016/j.pocean.2013.05.012 (2013).
26. Stratoudakis, Y. et al. Sardine (*Sardina pilchardus*) spawning seasonality in European waters of the northeast Atlantic. *Mar. Biol.* **152**, 201–212. https://doi.org/10.1007/s00227-007-0674-4 (2007).
27. Albo-Puigserver, M., Navarro, J., Coll, M., Layman, C. A. & Palomera, I. Trophic structure of pelagic species in the northwestern Mediterranean Sea. *J. Sea Res.* **117**, 27–35. https://doi.org/10.1016/j.seares.2016.09.003 (2016).
28. Garrido, S. et al. Diet and feeding intensity of sardine *Sardina pilchardus*: correlation with satellite-derived chlorophyll data. *Mar. Ecol. Prog. Ser.* **354**, 245–256. https://doi.org/10.3354/meps07201 (2008).
29. 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.* https://doi.org/10.1007/s11160-022-09704-z (2022).
30. Monteiro, P. V. The purse seine fishing of sardine in portuguese waters: A difficult compromise between fish stock sustainability and fishing effort. *Rev. Fish. Sci. Aquac.* **25**, 218–229. https://doi.org/10.1080/23308249.2016.1269720 (2017).
31. INE. *Statistical data for fisheries 1969–2022* (2024).
32. Braga, H. O., Azeiteiro, U. M., Oliveira, H. M. F. & Pardal, M. A. Evaluating fishermen's conservation attitudes and local ecological knowledge of the European sardine (*Sardina pilchardus*), Peniche, Portugal. *J. Ethnobiol. Ethnomed.* **13**, 25. https://doi.org/10.1186/s13002-017-0154-y (2017).
33. Ganias, K. *Biology and Ecology of Sardines and Anchovies*. 394 (CRC Press, 2014).
34. ICES. Working group on southern horse mackerel, anchovy and sardine (WGHANSA). *ICES Sci. Rep.* **5**, 578. https://doi.org/10.17895/ices.pub.23507922.v1 (2023).
35. Costalago, D. & Palomera, I. Feeding of European pilchard (*Sardina pilchardus*) in the northwestern Mediterranean: From late larvae to adults. *Sci. Mar.* **78**, 41–54. https://doi.org/10.3989/scimar.03898.06D (2014).
36. Dimarchopoulou, D. & Tsikliras, A. C. Linking growth patterns to sea temperature and oxygen levels across European sardine (*Sardina pilchardus*) populations. *Environ. Biol. Fishes* https://doi.org/10.1007/s10641-022-01229-5 (2022).
37. Silva, A. et al. Temporal and geographic variability of sardine maturity at length in the northeastern Atlantic and the western Mediterranean. *ICES J. Mar. Sci.* **63**, 663–676. https://doi.org/10.1016/j.icesjms.2006.01.005 (2006).
38. Stratoudakis, Y., Bernal, M., Borchers, D. L. & Borges, M. F. Changes in the distribution of sardine eggs and larvae off Portugal, 1985–2000. *Fish. Oceanogr.* **12**, 49–60. https://doi.org/10.1046/j.1365-2419.2003.00222.x (2003).
39. Jemaa, S. et al. What can otolith shape analysis tell us about population structure of the European sardine, *Sardina pilchardus*, from Atlantic and Mediterranean waters?. *J. Sea Res.* **96**, 11–17. https://doi.org/10.1016/j.seares.2014.11.002 (2015).
40. Neves, J. et al. Population structure of the European sardine *Sardina pilchardus* from Atlantic and Mediterranean waters based on otolith shape analysis. *Fish. Res.* **243**, 106050. https://doi.org/10.1016/j.fishres.2021.106050 (2021).
41. Silva, A. Morphometric variation among sardine (*Sardina pilchardus*) populations from the northeastern Atlantic and the western Mediterranean. *ICES J. Mar. Sci.* **60**, 1352–1360. https://doi.org/10.1016/S1054e3139(03)00141-3 (2003).
42. Chlaida, M., Kifani, S., Lenfant, P. & Ouragh, L. First approach for the identification of sardine populations *Sardina pilchardus* (Walbaum 1792) in the Moroccan Atlantic by allozymes. *Mar. Biol.* **149**, 169–175. https://doi.org/10.1007/s00227-005-0185-0 (2006).
43. Laurent, V., Caneco, B., Magoulas, A. & Planes, S. Isolation by distance and selection effects on genetic structure of sardines *Sardina pilchardus* Walbaum. *J. Fish Biol.* **71**, 1–17. https://doi.org/10.1111/j.1095-8649.2007.01450.x (2007).
44. Ramon, M. M. & Castro, J. A. Genetic variation in natural stocks of *Sardina pilchardus* (sardines) from the western Mediterranean Sea. *Heredity* **78**, 520–528. https://doi.org/10.1038/hdy.1997.81 (1997).
45. Louro, B. et al. A haplotype-resolved draft genome of the European sardine (*Sardina pilchardus*). *GigaScience*. https://doi.org/10.1093/gigascience/giz059 (2019).
46. Machado, A. et al. "Out of the can": A draft genome assembly, liver transcriptome, and nutrigenomics of the European Sardine Sardina pilchardus. *Genes* **9**, 485. https://doi.org/10.3390/genes9100485 (2018).
47. Ballard, J. W. & Whitlock, M. C. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744. https://doi.org/10.1046/j.1365-294x.2003.02063.x (2004).
48. Blair, C. Organellar DNA continues to provide a rich source of information in the genomics era. *Mol. Ecol.* **32**, 2144–2150. https://doi.org/10.1111/mec.16872 (2023).
49. Galtier, N., Nabholz, B., Glemin, S. & Hurst, G. D. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* **18**, 4541–4550. https://doi.org/10.1111/j.1365-294X.2009.04380.x (2009).
50. Towarnicki, S. G. & Ballard, J. W. O. Towards understanding the evolutionary dynamics of mtDNA. *Mitochon. DNA Part A DNA Map. Sequen. Anal.* **31**, 355–364. https://doi.org/10.1080/24701394.2020.1830076 (2020).
51. Sebastian, W., Sukumaran, S. & Gopalakrishnan, A. Comparative mitogenomics of Clupeoid fish provides insights into the adaptive evolution of mitochondrial oxidative phosphorylation (OXPHOS) genes and codon usage in the heterogeneous habitats. *Heredity* **128**, 236–249. https://doi.org/10.1038/s41437-022-00519-z (2022).
52. Baltazar-Soares, M., de Araújo Lima, A. R. & Silva, G. Targeted sequencing of mitochondrial genes reveals signatures of molecular adaptation in a nearly panmictic small pelagic fish species. *Genes* **12**, 91. https://doi.org/10.3390/genes12010091 (2021).
53. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (2010).
54. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354 (2016).
55. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890. https://doi.org/10.1093/bioinformatics/bty560 (2018).
56. Jin, J. J. et al. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241. https://doi.org/10.1186/s13059-020-02154-5 (2020).
57. Iwasaki, W. et al. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.* **30**, 2531–2540. https://doi.org/10.1093/molbev/mst141 (2013).

58. Sato, Y., Miya, M., Fukunaga, T., Sado, T. & Iwasaki, W. MitoFish and MiFish pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Mol. Biol. Evol.* **35**, 1553–1555. https://doi.org/10.1093/molbev/msy074 (2018).
59. Zhu, T., Sato, Y., Sado, T., Miya, M. & Iwasaki, W. MitoFish, MitoAnnotator, and MiFish pipeline: Updates in 10 years. *Mol. Biol. Evol.* https://doi.org/10.1093/molbev/msad035 (2023).
60. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. https://doi.org/10.1093/molbev/mst010 (2013).
61. Nylander, J. A. A. catfasta2phyml.pl - Concatenate FASTA alignments to PHYML, PHYLIP, or FASTA format. Retrieved from https://github.com/nylander/catfasta2phyml. (2022).
62. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x (2010).
63. Leigh, J. W., Bryant, D. & Nakagawa, S. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116. https://doi.org/10.1111/2041-210x.12410 (2015).
64. Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034 (2002).
65. Paradis, E. pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420. https://doi.org/10.1093/bioinformatics/btp696 (2010).
66. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129 (2008).
67. Jombart, T. & Ahmed, I. adegenet 1.3–1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521 (2011).
68. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528. https://doi.org/10.1093/bioinformatics/bty633 (2019).
69. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2022).
70. RStudio: Integrated Development Environment for R (RStudio, Inc., Boston, MA, 2016).
71. Goudet, J. & Jombart, T. hierfstat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.5–11, https://CRAN.R-project.org/package=hierfstat. (2022).
72. Wickham, H. ggplot2. *WIREs Comput. Stat.* **3**, 180–185. https://doi.org/10.1002/wics.147 (2011).
73. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. Partitionfinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701. https://doi.org/10.1093/molbev/mss020 (2012).
74. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773. https://doi.org/10.1093/molbev/msw260 (2017).
75. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. https://doi.org/10.1093/sysbio/syq010 (2010).
76. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. https://doi.org/10.1093/molbev/msu300 (2015).
77. Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650. https://doi.org/10.1371/journal.pcbi.1006650 (2019).
78. Allio, R., Donega, S., Galtier, N. & Nabholz, B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772. https://doi.org/10.1093/molbev/msx197 (2017).
79. Bagley, J. C. et al. Assessing species boundaries using multilocus species delimitation in a morphologically conserved group of neotropical freshwater fishes, the *Poecilia sphenops* species complex (Poeciliidae). *PLoS ONE* **10**, e0121139. https://doi.org/10.1371/journal.pone.0121139 (2015).
80. Watanabe, K., Sakai, H., Sanada, T. & Nishida, M. Comparative phylogeography of diadromous and freshwater daces of the genus *Tribolodon* (Cyprinidae). *Ichthyol. Res.* **65**, 383–397. https://doi.org/10.1007/s10228-018-0624-9 (2018).
81. Bloom, D. D. & Lovejoy, N. R. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proc. R. Soc. B Biol. Sci.* **281**, 20132081. https://doi.org/10.1098/rspb.2013.2081 (2014).
82. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010, New Orleans, LA, pp 1–8. (2010).
83. De Maio, N., Schlotterer, C. & Kosiol, C. Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Mol. Biol. Evol.* **30**, 2249–2262. https://doi.org/10.1093/molbev/mst131 (2013).
84. De Maio, N., Schrempf, D. & Kosiol, C. PoMo: An allele frequency-based approach for species tree estimation. *Syst. Biol.* **64**, 1018–1031. https://doi.org/10.1093/sysbio/syv048 (2015).
85. Murrell, B. et al. FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. *Mol. Biol. Evol.* **30**, 1196–1205. https://doi.org/10.1093/molbev/mst030 (2013).
86. Murrell, B. et al. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764. https://doi.org/10.1371/journal.pgen.1002764 (2012).
87. Kosakovsky Pond, S. L. & Frost, S. D. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222. https://doi.org/10.1093/molbev/msi105 (2005).
88. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679. https://doi.org/10.1093/bioinformatics/bti079 (2005).
89. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. https://doi.org/10.1093/molbev/msm088 (2007).
90. Shank, S. D., Weaver, S. & Kosakovsky Pond, S. L. phylotree.js - a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinf.* **19**, 276. https://doi.org/10.1186/s12859-018-2283-2 (2018).
91. Liu, L. et al. The complete mitochondrial genome of the *Plectorhinchus cinctus* (Teleostei, Haemulidae). *Mitochon. DNA Part A DNA Map. Sequenc. Anal.* **27**, 842–843. https://doi.org/10.3109/19401736.2014.919468 (2016).
92. Basilone, G. et al. Reproduction and sexual maturity of European Sardine (*Sardina pilchardus*) in the Central Mediterranean Sea. *Front. Mar. Sci.* https://doi.org/10.3389/fmars.2021.715846 (2021).
93. da Fonseca, R. R. et al. Population genomics reveals the underlying structure of the small pelagic European Sardine and suggests low connectivity within macaronesia. *Genes.* https://doi.org/10.3390/genes15020170 (2024).
94. McCormack, J. E., Heled, J., Delaney, K. S., Peterson, A. T. & Knowles, L. L. Calibrating divergence times on species trees versus gene trees: implications for speciation history of Aphelocoma jays. *Evolution* **65**, 184–202. https://doi.org/10.1111/j.1558-5646.2010.01097.x (2011).
95. Heled, J. & Drummond, A. J. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**, 570–580. https://doi.org/10.1093/molbev/msp274 (2010).
96. Steinthorsdottir, M. et al. The Miocene: The future of the past. *Paleoceanogr. Paleoclimatol.* https://doi.org/10.1029/2020pa004037 (2021).
97. Super, J. R. et al. Miocene evolution of North Atlantic sea surface temperature. *Paleoceanogr. Paleoclimatol.* https://doi.org/10.1029/2019pa003748 (2020).

98. Siesser, W. G. Late Miocene Origin of the Benguela Upswelling System off Northern Namibia. *Science* **208**, 283–285. https://doi.org/10.1126/science.208.4441.283 (1980).
99. Farrell, J. W. *et al.* Late Neogene Sedimentation Patterns in the Eastern Equatorial Pacific. *Proc. Ocean Drill. Prog.. Sci. Results* **138**, 717–756. https://doi.org/10.2973/odp.proc.sr.138.143.1995 (1995).
100. Pillot, Q., Suchéras-Marx, B., Sarr, A. C., Bolton, C. T. & Donnadieu, Y. A global reassessment of the spatial and temporal expression of the late miocene biogenic bloom. *Paleoceanogr. Paleoclimatol.* **38**, e2022PA004564. https://doi.org/10.1029/2022pa004564 (2023).
101. Lyle, M. & Baldauf, J. Biogenic sediment regimes in the Neogene equatorial Pacific, IODP Site U1338: Burial, production, and diatom community. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **433**, 106–128. https://doi.org/10.1016/j.palaeo.2015.04.001 (2015).
102. Lyle, M., Drury, A. J., Tian, J., Wilkens, R. & Westerhold, T. Late Miocene to Holocene high-resolution eastern equatorial Pacific carbonate records: stratigraphy linked by dissolution and paleoproductivity. *Clim. Past* **15**, 1715–1739. https://doi.org/10.5194/cp-15-1715-2019 (2019).
103. Diester-Haass, L., Billups, K. & Emeis, K. C. In search of the late Miocene–early Pliocene "biogenic bloom" in the Atlantic Ocean (Ocean Drilling Program Sites 982, 925, and 1088). *Paleoceanography* https://doi.org/10.1029/2005pa001139 (2005).
104. Herbert, T. D. et al. Late Miocene global cooling and the rise of modern ecosystems. *Nat. Geosci.* **9**, 843–847. https://doi.org/10.1038/ngeo2813 (2016).
105. Garrido, S. et al. Temperature and food-mediated variability of European Atlantic sardine recruitment. *Prog. Oceanogr.* **159**, 267–275. https://doi.org/10.1016/j.pocean.2017.10.006 (2017).
106. Haywood, A. M., Dowsett, H. J. & Dolan, A. M. Integrating geological archives and climate models for the mid-Pliocene warm period. *Nat. Commun.* **7**, 10646. https://doi.org/10.1038/ncomms10646 (2016).
107. Naafs, B. D. A. et al. Late Pliocene changes in the North Atlantic Current. *Earth Planet. Sci. Lett.* **298**, 434–442. https://doi.org/10.1016/j.epsl.2010.08.023 (2010).
108. Rahaman, W. et al. Reduced Arctic sea ice extent during the mid-Pliocene Warm Period concurrent with increased Atlantic-climate regime. *Earth Planet. Sci. Lett.* https://doi.org/10.1016/j.epsl.2020.116535 (2020).
109. Dowsett, H. J. & Cronin, T. M. High eustatic sea level during the middle Pliocene: Evidence from the southeastern US Atlantic Coastal Plain. *Geology* **18**, 435–438 (1990).
110. Miller, K. G. et al. High tide of the warm Pliocene: Implications of global sea level for Antarctic deglaciation. *Geology* **40**, 407–410. https://doi.org/10.1130/g32869.1 (2012).
111. Bartoli, G. et al. Final closure of Panama and the onset of northern hemisphere glaciation. *Earth Planet. Sci. Lett.* **237**, 33–44. https://doi.org/10.1016/j.epsl.2005.06.020 (2005).
112. Roveri, M. et al. The Messinian Salinity Crisis: Past and future of a great challenge for marine sciences. *Mar. Geol.* **352**, 25–58. https://doi.org/10.1016/j.margeo.2014.02.002 (2014).
113. Landini, W. & Menesini, E. Messinian marine fish communities of the Mediterranean Sea. *Atti della Societa Toscana di Scienze Naturali A* **91**, 279–290 (1984).
114. Carnevale, G., Longinelli, A., Caputo, D., Barbieri, M. & Landini, W. Did the Mediterranean marine reflooding precede the Mio-Pliocene boundary? Paleontological and geochemical evidence from upper Messinian sequences of Tuscany, Italy. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **257**, 81–105. https://doi.org/10.1016/j.palaeo.2007.09.005 (2008).
115. Carnevale, G., Landini, W. & Sarti, G. Mare versus Lago-mare: marine fishes and the Mediterranean environment at the end of the Messinian Salinity Crisis. *J. Geol. Soc.* **163**, 75–80. https://doi.org/10.1144/0016-764904-158 (2006).
116. Rodríguez-Climent, S. et al. Essential habitat for sardine juveniles in Iberian waters. *Sci. Mar.* **81**, 351. https://doi.org/10.3989/scimar.04554.07A (2017).
117. Beltran, C., Sicre, M.-A., Ohneiser, C. & Sainz, M. A composite Pliocene record of sea surface temperature in the central Mediterranean (Capo Rossello composite section – South Sicily). *Sediment. Geol.* https://doi.org/10.1016/j.sedgeo.2021.105921 (2021).
118. Pavlova, A. et al. Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. *Heredity* **118**, 466–476. https://doi.org/10.1038/hdy.2016.120 (2017).
119. Dowsett, H. J., Chandler, M. A. & Robinson, M. M. Surface temperatures of the Mid-Pliocene North Atlantic Ocean: implications for future climate. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **367**, 69–84. https://doi.org/10.1098/rsta.2008.0213 (2009).

Acknowledgements

This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT, Portugal) by the research contract CEECIND/01528/2017 (https://doi.org/10.54499/CEECIND/01528/2017/CP1387/CT0040) attributed to Ana Rita Vieira and the PhD grant 2021.08758.BD attributed to Mariana Bray Viegas. MARE (UIDB/04292/2020, https://doi.org/10.54499/UIDB/04292/2020; and UIDP/04292/2020, https://doi.org/10.54499/UIDP/04292/2020), ARNET (LA/P/0069/2020, https://doi.org/10.54499/LA/P/0069/2020), cE3c (UIDB/00329/2020, https://doi.org/10.54499/UIDB/00329/2020; and UIDP/00329/2020, https://doi.org/10.54499/UIDP/00329/2020) and CHANGE (LA/P/0121/2020, https://doi.org/10.54499/LA/P/0121/2020) are also supported by FCT through strategic projects. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The authors would like to thank colleagues who helped in the collection of the European sardine muscle samples from the Northeast Atlantic (Maria Manuel Angélico from IPMA, Celso Fariña from IEO A Coruña, Henrique Cabral from INRAE, Ciaran O'Donnell from Irish Marine Institute, Finlay Burns from Marine Scotland Science, Luís Almeida from Wageningen University & Research, Joana Vasconcelos from MARE-Madeira, Gonçalo Graça from Flying Sharks, and Hammou El Habouz from Agadir INRH) and the Mediterranean Sea (Leila Bouayad from Algeria HASAQ Laboratory and Dimitrios K. Moutopoulos from University of Patras).

Author contributions

ARV conceptualised the idea, acquired funding and resources, collected samples, optimized wet lab protocols, analysed data, prepared figures and tables, and wrote the main manuscript text. FS conceptualised the idea, analysed data, prepared figures and tables, and wrote the main manuscript text. JB analysed data and prepared figures and tables. MVB ran the wet lab protocols. RS was part of the sequencing process. LSG collected samples and acquired funding and resources. OSP conceptualised the idea, analysed data, and wrote the main manuscript text. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Ethics statement
Sampling for the present study focused on the European sardine from almost the entire species distribution range. Ethical or government approval, specific permissions or licenses for sample collection were not required for this study, as all specimens were collected as part of routine fishing procedures by fishermen of commercial fleets or scientific research campaigns. Fish are killed during the hauling of fishing gears due to differences in atmospheric pressure and fishing procedures. The experiments were conducted on dead animals, hence, no animal welfare or animal use permits were required for this study. *Sardina pilchardus* is not an endangered or protected species.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-82054-x.

Correspondence and requests for materials should be addressed to A.R.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.