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.

Running title: Population genomics of horse mackerel

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 12 Atlantic horse mackerel pool 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 Seas 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 massive chromosomal inversion (9.9 Mb) that underlies a climate-related latitudinal cline with a break near mid Portugal. Genome-environment association analysis indicated that sea water dissolved oxygen concentration and temperature are likely the main environmental drivers of local adaptation. Our genomic data broadly supports the current ICES stock divisions, but recommends revision of the western and southern stock boundaries. We developed a reduced SNP panel that genetically discriminate the North Sea 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.

Keywords

Whole-genome resequencing, genomics, adaptation, management, conservation, marine fish.

1. Introduction

To which extent 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). Resolving this question has been challenging because 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 (Han et al., 2020). Moreover, genomic divergence has been linked to ecological variation, for example, in migratory behavior (Kirubakaran et al., 2016), seasonal reproduction (Lamichhaney et al., 2017; Petrou et al., 2021), or local adaptation along environmental gradients (Fuentes-Pardo et al., 2019; 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 of fisheries resources 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 survey 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 an ideal species 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 often found from the surface to 100 m deep on the continental shelf or near the coast (Alvarez & Chifflet, 2012; Beveren, Klein, Serrão, Gonçalves, & Borges, 2016).

In European waters 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 multidisciplinary 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, they are less studied and monitored than those in the north. The discreteness of the stocks and the location and levels of mixing between them are 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 highly 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 fisheries surveys and from both target and non-target fisheries 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

124 of tissue was excised from the dorsal musculature of each specimen and stored at 4°C in absolute
125 ethanol. Total genomic DNA was extracted from the majority of samples by Weatherbys
126 Scientific Ltd, Ireland from 30 mg of tissue using sbeadex™ magnetic bead-based extraction
127 chemistry on the LGC Oktopure™ platform. The remaining samples were extracted using a
128 CTAB or Chelex and proteinase-K based extraction protocol (Table S1). The Chelex protocol
129 produced single-stranded DNA, whereas the other methods, double-stranded. DNA quantity was
130 measured with a NanoDrop ND-1000 spectrophotometer.
131
132 **2.2 Pool library preparation and sequencing**
133 We performed whole-genome sequencing of pooled DNA (pool-seq) to assess the genomic
134 variation among samples. This method provides population-level allele frequencies by
135 sequencing to a high depth a single-barcoded library prepared from a mixture (pool) of DNA of
136 individuals from a population (Schlötterer, Tobler, Kofler, & Nolte, 2014). DNA pools were
137 prepared by mixing equal amounts of DNA of 30 to 96 individuals collected in close spatial and
138 temporal proximity. DNA pools were quantified using a Qubit Fluorometer (Thermo Fischer
139 Scientific Inc) aiming to have at least 1.5 μg of DNA in 25-50 μL and were submitted to the
140 SNP&SEQ Technology Platform in Uppsala, Sweden, for library preparation and high-
141 throughput sequencing. A PCR-free Illumina TruSeq library with an insert size of 350 base pairs
142 (bp) was prepared for most pools, except for the ones extracted with Chelex, for which a Splinted
143 Ligation Adapter Tagging (SPLAT) library was used instead, as it is aimed for single-stranded
144 DNA (Raine, Manlig, Wahlberg, Syvänen, & Nordlund, 2017). Paired-end short reads (2x150
145 bp) were generated using an Illumina NovaSeq sequencer and S4 flow cells.
146
147 **2.3 Read mapping and variant calling**
148 Low quality bases, sequencing adapters, and short reads were removed from the raw reads using
149 *Trimmomatic* v.0.36 (Bolger, Lohse, & Usadel, 2014). Clean reads were mapped against the
150 *Trachurus trachurus* genome assembly (Accession: GCA_905171665.1, (Genner & Collins,
151 2022)) using *bwa-mem* 0.7.17 (Li, 2013). A pilot examination of read mapping statistics
152 indicated that the temporal replicates from northern and southern Portugal (NPT2, SPT2) might
153 be affected by technical artefacts (Figure S2). Hence, these samples were excluded from

154 genome-wide analyses, but were included in analyses focused on specific loci (e.g., in the
155 examination of allele frequency patterns).
156 Variant calling was performed using the algorithm *UnifiedGenotyper* of *GATK* v3.8
157 (McKenna et al., 2010). Biallelic SNPs were retained and various filters were applied to remove
158 spurious markers (see Supplementary Data for details). The resulting high-quality SNPs were
159 used in further analyses. A summary of the data generation steps is shown in Figure S3.
160
161 **2.4 Population genetic structure and genetic diversity**
162 We assessed population structure with pairwise pool-F_{ST} and principal components analysis
163 (PCA). We computed the pool-F_{ST} ($\hat{F}_{\mathrm{ST}}^{\mathrm{pool}}$) statistic using the R package *poolfstat* (Hivert,
164 Leblois, Petit, Gautier, & Vitalis, 2018), which is equivalent to the Weir & Cockerham (1984)
165 F_{ST} and accounts for random chromosome sampling in pool-seq.
166 To evaluate whether patterns of population structure are better explained by neutral of
167 selective processes, we separately performed PCA on two SNP subsets, one of undifferentiated
168 (assumed neutral) and the other of highly differentiated markers (outliers, assumed selective).
169 These marker sets were chosen based on the empirical distribution of allele frequencies and
170 standard deviation (SD) cut-off values (Figure S4) (see Supplementary Data for details). To
171 reduce redundancy and physical linkage among SNPs, we retained one marker every 1 kb in the
172 undifferentiated marker set, and one marker every 10 kb in the differentiated set, as it is expected
173 that linkage is more pronounced in regions under selection. PCA was performed using the R
174 package *prcomp.* In a pilot analysis one sample from the west of Ireland (WIE2) appeared as an
175 outlier (Figure S5). Since no biological reason can explain this behavior, this sample was
176 removed.
177 To examine the genome-wide variation of genetic diversity in each pool, we calculated
178 nucleotide diversity (π) per pool in 10 kb-sliding windows with a step size of 2 kb using
179 *PoPoolation 1.2.2* (Kofler et al., 2011) (see Supplementary Data for details). Plotting and
180 statistical testing was performed using the R environment (R Core Development Team, 2021).
181
182 Additionally, we evaluated whether population structure followed an isolation-by-
183 distance pattern with a Mantel test and 9999 permutations using the R package *ade4* (Dray &
184 Dufour, 2007). We compared the linearized genetic distances, calculated with the formula

185 linearized-$\hat{F}_{ST}^{pool} = \frac{\hat{F}_{ST}^{pool}}{1-\hat{F}_{ST}^{pool}}$ (François Rousset, 1997), and the geographic distances, estimated as
186 the least-coast oceanic distance in kilometers (km) considering land as barrier using the R
187 package *CartDist* (R. R. E. Stanley & Jeffery, 2017).
188
189 **2.5 Detection of loci under selection**
190 We applied the effective coverage correction (n_{eff}) to raw read counts to account for random
191 variation of read coverage and chromosome sampling during pooling and sequencing (Bergland,
192 Behrman, O'Brien, Schmidt, & Petrov, 2014; Feder, Petrov, & Bergland, 2012; Kolaczkowski,
193 Kern, Holloway, & Begun, 2011) (refer to Supplementary Data for details). Pool allele
194 frequencies were then computed using a python script.
195 To identify genomic regions with elevated differentiation respect to the genomic
196 background, we calculated the absolute delta allele frequency (dAF) per SNP between paired
197 contrasts of grouped pools, as *dAF = absolute(meanAF(group1) – meanAF(group2)*. The
198 contrasts evaluated were established based on geographic closeness, PCA clustering, and
199 biological knowledge (see details in Table S3). We also calculated the moving (rolling) average
200 of dAF values in 100-SNP windows to identify regions with consistent differentiation across
201 nearby markers, while ruling out single SNPs that could be influenced by random effects of pool-
202 seq. We further explored the allele frequency patterns of the most differentiated SNPs at each
203 locus and contrast for all the 12 pools. All the analyses were performed using R and plotting was
204 done with the R package *ggplot2* (Wickham, 2016).
205
206 **2.6 Validation of informative markers for genetic stock assessment**
207 To identify a reduced panel of highly informative SNPs for genetic stock identification and to
208 validate pool-seq findings, we obtained the genotypes of 160 individuals (20 fish each from eight
209 locations) in 100 SNPs (Table S4). The 100-SNPs panel consisted of 24 neutral markers and 76
210 putatively adaptive markers (see Supplementary Data for details). The final split of adaptive
211 markers was: North Sea (n=28), north-south break (n=13), west of Ireland (n=14), Alboran Sea
212 (n=13), southern Portugal (n=4), north Africa (n=4). Three to four individuals per location were
213 genotyped twice to assess genotyping error rate. DNA extraction and SNP genotyping were
214 undertaken by IdentiGEN, Ireland, using their IdentiSNP genotyping assay chemistry. The

215 protocol utilises target specific primers and universal hydrolysis probes. Following an end-point
216 PCR reaction, different genotypes are detected using a fluorescence reader.
217 Based on individual allele frequencies, we undertook a preliminary analysis of population
218 structure among the eight genotyped fish aggregations. It should be noted that sample sizes were
219 small and therefore the results of the population analyses should be viewed as preliminary until
220 further large-scale screening is undertaken. Only individuals and markers with >80% genotyping
221 success were retained. Deviations from Hardy–Weinberg equilibrium (HWE) and linkage
222 disequilibrium (LD) were assessed with *Genepop* 4.2 (Rousset, 2008). Six SNPs had indication
223 of deviation from HWE, two markers (12_3119866 and 17_972744) were not polymorphic and
224 one had evident scoring errors (24_5252083), thus these nine markers were excluded.
225 *Microsatellite Analyzer* (*MSA*) 4.05 was used to calculate pairwise F_{ST} estimates (Dieringer &
226 Schlötterer, 2003). In all cases with multiple tests, significance levels were adjusted using the
227 sequential Bonferroni technique (Rice, 1989). PCA was performed using the R function *prcomp*.
228 We estimated individual admixture coefficients for a given number of ancestral source
229 populations (K) using the *sNMF* algorithm (Frichot, Mathieu, Trouillon, Bouchard, & François,
230 2014a) of the R package *LEA* (Frichot, Francois, & Caye, 2017). We tested K = 1 to 5, with 10
231 repetitions and 200 iterations. The most likely K corresponds to the value where the cross-
232 entropy criterion (metric that evaluates the error of the ancestry prediction) plateaus or increases
233 (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014b). We plotted the average admixture
234 proportions per population sample over a map using the R package *ggplot* (Wickham, 2016) and
235 *ggOceanMaps* (Vihtakari, 2020).
236
237 **2.7 Characterization of a putative inversion on chromosome 21**
238 To assess the genetic diversity and spatial distribution of haplotypes of the putative inversion on
239 chromosome (chr) 21, we extracted the individual genotypes of 12 diagnostic SNPs within the
240 inversion from the 100-SNP dataset (Figure S11). To identify the inversion genotype of each
241 individual, we performed a PCA with the R function *prcomp*. Individuals were assigned to each
242 haplotype group using the first two eigenvectors of the PCA and the k-means clustering
243 algorithm implemented in the R function *kmeans*. We calculated observed heterozygosity for
244 each of the PCA clusters, with the expectation that the middle cluster will have the highest
245 heterozygosity of all. These analyses and correspondent graphics were performed using R.

246

247 **2.8 Genome-Environment Association**

248 To identify which environmental variables are strongly correlated with adaptive genetic variation
249 and local adaptation, we tested genome-environment association using redundancy analysis
250 (RDA). We retrieved environmental data from *Bio-Oracle* v.2.1 (Assis et al., 2018; Tyberghein
251 et al., 2012) for each sampled location using the *R* package *sdmpredictors* (Bosch, 2020). Data
252 corresponded to mean bottom depth layers of minimum, mean, maximum, and range values of
253 sea water temperature (°C), sea water salinity (PSS), and dissolved oxygen concentration
254 (μmol/m^3) for a 14-year period (from 2000 to 2014) and a spatial resolution of 0.25 arcdegree.
255 This data was derived from a numerical model developed by the Global Ocean Biogeochemistry
256 non assimilative Hindcast (PISCES) (Copernicus Programme of the European Union, n.d.).

257 Environmental data per location were standardized to zero mean and unit variance, and
258 collinear variables were removed (see Supplementary Data for details). To run RDA, we used the
259 R package *vegan* (Dixon, 2003), the non-redundant environmental dataset, and the pool-allele
260 frequencies of SNPs in outlier loci found with genome scans. The significance of the RDA
261 model and environmental variables was assessed with an analysis of variance (ANOVA) using
262 1000 permutations.

263

264 **2.9 Functional annotation of gene models**

265 The gene models of the Atlantic horse mackerel genome were developed by Ensembl (Howe et
266 al., 2021) and become available in an Ensembl Rapid release since March 2021 (Ensembl,
267 2021b). However, at the time this research was conducted, the gene models lacked gene symbols
268 (names) and GO terms (only Ensembl gene IDs were available). Given that these annotations are
269 relevant for interpretation of results, we ran the functional gene annotation pipeline developed by
270 the National Bioinformatics Infrastructure Sweden (NBIS) (Binzer-Panchal, Dainat, & Soler,
271 2021) to retrieve this information (see Supplementary Data for details).

272 Additionally, for the top 2% most differentiated SNPs within each divergent genomic
273 region detected with genome scans, we annotated the closest overlapping gene (up to ± 40 kb)
274 and the variant effect prediction (e.g., missense, synonymous, upstream, downstream, intergenic)
275 using snpEff v.4.1 (Cingolani et al., 2012).

276

3. Results

3.1 Sequencing and variant calling

We generated pooled DNA whole-genome sequence data of 12 Atlantic horse mackerel aggregations sampled across the species' range in the east Atlantic and the western Mediterranean Sea, Alboran Sea (Figure 1A, Table 1). Temporal replicates of four locations were collected one year apart, so the temporal stability of their genetic composition could be investigated (west of Ireland WIE1-WIE2, North Sea NOS1-NOS2, northern Portugal NPT1-NPT2, and southern Portugal SPT1- SPT2). Our sequencing effort yielded 490-764 million high-quality reads per pool. 98.2-99.2% of reads aligned to the chromosome-level genome assembly of *T. trachurus* (Genner & Collins, 2022), and each pool had a mean depth of coverage between 25.7x and 46.3x (Table S5). After variant calling with GATK, a total of ~12.8 million biallelic SNPs passed quality filters and were used in the population analysis.

3.2 Population genetic structure

Pairwise pool-F_{ST} estimates indicated low genome-wide differentiation among horse mackerel populations (global mean pool-F_{ST}=0.007, range 0.001-0.012) (Figure 1B), but revealed four subtle population structure patterns that were also supported by genome scans. First, the western Mediterranean Sea (MED) differentiates from Atlantic populations (mean pool-F_{ST}=0.011). Second, Atlantic populations differ along a latitudinal cline that distinguishes "northern" or "southern" groups respect to a pronounced genetic break near Lisbon, Portugal (~38.7-39.0°N). Third, populations within the "northern" group (North Sea–NOS1, NOS2, west of Ireland–WIE1, WIE3, northern Spanish Shelf–NES, northern Portugal–NPT1) are genetically more similar to each other (mean pool-F_{ST}=0.004) than populations within the "southern" group (southern Portugal–SPT1 and north Africa–NAF) (mean pool-F_{ST}=0.006). Forth, the two samples from the North Sea (NOS1 and NOS2) show the highest genetic similarity (pool-F_{ST}=0.001) of all and are distinguished from other "northern" populations.

While genetic differences within the "northern" group are small, they are not negligeable (pool-F_{ST} range 0.001-0.006) (Figure S8). A significant Mantel test assessing the degree of correlation between genetic and geographic distances (H_o: no association) between pairs of populations, indicated that subtle genetic differences within the northern group follow an

308 isolation-by-distance (IBD) pattern, meaning that they are shaped by effective gene flow
309 between neighbouring areas (Mantel test *p*-value=0.015; linear regression $R.adjusted^2$=0.9, *p*-
310 value=1.1^{-13}) (Figure 1C).
311
312 We further examined whether population structure patterns were primarily driven by
313 neutral or selection processes by separately performing PCA for two SNP subsets, of
314 undifferentiated or highly differentiated markers. These marker sets were chosen based on the
315 standard deviation in allele frequencies across populations (see Methods for details, Figure S4).
316 The undifferentiated markers were considered "neutral" (61,543 SNPs) (Figure 1D, S8A), and
317 the highly differentiated markers were assumed under selection (818 SNPs) (Figure 1E, S8B). In
318 both PCA plots the separation between the "northern" and "southern" samples was evident,
319 although the composition of the clusters slightly varied depending on the marker set. In both
320 cases the "northern" group appeared as a tight cluster of genetically similar samples mainly from
321 the North Sea, west of Ireland, the northern Spanish shelf, and northern Portugal. In stark
322 contrast, the "southern" samples were dissimilar to each other. In the PCA with neutral markers,
323 the first two axis explained 30.6% of the genetic variance. The mid-range samples from the
324 Mediterranean Sea (MED) and southern Portugal (SPT1) were the most genetically distinct,
325 while the north Africa (NAF) sample was genetically similar to the "northern" group. The North
326 Sea samples tightly clustered with others in the "northern" group, being almost
327 undistinguishable. In contrast, in the PCA with selective makers, the first two axis explained
328 twice as much genetic variance (60.8%). The Mediterranean Sea sample was genetically closer
329 to the "northern" group, while the other two "southern" samples, from southern Portugal and
330 north Africa, were the most differentiated. Notably, the two North Sea samples formed a distinct
331 cluster separated from other "northern" samples.
332
333 Taken together, these results indicate that the separation between the westernmost part of
334 the Mediterranean Sea and Atlantic populations is largely driven by neutral processes, while the
335 latitudinal genetic cline and separation of North Sea samples could be the result of selective
336 processes.
337
338 **3.3 Putative loci under selection**

339 We performed whole genome scans based on the absolute difference in allele frequencies (dAF)
340 for paired contrasts to identify outlier loci respect to the genomic background (presumed under
341 selection). The contrasts were chosen based on PCA clustering patterns (Figure 1D, 1E), and the
342 outlier regions were identified using a Bonferroni Z-score threshold of significance. Our analysis
343 revealed a number of genomic regions with elevated differentiation for three contrasts: (i) the
344 Mediterranean Sea vs. others, (ii) "northern" vs. "southern" groups, and (iii) the North Sea vs.
345 others (Figure 2-4). A summary of the top 2% most differentiated SNPs per region, their closest
346 gene, gene function, and relative position to genes (e.g., missense, synonymous, upstream,
347 downstream, intergenic) inferred with snpEff are compiled in Table S6.
348
349 **Two outlier loci distinguish the western Mediterranean from Atlantic populations**
350 The contrast between Mediterranean Sea and other samples uncovered two distinctive regions at
351 chromosomes (chr) 5 and 21 (Figure 2). In this case the signals of divergence were not as evident
352 as in other contrasts given the overall higher genome-wide differentiation of this sample. Thus,
353 we focused on the peaks revealed by the rolling average of dAF values with consistent allele
354 frequency differences for the contrast of interest. We considered candidate genes those
355 overlapping or located in the vicinity of the top 2% most differentiated SNPs in each locus. In
356 the region on chr 5 (Figure 2B), the Mediterranean Sea sample had intermediate allele
357 frequencies across loci and higher nucleotide diversity, while other samples tended to be fixed
358 for one haplotype and showed lower nucleotide diversity. The top candidate gene in this region is
359 *opn1mw4* (*green-sensitive opsin-4*). In the putative selection signal on chr 21, the Mediterranean
360 sample had one haplotype close to fixation, the opposite haplotype was in high frequency in the
361 North Sea, and intermediate frequencies were prevalent elsewhere (Figure 2C). This region
362 harbours a single gene, *taar7a* (*trace amine-associated receptor 7A*). The current knowledge of
363 potential function of candidate genes is presented in Table S6.
364
365 **A presumed 9.9 Mb inversion distinguishes populations along a latitudinal cline**
366 With the "northern" vs. "southern" contrast, we discovered that a single large locus on chr 21
367 underpins the latitudinal cline with a genetic break along mid Portugal (Figure 3). This locus
368 consists of a large block of SNPs with elevated dAF spanning 9.9 Mb (Figure 3A-B). The large
369 size and abrupt drop in allele frequency differences towards the edges are characteristic of

370 structural variants with suppressed recombination (e.g., inversions) that do not represent recent
371 selective sweeps (Han et al. 2020). An analysis of the pool-allele frequencies of the top 2% most
372 differentiated SNPs at this locus (Figure 3C) revealed striking allele frequency differences
373 between the "northern" group (including the Mediterranean Sea) and north Africa, while
374 intermediate allele frequencies were prevalent in southern Portugal.
375 To further characterize this structural variant, we leveraged genotype data of 160
376 individuals for 12 SNPs within the region. As expected for a chromosomal inversion, in a PCA
377 plot individuals separated into three clusters depending on their genotype (Figure 3D), and
378 individuals in the intermediate cluster exhibited the highest average observed heterozygosity
379 compared with the other two (Figure 3E). Overall, the individual genotype data supported the
380 observations made with pool-seq data regarding the spatial distribution of inversion haplotypes,
381 although it offered greater resolution. For instance, it revealed that the "northern" haplotype
382 dominates in the North Sea, west of Ireland, northern Spanish shelf, and northern Portugal, but
383 also occurs at an intermediate frequency in southern Portugal. The "southern" haplotype is
384 predominant in north Africa, and common in southern Portugal (Figure 1F). This putative
385 inversion contains about 1,077 genes, making it difficult to infer which particular genes are
386 under selection.
387
388 **A genetically distinct population in the North Sea**
389 The contrast between the North Sea and other samples revealed seven genomic regions with
390 elevated differentiation on chr 1, 4, 7, 11, 20, and 21 (Figure 4A). Further examination of allele
391 frequencies of top SNPs per locus indicated that the North Sea temporal replicates were very
392 genetically similar, and that their allele frequency patterns were distinctive of this location
393 (Figure 4B-E, Figure S9). For instance, in the selection signals on chr 1 and 21, North Sea
394 samples had variant alleles close to fixation and a slight reduction in nucleotide diversity (π),
395 suggesting the possibility of a soft selective sweep (Figure 4B, 4C, 4E). At the outlier loci on chr
396 4, 11, and 20 (Figure S9B, S9D, and Figure 4E, respectively), North Sea samples had
397 predominantly intermediate allele frequencies, with similar (chr 4 and 11, Figure S9B, S9D) or
398 higher (chr 20, Figure 4D) nucleotide diversity with respect to other pools. Additional divergent
399 regions on chr 7 and 11 showed a similar but noisier pattern, presumably due to the presence of
400 complex structural variants at these loci (Figure S9C, S9D).

401

402 The most striking signals were located on chr 1, followed by those on chr 20 and 21
403 (Figure 4). The signal on chr 1 encompasses two regions (chr1:17400000-17600000,
404 chr1:22020000-22100000). In the first region (Figure 4B), the most differentiated SNPs are in
405 the vicinity of the gene *gpr83* (G-protein coupled receptor 83). In the second region, the top
406 candidate gene is *sgms2* (*sphingomyelin synthase 2*) (Figure 4C). The signals on chr 20 (Figure
407 4D) and chr 21 (Figure 4E) contain a single gene each, *ncoa2* (*nuclear receptor coactivator 2*)
408 and *taar7a* (*trace amine-associated receptor 7A*). Note that the signal in chr 21 is the same as
409 one previously described distinctive in the Mediterranean. Additional signals were found on chr
410 4, 7 and 11 (Figure S9), however they were not as clear as the ones just described (e.g., smaller
411 allele frequency differences or inconsistency allele pattern). Therefore, it was difficult to identify
412 candidate genes for these regions. The current knowledge concerning the function of candidate
413 genes in the differentiated genomic regions is summarized in the Table S6.
414

415 **3.4 Genome-Environment Association**

416 To identify the environmental variables likely driving adaptive genetic variation, we retrieved
417 sea water environmental variables (temperature, dissolved oxygen, and salinity) and performed
418 genome–environment association (GEA) based on redundancy analysis (RDA). A total of
419 210,568 SNPs within the most divergent genomic regions in the North Sea, Mediterranean Sea,
420 and northern vs. southern contrasts were used in this analysis.

421 Mean sea water dissolved oxygen concentration has the strongest association with genetic
422 variation ($R^2 = 0.83$), followed by mean sea water temperature ($R^2 = 0.71$) (Figure 5) and, to a
423 lesser extent, with salinity ($R^2 = 0.34$). Notably, the latitudinal gradient shows an inverse
424 relationship with oxygen content, with the North Sea being the location with the highest oxygen
425 levels (~270-300 μmol/m^3), almost three times the oxygen content in north Africa (~100
426 μmol/m^3). The two southernmost populations studied (in north Africa and southern Portugal)—
427 which showed a high and intermediate frequency of the "southern" inversion haplotype,
428 respectively—are characterized by higher mean temperatures (18°C), with respect to locations
429 north of Lisbon, in which the "northern" haplotype dominates. The North Sea has the coldest
430 temperatures of all locations (< 10°C). In terms of salinity, as expected, the Mediterranean Sea
431 has the highest salinity followed by southern Portugal, while the lowest salinity is observed in

432 the North Sea. That being said, the salinity range between the northern- and southernmost
433 locations here analysed is minimal, between 35.0 and 36.5 PSS, which is considered a fully
434 marine environment.
435
436 **3.5 Validation of informative markers for genetic stock identification**
437 To confirm the selection signals found with pool-seq data and develop a tool for genetic stock
438 identification, we genotyped 160 individuals in 100 SNPs. Of these, 24 markers were neutral and
439 76 were among the most differentiated across contrasts (Figure S11). We found a strong
440 correlation between allele frequencies calculated from individual genotypes and pool-seq data
441 (mean $R^2 = 0.9 \pm 0.1$), supporting the findings of the pool-seq analysis (Figure S12). Of the 76
442 outlier loci, 72 had a genotyping success >80% (Table S7). After applying quality filters, the
443 final dataset had 63 SNPs and 157 out of 160 individuals had a genotyping success >80%
444 (dataset called the 63-SNP panel).
445
446 To minimise marker redundancy in the 63-SNP panel, we performed a linkage
447 disequilibrium (LD) analysis for all loci and samples. As expected, significant LD was found
448 between a number of SNPs located in close proximity on the same chromosome (Table S7).
449 Though LD was not statistically significant in some cases (e.g., SNPs on chr 5), these were
450 considered to be linked due to physical closeness. To identify the most informative SNPs for
451 discriminating the samples while reducing LD, we analyzed the F_{ST} by marker and by population
452 (Figure S13) for each genomic region. We retained the SNP with the highest average F_{ST} per
453 linkage group (assumed to be the most informative), yielding a 17 SNP dataset comprising 155
454 out of 160 individuals with a genotyping success >80% (dataset called the 17-SNP panel).
455 Further analyses were conducted with both the 63-SNP and the 17-SNP datasets (individual
456 genotypes are shown in Figure S14).
457 An examination of pairwise F_{ST} values indicated lack of significant genetic differences
458 between the North Sea temporal replicates or between the two west of Ireland samples (Table
459 S8). There was also no significant genetic differentiation between the individuals from the
460 northern Spanish shelf, northern Portugal, and the two samples from the west of Ireland (Table
461 S8).
462

463 To assess whether the same population structure is achieved with the 63- and 17-SNP
464 panels, we performed separate PCA and admixture analyses. We observed similar clustering
465 patterns with both panels (Figure 6A, 6B), consisting of four main groups: (1) North Sea (NOS1,
466 NOS2); (2) west of Ireland (WIE1, WIE2), northern Spanish shelf (NES), and northern Portugal
467 (NPT1); (3) southern Portugal (SPT1); and (4) north Africa (NAF). While the 63-SNP panel
468 provided better discrimination between southern Portugal (SPT1) and north Africa (NAF), both
469 marker sets separated most individuals from the North Sea, as well as the "northern" and
470 "southern" groups. It should be noted that some individuals clustered in different groups from
471 the ones expected, suggesting that gene flow occurs among geographic regions.
472 An individual admixture analysis with both datasets supported the same four major
473 genetic groups (Figure 6C, 6D) as the ones identified with PCA, showing various levels of
474 mixing among them. For example, a few individuals in the North Sea (n = 3-4) had a high
475 probability of originating from the western group (first-generation migrants). Similarly, southern
476 Portugal had a number of individuals that appear to originate from the western group (n = 3) or
477 from the African group (n = 3). In all four groups some individuals showed admixed ancestry,
478 indicating that they are probably F1-hybrids or backcrosses between local and migrants. Overall,
479 these results indicate that gene flow occurs more often between neighboring geographic areas,
480 following an isolation-by-distance pattern (e.g., between the North Sea and the west of Ireland;
481 between the northern Spanish shelf, northern Portugal and southern of Portugal; and between
482 southern Portugal and north Africa).
483
484 **4. Discussion**
485 This study represents the most comprehensive assessment of the genomic variation of Atlantic
486 horse mackerel populations to date (Figure 1). The combination of extensive geographic
487 sampling and genomic screening derived from whole-genome sequencing, provided a powerful
488 dataset that allowed us to discover genomic regions underlying population structure and local
489 adaptation. Moreover, we developed a genetic tool that can help distinguish some of the main
490 biological groups found.
491
492 ***Population structure and temporal stability of genetic groups***

493 Our genomic analysis indicated that very few genetic differences exist between horse mackerel
494 populations inhabiting the vast area from the North Sea to north Africa (Figure 1A), more than
495 4,000 km of coastline (global mean pool-F_{ST}=0.007, range: 0.001-0.012). This result is in
496 agreement with previous studies using a few genetic markers (Cimmaruta et al., 2008; Comesaña
497 et al., 2008; Healey et al., 2020; Kasapidis & Magoulas, 2008), and implies that gene flow is
498 pervasive. Low genetic differentiation can be explained by minimal barriers to gene flow and
499 large population sizes, commonly observed in migratory marine fish. We infer that in the horse
500 mackerel, since the pelagic larval stage lasts about one month (Rusell, 1976), gene flow could
501 predominantly occur between neighbouring areas through passive transport of eggs and larvae by
502 ocean currents. Moreover, the large population sizes of this species imply a minimal role of
503 genetic drift in shaping patterns of genetic diversity.
504
505 Despite low differentiation at neutral loci, we uncovered patterns of population structure
506 (Figure 1B, D, E) mainly at loci putatively under selection (Figures 2-4). Below, we will touch
507 on the summary statistics results, and in the next section we will discuss the genome scan results.
508 The greatest genetic differences found occurred primarily at neutral markers between the
509 Alboran Sea (the westernmost part of the Mediterranean) and the Atlantic. This separation was
510 already proposed in earlier studies using body morphometrics, otolith shape, and parasitofauna
511 (Abaunza et al., 2008). However, this is the first time that genetic data supports it, as previous
512 studies using a few markers were inconclusive (Cimmaruta et al., 2008; Comesaña et al., 2008;
513 Healey et al., 2020; Kasapidis & Magoulas, 2008). The Mediterranean-Atlantic divide has also
514 been described in several marine species but not in all (Patarnello, Volckaert, & Castilho, 2007),
515 likely because of differences in species' life history or low resolution of the genetic markers
516 used. Previous research based on parasite composition indicated that the Alboran Sea is a mixing
517 area of horse mackerel from the Mediterranean and the Atlantic (Abaunza et al., 2008; Mattiucci
518 et al., 2008). If this were the case, we would expect to see both inversion haplotypes at this
519 location, but what we observed is that the northern haplotype prevails. This, in addition to the
520 fact that none of the individuals collected were actively spawning (Table S1), let us infer that our
521 sample is probably not representative of this area, and likely comes from a region with colder
522 waters (e.g. Gulf of Lyon, Balearic Sea) (Figure S10). A possible neutral mechanism for reduced
523 gene flow between Atlantic and western Mediterranean populations is the presence of retentive

524 currents in the Almeria-Oran front, which have been proposed to act as barriers for gene flow for
525 various marine species (Patarnello et al., 2007).
526
527 Our genomic data also revealed an undescribed latitudinal genetic break in mid Portugal,
528 near Lisbon (~38.7-39.0°N), distinguishing populations from the north or south of this area. This
529 genetic separation was noticeable in pairwise-F_{ST} estimates, but it was more evident in the PCA
530 based on outlier SNPs, suggesting a putative role of natural selection. Interestingly, the inferred
531 location of the north-south genetic break coincides with a major biogeographical transition zone
532 between temperate and subtropical waters off the coast of Portugal, which have differential
533 species composition (Cunha, 2001; Santos et al., 2007). Moreover, a recent genetic study in the
534 boarfish (*Capros aper*), a pelagic fish with similar distribution and life history as the horse
535 mackerel, reported a comparable latitudinal pattern (Farrell, Carlsson, & Carlsson, 2016). Thus,
536 it is possible that the environmental transition zone in Portugal is a major driver of population
537 structuring of many marine species inhabiting this area, including the horse mackerel.
538 Within the "northern" group—North Sea, west of Ireland, the northern Spanish shelf
539 (Bay of Biscay), and northern Portugal—genetic differences are small at both, neutral and
540 selective markers (Figure 1B, 1D). Such genetic variability resembles an isolation-by-distance
541 pattern (Mantel test p-value=0.015), which was also observed in a previous study with allozymes
542 (Cimmaruta et al., 2008). Therefore, genetic differences between populations within the
543 "northern" group seem to be largely determined by the effective gene flow via passive larval
544 dispersal by ocean currents. The North Sea samples though, were slightly more differentiated at
545 outlier SNPs (Figure 1E), suggesting a putative role of natural selection.
546 In contrast, populations within the "southern" group—inhabiting waters in southern
547 Portugal and north Africa, Mauritania—are genetically more differentiated between them and
548 from all other populations. Interestingly, the population in the south of Portugal was different at
549 both, neutral and selective markers, whereas the population from north Africa was more
550 genetically distinct at outlier markers. We will discuss these sub-structuring patterns in more
551 detail in the genome scan section.
552
553 In summary, our genomic data indicates five possible genetic groups: (1) the western
554 Mediterranean Sea; (2) the North Sea; (3) the west of Ireland, northern Spanish shelf (Bay of

555 Biscay), and northern Portugal; (4) southern Portugal; and (5) north Africa, Mauritania. These
556 groups seem to be temporally stable, as indicated by the similar allele frequencies at outlier loci
557 observed between temporal replicates (west of Ireland WIE1-WIE2, North Sea NOS1-NOS2,
558 northern Portugal NPT1-NPT2, and southern Portugal SPT1- SPT2) (Figures 2-4).
559
560 ***Genomic regions and environmental factors involved in local adaptation***
561 Genome scans uncovered a number of regions with elevated differentiation that support the
562 subtle patterns of population sub-structuring identified with summary statistics. The informative
563 contrasts were: (i) the Mediterranean Sea vs. others, (ii) "northern" vs. "southern" groups, (iii)
564 the North Sea vs. others. The divergent genomic regions varied in size (70-600 kb, and 9.9 Mb
565 for a structural variant) and harboured numerous candidate genes.
566 The test for genome-environment association (GEA) indicated that genetic differences at
567 putatively adaptive loci are primarily driven by differences in sea water oxygen content and
568 temperature between locations (R^2=0.83 and R^2=0.71, respectively) (Figure 5). Salinity seemed
569 to be less important, reflecting the minute differences in salinity between areas (R^2=0.34, range
570 35-36.5 PSS). Notably, there was an inverse relationship between sea water oxygen, temperature,
571 and the latitudinal cline, where the North Sea had the highest oxygen concentration (~270-300
572 μmol/m^3), almost three times the oxygen content in north Africa (~100 μmol/m^3). Moreover, the
573 North Sea also exhibited the coldest sea water temperature of all locations analyzed (< 10°C),
574 strikingly different from the two southernmost populations sampled (north Africa and south
575 Portugal) (18°C).
576
577 We found that two outlier loci distinguish the western Mediterranean Sea, one on chr 5
578 and the other on chr 21 (Figure 2, Figure S10). The signal on chr 21 harbours a single gene,
579 *taar7a* (*trace amine-associated receptor 7A*), which encodes a receptor involved in olfactory
580 sensing of amines (Hashiguchi & Nishida, 2007). The signal on chr 5 includes three genes that
581 encode proteins with crucial roles in synapsis (*cadps, synpr, slc6a13*), and five in tandem copies
582 of the green-sensitive opsin-4 (*opn1mw4*) gene. This gene organization (*opn1mw4* copies flanked
583 by *synpr* and *slc6a13*) is highly conserved among several Clupeocephala species (e.g., Atlantic
584 herring, zebrafish, medaka, stickleback, (Lin, Wang, Li, & Wang, 2017; Nakamura et al., 2013)),
585 suggesting that it has been maintained by natural selection.

586 The top candidate gene at this locus is *opn1mw4*, a paralog of the *opn1mw* (RH2) gene,
587 which encodes cone photopigments that are essential for vision of blue-green light (medium
588 wavelength). This gene contains two missense mutations (c.850G>A and c.670G>A, Figure S10
589 A, B), one more divergent than the other (dAF=0.53 and dAF=0.40, respectively). The most
590 divergent variant results in an amino acid change from alanine to tryptophan at residue 284
591 (p.Ala284Thr), and the less divergent variant, from valine to isoleucine at residue 224
592 (p.Val224Ile). While all four amino acids are hydrophobic, tryptophan is more hydrophobic and
593 larger in volume than alanine, and the equivalent occurs between isoleucine respect to valine,
594 suggesting that these amino acid changes could potentially result in structural protein
595 modifications (Behmard, Abdolmaleki, & Asadabadi, 2012). Therefore, it is possible that the
596 missense mutations in *opn1mw4* are causal, as they could potentially affect the structure of the
597 photopigments and, thus, generate a shift in spectral sensitivity.
598 A change in visual sensitivity could be an adaptive response to contrasting light
599 environments between the Mediterranean and the Atlantic. In clear waters the blue-green light
600 spectrum dominates, especially as depth increases, while in shallower or turbid waters red-
601 shifted light conditions prevail. Comparative genomic studies indicate that fish living in marine
602 habitats dominated by green-blue light (e.g., in the deep sea, the pelagic open ocean, or in coral
603 reefs at night) commonly show gene copy expansions of *opn1mw* (RH2) respect to fish living in
604 shallower (<30 m) or more turbid waters (Lin et al., 2017; Musilova & Cortesi, 2021). Indeed,
605 the horse mackerel has five tandem copies of green-light opsins, which might relate to their
606 preferred habitat at deeper waters (100–200 m depth) (Brunel et al., 2016). Sea water turbidity in
607 the Mediterranean Sea is generally lower than in the Atlantic Ocean (Figure S10C), suggesting
608 that more green light might be transmitted to deeper layers of the water column.
609 Taken together, the probable adaptive significance of the missense mutations in *opn1mw4*
610 is that they might be involved in spectral fine-tuning of green-opsins to downwelling light
611 detection under contrasting light conditions between the Mediterranean Sea and the Atlantic. The
612 former likely dominated by blue-green light conditions and the latter, by a red-shifted light
613 environment. This visual adaptation could confer survival advantages related to feeding,
614 recognition of conspecifics, and escape from predators.
615

616 We discovered that the latitudinal genetic cline is underlined by a large (9.9 Mb) putative
617 chromosomal inversion on chr 21 (Figure 3). The northern haplotype shows high frequency in
618 the "northern" group and in the western Mediterranean Sea, while the alternative haplotype
619 dominates in north Africa, and intermediate frequencies of the two haplotypes occur in the south
620 of Portugal. This chromosomal inversion harbours thousands of genes, for which it is impossible
621 to infer its role without further studies. However, considering the strong genetic differentiation it
622 drives along an important oceanographic transition zone, we infer this inversion may play a
623 fundamental role in local adaptation to contrasting thermal regimes: a temperate (cooler) climate
624 in the north, tropical (warmer) in the south, and variable/intermediate at the transition zone in
625 mid Portugal. A possible explanation that the northern haplotype is prevalent in the western
626 Mediterranean Sea is that this sample might actually spawn in a different location with cooler
627 waters than where it was collected (near the Gibraltar strait, warmer temperatures similar to
628 those in southern Portugal, ~16 °C). Within the Mediterranean Sea, comparable temperatures to
629 those in the northern region (~11-12 °C) occur in the winter season in coastal waters of the
630 Balearic Sea near Catalonia (Spain), the Gulf of Lyon (France), and the Ligurian Sea (northeast
631 Italy) (Pastor, Valiente, & Khodayar, 2020, see Figure 3). Indeed, previous studies have reported
632 spawning events of horse mackerel near Catalonia during the winter months (Andreu &
633 Rodriguez-Roda, 1953; Planas & Vives, 1953). Given that the individuals collected in the
634 Alboran Sea were not actively spawning, we infer that they likely come from a region further
635 north, near the Gulf of Lyon.
636
637 We uncovered a genetically distinct population in the North Sea. We reached this
638 conclusion because seven genomic regions clearly distinguish samples from this location, and
639 the temporal replicates showed nearly identical allele frequencies at outlier loci. This observation
640 adds to previous morphometric and parasite data suggesting that horse mackerel from the North
641 Sea differs from nearby Atlantic populations (Abaunza et al., 2008).
642 Some of the top candidate genes likely involved in local adaptation to the North Sea are
643 *gpr83*, *sgms2*, *ncoa2*, and *taar7a*. *gpr83* (G-protein coupled receptor 83) encodes a receptor that
644 play an important role in the regulation of energy metabolism, feeding, reward pathway, and
645 stress/anxiety responses in mice, for which it has been linked with the control of body weight
646 (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 is likely favoring alleles that result in increased energy metabolism required for survival in the colder environment of the North Sea.

sgms2 (*sphingomyelin synthase 2*) encodes a protein that is involved in the synthesis of sphingomyelin, a major component of cell and Golgi apparatus membrane. Previous studies in fish indicate that this protein is crucial to maintain cell membrane structure and fluidity at low temperatures (Wang et al., 2014; Windisch, Kathöver, Pörtner, Frickenhaus, & Lucassen, 2011). Thus, we infer 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 that are proposed to play a critical role in intra- and interspecific communication, for example, in sexual attraction or in avoidance of predators or rotting food (Dewan, 2021). A previous study on two goatfish species with contrasting bottom habitat preferences (*Mullus surmuletus* lives on sandy and rocky bottoms, and *Mullus barbatus* lives at deeper waters on muddy bottoms) 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 water, 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 (Shi & Wang, 2010)(Figure S10). Thus, we infer that natural selection is likely favoring alleles in the gene *taar7a* 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*, sex determination-*hsd11b1l*, 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

678 temperatures (*hsd11b1l*). The complete list of candidate genes and their putative function are
679 presented in Table S6.
680
681 Interestingly, the number of divergent loci and their degree of differentiation in the
682 Atlantic horse mackerel are relatively small compared with those in the Atlantic herring (Han et
683 al., 2020), but is rather intermediate to the Atlantic herring and the European eel, which in fact
684 constitutes a single panmictic population (Enbody et al., 2021). We propose that the most
685 important explanation for the differences in genetic structuring between these three species is the
686 striking diversity of ecological conditions during spawning and early development, the most
687 sensitive period of life for a fish and a period characterized by high mortality (Dahlke, Wohlrab,
688 Butzin, & Pörtner, 2020). The Atlantic herring is a demersal spawner that breeds close to the
689 coast in areas with marked environmental differences as regards to temperature, salinity, depth
690 and biotic conditions (plankton production, predators, etc.). It also presumably shows homing
691 behaviour, that is, individuals return for spawning to the locations where they were hatched. In
692 contrast, the Atlantic horse mackerel is a benthopelagic fish that spawns at deeper waters, near
693 the shelf edge (100-200 m deep) at not well-defined spawning areas, implying that the range of
694 environmental conditions in which breeding occurs are comparatively less diverse. Finally, to the
695 best of our knowledge, all European eels spawn in the Sargasso Sea under similar environmental
696 conditions.
697
698 In summary, we revealed a number of highly divergent genomic regions encompassing
699 genes that may be involved in adaptation to local environmental conditions in the North Sea, the
700 western Mediterranean Sea, and along a climate-related latitudinal cline in the Atlantic. In
701 addition, our results highlight the importance of chromosomal rearrangements for local
702 adaptation with gene flow.
703
704 ***Implications for fisheries management***
705 The genetic-based groups identified here are largely in agreement with the current horse
706 mackerel stock divisions, informed by the results of the 2000-2003 HOMSIR project (Abaunza
707 et al., 2008). However, our data do not support the current definition of the "southern stock" in
708 Portuguese waters (Figure S1), and of the southern boundary of the "western stock".

709 Our genomic data indicates that the "southern stock" actually consists of two biological
710 units (Figures 1, 2, 5), not one as previously assumed. Samples from northern Portugal (north of
711 Lisbon, ~38.7-39.0°N) appear to be genetically closer to the "western stock", while samples
712 from southern Portugal (south of Lisbon) form their own group but are genetically closer to the
713 samples from the "Saharo-Mauritanian stock", in north Africa. To confirm these findings and
714 assess the spatial and temporal trends of mixing between these areas, further studies are required,
715 including a finer geographic sampling and screening of informative genetic variants in a large
716 number of individuals throughout this area.
717 We did not find significant genetic differences between northern Portugal (currently
718 considered part of the "southern stock"), northern Spanish shelf (Bay of Biscay), and the west of
719 Ireland, implying that the southern boundary of the "western stock" could possibly be extended
720 down to the north of Portugal. However, the minute genetic differentiation does not exclude the
721 possibility that isolation in an ecologically-relevant timescale of interest for fisheries
722 management might occur (Hauser & Carvalho, 2008b).
723 Finally, our genomic data support the consideration of the Mediterranean Sea as a
724 separate stock, as proposed by the HOMSIR project (Abaunza et al., 2008). While a single
725 sample from the westernmost part of the Mediterranean was studied, its genetic distinctiveness
726 suffices to infer that the Mediterranean horse mackerel likely constitutes a separate population
727 from those in the Atlantic. Wide-scale sampling within the Mediterranean Sea is required to
728 further explore the population structure in this region.
729 This study disclosed a number of genetic markers (SNPs) that can be used as a genetic
730 tool for fisheries stock assessment. A panel of only 63 markers suffice to identify the main
731 genetic subdivisions. In fact, using a reduced panel of only 17 markers, it is possible to
732 differentiate individuals collected in "the North Sea stock" and "western stock", and from the
733 "northern" (North Sea, west of Ireland, northern Spanish shelf, northern Portugal, and the
734 Mediterranean) and the "southern" (southern Portugal, north Africa) groups with respect to the
735 genetic break along Lisbon. Thus, these markers can help, for instance, to elucidate the extent of
736 mixing between the Western and North Sea stocks in the English Channel (ICES Divisions 7.e
737 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

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

Behmard, E., Abdolmaleki, P., & Asadabadi, E. B. (2012). Mutation in a valine residue induces drastic changes in 3D structure of human prion protein. *Frontiers in Life Science*, *6*(1–2), 47–51. doi: 10.1080/21553769.2013.775078

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

Fuentes-Pardo, A. P., Bourne, C., Singh, R., Emond, K., Pinkham, L., Mcdermid, J. L., … Ruzzante, D. E. (2019). Adaptation to seasonal reproduction and thermal minima-related factors drives fine-scale divergence despite gene flow in Atlantic herring populations. *BioRxiv Preprint*. doi: 10.1101/578484

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

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*.

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 contributed to 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.

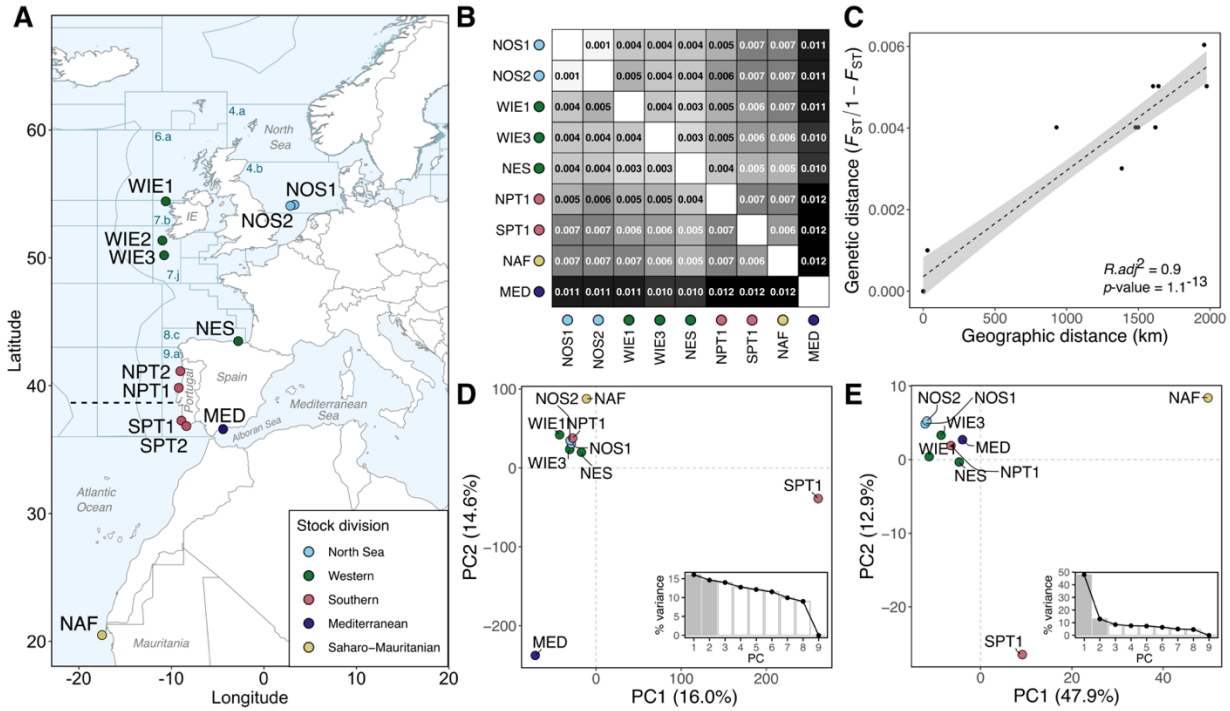

Figure 1. Population structure of the Atlantic horse mackerel. (**A**) Map depicting the 12 sampling sites in the east Atlantic Ocean. ICES fishing divisions are denoted in light blue, same as their alphanumerical code. The approximate location of a biogeographical transition zone north of Lisbon, in 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 plot 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 populations in the "northern" group. This analysis was performed excluding the Mediterranean sample (MED). 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 shows 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 name abbreviations as in Table 1.

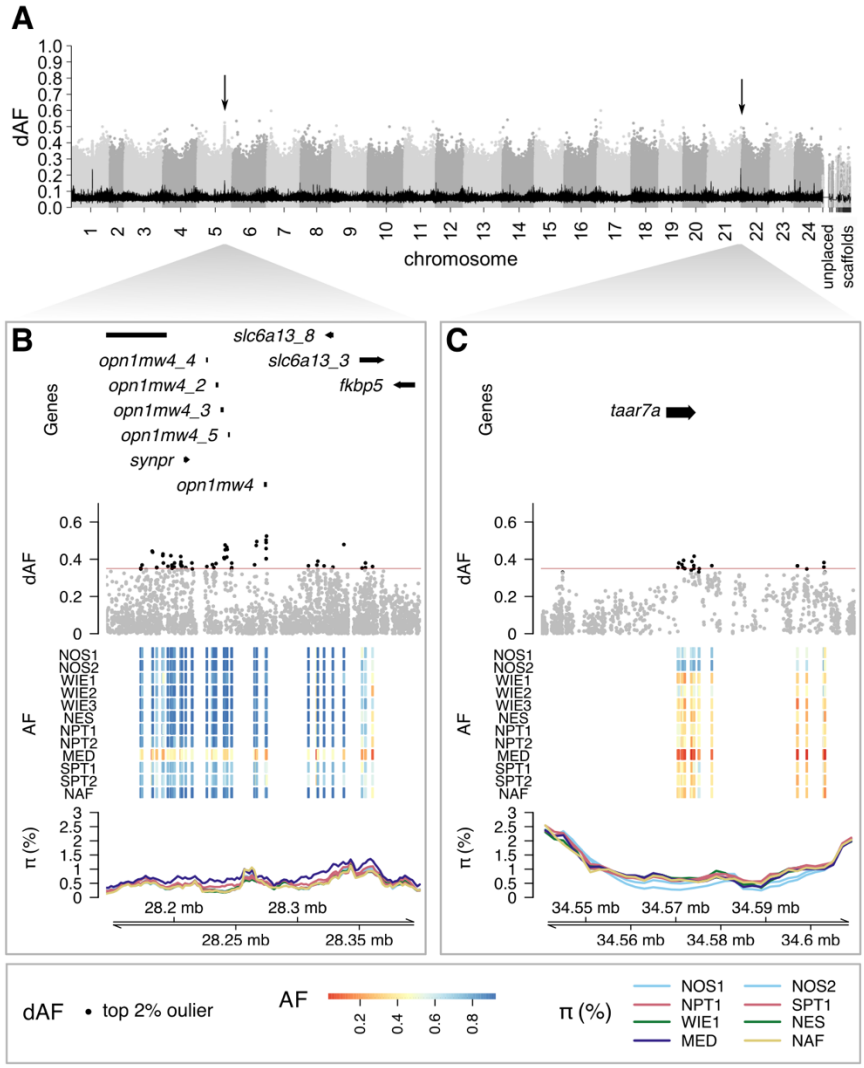

Figure 2. Two selective regions distinctive of the western Mediterranean Sea. (**A**) Manhattan plot depicting the dAF per SNP across the genome for the contrast between the Alboran Sea vs. all other samples. Each dot is a single SNP, and an alternating color (gray, blue) was used to differentiate SNPs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

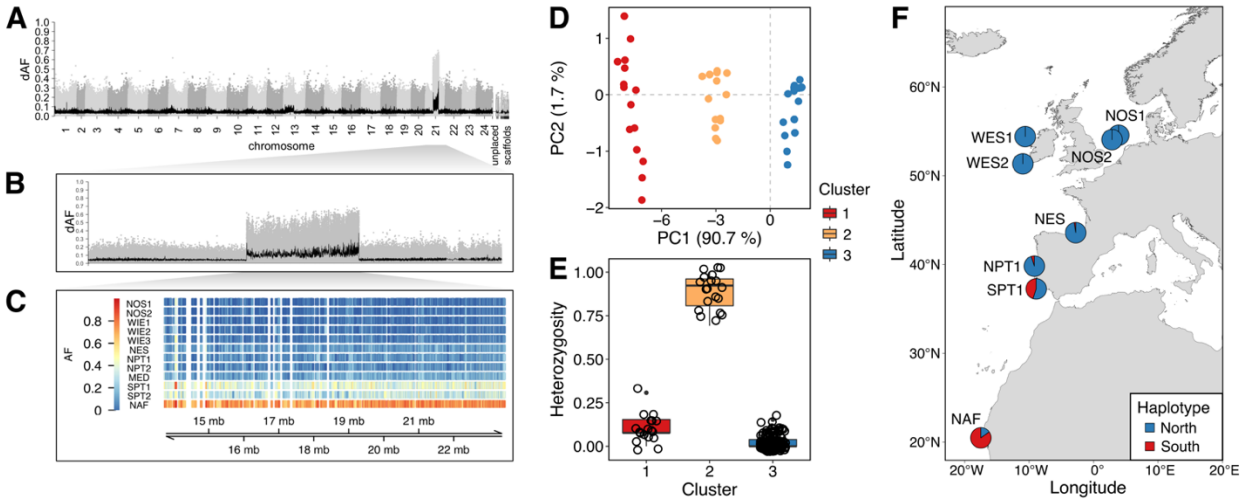

Figure 3. Putative chromosomal inversion on chromosome 21 that underlies a steep north-south genetic break in Atlantic horse mackerel. (**A**) Manhattan plot representing the absolute difference in pool-allele frequencies (dAF) of each SNP along the genome for the contrast between north vs. south groups (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 colors, gray and blue. The black line towards the bottom of the plot corresponds to the 100 SNPs-rolling average of dAF values. (**B**) Close-up Manhattan plot for chromosome 21. (**C**) Pool-allele frequency per population sample (rows) of the top 2% markers (columns) within the putative inversion. (**D-F**) Haplotype 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, as indicated by their higher average observed heterozygosity respect to the other two clusters. (**F**) Map showing the geographic distribution of inversion haplotypes across sampling locations. Pie charts represent the frequency of haplotypes "North" or "South" per location. Sample name abbreviations as in Table 1.

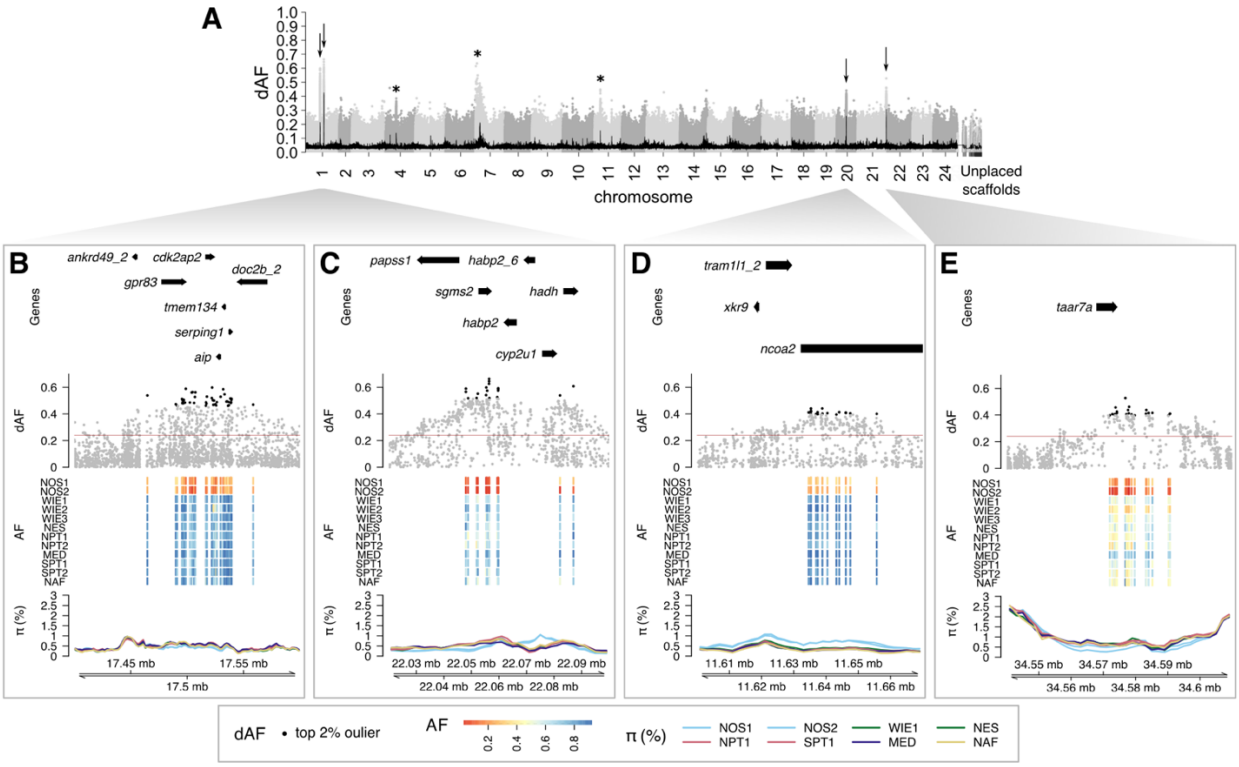

Figure 4. Genomic regions characteristic of the North Sea populations. (**A**) Manhattan plot representing the dAF of each SNP along the genome for the contrast between the North Sea vs. all other samples. Each dot is a single SNP. The black line towards the bottom is the rolling mean of dAF over 100 SNPs. Close-up plots of the four most divergent regions in (**B-C**) chr 1, (**D**) chr 20, and (**E**) chr 21 (highlighted with arrows). Plots of the other differentiated regions on chr 4, 7 and 11 are shown in the Supplementary Figure S9 (highlighted with an asterisk). Each close-up plot consists of four sections (from top to bottom). The first section illustrates the gene models. The second section 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 section is a heatmap plot depicting the pool-allele frequency of the top 2% SNPs, where the rows are pool samples and the columns are SNPs. The fourth section is the nucleotide diversity (π) for each pool sample calculated over 10 Kbp sliding windows with a step size of 2 Kbp. 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.

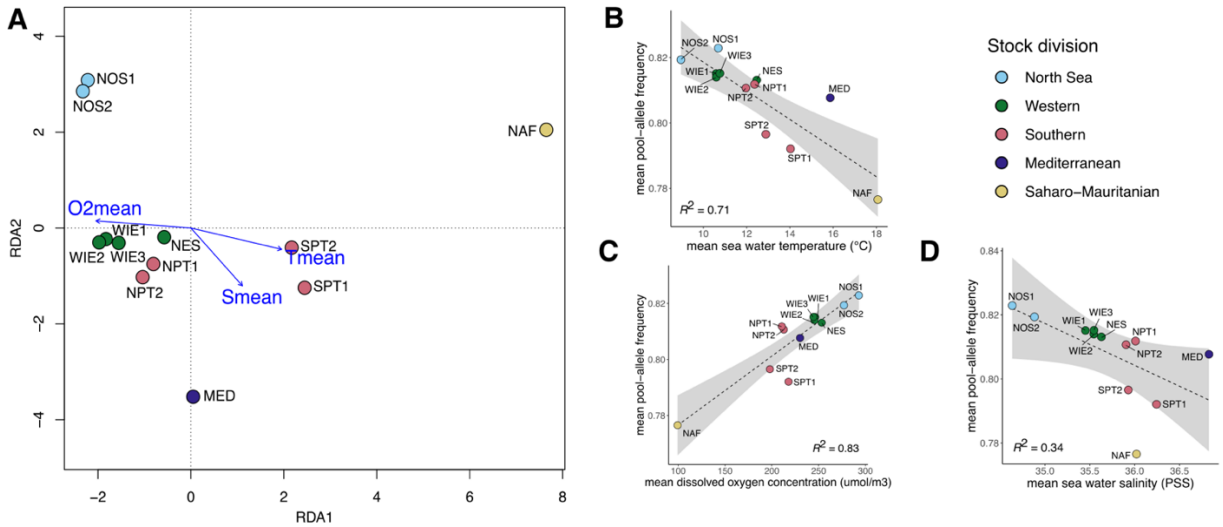

Figure 5. Association between genetic variants and environmental variables. Analysis based on the pool-allele frequencies of 210,568 SNPs within the most divergent genomic regions in the North Sea and the Mediterranean Sea, and underlying the north-south genetic break. (**A**) Redundancy Analysis (RDA) plot showing the association between pool-allele frequencies of SNPs and three uncorrelated environmental variables: Tmean, mean sea water temperature (°C); O_2mean, mean dissolved oxygen concentration (μmol/m^3); and Smean, mean sea water salinity (PSS). Blue arrows represent the 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**) sea water temperature, (**C**) mean dissolved oxygen concentration, and (**D**) mean sea water 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.

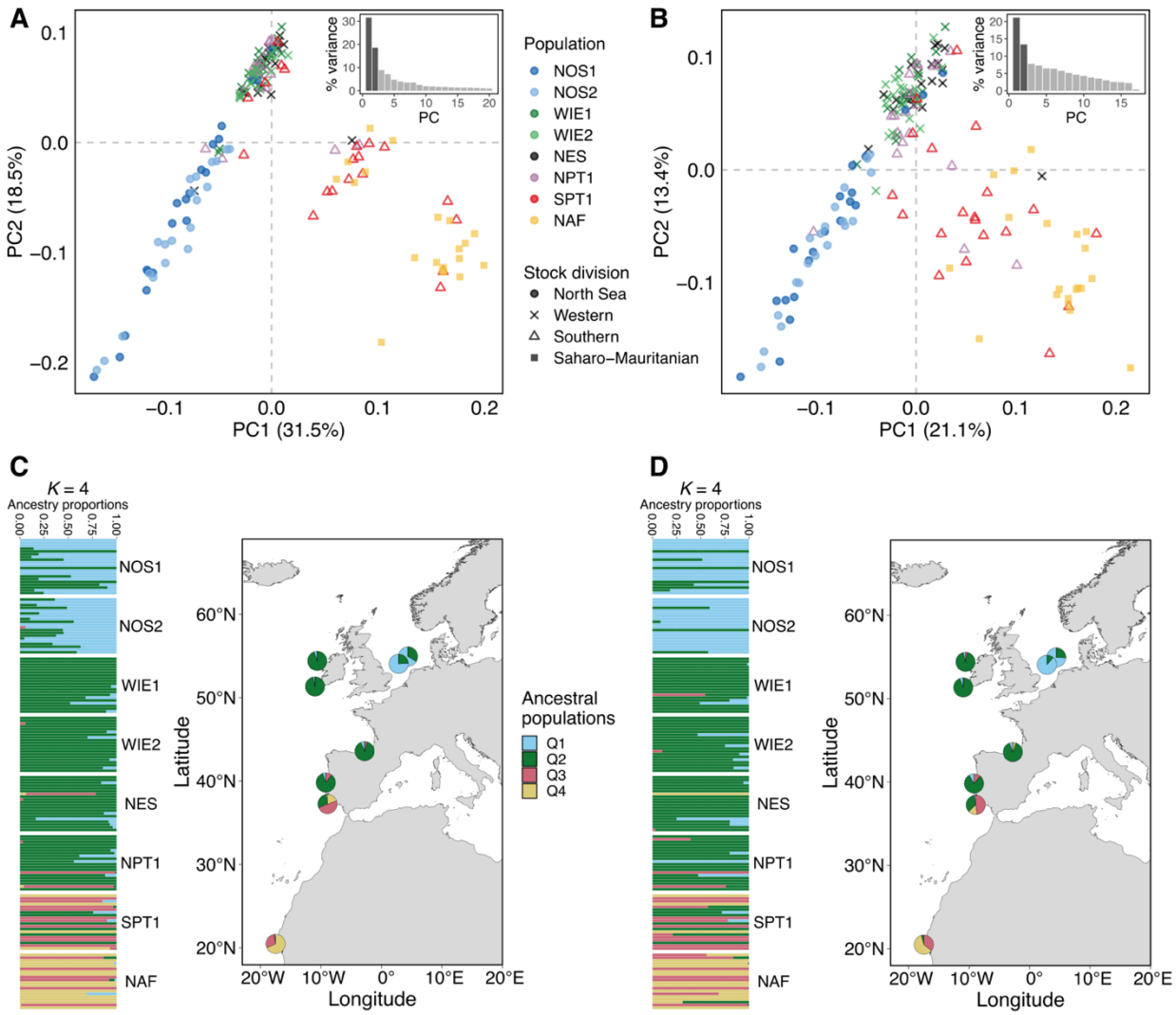

Figure 6. Population structure based on individual SNP genotyping. The two SNP datasets consisted on the genotype of 157 and 155 individuals screened using 63 and 17 markers, respectively. Principal components analysis plot based on (**A**) the 63-SNP dataset or the (**B**) the 17-SNP dataset. Only the 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 the 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 the (**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 name abbreviations as in Table 1.

1142 **Table 1.** Collection details of the 12 Atlantic horse mackerel samples included in this study.
1143

Pool ID	Stock (*)	Area	Year	Latitude	Longitud	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

1144 (*) Stock divisions after Abaunza et al. (2008).
1145

1146