The genomic basis and environmental correlates of local adaptation in the Atlantic horse mackerel (*Trachurus trachurus*)

Angela P. Fuentes-Pardo[1], Edward D. Farrell[2,3], Mats E. Pettersson[1], C. Grace Sprehn[1], Leif Andersson[1,4,5]

[1] Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden.

[2] EDF Scientific, Rathaha, Ladysbridge, Cork, Ireland.

[3] Killybegs Fishermen's Organisation, Bruach na Mara, St. Catherine's Rd, Killybegs, Co. Donegal, Ireland.

[4] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

[5] Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, USA.

Corresponding authors: Angela P. Fuentes-Pardo, e-mail: apfpgen@gmail.com, Leif Andersson, e-mail: leif.andersson@imbim.uu.se.

Abstract

Understanding how populations adapt to local environments is increasingly important to prevent biodiversity loss due to climate change. Here we examined whole-genome variation of twelve Atlantic horse mackerel samples from the North Sea to North Africa, and the western Mediterranean Sea. This marine migratory benthopelagic fish is one of the most widely distributed and commercially important species in the eastern Atlantic. We found low population structure at neutral loci, but high differentiation at adaptive loci distinguishing the western Mediterranean and the North Sea populations from other Atlantic locations. Candidate genes distinctive of the Mediterranean include a green-sensitive-opsin harbouring two missense mutations that might fine-tune the spectral sensitivity to blue-green light conditions. Candidate genes characteristic of the North Sea could play a critical role in cold tolerance (energy metabolism and cell membrane structure) and increased sensitivity to odours, presumably to compensate reduced visibility in turbid waters. We also discovered a

putative chromosomal inversion (9.9 Mb) that follows a climate-related latitudinal cline with a break near mid Portugal. Genome-environment association analysis indicated that seawater-dissolved oxygen concentration and temperature are likely the main environmental drivers of local adaptation. Our genomic data broadly supports the current stock divisions, but recommends revision of the western and southern stock boundaries. We developed a reduced SNP panel that genetically discriminate the North Sea and North Africa from neighbouring populations. Our study highlights the importance of life history and chromosomal inversions in adaptation with gene flow, and the complexity of evolutionary and ecological processes involved in local adaptation.

1. Introduction

The extent to which marine species with no evident barriers restricting gene flow show genetic differentiation and local adaptation across geographic areas is a question of considerable interest in evolutionary biology, conservation, and management (Palumbi, 1994). Several marine species exhibit large population sizes, high gene flow, and minute genetic drift, resulting in low genetic differentiation that has been difficult to resolve with neutral genetic markers (Hauser & Carvalho, 2008a).

Owing to advances in high-throughput sequencing, recent genomic studies screening thousands to millions of genetic markers across the genome have revealed population structure and selection signatures in species previously assumed to be panmictic [e.g., Atlantic herring (Han et al., 2020)] or lowly structured [e.g., Atlantic cod (Barth et al., 2017), Atlantic halibut (Kess et al., 2021)]. Population structure in marine fish has been characterized by shifts in allele frequencies at many small effect loci or fewer large effect loci (Gagnaire & Gaggiotti, 2016) and in chromosomal rearrangements (Akopyan et al., 2022; Han et al., 2020; Matschiner et al., 2022). Moreover, genomic divergence has been linked to ecological diversity, for example, in migratory behavior (Kirubakaran et al., 2016), seasonal reproduction (Lamichhaney et al., 2017; Petrou et al., 2021), or along environmental gradients (Han et al., 2020; Stanley et al., 2018). Therefore, a thorough examination of genomic variation, including neutral and adaptive loci, can help identify distinct biological units and genetic variants associated with adaptation to local environments, knowledge of great interest in conservation and management, especially in the face of climate change.

Fish stock identification is an important prerequisite for fisheries management (Cadrin & Secor, 2009), however many exploited stocks have traditionally been defined according to geographical and political features or regions rather than on a biological basis. As more information becomes available, it is evident that the temporal and spatial distributions of most fisheries resources are not aligned to these artificial divisions (Kerr et al., 2017) and that biological populations are more dynamic and complex (Reiss, Hoarau, Dickey-Collas, & Wolff, 2009; Stephenson, 2002). Therefore, it is critical to identify the underlying population structure and use this information to identify the appropriate level at which to define assessment and management units. It is also important to be able to assign individuals in mixed surveys and commercial catches to the population or assessment unit to which they belong (Casey, Jardim, & Martinsohn, 2016; Hintzen et al., 2015).

The Atlantic horse mackerel (*Trachurus trachurus* Linnaeus, 1758) is a marine benthopelagic shoaling species widely distributed in the east Atlantic, from Norway to southern Africa (FAO Major Fishing areas 27, 34 and 47) and in the Mediterranean Sea (FAO Major Fishing area 37) (Froese & Pauly, 2021). It inhabits a variety of environments and different populations may be under contrasting selection pressures, making this species ideal for the study of local adaptation. Horse mackerel are generally found in continental shelf waters (100-200 m depth), but have also been observed in deeper (~1000 m) or near-shore waters. The species undertakes annual migrations between spawning, feeding and over-wintering areas (Abaunza et al., 2003). Horse mackerel are considered to be asynchronous batch spawners with indeterminate fecundity and are not known to be faithful to their original spawning grounds (Gordo et al., 2008; Ndjaula, Hansen, Krüger-Johnsen, & Kjesbu, 2009). Eggs and larvae are pelagic and are typically either found over the continental shelf, from the surface to 100 m depth, or near the coast (Alvarez & Chifflet, 2012; Beveren, Klein, Serrão, Gonçalves, & Borges, 2016).

In the northeast Atlantic, horse mackerel are assessed and managed as three separate stocks: the Western, the North Sea, and the Southern (Figure S1), which were largely defined based on the results of the HOMSIR project (Abaunza et al., 2008). Populations inhabiting coastal waters along Morocco and Mauritania, in northwest Africa, are considered a separate group, denominated the “Saharo-Mauritanian stock”. However, these populations are less studied and monitored than those in the north. The discreteness of the stocks, as well as the location and levels of mixing between them, is unknown, which leads to uncertainty in the

input data for stock assessments. Previous genetic studies have indicated low genetic differentiation between populations, and have provided inconclusive results in regards to population sub-structuring beyond the three main stock divisions identified using traditional methods (Brunel et al., 2016; Cimmaruta, Bondanelli, Ruggi, & Nascetti, 2008; Comesaña, Martínez-Areal, & Sanjuan, 2008; Farrell & Carlsson, 2018; Healey et al., 2020; Kasapidis & Magoulas, 2008; Mariani, 2012; Sala Bozano et al., 2015).

Given the elusive nature of the population structure of the Atlantic horse mackerel and its ecological and commercial importance in the east Atlantic, the aims of this study were four-fold: (i) Estimate the extent of genetic differentiation between populations based on whole-genome sequencing; (ii) Identify the genetic basis of local adaptation; (iii) Identify which ecological and evolutionary processes may have shaped patterns of population subdivision; (iv) Design a cost-effective and informative SNP panel that can be used for future population studies and genetic stock identification.

2. Materials and Methods

2.1 Sampling and DNA isolation

Samples were collected opportunistically between 2015 and 2017 through existing fishery surveys, fisheries targeted to horse mackerel, and as bycatch at 12 locations across the east Atlantic and the western Mediterranean Sea (Figure 1). We aimed to collect spawning fish to ensure that samples could provide a valid baseline (Figure S1). However, due to the opportunistic nature of sampling this was not always possible (Table S1). Maturity stages were recorded by sample collectors using a number of different maturity keys. Therefore, these were standardised to the six-point international horse mackerel maturity scale (Table S2; (ICES, 2015)). A 0.5 cm^3 piece of tissue was excised from the dorsal musculature of each specimen and stored at 4°C in absolute ethanol. Total genomic DNA was extracted from the majority of samples by Weatherbys Scientific Ltd, Ireland from 30 mg of tissue using sbeadex™ magnetic bead-based extraction chemistry on the LGC Oktopure™ platform. The remaining samples were extracted using a CTAB or Chelex and proteinase-K based extraction protocol (Table S1). The Chelex protocol produced single-stranded DNA, whereas the other methods, double-stranded. DNA quantity was measured with a NanoDrop ND-1000 spectrophotometer.

128 **2.2 Pool library preparation and sequencing**

129 We performed whole-genome sequencing of pooled DNA (pool-seq) to assess the genomic
130 variation among samples. This method provides population-level allele frequencies by
131 sequencing to a high depth a single-barcoded library prepared from a mixture (pool) of DNA
132 of individuals from a population (Schlötterer, Tobler, Kofler, & Nolte, 2014). DNA pools
133 were prepared by mixing equal amounts of DNA of 30 to 96 individuals collected in close
134 spatial and temporal proximity. DNA pools were quantified using a Qubit Fluorometer
135 (Thermo Fischer Scientific Inc) aiming to have at least 1.5 μg of DNA in 25-50 μL and were
136 submitted to the SNP&SEQ Technology Platform in Uppsala, Sweden, for library preparation
137 and high-throughput sequencing. A PCR-free Illumina TruSeq library with an insert size of
138 350 base pairs (bp) was prepared for most pools, except for the ones extracted with Chelex,
139 for which a Splinted Ligation Adapter Tagging (SPLAT) library was used instead, as it is
140 aimed for single-stranded DNA (Raine, Manlig, Wahlberg, Syvänen, & Nordlund, 2017).
141 Paired-end short reads (2x150 bp) were generated using an Illumina NovaSeq sequencer and
142 S4 flow cells.

143

144 **2.3 Read mapping and variant calling**

145 Low quality bases, sequencing adapters, and short reads were removed from the raw reads
146 using *Trimmomatic* v.0.36 (Bolger, Lohse, & Usadel, 2014). Clean reads were mapped
147 against the *Trachurus trachurus* genome assembly (Accession: GCA_905171665.1, (Genner
148 & Collins, 2022)) using *bwa-mem* 0.7.17 (Li, 2013). A pilot examination of read mapping
149 statistics indicated that the temporal replicates from northern and southern Portugal (NPT2,
150 SPT2) might be affected by technical artefacts (Figure S2). Hence, these samples were
151 excluded from genome-wide analyses, but were included in analyses focused on specific loci
152 (e.g., in the examination of allele frequency patterns).

153 Variant calling was performed using the algorithm *UnifiedGenotyper* of *GATK* v3.8
154 (McKenna et al., 2010). Biallelic SNPs were retained and various quality filters were applied
155 to remove spurious markers (see Supporting Information for details, Figure S3). The resulting
156 high-quality SNPs were used in further analyses. A summary of the data generation steps is
157 shown in Figure S4.

158

159 **2.4 Population genetic structure and genetic diversity**

We assessed population structure using pairwise pool-F_{ST} and principal components analysis (PCA). We computed the pool-F_{ST} ($\hat{F}_{ST}^{pool}$) statistic using the R package *poolfstat* (Hivert, Leblois, Petit, Gautier, & Vitalis, 2018), which is equivalent to the Weir & Cockerham (1984) F_{ST} and accounts for random chromosome sampling in pool-seq.

To evaluate whether neutral or selective processes better explain the observed patterns of genetic differentiation, we separately performed PCA on two SNP subsets, one of undifferentiated (i.e., presumably neutral) and the other of highly differentiated markers (outliers, assumed to have been subject to selection). These marker sets were chosen based on the empirical distribution of allele frequencies and standard deviation (SD) cut-off values (Figure S5) (see Supporting Information for details). To reduce redundancy and physical linkage among SNPs, we retained, at most, one marker every 1 kb in the undifferentiated marker set, and one marker every 10 kb in the differentiated set, as it is expected that linkage is more pronounced in regions under selection. PCA was performed using the R package *prcomp.* In a pilot analysis one sample from the west of Ireland (WIE2) appeared as an outlier (Figure S6). In the absence of any plausible biological reason for this observation, this sample was removed.

To examine the genome-wide variation of genetic diversity in each pool, we calculated nucleotide diversity (π) per pool in 10 kb-sliding windows with a step size of 2 kb using *PoPoolation 1.2.2* (Kofler et al., 2011) (see Supporting Information for details). Plotting and statistical testing was performed using the R environment (R Core Development Team, 2021).

Additionally, we evaluated whether population structure followed an isolation-by-distance pattern with a Mantel test and 9999 permutations using the R package *ade4* (Dray & Dufour, 2007). We compared the linearized genetic distances, calculated with the formula linearized-$\hat{F}_{ST}^{pool} = \frac{\hat{F}_{ST}^{pool}}{1-\hat{F}_{ST}^{pool}}$ (François Rousset, 1997), and the geographic distances, estimated as the least-coast oceanic distance in kilometers (km) considering land as barrier using the R package *CartDist* (Stanley & Jeffery, 2017).

2.5 Detection of loci under selection

We applied effective coverage correction (n_{eff}) to the raw read counts, in order to account for random variation of read coverage and chromosome sampling during pooling and sequencing

191 (Bergland, Behrman, O'Brien, Schmidt, & Petrov, 2014; Feder, Petrov, & Bergland, 2012;
192 Kolaczkowski, Kern, Holloway, & Begun, 2011) (see Supporting Information for details).
193 Pool allele frequencies were then computed using a custom python script.
194 To identify genomic regions with elevated differentiation respect to the genomic
195 background, we calculated the absolute delta allele frequency (dAF) per SNP between paired
196 contrasts of grouped pools, as *dAF = absolute(meanAF(group1) – meanAF(group2)*. The
197 contrasts evaluated were established based on geographic closeness, PCA clustering, and
198 biological knowledge (Table S3). We also calculated the moving average of dAF values in
199 100-SNP windows to identify regions with consistent differentiation across nearby markers,
200 while ruling out single SNPs that may be caused by random effects of pool-seq. We further
201 explored the allele frequency patterns of the most differentiated SNPs at each locus and
202 contrast for all the 12 pools. All the analyses were performed using R and plotting was done
203 with the R package *ggplot2* (Wickham, 2016).
204
205 **2.6 Validation of informative markers for genetic stock assessment**
206 To identify a panel of highly informative SNPs for genetic stock identification and to validate
207 pool-seq findings, we obtained the genotypes of 160 individuals (20 fish each from eight
208 locations) in 100 SNPs (Table S4). The 100-SNPs panel consisted of 24 neutral markers and
209 76 putatively adaptive markers (see Supporting Information for details). The split of adaptive
210 markers, in terms of observed association, was: North Sea (n=28), north-south break (n=13),
211 west of Ireland (n=14), Alboran Sea (n=13), southern Portugal (n=4), north Africa (n=4).
212 Three to four individuals per location were genotyped twice to assess genotyping error rate.
213 DNA extraction and SNP genotyping were undertaken by IdentiGEN, Ireland, using their
214 IdentiSNP genotyping assay chemistry. The protocol utilises target specific primers and
215 universal hydrolysis probes. Following an end-point PCR reaction, different genotypes are
216 detected using a fluorescence reader.
217 Based on individual allele frequencies, we undertook a preliminary analysis of
218 population structure among the eight genotyped fish aggregations. It should be noted that
219 sample sizes were small and therefore the results of the population analyses should be viewed
220 as preliminary until further large-scale screening is undertaken. Only individuals and markers
221 with >80% genotyping success were retained. Deviations from Hardy–Weinberg equilibrium
222 (HWE) and linkage disequilibrium (LD) were assessed with *Genepop* 4.2 (Rousset, 2008).

223 Six SNPs had indication of deviation from HWE, two markers (12_3119866 and 17_972744)
224 were not polymorphic and one had evident scoring errors (24_5252083), thus these nine
225 markers were excluded. *Microsatellite Analyzer* (*MSA*) 4.05 was used to calculate pairwise
226 F_{ST} estimates (Dieringer & Schlötterer, 2003). In all cases with multiple tests, significance
227 levels were adjusted using the sequential Bonferroni technique (Rice, 1989). PCA was
228 performed using the R function *prcomp*.
229 We estimated individual admixture coefficients for a given number of ancestral source
230 populations (*K*) using the *sNMF* algorithm (Frichot, Mathieu, Trouillon, Bouchard, &
231 François, 2014a) of the R package *LEA* (Frichot, Francois, & Caye, 2017). We tested $K = 1$ to
232 5, with 10 repetitions and 200 iterations. The most likely *K* corresponds to the value where
233 the cross-entropy criterion (metric that evaluates the error of the ancestry prediction) plateaus
234 or increases (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014b). We plotted the
235 average admixture proportions per population sample over a map using the R packages *ggplot*
236 (Wickham, 2016) and *ggOceanMaps* (Vihtakari, 2020).
237
238 **2.7 Characterization of a putative inversion on chromosome 21**
239 To assess and compare the genetic diversity and spatial distribution of haplotypes of the
240 putative inversion on chromosome (chr) 21, we extracted the individual genotypes of 12
241 diagnostic SNPs within the inversion from the 100-SNP dataset (Figure S7). We performed a
242 PCA with the R function *prcomp* to identify the inversion genotype of each individual.
243 Individuals were assigned to a haplotype group using the first two eigenvectors of the PCA
244 and the k-means clustering algorithm implemented in the R function *kmeans*. We calculated
245 observed heterozygosity for each of the PCA clusters, with the expectation that the middle
246 cluster, presumably corresponding to inversion-level heterozygotes, will have the highest
247 heterozygosity. These analyses and correspondent graphics were performed using R.
248
249 **2.8 Genome-Environment Association**
250 To identify which environmental variables are strongly correlated with adaptive genetic
251 variation and local adaptation, we tested genome-environment association using redundancy
252 analysis (RDA). We retrieved environmental data from *Bio-Oracle* v.2.1 (Assis et al., 2018;
253 Tyberghein et al., 2012) for each sampled location using the *R* package *sdmpredictors*
254 (Bosch, 2020). Data corresponded to mean bottom depth layers of minimum, mean,

255 maximum, and range values of sea water temperature (°C), sea water salinity (PSS), and
256 dissolved oxygen concentration (μmol/m^3) for a 14-year period (from 2000 to 2014) and a
257 spatial resolution of 0.25 arcdegrees. The data was derived from a numerical model developed
258 by the Global Ocean Biogeochemistry non-assimilative Hindcast (PISCES) (Copernicus
259 Programme of the European Union, n.d.).
260 Environmental data per location were standardized to zero mean and unit variance,
261 and collinear variables were removed (see Supporting Information for details, Figure S8). To
262 run RDA, we used the R package *vegan* (Dixon, 2003), the non-redundant environmental
263 dataset, and the pool-allele frequencies of SNPs in outlier loci found with genome scans. The
264 significance of the RDA model and environmental variables was assessed with an analysis of
265 variance (ANOVA) using 1000 permutations.
266
267 **2.9 Functional annotation of gene models**
268 The gene models of the Atlantic horse mackerel genome were developed by Ensembl (Howe
269 et al., 2021) and become available in an Ensembl Rapid release in March 2021 (Ensembl,
270 2021b). However, at the time this research was conducted, the gene models lacked gene
271 symbols (names) and GO terms (only Ensembl gene IDs were available). Given that these
272 annotations are relevant for interpretation of results, we ran the functional gene annotation
273 pipeline developed by the National Bioinformatics Infrastructure Sweden (NBIS) (Binzer-
274 Panchal, Dainat, & Soler, 2021) to retrieve this information (see Supporting Information for
275 details).
276 Additionally, for the top 2% most differentiated SNPs within each divergent genomic
277 region detected with genome scans, we annotated the closest overlapping gene (up to ± 40 kb)
278 and the variant effect prediction (e.g., missense, synonymous, upstream, downstream,
279 intergenic) using snpEff v.4.1 (Cingolani et al., 2012).
280
281 **3. Results**
282 **3.1 Sequencing and variant calling**
283 We generated pooled DNA whole-genome sequence data of 12 Atlantic horse mackerel
284 aggregations sampled across the species' range in the east Atlantic and the western
285 Mediterranean Sea, Alboran Sea (Figure 1A, Table 1). Temporal replicates of four locations
286 were collected one year apart, so the temporal stability of their genetic composition could be

287 investigated (west of Ireland WIE1-WIE2, North Sea NOS1-NOS2, northern Portugal NPT1-
288 NPT2, and southern Portugal SPT1- SPT2). Our sequencing effort yielded 490-764 million
289 high-quality reads per pool. 98.2-99.2% of reads aligned to the genome assembly of *T.*
290 *trachurus* (Genner & Collins, 2022), and each pool had a mean depth of coverage between
291 25.7x and 46.3x (Table S5). After variant calling with GATK, a total of ~12.8 million
292 biallelic SNPs passed quality filters and were used in the population analysis.
293
294 **3.2 Population genetic structure**
295 Pairwise pool-F_{ST} estimates indicated low genome-wide differentiation among horse
296 mackerel populations (global mean pool-F_{ST}=0.007, range 0.001-0.012) (Figure 1B), but
297 revealed four subtle population structure patterns that were also supported by genome scans.
298 First, the western Mediterranean Sea (MED) differentiates from Atlantic populations (mean
299 pool-F_{ST}=0.011). Second, Atlantic populations differ along a latitudinal cline that
300 distinguishes "northern" or "southern" groups respect to a pronounced genetic break near
301 Lisbon, Portugal (~38.7-39.0°N). Third, populations within the "northern" group (North Sea–
302 NOS1, NOS2, west of Ireland–WIE1, WIE3, northern Spanish Shelf–NES, northern
303 Portugal–NPT1) are genetically more similar to each other (mean pool-F_{ST}=0.004) than
304 populations within the "southern" group (southern Portugal–SPT1 and north Africa–NAF)
305 (mean pool-F_{ST}=0.006). Forth, the two samples from the North Sea (NOS1 and NOS2) show
306 the highest genetic similarity (pool-F_{ST}=0.001) of all and are distinguished from other
307 "northern" populations.
308 While genetic differences within the "northern" group are small, they are not
309 negligible (pool-F_{ST} range 0.001-0.006) (Figure S9). A Mantel test assessing the degree of
310 correlation between genetic and geographic distances (H_o: no association) between pairs of
311 populations indicated that subtle genetic differences within the northern group follow an
312 isolation-by-distance (IBD) pattern, suggesting that they might be shaped by effective gene
313 flow between neighbouring areas (Mantel test *p*-value=0.015; linear regression $R.adjusted^2$ =
314 0.9, *p*-value = 1.1^{-13}) (Figure 1C).
315 We further examined whether population structure patterns were driven primarily by
316 neutral or selective processes by separately performing PCA for two SNP subsets, of
317 undifferentiated or highly differentiated markers. These marker sets were chosen based on the
318 standard deviation in allele frequencies across populations (Figure S5). The undifferentiated

319 markers were considered "neutral" (61,543 SNPs) (Figure 1D, S8A), and the highly
320 differentiated markers were assumed under selection (818 SNPs) (Figure 1E, S8B). In both
321 PCA plots the separation between the "northern" and "southern" samples was evident,
322 although the composition of the clusters slightly varied depending on the marker set. In both
323 cases the "northern" group appeared as a tight cluster of genetically similar samples mainly
324 from the North Sea, west of Ireland, the northern Spanish shelf, and northern Portugal. In
325 stark contrast, the "southern" samples were dissimilar to each other. In the PCA with neutral
326 markers, the first two axes explained 30.6% of the genetic variance. The mid-range samples
327 from the Mediterranean Sea (MED) and southern Portugal (SPT1) were the most genetically
328 distinct, while the North Africa (NAF) sample was genetically similar to the "northern"
329 group. The North Sea samples tightly clustered with others in the "northern" group, being
330 almost undistinguishable. In contrast, in the PCA with selective makers, the first two axes
331 explained twice as much genetic variance (60.8%). The Mediterranean Sea sample was
332 genetically closer to the "northern" group, while the other two "southern" samples, from
333 southern Portugal and North Africa, were the most differentiated. Notably, the two North Sea
334 samples formed a distinct cluster separated from other "northern" samples.
335 Taken together, these results indicate that the separation between the western-most
336 part of the Mediterranean Sea and Atlantic populations is largely driven by neutral processes,
337 while the latitudinal genetic cline and separation of North Sea samples could be the result of
338 selective processes.
339
340 **3.3 Putative loci under selection**
341 We performed whole genome scans based on the absolute difference in allele frequencies
342 (dAF) for paired contrasts to identify outlier loci respect to the genomic background
343 (presumed under selection). The contrasts were chosen based on PCA clustering patterns
344 (Figure 1D, 1E), and the outlier regions were identified using a Bonferroni Z-score threshold
345 of significance. Our analysis revealed a number of genomic regions with elevated
346 differentiation for three contrasts: (i) the Mediterranean Sea vs. others, (ii) "northern" vs.
347 "southern" groups, and (iii) the North Sea vs. others (Figures 2-4). A summary of the top 2%
348 most differentiated SNPs per region, their closest gene, gene function, and relative position to
349 genes (e.g., missense, synonymous, upstream, downstream, intergenic) inferred with snpEff
350 are compiled in Table S6.

351

352 **Two outlier loci distinguish the western Mediterranean from Atlantic populations**

353 The contrast between Mediterranean Sea and other samples revealed two distinctive regions at
354 chromosomes (chr) 5 and 21 (Figure 2). In this case the signals of divergence were not as
355 evident as in other contrasts given the overall higher genome-wide differentiation of this
356 sample. Thus, we focused on the peaks in the rolling average of dAF values (Figure 2A), as it
357 requires consistent allele frequency differences for the contrast of interest. We considered
358 candidate genes those overlapping or located in the vicinity of the top 2% most differentiated
359 SNPs in each locus. In the region on chr 5 (Figure 2B), the Mediterranean Sea sample had
360 intermediate allele frequencies across loci and higher nucleotide diversity, while other
361 samples tended to be fixed for one haplotype and showed lower nucleotide diversity. The top
362 candidate gene in this region is *opn1mw4* (*green-sensitive opsin-4*). In the putative selection
363 signal on chr 21 (Figure 2C), the Mediterranean sample had one haplotype close to fixation,
364 the opposite haplotype was in high frequency in the North Sea, and intermediate frequencies
365 were prevalent elsewhere. This region harbours a single gene, *taar7a* (*trace amine-associated*
366 *receptor 7A*). The current knowledge of known functions of candidate genes are presented in
367 Table S6.

368

369 **A presumed 9.9 Mb inversion distinguishes populations along a latitudinal cline**

370 In the "northern" vs. "southern" contrast, we discovered that a single large region on chr 21
371 underpins the latitudinal cline with a genetic break along mid Portugal (Figure 3). This locus
372 consists of a large block of SNPs with elevated dAF spanning 9.9 Mb (Figure 3A-B). The
373 large size and abrupt drop in allele frequency differences towards the edges are characteristic
374 of structural variants with suppressed recombination (e.g., inversions) that do not represent
375 recent selective sweeps (Han et al. 2020). An analysis of the pool-allele frequencies of the top
376 2% most differentiated SNPs at this locus (Figure 3C) revealed striking allele frequency
377 differences between the "northern" group (including the Mediterranean Sea) and North
378 Africa, while intermediate allele frequencies were prevalent in southern Portugal.

379 To further characterize this structural variant, we leveraged genotype data of 160
380 individuals for 12 SNPs within the region. As expected for a chromosomal inversion, in a
381 PCA plot individuals separated into three clusters depending on their genotype (Figure 3D),
382 and individuals in the intermediate cluster exhibited the highest average observed

383 heterozygosity (Figure 3E). Overall, the individual genotype data supported the observations
384 made with pool-seq data regarding the spatial distribution of inversion haplotypes, although it
385 offered greater resolution. For instance, it revealed that the "northern" haplotype dominates in
386 the North Sea, west of Ireland, northern Spanish shelf, and northern Portugal, but also occurs
387 at an intermediate frequency in southern Portugal. The "southern" haplotype is predominant
388 in North Africa, and common in southern Portugal (Figure 1F). This putative inversion
389 contains about 1,077 genes, making it difficult to infer which particular genes are under
390 selection.
391
392 **A genetically distinct population in the North Sea**
393 The contrast between the North Sea and other samples revealed seven genomic regions with
394 elevated differentiation on chr 1, 4, 7, 11, 20, and 21 (Figure 4A). Further examination of
395 allele frequencies of top SNPs per locus indicated that the North Sea temporal replicates were
396 very similar genetically, and that their allele frequency patterns were distinctive of this
397 location (Figure 4B-E, Figure S10). For instance, in the selection signals on chr 1 and 21,
398 North Sea samples had variant alleles close to fixation and a slight reduction in nucleotide
399 diversity (π) (Figure 4B, 4C, 4E). At the outlier loci on chr 4, 11 and 20, North Sea samples
400 had predominantly intermediate allele frequencies, with similar (chr 4 and 11, Figure S10B,
401 S10D) or higher (chr 20, Figure 4D) nucleotide diversity with respect to other pools.
402 Additional divergent regions on chr 7 and 11 showed a similar but noisier pattern, presumably
403 due to the presence of complex structural variants at these loci (Figure S10C, S10D).
404 The most striking signals were located on chr 1, followed by those on chr 20 and 21
405 (Figure 4). The signal on chr 1 encompasses two regions (chr1:17.4-17.6 Mb, chr1:22.0-22.1
406 Mb). In the first region (Figure 4B), the most differentiated SNPs are in the vicinity of the
407 gene *gpr83* (G-protein coupled receptor 83). In the second region (Figure 4C), the top
408 candidate gene is *sgms2* (*sphingomyelin synthase 2*). The signals on chr 20 (Figure 4D) and
409 chr 21 (Figure 4E) contain a single gene each, *ncoa2* (*nuclear receptor coactivator 2*) and
410 *taar7a* (*trace amine-associated receptor 7A*). Note that the signal on chr 21 is the same as the
411 one noted in the Mediterranean sample (see above). Additional signals on chr 4, 7 and 11
412 (Figure S10) were not as clear as the ones just described (e.g., smaller allele frequency
413 differences and/or inconsistent allele patterns). Therefore, it was difficult to identify candidate

genes for these regions. The current knowledge concerning the function of candidate genes in the differentiated genomic regions is summarized in the Table S6.

3.4 Genome-Environment Association

To identify the environmental variables likely driving adaptive genetic variation, we retrieved seawater-related environmental variables (temperature, dissolved oxygen and salinity) and examined genome-environment association (GEA) based on redundancy analysis (RDA) (Figure 5A). A total of 210,568 SNPs within the most divergent genomic regions in the North Sea, Mediterranean Sea and northern vs. southern contrasts were used in this analysis.

Mean seawater dissolved oxygen concentration has the strongest association with genetic variation ($R^2 = 0.83$), followed by mean seawater temperature ($R^2 = 0.71$) and, to a lesser extent, with salinity ($R^2 = 0.34$) (Figure 5B-D). Notably, the latitudinal genetic gradient discovered shows an inverse relationship with oxygen content, with the North Sea being the location with the highest oxygen levels (~270-300 μmol/m^3), almost three times the oxygen content in North Africa (~100 μmol/m^3) (Figure 5C). The two southernmost populations studied (in North Africa and southern Portugal)—which showed a high and intermediate frequency of the "southern" inversion haplotype, respectively—are from waters with higher mean temperatures (18°C), compared to locations north of Lisbon, in which the "northern" haplotype dominates. The North Sea has the coldest temperatures of all locations (< 10°C) (Figure 5B). In terms of salinity, as expected, the Mediterranean Sea has the highest salinity followed by southern Portugal, while the lowest salinity is observed in the North Sea (Figure 5D). That being said, the salinity range between the northern- and southernmost locations here analysed is minimal, between 35.0 and 36.5 PSS, which is considered a fully marine environment.

3.5 Validation of informative markers for genetic stock identification

To confirm the selection signals found with pool-seq data and develop a tool for genetic stock identification, we genotyped 160 individuals using 100 SNPs. Of these, 24 markers were considered neutral and 76 were among the most differentiated across contrasts (Figure S7). We found a strong correlation between allele frequencies calculated from individual genotypes and pool-seq data (mean $R^2 = 0.9 \pm 0.1$), supporting the findings of the pool-seq analysis (Figure S11). Of the 76 outlier loci, 72 had a genotyping success >80% (Table S7).

446 After applying quality filters, the final dataset had 63 SNPs, while 157 out of 160 individuals
447 had genotyping success >80%. Henceforth, this dataset will be referred to as the 63-SNP
448 panel.
449 To minimise marker redundancy in the 63-SNP panel, we performed a linkage
450 disequilibrium (LD) analysis for all loci and samples. As expected, significant LD was found
451 between a number of SNPs located in close proximity on the same chromosome (Table S7).
452 Though LD was not statistically significant in some cases (e.g., SNPs on chr 5), these were
453 considered linked due to their physical closeness. To identify the most informative SNPs for
454 discriminating the samples while reducing LD, we analyzed F_{ST} by marker and by population
455 (Figure S12) for each genomic region. We retained the SNP with the highest average F_{ST} per
456 linkage group (assumed to be the most informative), yielding a 17 SNP dataset comprising
457 155 out of 160 individuals with a genotyping success >80% (henceforth: the 17-SNP panel).
458 Further analyses were conducted with both the 63-SNP and the 17-SNP datasets (individual
459 genotypes are shown in Figure S13).
460 An examination of pairwise F_{ST} values indicated lack of significant genetic
461 differences between the North Sea temporal replicates or between the two west of Ireland
462 samples (Table S8). There was also no significant genetic differentiation between the
463 individuals from the northern Spanish shelf, northern Portugal, and the two samples from the
464 west of Ireland (Table S8).
465 To assess whether the same population structure is recovered with both the 63- and
466 17-SNP panels, we performed separate PCA and admixture analyses. We observed similar
467 clustering patterns with both panels (Figure 6A, 6B), consisting of four main groups: (i) North
468 Sea (NOS1, NOS2); (ii) west of Ireland (WIE1, WIE2), northern Spanish shelf (NES), and
469 northern Portugal (NPT1); (iii) southern Portugal (SPT1); and (iv) north Africa (NAF). While
470 the 63-SNP panel provided better discrimination between southern Portugal (SPT1) and North
471 Africa (NAF), both marker sets separated most individuals from the North Sea, as well as the
472 "northern" and "southern" groups. It should be noted that some individuals clustered in
473 different groups from the ones expected, suggesting that gene flow occurs among geographic
474 regions.
475 An individual admixture analysis with both datasets supported the same four major
476 genetic groups (Figure 6C, 6D) as the ones identified with PCA, showing various levels of
477 mixing among them. For example, a few individuals in the North Sea (n = 3-4) had a high

probability of originating from the western group (first-generation migrants). Similarly, southern Portugal had a number of individuals that appear to originate from the western group (n = 3) or from the African group (n = 3). In all four groups some individuals showed admixed ancestry, indicating that they are probably F1-hybrids or backcrosses between local and migrants. Overall, these results indicate that gene flow occurs more often between neighboring geographic areas, following an isolation-by-distance pattern (e.g., between the North Sea and the west of Ireland; between the northern Spanish shelf, northern Portugal and southern of Portugal; and between southern Portugal and north Africa).

4. Discussion

This study represents the most comprehensive assessment of the genomic variation of Atlantic horse mackerel populations to date (Figure 1). The combination of extensive geographic sampling and genomic screening derived from whole-genome sequencing, provided a powerful dataset that allowed us to discover genomic regions underlying population structure and local adaptation. Moreover, we developed a genetic tool that can distinguish the main biological groups.

Population structure and temporal stability of genetic groups

Our genomic analysis indicated low genetic differentiation among horse mackerel populations inhabiting the area from the North Sea to north Africa (Figure 1A), more than 4,000 km of coastline (global mean pool-F_{ST}=0.007, range: 0.001-0.012). This result is in agreement with previous studies using few genetic markers (Cimmaruta et al., 2008; Comesaña et al., 2008; Healey et al., 2020; Kasapidis & Magoulas, 2008), and implies that gene flow is pervasive. We infer that in the horse mackerel, since the larval stage lasts about one month (Rusell, 1976), gene flow could predominantly occur between neighbouring areas through passive transport of pelagic eggs and larvae by ocean currents. Moreover, the large population sizes of this species imply a minimal role of genetic drift in shaping patterns of genetic diversity.

Despite low differentiation at neutral loci, we uncovered patterns of population structure (Figure 1B, D, E) mainly at loci putatively under selection (Figures 2-4). Below, we will touch on the summary statistics results, and in the next section we will discuss the genome scan results.

509 Genetic differences between the Alboran Sea (the westernmost part of the
510 Mediterranean) and the Atlantic occurred across the genome, implying that genetic drift
511 contributed to the genetic differentiation. This separation was already proposed in earlier
512 studies using body morphometrics, otolith shape, and parasitofauna (Abaunza et al., 2008).
513 However, this is the first presentation of genetic data supporting this, as previous studies were
514 inconclusive (Cimmaruta et al., 2008; Comesaña et al., 2008; Healey et al., 2020; Kasapidis
515 & Magoulas, 2008). A Mediterranean-Atlantic divide has also been described in several
516 marine species, but is not observed in all (Patarnello, Volckaert, & Castilho, 2007), likely
517 because of differences in species' life history or low resolution of the genetic markers used.
518 Previous research based on parasite composition indicated that the Alboran Sea is a mixing
519 area of horse mackerel from the Mediterranean and the Atlantic (Abaunza et al., 2008;
520 Mattiucci et al., 2008). If this were the case, we would expect to see both inversion haplotypes
521 at this location, but what we observed is that the northern haplotype prevails. This, in addition
522 to the fact that none of the individuals collected were actively spawning (Table S1), let us
523 infer that our sample is probably not representative of this area, and likely comes from a
524 region with colder waters (e.g. Gulf of Lyon, Balearic Sea) (Figure S14). A possible neutral
525 mechanism for reduced gene flow between Atlantic and western Mediterranean populations is
526 the presence of retentive currents in the Almeria-Oran front, which have been proposed to act
527 as barriers for gene flow for various marine species (Patarnello et al., 2007).
528 Our genomic data also revealed a hitherto undescribed latitudinal break off mid
529 Portugal, near Lisbon (~38.7-39.0°N), distinguishing populations from north and south of this
530 area. The genetic separation was noticeable in pairwise-F_{ST} estimates, but more evident in the
531 PCA based on outlier SNPs, suggesting a putative role of natural selection. Interestingly, the
532 inferred location of the north-south genetic break coincides with a major biogeographical
533 transition zone between temperate and subtropical waters off the coast of Portugal (Cunha,
534 2001; Santos et al., 2007). Moreover, a recent genetic study in the boarfish (*Capros aper*), a
535 pelagic fish with similar distribution and life history as the horse mackerel, reported a
536 comparable latitudinal pattern (Farrell, Carlsson, & Carlsson, 2016). Thus, it is possible that
537 the environmental transition zone in Portugal is a major driver of population structuring of
538 many marine species inhabiting this area, including the horse mackerel.
539 Within the "northern" group—North Sea, west of Ireland, the northern Spanish shelf
540 (Bay of Biscay), and northern Portugal—genetic differences are small at both, neutral and

541 selective markers (Figure 1D, E). Such genetic variability resembles an isolation-by-distance
542 pattern (Mantel test *p*-value=0.015), which was also observed in a previous allozyme-based
543 study (Cimmaruta et al., 2008). Thus, genetic differences between populations within the
544 "northern" group seem to be largely determined by the effective gene flow via passive larval
545 dispersal by ocean currents. The North Sea samples, though, were slightly more differentiated
546 at outlier SNPs (Figure 1E), suggesting a putative role of natural selection.
547 In contrast, populations within the "southern" group—inhabiting waters in southern
548 Portugal and North Africa, Mauritania—are genetically more differentiated between them and
549 from all other populations. Interestingly, the population in the south of Portugal was different
550 at both, neutral and selective markers, whereas the population from North Africa was more
551 genetically distinct at outlier markers.
552 In summary, our genomic data indicates five possible genetic groups: (1) the western
553 Mediterranean Sea; (2) the North Sea; (3) the west of Ireland, northern Spanish shelf (Bay of
554 Biscay), and northern Portugal; (4) southern Portugal; and (5) north Africa, Mauritania. These
555 groups seem to be temporally stable, as indicated by similar allele frequencies at outlier loci
556 observed between temporal replicates (west of Ireland WIE1-WIE2, North Sea NOS1-NOS2,
557 northern Portugal NPT1-NPT2, and southern Portugal SPT1- SPT2) (Figures 2-4).
558
559 ***Genomic regions and environmental factors involved in local adaptation***
560 Genome scans uncovered a number of regions with elevated differentiation that support the
561 subtle patterns of population sub-structuring identified with summary statistics. The
562 informative contrasts were: (i) the Mediterranean Sea vs. others, (ii) "northern" vs. "southern"
563 groups, (iii) the North Sea vs. others. The divergent genomic regions varied in size (70-600
564 kb, and 9.9 Mb for a structural variant) and harboured numerous candidate genes.
565 The test for genome-environment association (GEA) indicated that genetic differences
566 at putatively adaptive loci are primarily driven by differences in seawater oxygen content and
567 temperature between locations (R^2=0.83 and R^2=0.71, respectively) (Figure 5). Notably, there
568 was an inverse relationship between sea water oxygen, temperature, and the latitudinal cline,
569 where the North Sea had the highest oxygen concentration (~270-300 μmol/m^3), almost three
570 times the oxygen content in north Africa (~100 μmol/m^3). Moreover, the North Sea also
571 exhibited the coldest seawater temperature of all locations analyzed (< 10°C), strikingly

572 different from the two southernmost populations sampled (north Africa and south Portugal)
573 (18°C).
574 We found that two outlier loci distinguish the western Mediterranean Sea, one on chr
575 5 and the other on chr 21 (Figure 2). The signal on chr 21 harbours a single gene, *taar7a*
576 (*trace amine-associated receptor 7A*), which encodes a receptor involved in olfactory sensing
577 of amines (Hashiguchi & Nishida, 2007). The signal on chr 5 includes three genes that encode
578 proteins with crucial roles in synapsis (*cadps, synpr, slc6a13*), and five in-tandem copies of
579 the green-sensitive opsin-4 (*opn1mw4*) gene. This gene organization (*opn1mw4* copies
580 flanked by *synpr* and *slc6a13*) is highly conserved among several Clupeocephala species
581 (e.g., Atlantic herring, zebrafish, medaka, stickleback (Lin, Wang, Li, & Wang, 2017;
582 Nakamura et al., 2013)).
583 The top candidate gene at this locus is *opn1mw4*, a paralog of the *opn1mw* (RH2)
584 gene, which encodes cone photopigments essential for vision of blue-green light. This gene
585 contains two missense mutations (c.850G>A and c.670G>A, Figure S14A, B), one more
586 divergent than the other (dAF=0.53 and dAF=0.40, respectively). The most divergent variant
587 results in an amino acid change from alanine to tryptophan at residue 284 (p.Ala284Thr), and
588 the less divergent variant, from valine to isoleucine at residue 224 (p.Val224Ile). While all
589 four amino acids are hydrophobic, tryptophan is more hydrophobic and occupies a larger in
590 volume than alanine, and the same is true for isoleucine in respect to valine, suggesting that
591 these amino acid changes could potentially affect protein structure. Therefore, it is possible
592 that the missense mutations in *opn1mw4* are causal, as affecting the structure of the
593 photopigments could generate a shift in spectral sensitivity similar to the Phe261Tyr
594 substitution in rhodopsin present in many fish species that live in brackish or freshwater (Hill
595 et al., 2019).
596 A change in visual sensitivity could be an adaptive response to contrasting light
597 environments between the Mediterranean and the Atlantic. In clear waters the blue-green light
598 spectrum dominates, especially as depth increases, while in shallower or turbid waters red-
599 shifted light conditions prevail. Comparative genomic studies indicate that fish living in
600 marine habitats dominated by green-blue light (e.g., in the deep sea, the pelagic open ocean,
601 or in coral reefs at night) commonly show gene copy expansions of *opn1mw* (RH2) respect to
602 fish living in shallower (<30 m) or more turbid waters (Lin et al., 2017; Musilova & Cortesi,
603 2021). Indeed, the horse mackerel has five tandem copies of green-light opsins, which might

604 relate to their preferred habitat (100–200 m depth) (Brunel et al., 2016). Seawater turbidity in
605 the Mediterranean Sea is generally lower than in the Atlantic Ocean (Figure S14C),
606 suggesting that more green light might be transmitted to deeper layers of the water column.
607 Taken together, the probable adaptive significance of the missense mutations in
608 *opn1mw4* is that they might be involved in spectral fine-tuning of green-opsins to
609 downwelling light detection under contrasting light conditions between the Mediterranean
610 Sea and the Atlantic. The former is likely dominated by blue-green light conditions, while in
611 the latter conditions are comparatively red-shifted. Visual adaptation could confer survival
612 advantages related to feeding, recognition of conspecifics, and escape from predators.
613 We discovered that the latitudinal genetic cline is underlined by a large (9.9 Mb)
614 putative chromosomal inversion on chr 21 (Figure 3). The northern haplotype shows high
615 frequency in the "northern" group and in the western Mediterranean Sea, while the alternative
616 haplotype dominates in north Africa, and occurs at an intermediate frequency south of
617 Portugal. This chromosomal inversion harbours thousands of genes, the roles of which cannot
618 be resolved without further studies. However, considering the strong genetic differentiation it
619 drives along an important oceanographic transition zone, we infer this inversion may play a
620 fundamental role in local adaptation to contrasting thermal regimes: a temperate (cooler)
621 climate in the north, tropical (warmer) in the south, and variable/intermediate at the transition
622 zone in mid Portugal. A possible explanation for the fact that the "northern" haplotype is
623 prevalent in the western Mediterranean Sea is that this sample might actually spawn in a
624 different location with cooler waters than where it was collected (near the Gibraltar strait, ~16
625 °C). Within the Mediterranean Sea, comparable temperatures to those in the northern region
626 (~11-12 °C) occur in the winter season in coastal waters of the Balearic Sea near Catalonia
627 (Spain), the Gulf of Lyon (France), and the Ligurian Sea (northeast Italy) (Pastor, Valiente, &
628 Khodayar, 2020). Indeed, previous studies have reported spawning events of horse mackerel
629 near Catalonia during the winter months (Andreu & Rodriguez-Roda, 1953; Planas & Vives,
630 1953). Given that the individuals collected in the Alboran Sea were not actively spawning, it
631 is plausible that they come from a region further north.
632 We show that the population in the North Sea is genetically distinct. We reached this
633 conclusion because seven genomic regions clearly distinguish samples from this location
634 from all others, and the temporal replicates showed nearly identical allele frequencies at
635 outlier loci. This observation adds to previous morphometric and parasite data suggesting that

horse mackerel from the North Sea differs from nearby Atlantic populations (Abaunza et al., 2008).

Some of the top candidate genes likely involved in local adaptation to the North Sea are *gpr83*, *sgms2*, *ncoa2*, and *taar7a*. *gpr83* (G-protein coupled receptor 83) encodes a receptor that plays an important role in the regulation of energy metabolism, feeding, reward pathway, and stress/anxiety responses in mice, for which it has been linked with the control of body weight (Gomes et al., 2016; Lueptow, Devi, & Fakira, 2018). *ncoa2* (*nuclear receptor coactivator 2*) encodes a transcriptional coactivator for steroid receptors that is presumably involved in glucose metabolism regulation (Bateman et al., 2021). Previous experimental studies indicate that fish adapted to cold climates often have higher metabolic rates than those adapted to warm climates (Wang, Tan, Jiao, You, & Zhang, 2014; White, Alton, & Frappell, 2012). Therefore, we infer that selection may favor alleles that result in increased energy metabolism required for adaptation to the colder environment of the North Sea.

sgms2 (*sphingomyelin synthase 2*) encodes a protein involved in the synthesis of sphingomyelin, a major component of cell and Golgi apparatus membrane. Previous studies indicate that this protein is crucial to maintain cell membrane structure and fluidity at low temperatures in fish (Wang et al., 2014; Windisch, Kathöver, Pörtner, Frickenhaus, & Lucassen, 2011). Thus, we propose that genetic variants in this gene could facilitate cold tolerance in the North Sea.

taar7a (*trace amine-associated receptor 7A*) codifies an olfactory receptor specific for sensing amines in vertebrates (Hashiguchi & Nishida, 2007; Hussain et al., 2009; Tessarolo et al., 2014; Yamamoto et al., 2010). Amines are behaviorally relevant odorants proposed to play a critical role in intra- and interspecific communication in, for example, sexual attraction or avoidance of predators or rotting food (Dewan, 2021). A previous study by Shi & Wang (2010) on two goatfish species with contrasting bottom habitat preferences (*Mullus surmuletus* and *Mullus barbatus*) reported significant differences in the morphology of chemoreceptors. Such differences are proposed to be associated with increased sensitivity to chemical stimuli in the species living in muddy and deeper waters, which has reduced visual capabilities. Among the locations in the east Atlantic included in this study, the North Sea has the highest water turbidity of all (Figure S14). Thus, natural selection may favor *taar7a* alleles that confer an enhanced sense of smell under the reduced visibility in the North Sea.

Other candidate genes distinctive of the North Sea might play critical roles in reproduction (spermatogenesis-*ankrd49*, gametogenesis-*spin1*, fertilization-*cdk2ap2*); development of bone (*papss1*), skeletal muscle (*spin1*) and skin (*ptk6*); cell processes (*aip*, apoptosis-*nol4-nol4l*, nutrient-dependant growth-*srms*); energy metabolism and body weight (glucose-*eno2-ncoa2*, lipids-*cyp2u1*, nitrogen-*slc38a3*, weight maintenance-*hadh*); nervous and olfactory system (neurotransmitter release, *doc2b*; synaptic transmission, *slc38a3*); immune response (*tmem134, serping1*, *clec4e*); and osmolarity regulation at low temperatures (*hsd11b1l*). The complete list of genes located within regions of genetic differentiation and their putative function are presented in Table S6.

Interestingly, the number of loci involved and their degree of genetic differentiation in the Atlantic horse mackerel are relatively small compared with those in the Atlantic herring (Han et al., 2020), but is rather intermediate to the Atlantic herring and the European eel, as the latter constitutes a single panmictic population (Enbody et al., 2021). We propose that the most important explanation for the differences in genetic structuring between these three species is related to their respective spawning strategies, since spawning and early development constitute the most sensitive period of life for a fish, characterized by high mortality (Dahlke, Wohlrab, Butzin, & Pörtner, 2020) and thus strong selection. The Atlantic herring is a demersal spawner that breeds close to the coast in areas with marked environmental differences between subpopulations as regards to temperature, salinity, depth and biotic conditions (plankton production, predators, etc.). It is also presumed to show prevalent homing behaviour, that is, individuals return for spawning to the locations where they were hatched. In contrast, the Atlantic horse mackerel is a benthopelagic fish that spawns at deeper waters, near the shelf edge (100-200 m deep), at not well-defined spawning areas. This implies that the environmental conditions in which breeding occurs are comparatively less diverse. On the other extreme, to the best of our knowledge, all European eels spawn in the Sargasso Sea under similar environmental conditions.

In summary, we revealed a number of highly divergent genomic regions encompassing genes that may be involved in adaptation to local environmental conditions in the North Sea, the western Mediterranean Sea, and along a climate-related latitudinal cline in the Atlantic. In addition, our results highlight the importance of chromosomal rearrangements for local adaptation with gene flow.

699 ***Implications for fisheries management***

700 The genetic-based groups identified here are largely in agreement with the current horse
701 mackerel stock divisions, informed by the results of the 2000-2003 HOMSIR project
702 (Abaunza et al., 2008). However, our data do not support the current definition of the
703 "southern stock" in Portuguese waters (Figure S1), and of the southern boundary of the
704 "western stock".

705 Our genomic data indicates that the "southern stock" actually consists of two
706 biological units (Figures 1, 3, 6), not one as previously assumed. Samples from northern
707 Portugal (north of Lisbon, ~38.7-39.0°N) appear to be genetically closer to the "western
708 stock", while samples from southern Portugal (south of Lisbon) form their own group but are
709 genetically closer to the samples from the "Saharo-Mauritanian stock", in North Africa. To
710 confirm these findings and assess the spatial and temporal trends of mixing between these
711 areas, further studies are required, including a finer geographic sampling and screening of
712 informative genetic variants in a large number of individuals throughout this area.

713 We did not find significant genetic differences between northern Portugal (currently
714 considered part of the "southern stock"), northern Spanish shelf (Bay of Biscay), and the west
715 of Ireland, implying that the southern boundary of the "western stock" could possibly be
716 extended down to the north of Portugal. However, the minute genetic differentiation does not
717 exclude the possibility that isolation in an ecologically-relevant timescale of interest for
718 fisheries management might occur (Hauser & Carvalho, 2008).

719 Finally, our genomic data support the consideration of the Mediterranean Sea as a
720 separate stock, as proposed by the HOMSIR project (Abaunza et al., 2008). While a single
721 sample from the westernmost part of the Mediterranean was studied, its genetic
722 distinctiveness suffices to infer that the Mediterranean horse mackerel likely constitutes a
723 separate population from those in the Atlantic. Wide-scale sampling within the Mediterranean
724 Sea is required to further explore the population structure in this region.

725 This study identified a number of genetic markers (SNPs) that can be used as a genetic
726 tool for fisheries stock assessment. A panel of only 63 markers suffice to identify the main
727 genetic subdivisions. In fact, using a reduced panel of only 17 markers, it is possible to
728 differentiate individuals collected in "the North Sea stock" and "western stock", and from the
729 "northern" (North Sea, west of Ireland, northern Spanish shelf, northern Portugal, and the
730 Mediterranean) and the "southern" (southern Portugal, north Africa) groups with respect to

the genetic break outside Lisbon. Thus, these markers can help, for instance, to elucidate the extent of mixing between the Western and North Sea stocks in the English Channel (ICES Divisions 7.e and 7.d) and in ICES area 4.a in the northern North Sea.

Acknowledgements

Authors want to specially thank the members of the Northern Pelagic Working Group of the European Association of Fish Producers Organisation (EAPO) and the Pelagic Advisory Council (PelAC) for funding this study, and to Martin Pastoors, Maria Manuel Angélico, Finlay Burns, Gersom Costas, Cindy Van Damme, Cristina Nunes, Brendan O'Hea, Ciaran O'Donnell, Michael O'Malley, Jan Beintema and all IEO, IPMA, MI, MSS, WUR scientists and crew on the survey and commercial vessels involved in sampling. We also thank Jens Carlsson and the EU Atlas Project for providing the Mediterranean sample. The project was also financially supported by Vetenskapsrådet (2017-02907) and Knut and Alice Wallenberg Foundation (KAW 2016.0063). The National Genomics Infrastructure (NGI)/Uppsala Genome Center provided service in massive parallel sequencing and the computational infrastructure was provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Thanks also to the National Bioinformatics Infrastructure Sweden (NBIS) for the development of the pipeline used for functional gene annotation. Thanks to Scott Campbell for enriching discussions.

5. References

Abaunza, P., Gordo, L., Karlou-Riga, C., Murta, A., Eltink, A. T. G. W., García Santamaría, M. T., … Gallo, E. (2003). Growth and reproduction of horse mackerel, Trachurus trachurus (carangidae). *Reviews in Fish Biology and Fisheries*, *13*(1), 27–61. doi: 10.1023/A:1026334532390

Abaunza, P., Murta, A. G., Campbell, N., Cimmaruta, R., Comesaña, A. S., Dahle, G., … Zimmermann, C. (2008). Stock identity of horse mackerel (Trachurus trachurus) in the Northeast Atlantic and Mediterranean Sea: Integrating the results from different stock identification approaches. *Fisheries Research*, *89*(2), 196–209. doi: 10.1016/j.fishres.2007.09.022

Akopyan, M., Tigano, A., Jacobs, A., Wilder, A. P., Baumann, H., & Therkildsen, N. O. (2022). Comparative linkage mapping uncovers recombination suppression across massive chromosomal inversions associated with local adaptation in Atlantic silversides. *Molecular Ecology*. doi: 10.1111/mec.16472

Alvarez, P., & Chifflet, M. (2012). The fate of eggs and larvae of three pelagic species, mackerel (<i>Scomber scombrus</i>), horse mackerel (<i>Trachurus trachurus</i>) and sardine (<i>Sardina pilchardus</i>) in relation to prevailing currents in the Bay of Biscay: Could they affect larval survival? *Scientia Marina*, *76*(3), 573–586. doi: 10.3989/scimar.03298.07H

Andreu, B., & Rodriguez-Roda, J. (1953). La pesca marítimaen Castellón. Rendimento por unidad de esfuerzo (1945–49) y consideraciones biometricas de las especies de interés comercial. *Publicaciones Del Instituto de Biología Aplicada*, *8*, 223–277.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., de Clerck, O., & Tittensor, D. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3). doi: 10.1111/geb.12693

Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., … André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, *26*(17), 4452–4466. doi: 10.1111/mec.14207

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., … Teodoro, D. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. doi: 10.1093/nar/gkaa1100

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genetics*, *10*(11), e1004775. doi: 10.1371/journal.pgen.1004775

Beveren, E. van, Klein, M., Serrão, E. A., Gonçalves, E. J., & Borges, R. (2016). Early life history of larvae and early juvenile Atlantic horse mackerel Trachurus trachurus off the Portuguese west coast. *Fisheries Research*, *183*, 111–118. doi: 10.1016/j.fishres.2016.05.018

Binzer-Panchal, M., Dainat, J., & Soler, L. (2021). NBIS Genome Annotation Workflows (Version v1.0.0) [Computer software]. Retrieved July 12, 2021, from https://github.com/NBISweden/pipelines-nextflow

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bosch, S. (2020). sdmpredictors: Species Distribution Modelling Predictor Datasets. R package version 0.2.9. Retrieved from https://cran.r-project.org/package=sdmpredictors

Brunel, T., Farrell, E. D., Kotterman, M., Kwadijk, C., Verkempynck, R., Chen, C., & Miller, D. (2016). *Improving the knowledge basis for advice on North Sea horse mackerel.*

Developing new methods to get insight on stock boundaries and abundance. Wageningen, IMARES Wageningen UR (University & Research centre), Wageningen Marine Research report C092/16.

Cadrin, S. X., & Secor, D. H. (2009). Accounting for Spatial Population Structure in Stock Assessment: Past, Present, and Future. In R. B. J. Beamish R. J. (Ed.), *The Future of Fisheries Science in North America* (pp. 405–426). Dordrecht: Springer Netherlands. doi: 10.1007/978-1-4020-9210-7_22

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

Cimmaruta, R., Bondanelli, P., Ruggi, A., & Nascetti, G. (2008). Genetic structure and temporal stability in the horse mackerel (Trachurus trachurus). *Fisheries Research*, *89*(2), 114–121. doi: 10.1016/j.fishres.2007.09.030

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. doi: 10.4161/fly.19695

Comesaña, A. S., Martínez-Areal, M. T., & Sanjuan, A. (2008). Genetic variation in the mitochondrial DNA control region among horse mackerel (Trachurus trachurus) from the Atlantic and Mediterranean areas. *Fisheries Research*, *89*(2), 122–131. doi: 10.1016/j.fishres.2007.09.014

Copernicus Programme of the European Union. (n.d.). Copernicus Marine Service. Retrieved from http://marine.copernicus.eu/

Cunha, M. E. (2001). *Physical Control of Biological Processes in a Coastal Upwelling System: Comparison of the Effects of Coastal Topography, River Run-off and Physical Oceanography in the Northern and Southern Parts of Western Portuguese Coastal Waters* (Faculdade de Ciências da Universidade de Lisboa). Faculdade de Ciências da Universidade de Lisboa, Lisboa. Retrieved from https://www.researchgate.net/publication/237418870_Physical_control_of_biological_processes_in_a_coastal_upwelling_system_comparison_of_the_effects_of_coastal_topography_river_run-off_and_physical_oceanography_in_the_northern_and_southern_parts_of_Por

Dahlke, F. T., Wohlrab, S., Butzin, M., & Pörtner, H.-O. (2020). Thermal bottlenecks in the life cycle define climate vulnerability of fish. *Science*, *369*(6499), 65–70. doi: 10.1126/science.aaz3658

Dewan, A. (2021). Olfactory signaling via trace amine-associated receptors. *Cell and Tissue Research*, *383*(1), 395–407. doi: 10.1007/s00441-020-03331-5

Dieringer, D., & Schlötterer, C. (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, *3*(1), 167–169. doi: 10.1046/j.1471-8286.2003.00351.x

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*(6), 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x

Dray, S., & Dufour, A.-B. (2007). The **ade4** Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4). doi: 10.18637/jss.v022.i04

Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Palm, S., Wickström, H., & Andersson, L. (2021). Ecological adaptation in European eels is based on phenotypic plasticity. *Proceedings of the National Academy of Sciences*, *118*(4), e2022620118. doi: 10.1073/pnas.2022620118

Farrell, E. D., & Carlsson, J. (2018). *Genetic stock Identification of Northeast Atlantic Horse mackerel, Trachurus trachurus*.

Farrell, Edward D., Carlsson, J. E. L., & Carlsson, J. (2016). Next Gen Pop Gen: implementing a high-throughput approach to population genetics in boarfish (Capros aper). *Royal Society Open Science*, *3*(12), 160651. doi: 10.1098/rsos.160651

Feder, A. F., Petrov, D. A., & Bergland, A. O. (2012). LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data. *PLoS ONE*, *7*(11), e48588. doi: 10.1371/journal.pone.0048588

Frichot, E., Francois, O., & Caye, K. (2017). LEA: an R package for Landscape and Ecological Association studies. version 2.0. Retrieved from http://membres-timc.imag.fr/Olivier.Francois/LEA/software.htm

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014a). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983. doi: 10.1534/genetics.113.160572

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014b). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983. doi: 10.1534/genetics.113.160572

Froese, R., & Pauly, D. (2021). Atlantic horse mackerel. FishBase. Retrieved June 7, 2021, from www.fishbase.org

Gagnaire, P.-A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, *62*(6), 603–616. doi: 10.1093/cz/zow088

Genner, M., & Collins, R. (2022). The genome sequence of the Atlantic horse mackerel, Trachurus trachurus (Linnaeus 1758). *Wellcome Open Research*, *7*, 118. doi: 10.12688/wellcomeopenres.17813.1

Gomes, I., Bobeck, E. N., Margolis, E. B., Gupta, A., Sierra, S., Fakira, A. K., … Devi, L. A. (2016). Identification of GPR83 as the receptor for the neuroendocrine peptide PEN. *Science Signaling*, *9*(425). doi: 10.1126/scisignal.aad0694

Gordo, L. S., Costa, A., Abaunza, P., Lucio, P., Eltink, A. T. G. W., & Figueiredo, I. (2008). Determinate versus indeterminate fecundity in horse mackerel. *Fisheries Research*, *89*(2), 181–185. doi: 10.1016/j.fishres.2007.09.024

Han, F., Jamsandekar, M., Pettersson, M. E., Su, L., Fuentes-Pardo, A. P., Davis, B. W., … Andersson, L. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *ELife*, *9*, 1–20. doi: 10.7554/eLife.61076

Hashiguchi, Y., & Nishida, M. (2007). Evolution of Trace Amine–Associated Receptor (TAAR) Gene Family in Vertebrates: Lineage-Specific Expansions and Degradations of a Second Class of Vertebrate Chemosensory Receptors Expressed in the Olfactory Epithelium. *Molecular Biology and Evolution*, *24*(9), 2099–2107. doi: 10.1093/molbev/msm140

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

Hauser, L., & Carvalho, G. R. (2008b). Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, *9*, 333–362.

Healey, A. J. E., Farthing, M. W., Nunoo, F. K. E., Potts, W. M., Sauer, W. H. H., Skujina, I., … McKeown, N. J. (2020). Genetic analysis provides insights into species distribution and population structure in East Atlantic horse mackerel (Trachurus trachurus and T. capensis). *Journal of Fish Biology*, *96*(3), 795–805. doi: 10.1111/jfb.14276

Hill, J., Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Bekkevold, D., Folkvord, A., … Andersson, L. (2019). Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(37), 18473–18478. doi: 10.1073/pnas.1908332116

Hintzen, N. T., Roel, B., Benden, D., Clarke, M., Egan, A., Nash, R. D. M., … Hatfield, E. M. C. (2015). Managing a complex population structure: exploring the importance of information from fisheries-independent sources. *ICES Journal of Marine Science*, *72*(2), 528–542. doi: 10.1093/icesjms/fsu102

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring Genetic Differentiation from Pool-seq Data. *Genetics*, *210*(1), genetics.300900.2018. doi: 10.1534/genetics.118.300900

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. doi: 10.1093/nar/gkaa942

Hussain, A., Saraiva, L. R., & Korsching, S. I. (2009). Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts. *Proceedings of the National Academy of Sciences*, *106*(11), 4313–4318. doi: 10.1073/pnas.0803229106

ICES. (2005). *Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine, and Anchovy (WGMHSA). 7–16 September 2004, ICES Headquarters, Copenhagen. ICES CM 2005/ACFM:08: 477pp*. Retrieved from https://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acfm/2004/wgmhsa/WGMHSA05.pdf

ICES. (2015). *Report of the Workshop on Maturity Staging of Mackerel and Horse Mackerel (WKMSMAC2), 28 September–2 October 2015, Lisbon, Portugal* (p. 93). p. 93. ICES CM 2015/SSGIEOM.

Kasapidis, P., & Magoulas, A. (2008). Development and application of microsatellite markers to address the population structure of the horse mackerel Trachurus trachurus. *Fisheries Research*, *89*(2), 132–135. doi: 10.1016/j.fishres.2007.09.015

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

Kess, T., Einfeldt, A. L., Wringe, B., Lehnert, S. J., Layton, K. K. S., McBride, M. C., … Bradbury, I. R. (2021). A putative structural variant and environmental variation associated with genomic divergence across the Northwest Atlantic in Atlantic Halibut. *ICES Journal of Marine Science*, *78*(7), 2371–2384. doi: 10.1093/icesjms/fsab061

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., … Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, *25*(10), 2130–2143. doi: 10.1111/mec.13592

Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Nolte, V., Futschik, A., … Schlötterer, C. (2011). PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE*, *6*(1), e15925. doi: 10.1371/journal.pone.0015925

Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic Differentiation Between Temperate and Tropical Australian Populations of Drosophila melanogaster. *Genetics*, *187*(1), 245–260. doi: 10.1534/genetics.110.123059

Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., … Andersson, L. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proceedings of the National Academy of Sciences*, *114*(17), E3452–E3461. doi: 10.1073/pnas.1617728114

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv*, *00*(00), 3. doi: arXiv:1303.3997 [q-bio.GN]

Lin, J.-J., Wang, F.-Y., Li, W.-H., & Wang, T.-Y. (2017). The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. *Scientific Reports*, *7*(1), 15568. doi: 10.1038/s41598-017-15868-7

Lueptow, L. M., Devi, L. A., & Fakira, A. K. (2018). *Targeting the Recently Deorphanized Receptor GPR83 for the Treatment of Immunological, Neuroendocrine and Neuropsychiatric Disorders*. doi: 10.1016/bs.pmbts.2018.07.002

Mariani, S. (2012). *Genetic identification of horse mackerel (Trachurus trachurus) in the North-East Atlantic. Results of the service provided to the Pelagic Freezer-Trawler Association, between 2010 and 2012. Work carried out by the Mariani labs at UCD Dublin and Salford Un*.

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., … Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution*, *6*(4), 469–481. doi: 10.1038/s41559-022-01661-x

Mattiucci, S., Farina, V., Campbell, N., MacKenzie, K., Ramos, P., Pinto, A. L., … Nascetti, G. (2008). Anisakis spp. larvae (Nematoda: Anisakidae) from Atlantic horse mackerel: Their genetic identification and use as biological tags for host stock characterization. *Fisheries Research*, *89*(2), 146–151. doi: 10.1016/j.fishres.2007.09.032

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. doi: 10.1101/gr.107524.110

Musilova, Z., & Cortesi, F. (2021). Multiple ancestral and a plethora of recent gene duplications during the evolution of the green sensitive opsin genes (RH2) in teleost fishes. *BioRxiv 2021.05.11.443711*.

Nakamura, Y., Mori, K., Saitoh, K., Oshima, K., Mekuchi, M., Sugaya, T., … Inouye, K. (2013). Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proceedings of the National Academy of Sciences*, *110*(27), 11061–11066. doi: 10.1073/pnas.1302051110

Ndjaula, H. O. N., Hansen, T., Krüger-Johnsen, M., & Kjesbu, O. S. (2009). Oocyte development in captive Atlantic horse mackerel Trachurus trachurus. *ICES Journal of Marine Science*, *66*(4), 623–630. doi: 10.1093/icesjms/fsp032

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, *25*, 547–572. doi: 10.1146/annurev.es.25.110194.002555

Pastor, F., Valiente, J. A., & Khodayar, S. (2020). A Warming Mediterranean: 38 Years of Increasing Sea Surface Temperature. *Remote Sensing*, *12*(17), 2687. doi: 10.3390/rs12172687

Patarnello, T., Volckaert, F. A. M. J., & Castilho, R. (2007). Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, *16*(21), 4426–4444. doi: 10.1111/j.1365-294X.2007.03477.x

Planas, A., & Vives, F. (1953). Contribución al estudio del jurel (Trachurus tracgurus L.) del Mediterráneo occidental (Sectores de Vinaroz e islas Columbretes). *Publicaciones Del Instituto de Biología Aplicada*, *13*, 155–186.

R Core Development Team. (2021). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raine, A., Manlig, E., Wahlberg, P., Syvänen, A.-C., & Nordlund, J. (2017). SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Research*, *45*(6), e36–e36. doi: 10.1093/nar/gkw1110

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

Rice, W. R. (1989). Analyzing Tables of Statistical Tests. *Evolution*, *43*(1), 223. doi: 10.2307/2409177

Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Rousset, François. (1997). Genetic Differentiation and Estimation of Gene Flow from *F* -Statistics Under Isolation by Distance. *Genetics*, *145*(4), 1219–1228. doi: 10.1093/genetics/145.4.1219

Rusell, F. S. (1976). *The Eggs and Planktonic Stages of British Marine Fishes*. Academic Press, London.

Sala Bozano, M., Mariani, S., Barratt, C. D., Sacchi, C., Boufana, B., & Coscia, I. (2015). Spatio-temporal variability in the population structure in North-east Atlantic stocks of horse mackerel (Trachurus trachurus). *Biology and Environment: Proceedings of the Royal Irish Academy*, *115B*(3), 211. doi: 10.3318/bioe.2015.20

Santos, A. M. P., Chícharo, A., dos Santos, A., Moita, T., Oliveira, P. B., Peliz, Á., & Ré, P. (2007). Physical-biological interactions in the life history of small pelagic fish in the Western Iberia Upwelling Ecosystem. *Progress in Oceanography*, *74*(2–3), 192–209. doi: 10.1016/j.pocean.2007.04.008

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763. doi: 10.1038/nrg3803

Shi, W., & Wang, M. (2010). Characterization of global ocean turbidity from Moderate Resolution Imaging Spectroradiometer ocean color observations. *Journal of Geophysical Research*, *115*(C11), C11022. doi: 10.1029/2010JC006160

Stanley, R. R. E., & Jeffery, N. W. (2017). CartDist: Re-projection tool for complex marine systems.

Stanley, Ryan R. E., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., van Wyngaarden, M., … Bradbury, I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, *4*(3), eaaq0929. doi: 10.1126/sciadv.aaq0929

Stephenson, R. L. (2002). Stock structure management and structure: an ongoing challenge for ICES. *ICES Marine Science Symposium*, *215*, 305–314.

Tessarolo, J. A., Tabesh, M. J., Nesbitt, M., & Davidson, W. S. (2014). Genomic Organization and Evolution of the Trace Amine-Associated Receptor (TAAR) Repertoire in Atlantic Salmon (Salmo salar). *G3 Genes|Genomes|Genetics*, *4*(6), 1135–1141. doi: 10.1534/g3.114.010660

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & de Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*(2). doi: 10.1111/j.1466-8238.2011.00656.x

Vihtakari, M. (2020). ggOceanMaps: Plot Data on Oceanographic Maps using ggplot2. R package version 1.1.10. Retrieved from https://mikkovihtakari.github.io/ggOceanMaps/

Wang, Q., Tan, X., Jiao, S., You, F., & Zhang, P.-J. (2014). Analyzing Cold Tolerance Mechanism in Transgenic Zebrafish (Danio rerio). *PLoS ONE*, *9*(7), e102492. doi: 10.1371/journal.pone.0102492

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358. doi: 10.2307/2408641

White, C. R., Alton, L. A., & Frappell, P. B. (2012). Metabolic cold adaptation in fishes occurs at the level of whole animal, mitochondria and enzyme. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1734), 1740–1747. doi: 10.1098/rspb.2011.2060

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Windisch, H. S., Kathöver, R., Pörtner, H.-O., Frickenhaus, S., & Lucassen, M. (2011). Thermal acclimation in Antarctic fish: transcriptomic profiling of metabolic pathways. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *301*(5), R1453–R1466. doi: 10.1152/ajpregu.00158.2011

Yamamoto, Y., Hino, H., & Ueda, H. (2010). Olfactory Imprinting of Amino Acids in Lacustrine Sockeye Salmon. *PLoS ONE*, *5*(1), e8633. doi: 10.1371/journal.pone.0008633

Data Accessibility

Sequence data are available in NCBI under the bioproject XX, Genbank accession XX

Custom scripts are available in the Github repository: https://github.com/LeifAnderssonLab/2022_Horse_mackerel_popgen.

Author contributions. EDF and LA conceived the study and obtained funds. APF-P performed the bioinformatic data analysis. EDF performed the genotype-based data analysis. MEP contributed to data analysis. CGS performed laboratory work. APF-P, EDF, and LA wrote the paper with contributions from all other authors. All authors approved the manuscript before submission.

Competing interest statement. The authors declare no competing interest.

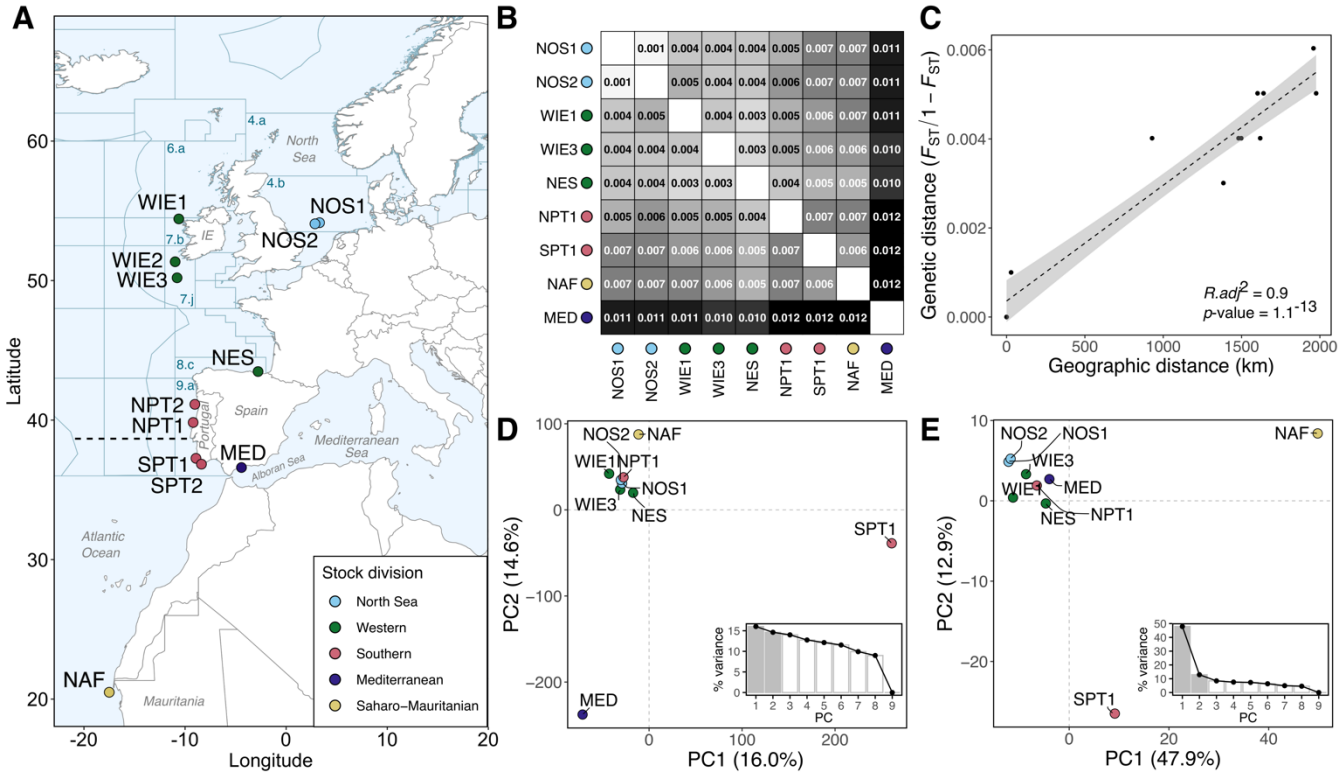

Figure 1. Sampling sites and population structure of the Atlantic horse mackerel. (**A**) Map depicting the 12 sampling sites in the east Atlantic Ocean. ICES fishing divisions are denoted with dark blue lines, same as their alphanumerical code. The approximate location of a biogeographical transition zone north of Lisbon, in Portugal (~38.7-39.0°N), is denoted with a horizontal dashed line (Cunha, 2001; Santos et al., 2007). In all plots, each dot represents a sampling location and its color indicates the corresponding ICES stock division (ICES, 2005) after the HOMSIR project (Abaunza et al., 2008). (**B**) Heatmap representing pairwise pool-F_{ST} values based on ~12 million SNPs. (**C**) Isolation by distance (IBD) plot illustrating the relationship between genetic distances (linearized pairwise pool-F_{ST}) and geographic distances in km for the Atlantic populations in the "northern" group. The dashed line indicates the linear relationship between variables. (**D-E**) Principal component analysis (PCA) plot based on (**D**) undifferentiated (61,543 SNPs) and (**E**) highly differentiated (818 SNPs) markers. The first two axes are shown. Inset bar plots in each PCA plot show the percentage (%) of genetic variance explained by the first nine principal components (PC). Note that samples WIE3, NPT2, and SPT2 were excluded from analyses B to E as they behaved as outliers likely due to technical reasons. Sample names are abbreviated as in Table 1.

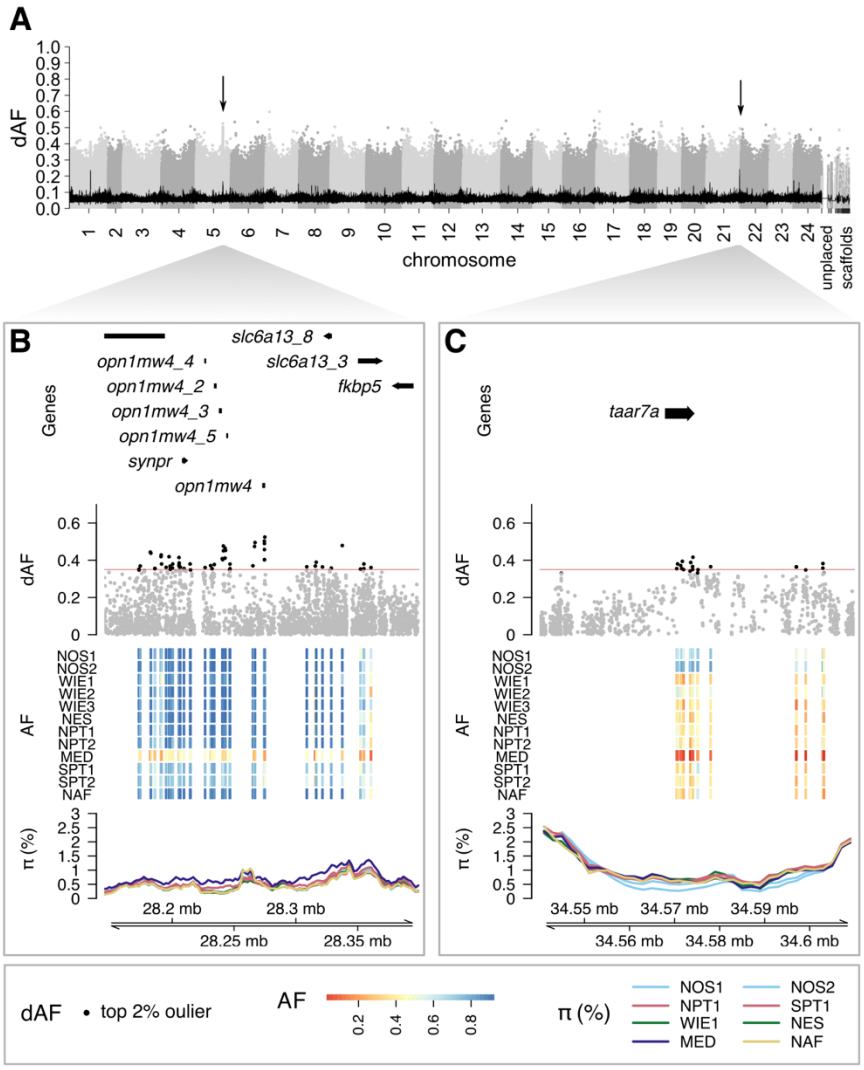

Figure 2. Two selective regions distinctive of the western Mediterranean Sea. (**A**) Manhattan plot depicting the dAF per SNP across the genome for the contrast between the Alboran Sea and all other samples. Each dot is a single SNP, and alternating gray tones were used to differentiate SNPs in consecutive chromosomes. The black line is the rolling average of dAF over 100 SNPs. Regions of interest are indicated with an arrow. (**B-C**) Zoom-in plots for the signals on (**B**) chr 5 and (**C**) chr 21. Each zoom-in plot consists of four tracks: the first, a representation of the gene models; the second, the dAF of SNPs, with the top 2% most differentiated SNPs denoted in black; the third, pool minor allele frequency of the top 2% SNPs in a heatmap plot, in which each row is one sample and each column is one SNP; and the fourth, the percentage of nucleotide diversity (π) calculated in 10 kb windows with 2 kb step size along the region with a separate line for each sample, colored based on the ICES stock divisions. Sample names are abbreviated as in Table 1.

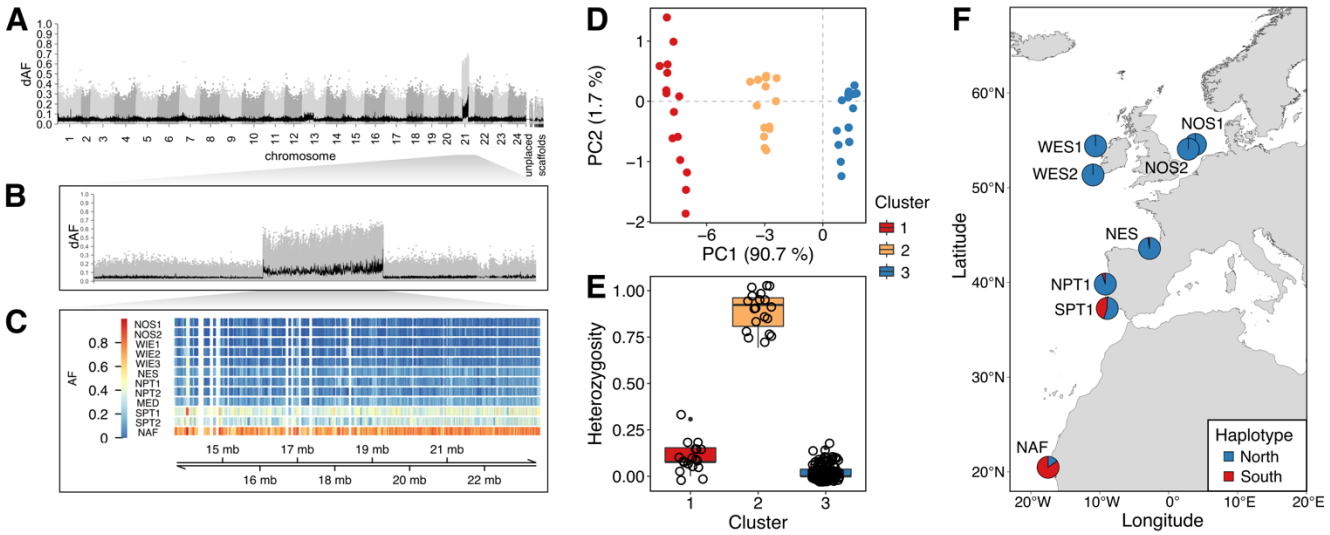

Figure 3. Putative chromosomal inversion on chromosome 21 underlies a latitudinal genetic cline. (**A**) Manhattan plot representing the absolute difference in pool-allele frequencies (dAF) of each SNP along the genome for the contrast between northern vs. southern groups (see Methods for details). The x-axis shows the genomic position of each SNP and the y-axis, its dAF value. SNPs in consecutive chromosomes are distinguished with alternating gray tones. The black line towards the bottom of the plot corresponds to the 100 SNPs-rolling average of dAF values. (**B**) Close-up Manhattan plot to chromosome 21. (**C**) Pool-allele frequency per sample (rows) of the top 2% markers (columns) within the putative inversion. (**D-F**) Analysis of inversion frequency based on the genotype of 15 SNPs screened in 160 individuals. (**D**) PCA plot showing individual clustering respect to the inversion genotype. (**E**) Individual observed heterozygosity in each PCA cluster. Clusters "1" and "3" correspond to homozygous "southern" and "northern" individuals, respectively, while cluster "2" are heterozygous individuals. (**F**) Map showing the geographic distribution of inversion haplotypes across sampled locations. Sample names are abbreviated as in Table 1.

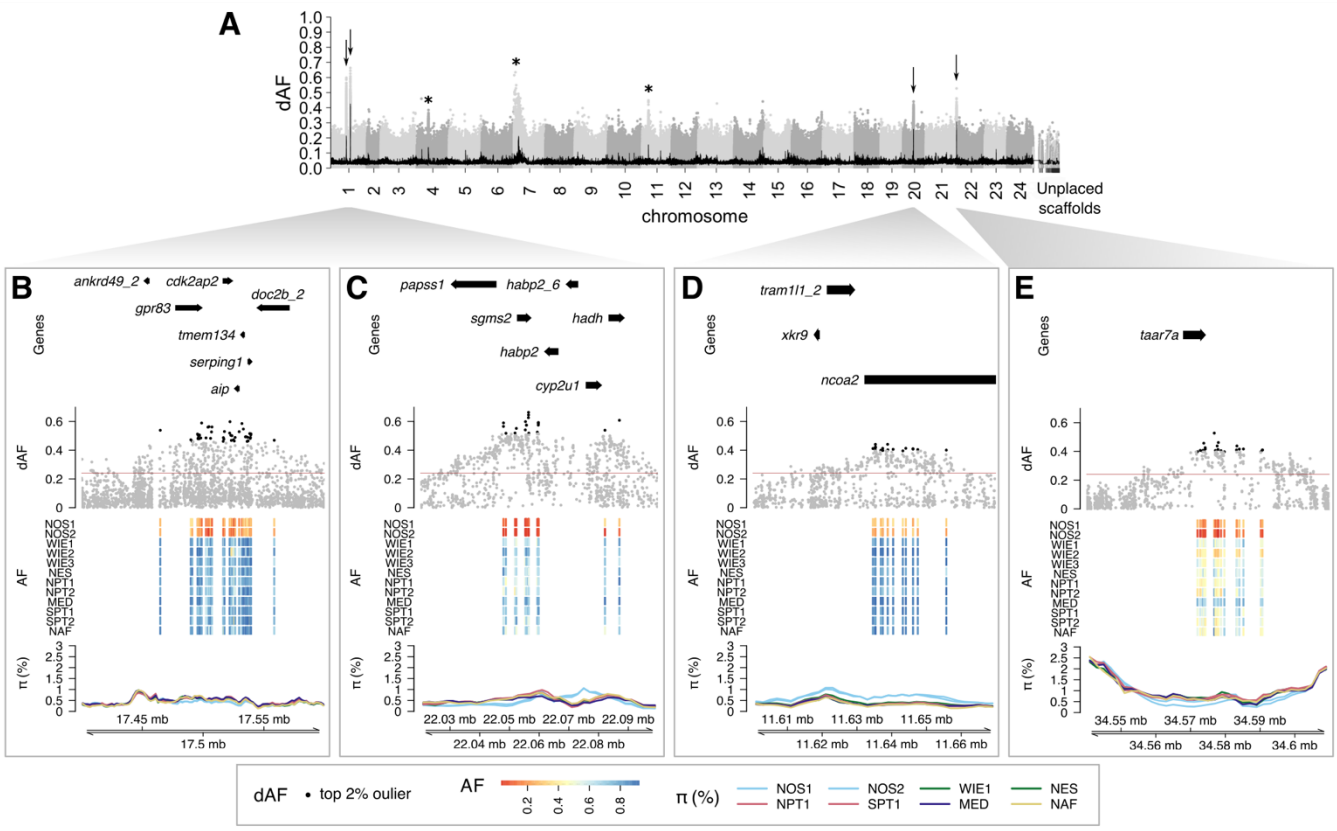

Figure 4. Genomic regions characteristic of the North Sea. (**A**) Manhattan plot representing the dAF of each SNP along the genome for the contrast between the North Sea and all other samples. Each dot is a single SNP, and alternating gray tones were used to differentiate SNPs in consecutive chromosomes. The black line towards the bottom is the rolling mean of dAF over 100 SNPs. (**B-E**) Close-up plots of the four most divergent regions (highlighted with arrows in A) on (**B-C**) chr 1, (**D**) chr 20, and (**E**) chr 21. Plots of the other differentiated regions on chr 4, 7 and 11 are shown in the Supplementary Figure S10 (highlighted with an asterisk in A). Each close-up plot consists of four tracks (from top to bottom): The first, illustrates gene models; the second, corresponds to the dAF of SNPs, in which the top 2% of markers are denoted in black. The horizontal red line indicates the Bonferroni Z-score threshold of significance; the third track, is a heatmap plot depicting the pool-allele frequency of the top 2% SNPs, where the rows are samples and the columns are SNPs; the fourth track, is the percentage of nucleotide diversity (π) for each sample calculated over 10 kb sliding windows with a step size of 2 kb. The color of each line indicates the designated ICES stock division of each pool prior to the HOMSIR project (Abaunza et al., 2008). Sample name abbreviations as in Table 1.

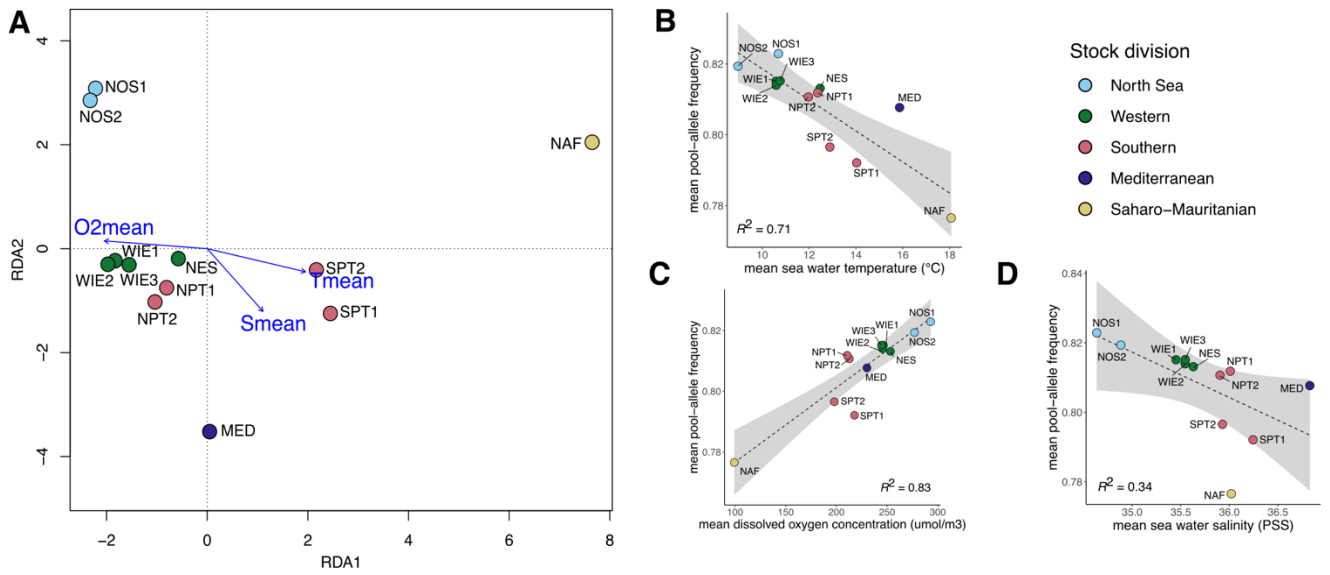

Figure 5. Association between putatively adaptive genetic variants and environmental variables. Analysis based on the pool-allele frequencies of 210,568 SNPs within the most divergent genomic regions in the North Sea and the Mediterranean Sea, and underlying the north-south genetic break. (**A**) Redundancy Analysis (RDA) plot showing the association between pool-allele frequencies of SNPs and three uncorrelated environmental variables: T_{mean}, mean sea water temperature (°C); $O2_{mean}$, mean dissolved oxygen concentration (µmol/m^3); and S_{mean}, mean sea water salinity (PSS). Blue arrows represent loadings of the environmental variables that contribute the most to the first two axes. Each dot represents a single pool sample, and its color indicates its ICES stock division. (**B-D**) Scatter plots showing the linear relationship (dotted line) and 95% confidence interval between average pool-allele frequencies and (**B**) seawater temperature, (**C**) mean dissolved oxygen concentration, and (**D**) mean seawater salinity (PSS). Sample name abbreviations, NOS1-2: southern North Sea, WIE1-3: western Ireland, NPT1-2: northern Portugal, SPT1-2: southern Portugal, MED: Mediterranean, NAF: North Africa.

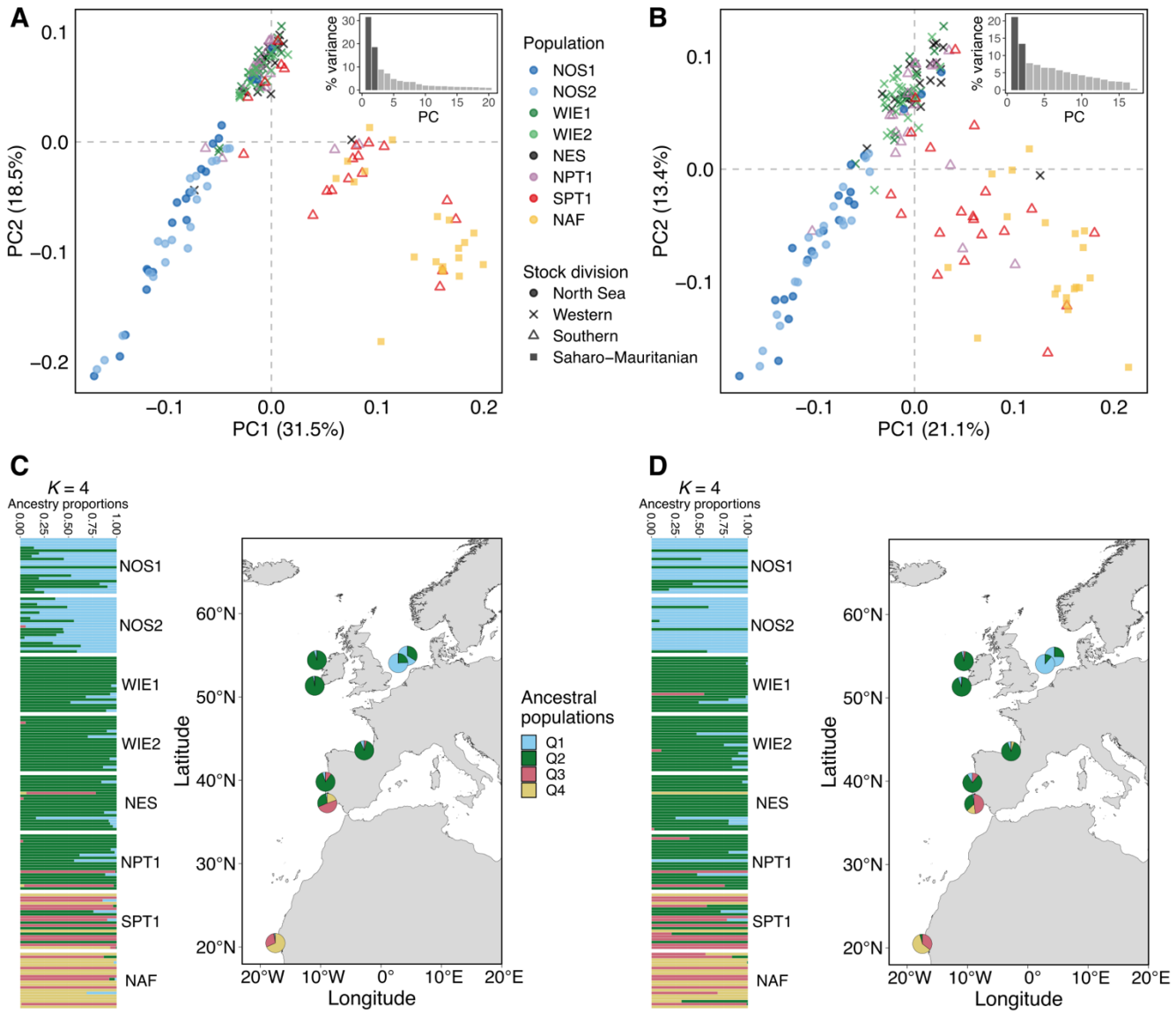

Figure 6. Population structure based on individual genotypes of reduced SNP panels. The two SNP datasets consisted on the genotype of 157 and 155 individuals screened using 63 and 17 markers, respectively. (**A-B**) Principal components analysis plot based on (**A**) the 63-SNP dataset, or (**B**) the 17-SNP dataset. Only the first two axes are shown. Each dot corresponds to an individual, the dot color indicates sampling location and the point shape, the designated stock based on the ICES stock divisions. Inset bar plots show the percentage (%) of genetic variance explained by each of the first nine principal components (PCs). (**C-D**) Admixture bar plot (left) and map showing the mean ancestry proportion per pool sample (right) for a number of ancestral populations (*K*), here equals to four, based on (**C**) the 63-SNP dataset, or (**D**) the 17-SNP dataset. In the admixture bar plot, each row is one individual, and the different colors per individual represent the probability of ancestry for a given *K*. Individuals collected in the same site are grouped as a block. Sample names are abbreviated as in Table 1.

Table 1. Collection details of the 12 Atlantic horse mackerel samples included in this study.

Pool ID	Stock (*)	Area	Year	Latitude	Longitude	Sample size
WIE1	Western	West of Ireland	2016	54.42	-10.62	51
WIE2	Western	Southwest of Ireland	2016	51.35	-10.98	44
WIE3	Western	West of Ireland	2017	52.76	-11.00	62
NOS1	North Sea	Southern North Sea	2016	54.15	3.30	96
NOS2	North Sea	Southern North Sea	2017	54.07	2.85	70
NPT1	Southern	Northern Portugal	2016	39.83	-9.20	64
SPT1	Southern	Southern Portugal	2016	37.26	-8.92	30
NPT2	Southern	Northern Portugal	2017	41.14	-9.03	47
SPT2	Southern	Southern Portugal	2017	36.84	-8.38	48
NAF	Saharo-Mauritanian	North Africa	2016	20.20	-17.50	57
NES	Western	Northern Spanish shelf	2016	43.59	-2.78	96
MED	Mediterranean	Alboran Sea	2018	36.49	-4.42	49

(*) Stock divisions after Abaunza et al. (2008).