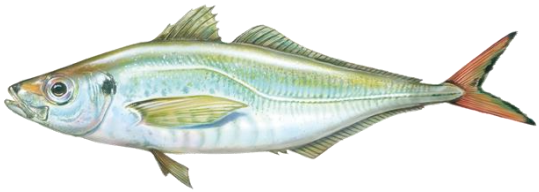
Population structure of the Atlantic horse mackerel (*Trachurus trachurus*) revealed by whole-genome sequencing



A report prepared for the members of the

Northern Pelagic Working Group

and

the Pelagic Advisory Council

by

Angela P. Fuentes-Pardo, Mats Pettersson, C. Grace Sprehn, Leif Andersson

Uppsala University, Sweden

&

Edward Farrell

EDF Scientific Limited, Ireland

July 2020

Executive Summary

The Atlantic horse mackerel, *Trachurus trachurus* (Linnaeus, 1758) is a species of jack mackerel distributed in the East Atlantic, from Norway to west Africa and the Mediterranean Sea. It is a pelagic shoaling species found on the continental shelf and it is one of the most widely distributed species in shelf waters in the northeast Atlantic, where it is targeted in pelagic fisheries. In the northeast Atlantic region, the species is assessed and managed as three stocks: the Western, the North Sea and the Southern. Despite the commercial importance of the horse mackerel, the accuracy of alignment of these stock divisions with biological units is still uncertain.

The aims of this study were to identify informative genetic markers for the stock identification of horse mackerel and to estimate the extent of genetic differentiation among populations distributed across the distribution range of the species. For this we used modern sequencing techniques that allowed us to assess genetic variants in the entire genome. We discovered that while the populations differ in a small fraction of their DNA (< 1.5%), such genetic differences are significant as they likely represent natural selection and might be involved in local adaptation. We validated a small fraction of these highly differentiated genetic variants by a SNP assay and demonstrated that they can be used as informative molecular markers for the genetic identification of the main stock divisions of the Atlantic horse mackerel.

The results, based on the analysed samples, indicated that the North Sea horse mackerel are a separate and distinct population. The samples from the Western stock, west of Ireland and the northern Spanish shelf, and the northern part of the Southern stock, northern Portugal, appear to form a genetically close group. There was significant genetic differentiation between the northern Portuguese samples and those collected in Southern Portuguese waters, with those in the south representing a separate population. The North African and Alboran Sea samples were distinct from each other and from all other samples.

These results indicate that a further large-scale analysis of samples, with a greater temporal and spatial coverage, with the newly identified molecular markers is required to test and reassess the current stock delineations.

Table of Contents

1. Background

1.1 Biology

The horse mackerel, *Trachurus trachurus* (Linnaeus, 1758) is a species of jack mackerel from the Carangidae family and is distributed in the East Atlantic from Norway to Western Africa and the Mediterranean Sea (Froese and Pauly, 2015). It is a pelagic shoaling species found on the continental shelf and is one of the most widely distributed species in shelf waters in the northeast Atlantic. The range of horse mackerel partially overlaps with four other *Trachurus* spp; *Trachurus picturatus* (Bowdich, 1825) and *Trachurus mediterraneus* (Steindachner, 1868) in Iberian, North African and Mediterranean waters, *Trachurus trecae* (Cadenat, 1949) in West African waters and the very closely related *Trachurus capensis* (Castelnau, 1861) in west and southwest African waters.

Horse mackerel are estimated to mature at c.20 cm total length and between 2 and 4 years of age (Abaunza et al., 2003). Waldron and Kerstan (2001) validated the age determination of horse mackerel otoliths, through marginal increment analysis of whole otoliths, up to age four. However, examination of subsequent growth zones indicated that false rings and annuli are often of a similar appearance and as such accurate ageing beyond four years of age year is difficult. Horse mackerel grow rapidly during the first years of life and more slowly after three years of age. The maximum estimated age is reported as 40 years (Abaunza et al., 2003). Both growth and age at maturity fluctuate, which is suggested to be a density-dependent response to the extremely large fluctuations in year-class strength (ICES, 1991).

Horse mackerel is considered to be an asynchronous batch spawner with an indeterminate fecundity (Gordo et al., 2008; Ndjaula et al., 2009). In the northeast Atlantic area, the horse mackerel population has an 8-month long spawning season (Abaunza et al., 2003; Dransfeld et al., 2005), although the duration of an individual's spawning period is unknown (Van Damme et al., 2014). Horse mackerel appear to undertake annual migrations to spawning, feeding and over-wintering area (Abaunza et al., 2003). The peak spawning in the northeast Atlantic west of Britain and Ireland is in June in shelf waters (ICES, 2017; van Damme et al., 2014). Peak spawning in the North Sea occurs in May and June (Macer, 1974), and spawning occurs in the coastal regions of the southern North Sea along the coasts of Belgium, the Netherlands, Germany, and Denmark. Peak spawning in Portuguese waters is earlier than the other regions being in February in shelf waters (Borges & Gordo, 1991), though it should be noted that there is significant overlap between these areas. In winter the North Sea spawning horse mackerel are believed to migrate to the Western English Channel, whilst those that spawn west of Ireland and Britain migrate from feeding grounds off Norway and the northern North Sea to the continental slope southwest of Ireland (Heessen et al., 2015).

1.2 Stock Identification

ICES has long considered horse mackerel in the northeast Atlantic to consist of three stocks (Figure 1). The southern stock was defined as that found in the Atlantic waters of the Iberian Peninsula (Division 9a), the North Sea stock in the eastern English Channel and southern North Sea area (Divisions 3a, 4b,c, and 7d), and the western stock on the northeast continental shelf of Europe, stretching from the Bay of Biscay in the south to Norway in the north (Subarea 8 and Divisions 2a, 4a, 5b, 6a, and 7a–c, e–k). This separation of horse mackerel was based on a variety of factors including the temporal and spatial distribution of the fishery, the observed egg and larval distributions, information from acoustic and trawl surveys and from parasite infestation rates (see ICES, 2015). A tagging programme was established in 1994 (ICES, 1995) and further studies based on genetic (allozyme) population structure and morphometric characteristics, were conducted in 1997 (ICES, 1998). Tagging studies failed to recover any tagged fish, and neither the genetic nor morphometric studies provided a basis for changing the stock separation as previously defined.

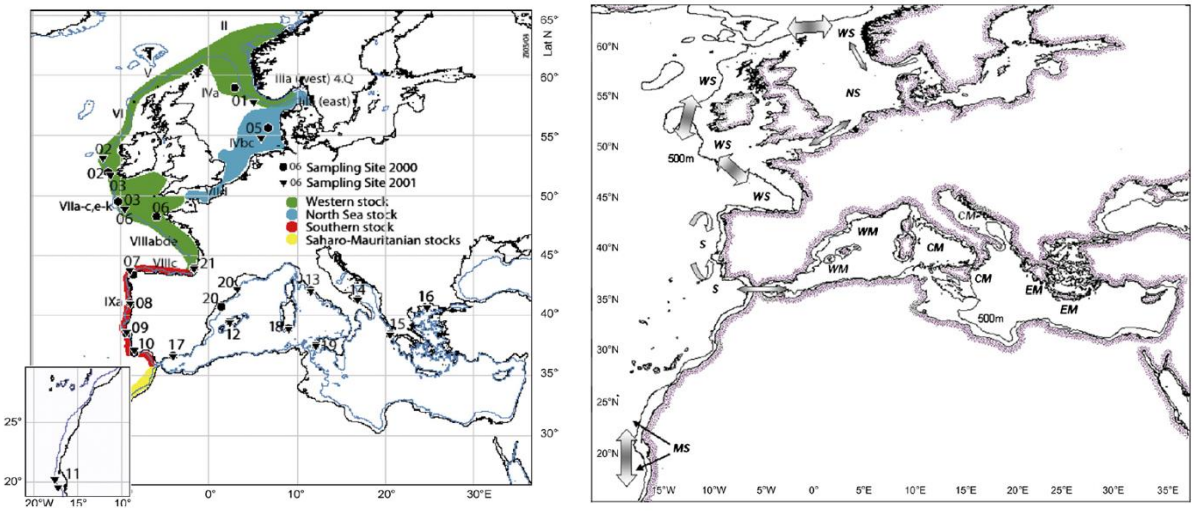


Figure 1. (Left panel) The suggested stocks of horse mackerel prior to the HOMSIR project. The sampling sites in the HOMSIR project in 2000 (circles) and 2001 (triangles). (Right panel) Proposed horse mackerel stocks according to the HOMSIR project. The arrows indicate possible migratory movements. WS: western stock; NS: North Sea stock; S: southern stock; MS: Saharo-Mauritanian stock; WM: western Mediterranean stock; CM: central Mediterranean stock; EM: eastern Mediterranean stock. From Abaunza *et al*. (2008).

Further refinements of the definitions of stock units were based on the results from the EU-funded HOMSIR project (2000-2003), which utilised a multidisciplinary approach including various genetic approaches (allozymes, mitochondrial DNA and microsatellites), the use of parasites as biological tags, body morphometrics, otolith shape analysis and the comparative study of life history traits (growth, reproduction and distribution) (Abaunza et al., 2008). The resulting stock structure was broadly similar to that previously considered by ICES (Figure 1). However, it was observed that the population structure in the western European coasts could be more complicated and that more research was needed to clarify the migration patterns within the Northeast Atlantic Ocean. This was especially relevant to the mixing areas between the North Sea stock and the Western stock (Northern North Sea and English Channel). The sampling in this region was relatively sparse whereas the southern regions had significantly better coverage (Figure 2). The genetic components of the project failed to resolve stock structure largely due to the low number (four microsatellites) and low power of the genetic markers employed (Kasapidis and Magoulas, 2008).

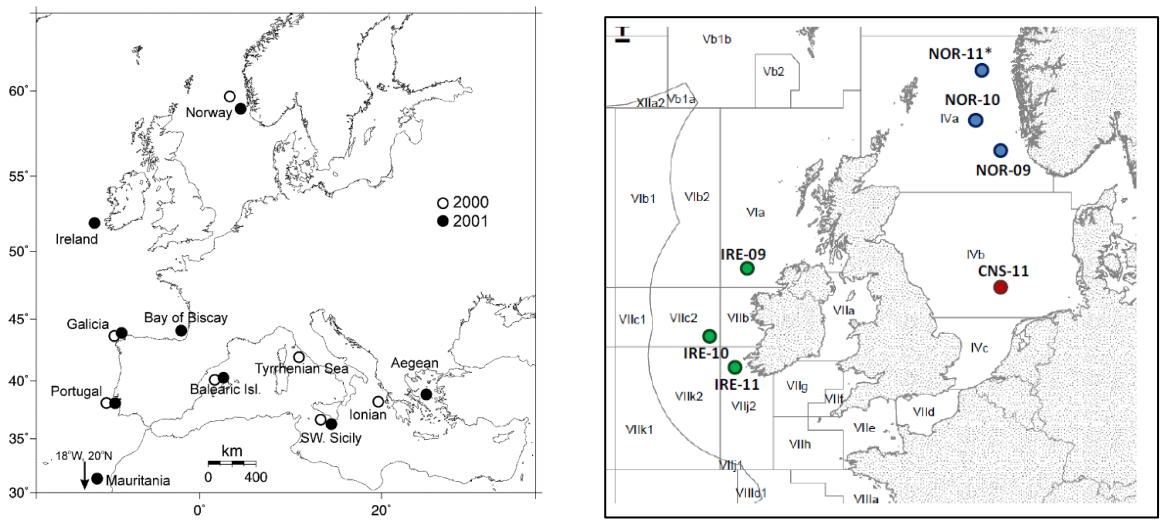


Figure 2. (Left Panel) The genetic samples collected and analysed in the Kasapidis and Magoulas (2008) study which was part of HOMSIR. (Right Panel) The genetic samples collected and analysed in the Mariani (2012) pilot study.

A recent preliminary study on western and North Sea horse mackerel employed 12 microsatellites (4 from horse mackerel, *Trachurus trachurus* and 8 from Chilean jack mackerel, *Trachurus murphyi* Nichols, 1920) to screen a small number of samples (n = 7 samples/339 individuals) from both putative stocks (Figure 2). The results indicated significant population structure within the samples from the western stock while no significant structure was observed between the samples collected west of Ireland and those collected in the central North Sea (Mariani, 2012). However, there were a number of issues related to the genetic markers employed being non species-specific and also the samples screened not being from spawning individuals.

The degree of separateness of the western and North Sea stocks is uncertain. It is known that the western stock spawns west of Ireland while the North Sea stock has a separate spawning ground in the North Sea. However, it is unclear if these spawning grounds are used interchangeably. Unlike herring (*Clupea harengus* Linnaeus, 1758), horse mackerel are not known to be faithful to their original spawning grounds. Therefore, without strong evidence to the contrary, it cannot be assumed that the two stocks are indeed separate. Treating these stocks as separate, if indeed they are not, is dangerous from a precautionary management perspective. Further research is needed to clarify the level of differentiation between the North Sea and Western stocks and also to define the boundary areas, if any, between them. The levels of mixing in the northern North Sea (area 4a) are also unclear and catches and survey data from this area are currently allocated to the North Sea stock in quarters 1 and 2 and to the western stock in quarters 3 and 4, highlighting the uncertainty in the assessments for these stocks.

1.3 Stage 1 - PFA/IMARES pilot study

In 2015 the Pelagic Freezer Trawler Association (PFA) contracted the Wageningen UR, Institute for Marine Resources and Ecosystem Studies, IJmuiden (IMARES) to undertake a study on North Sea Horse Mackerel (Brunel et al., 2016). The primary aim of the study was to improve the data quality used for an analytical stock assessment model of North Sea horse mackerel. The stock is currently classified by ICES as a data poor stock, for which the catch advice is based on the trend in an abundance index.

The management boundary between the western and North Sea stocks in the English Channel (corresponding to the separation between areas 7e, western Channel and 7d, eastern Channel) does not correspond to a real biological boundary, as mixing of the two stocks is known to occur in area 7d in autumn and winter (Brunel et al., 2016). The catches taken in 7d are officially considered as being North Sea horse mackerel and represent c.80% of the catches from this stock. An unknown proportion of this catch is likely from the western stock, which interferes with the cohort signal in the catch at age matrix, hampering the development of an age-structured assessment model for the North Sea stock. Developing methods to separate catches from the western stock from catches from the North Sea stock in area 7d are therefore necessary to improve the quality of the catch information for the North Sea stock. Within the project, two pilot studies, based on chemical fingerprint and genetics, were conducted to investigate new methods to determine stock structure and to develop techniques to identify the stock origin of the catches taken in the eastern English Channel.

The chemical fingerprint analysis was carried out by IMARES using two-dimensional gas chromatography (GCxGC-MS), in order to establish a full chemical fingerprint of the horse mackerel samples from both the western and North Sea stocks. Results were inconclusive but suggested that the chemical fingerprint approach was a potential tool to determine stock of origin, with a moderate risk of misclassification. However, more insight on the sources of variation of compound concentrations (seasonal changes, influence of sex, length, age, reproducibility of the results from year to year) is required before this method can be further developed.

IMARES, contracted University College Dublin (UCD) to undertake a pilot study to develop a method of genetic stock identification for discriminating North Sea and Western Horse mackerel (Brunel et al.,

2016). The aims of the pilot study were to firstly develop and validate at least 24 polymorphic microsatellites markers in horse mackerel and secondly to screen spawning fish collected in 2015 from the Western and North Sea stocks to establish a genetic baseline of the spawning stocks and test the presence of population structure. Recently developed Next Generation Sequencing (NGS) and Genotyping by Sequencing (GBS) based approaches, which were developed on cod (*Gadus morhua* Linnaeus, 1758), boarfish (*Capros aper* Lacépède, 1802) and 6a/7bc herring were used for marker development and screening of spawning samples (Carlsson *et al.*, 2013; Farrell *et al.*, 2016; Vartia *et al.*, 2014 & 2016). The pilot study successfully identified a large number of novel microsatellites, however initial data analyses were confounded by a poor-quality sequencing run and as such the discrimination power between the western and North Sea sample was low. This resulted in the pilot study being unable to separate the two stocks conclusively and unequivocally.

1.4 Stage 2 – Northern Pelagic Working Group (NPWG) genetic baseline project

In an effort to resolve these uncertainties the Northern Pelagic Working Group contracted EDF Scientific Limited and Jens Carlsson to undertake a comprehensive genetic stock identification study on Atlantic horse mackerel (Farrell & Carlsson, 2018). Sampling was conducted over three consecutive years and three spawning seasons and covered a large area of the distribution of the species including the Western, North Sea and Southern stock areas and also West African waters. In total 33 population samples, comprising 2,295 individual fish were collected from 2015 to 2017 across the study area (Figure 3). Total genomic DNA was extracted from 2,208 of these specimens. Spawning samples were analysed with a panel of 37 novel, putatively neutral microsatellite markers and statistical analyses (F_{ST}, structure, assignment testing, mixed stock analyses and FCA analyses) indicated that horse mackerel in the northeast Atlantic region does not represent a single biological unit. A high level of species misidentification in the West African samples was also observed. On the highest level there are mixed species catches in African waters, a clear separation of the southern North Sea from other regions and further, less pronounced, structure along the northeast Atlantic continental shelf. Exploratory assignment testing and mixed stock analysis of the western and North Sea baselines indicated a success rate of c.60-65% for self- assignment. This was considered relatively low and is due to the relatively low genetic differentiation between the populations at putatively neutral loci. Despite this, further exploratory assignment testing and mixed stock analysis of the fish caught outside spawning time in the northern North Sea and western English Channel (Figure 3) indicated that a large component of these fish belonged to the Western stock. No samples from the eastern English Channel were available for testing.

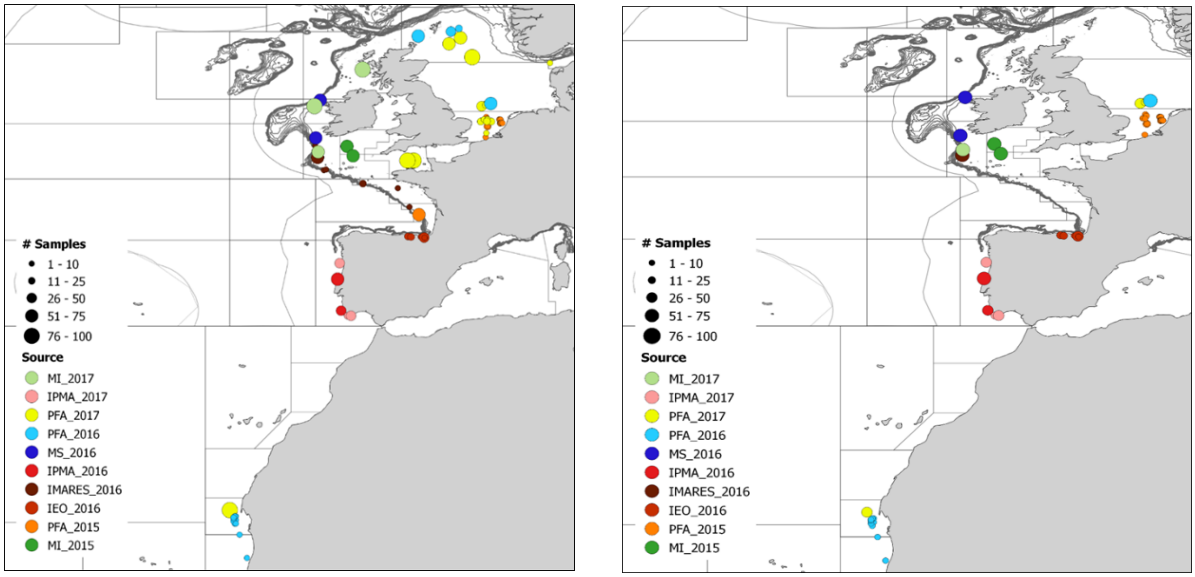


Figure 3. (Left Panel) The horse mackerel samples collected from 2015 to 2017 and (right panel) those included in the baseline dataset.

The results showed that the genetic information produced in the stage 2 study could be used for mixed stock analyses and that the information could be used to delineate the range of the North Sea stock – information that could be taken into account by fisheries management. However, it was suggested in the project report that further genetic analyses were warranted (full genome, RNA and RAD sequencing-based approaches) to increase the numbers and types of genetic markers available for this species. This would improve stock discrimination, mixed stock analyses and individual assignment capacity, similar to the approaches deployed for Baltic and Atlantic herring and other commercial fisheries species. This proposal by Dr Edward Farrell of EDF Scientific Limited, Ireland and Professor Leif Andersson, Uppsala University outlines one such approach.

1.5 Stage 3 & Stage 4 - Population genomics of horse mackerel and SNP validation

The current report presents the results of stages 3 and 4 of the horse mackerel project. To improve our ability to identify informative genetic markers, Dr Edward Farrell of EDF Scientific Limited, Ireland, and Professor Leif Andersson of Uppsala University, Sweden, proposed to undertake full genome sequencing of horse mackerel. This method provides the highest resolution of genetic variants with respect to the reference genome of the species', which was recently assembled by the Wellcome Sanger Institute, UK (website: https://vgp.github.io/genomeark/Trachurus_trachurus/). The Northern Pelagic Working Group funded stage 3, which involved the whole-genome pooled DNA sequencing of a subset of the populations sampled in stage 2 to identify population specific genetic markers. Further validation of potentially informative SNPs was undertaken as stage 4 and was funded by the Pelagic Advisory Council.

2. Materials and Methods

2.1 Sampling and DNA isolation

The samples included in the current study were a subset of the baseline samples analysed in stage 2 (Farrell and Carlsson, 2018). Sampling was organised by EDF Scientific and the Pelagic Freezer Trawler association (PFA). Samples were collected opportunistically, from 2015-2017, through existing fisheries surveys and from both target and non-target fisheries. One additional sample from the Alboran Sea in the Mediterranean Sea was provided by Dr Jens Carlsson from the ATLAS Project (https://www.eu-atlas.org/). The primary focus of sampling for the genetic analysis was collection of spawning fish, in order to ensure that samples could be considered to provide a valid baseline. However, due to the opportunistic nature of the sampling programme this was not always possible. 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 (see Annex 1 Table S1; ICES, 2015). Each fish was measured for total length (TL) to the 0.5 cm below and total body weight (TW) to the nearest 1.0 g. Sex and maturity were also assessed and a 0.5 cm^3 piece of tissue was excised from the dorsal musculature of each specimen and stored at 4°C in absolute ethanol. Total genomic DNA (gDNA) was extracted from the majority of samples by Weatherbys Scientific Ltd, from c.30 mg of tissue from each fish using sbeadex™ magnetic bead-based extraction chemistry on the LGC Oktopure™ platform. The remaining samples were extracted using a Chelex and proteinase-K or CTAB based extraction protocol (Table 1). Extracted DNA was quantified on a NanoDrop® ND-1000 spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA) and laid out on 96-well PCR plates.

2.2 High-throughput sequencing, QC of raw reads, and read mapping

We performed whole-genome resequencing of pooled DNA (Pool-Seq) to assess the population-level genomic variation of the 12 fish aggregates sampled in this study. For this, individual DNA samples were combined into 12 pools by location and year in equal quantity to obtain at least 1.5 μg in 25-50 μL (Table 1). Between 30 and 96 individuals were included in each pool (Table 1). Pools were quantified in ng/μL using a Qubit Fluorometer (Thermo Fischer Scientific Inc) prior to submission to the SNP&SEQ Technology Platform in Uppsala, Sweden for library preparation and high-throughput sequencing. A PCR-free Illumina TruSeq library kit with a target insert size of 350 base pairs (bp) (Illumina Inc) was used for most pools, except for 6a and 6b, for which a Splinted Ligation Adapter Tagging (SPLAT) library preparation was used because their DNA was single-stranded (Raine et al., 2016). All libraries were paired-end sequenced on Illumina NovaSeq S4 flowcells with a read sequence length of 150 bp.

The quality of raw sequence reads for each pool was examined with *FastQC* v0.11.8 (Andrews, 2010), and jointly analysed in a single report with *MultiQC* v.1.7 (Ewels et al., 2016). Based on this initial sequence quality assessment, we removed low quality bases (Phred score < 15), Illumina adapters, and short reads (< 36 bp) with *Trimmomatic* v.0.36 (Bolger et al., 2014) (parameters: ILLUMINACLIP:adapters.fa:2:40:15:8:true SLIDINGWINDOW:4:15 LEADING:15 TRAILING:15 MINLEN:36). The quality of the resulting trimmed reads was assessed again with *FastQC* before further analysis.

Reads were mapped against the Atlantic horse mackerel (*Trachurus trachurus*) genome using *bwa-mem* 0.7.17 (Li, 2013) and default parameters. Read mapping quality statistics, including the number of aligned reads and the average read depth of coverage, were generated with *QualiMap* v.2.2.1 (Okonechnikov et al., 2015). Prior to variant calling, mapped reads were sorted using *SAMtools* v.1.10

(Li et al., 2009), duplicated reads were marked and read groups were added, both with *Picard* v2.20.4 (Broad Institute, 2018), and an index file was created with *SAMtools*.

2.3 Variant calling and filtering

Variant calling was performed with *GATK-UnifiedGenotyper* v3.8 (McKenna et al., 2010) because, in our experience, this algorithm works well and produces less false positives than the *GATK-HaplotypeCaller* when analysing pooled samples. The *GATK-UnifiedGenotyper* is a single-base caller that simultaneously identifies Single Nucleotide Polymorphisms (SNPs) and small indels (insertions and deletions). Since we aimed to characterize genome-wide variation based on biallelic SNPs, we extracted these genetic markers from the raw variant set using *GATK*.

To remove spurious markers and thus, retain the best quality ones for further analysis, we applied various filters to the raw SNP set. First, we performed hard-filtering by retaining SNPs that passed cut-off values that were set based on the genome-wide distribution of GATK variant quality annotations. The cut-off values used were: FisherStrand (FS) > 60.0, StrandOddsRatio (SOR) > 3.0, RMSMappingQuality (MQ) < 40.0, MappingQualityRankSumTest (MQRankSum) < -12.5, and ReadPosRankSumTest (ReadPosRankSum) < -8.0 (for more details on the GATK quality annotations, see *https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants*). Next, we retained SNPs with a genotype quality (GQ) greater than 20, allowed for a missing rate per locus of a maximum of 20%, kept loci with a minor allele count of at least 3 reads (MAC), and removed monomorphic loci with *BCFtools* v.1.10 (Li et al., 2009). Lastly, we applied a depth of coverage filter as follows. Based on the total read depth (DP) per locus and pool, we generated depth of coverage distributions for each pool with *R* (R Core Development Team, 2020) and the *R* package *ggplot2* (Wickham, 2016). We evaluated three different cut-off value ranges (listed from the most to the least stringent filter): mean $\pm$ 1 standard deviation, mode $\pm$ ½ the mode, and between 20x and 300x (300x corresponds to three times the mean coverage for all pools). We retained SNPs that fulfilled the depth of coverage requirement for all pools while excluding samples 6a, 6b and 7 (see results for details). The resulting high-quality SNP set was used in further analysis. A schematic summary of the data generation steps is illustrated in Figure S1.

2.4 Population genetic structure

The population-level allele frequencies computed from Pool-Seq data are derived from the read counts of a variant site. To control for potential technical artifacts inherent to Pool-Seq that could bias the allele frequency calculation, such random variation in read coverage and in chromosome representation across pools (Dohm et al., 2008; Kolaczkowski et al., 2011), we applied the n_{eff} allele count correction (Feder et al., 2012; Kolaczkowski et al., 2011) to the read counts of each SNP using a custom script implementing this formula $n_{\text{eff}} = \frac{(n*CT)-1}{n+CT}$ where *CT* corresponds to read depth and *n* to the number of chromosomes in a pool, being equal to 2*N* for diploid species like herring. Population allele frequencies were then calculated based on the n_{eff} corrected read counts and constituted the basis of subsequent population analysis.

To estimate the level of genetic differentiation among pools, we computed the unbiased pool-F_{ST} statistic ($\hat{F}_{\text{ST}}^{\text{pool}}$) for all possible paired comparisons with the R package *poolfstat* (Hivert et al., 2018). This statistic is equivalent to the (Weir & Cockerham, 1984) F_{ST} and accounts for random chromosome sampling characteristic of Pool-Seq experiments. The pool-F_{ST} statistic ranges between 0 and 1, where a value of 0 indicates no genetic differences exists between populations, while a value of 1 means complete genetic differentiation between populations. In addition, to assess clustering patterns of pool samples, we performed Principal Component Analysis (PCA) using the whole SNP set. In a pilot analysis samples 1b, 6a, and 6b appeared as outliers (Figure S4). Considering that technical biases might have affected these samples, they were excluded from subsequent analyses.

Table 1. Collection details of the Atlantic horse mackerel samples analysed in the current project.
Abbreviations: N: North, S: South, W: West, SW: Southwest, *N*: Number of individuals, Mag: Magnetic, Med: Mediterranean.

											Maturity Stage						
Stock	**Area**	**Sample**	**Year**	***N***	**Latitude**	**Longitude**	**Extraction method**	**Pool**	***N* per**	**Pool ID**	**1**	**2**	**3**	**4**	**5**	**6**	**NA**
Western	W Ireland	1a	2016	51	54.42	-10.62	Mag Bead	1a	51	1a-WIR-2016		31	19	1			
Western	SW Ireland	1b	2016	44	51.35	-10.98	Mag Bead	1b	44	1b-WIR-2016		32	12				
Western	SW Ireland	2a	2017	46	50.20	-10.79	Mag Bead	2	62	2-WIR-2017			44	2			
Western	W Ireland	2b	2017	16	53.93	-11.09	Mag Bead	2					16				
N Sea	S North Sea	3	2016	96	54.15	3.30	Mag Bead	3	96	3-SNS-2016		88		8			
N Sea	S North Sea	4a	2017	18	54.07	2.85	Mag Bead	4	70	4-SNS-2017				18			
N Sea	S North Sea	4b	2017	21	54.03	2.90	Mag Bead	4						21			
N Sea	S North Sea	4c	2017	31	53.93	2.55	Mag Bead	4						31			
Southern	N Portugal	5a	2016	64	39.83	-9.20	Mag Bead	5a	64	5a-NPT-2016		64					
Southern	S Portugal	5b	2016	30	37.26	-8.92	Mag Bead	5b	30	5b-SPT-2016	22	5	3				
Southern	N Portugal	6a	2017	48	41.14	-9.03	Chelex	6a	47	6a-NPT-2017		47	1				
Southern	S Portugal	6b	2017	23	36.84	-8.38	Chelex	6b	48	6b-SPT-2017		18	2	3			
Southern	S Portugal	6c	2017	25	36.84	-8.10	Chelex	6b				19	6				
N African	Mauritania	7a	2016	4	20.20	-17.50	Mag Bead	7	57	7-NAF-2016		1		3			
N African	Mauritania	7b	2016	4	19.00	-17.20	Mag Bead	7						4			
N African	Mauritania	7c	2016	8	19.90	-17.60	Mag Bead	7				1		7			
N African	Mauritania	7d	2016	1	17.10	-16.60	Mag Bead	7				1					
N African	Mauritania	7e	2016	7	20.10	-17.70	Mag Bead	7					1	6			
N African	Mauritania	7f	2016	4	20.40	-17.70	Mag Bead	7				1		3			
N African	Mauritania	7g	2016	8	20.50	-17.50	Mag Bead	7				1		7			
N African	Mauritania	7h	2016	9	20.50	-17.6	Mag Bead	7				4		5			
N African	Mauritania	7j	2016	7	20.30	-17.7	Mag Bead	7						7			
N African	Mauritania	7k	2016	5	20.40	-17.7	Mag Bead	7				1		4			
Western	N Spanish Shelf	8a	2016	22	43.31	-3.46	Mag Bead	8	96	8-NSP-2016		9	12				1
Western	N Spanish Shelf	8b	2016	23	43.27	-3.21	Mag Bead	8				5	18				
Western	N Spanish Shelf	8c	2016	3	43.27	-2.42	Mag Bead	8					3				
Western	N Spanish Shelf	8d	2016	44	43.22	-2.14	Mag Bead	8				15	28	1			
Western	N Spanish Shelf	8e	2016	4	43.20	-2.10	Mag Bead	8					4				
Med	Alboran Sea	9a	2018	10	36.36	-5.12	CTAB	9	49	9-MED-2018				10			
Med	Alboran Sea	9b	2018	10	36.56	-4.55	CTAB	9						10			
Med	Alboran Sea	P9c	2018	10	36.49	-4.42	CTAB	9						10			
Med	Alboran Sea	P9d	2018	10	36.6865	-4.28	CTAB	9						10			
Med	Alboran Sea	P9e	2018	10	36.70	-3.56	CTAB	9						10			

2.5 Detection of loci putatively under selection

To identify regions of the genome with elevated genetic differences, generally interpreted as candidate signatures of natural selection, we calculated the absolute delta allele frequency (dAF) of each SNP between paired contrasts of single or grouped pools. In specific, we first calculated the mean allele frequency per SNP within each proposed group, and after, the absolute difference between the two groups. The contrasts and groupings examined were established taking in consideration geographic closeness, PCA clustering patterns, and stock divisions. The paired contrasts evaluated were:

- Each pool against all other samples
- Southern North Sea (3 and 4) vs. others (1a, 8, 5a, 5b, 9)
- Western Ireland (1a) vs. other northern samples (2, 3, 4, 8, 5a)
- Western Ireland (1a, 2) vs. other northern samples (3, 4, 8, 5a)
- Northern Spanish shelf (8) vs. other northern samples (2, 3, 4, 8, 5a)
- Southern Portugal and Alboran Sea (5b, 9) vs. all others (1a, 3, 4, 8, 5a)
- Southern Portugal and northern Africa (5b, 7) vs. all others (1a, 3, 4, 8, 5a, 9)
- "North" (1a, 2, 3, 4, 8, 5a) vs. "South" (5b, 7) groupings
- Northern Africa (7) vs. others (1a, 3, 4, 8, 5a, 5b, 9)

To identify genomic regions with consistent differentiation across various markers, we also calculated the moving (or rolling) mean of dAF values in windows of 100 SNPs for each contrast. In this way, we ruled out single SNPs that could be influenced by random effects of Pool-Seq experiments. We further explored the allele frequency pattern of the most highly differentiated SNPs at each locus and contrast across the 12 pool samples. We included here samples 1b, 6a, and 6b as it was focused on loci that were well supported in other samples. All the analyses were performed using *R* and plotting was done with the *ggplot2* package.

2.6 Individual validation of informative markers for stock assessment

The primary aim of this study was to identify a reduced and highly informative set of SNP markers that could be used for genetic stock identification. For this purpose and to validate the main findings with the Pool-Seq data, we screened a subset of the 100 most differentiated SNPs in a total of 160 individuals. In addition to confirming the allele frequencies observed in the Pool-Seq data it was also possible to undertake a preliminary analyses of population structure between the main sampling areas.

The loci included in the SNP panel were selected as follows. We started from a list of candidate SNPs with the highest dAF values from the major genomic regions of divergence in each of the main contrasts. In most cases we selected SNPs with dAF $\geq$ 0.35, but when a large number of SNPs passed this threshold we set a higher cut-off value, so we could obtain a reduced number of SNPs representative of that locus. We required that SNPs had a coverage $\geq$ 20x, a base quality $\geq$ 20, a mapping quality $\geq$ 20; that they were at least 10 bp away from an indel, more than 100 bp far from repetitive sequences, and more than 1 kb from the closest informative SNP; that alleles were equally supported by forward and reverse reads (no strand bias); that several chromosomes would be represented when that was the case; and that enough flanking sequence of good quality was available for primer design ($\pm$ 120 bp). The genomic context of target SNPs was further examined using the genome browser *IGV* (Robinson et al., 2011; Thorvaldsdóttir et al., 2013). We additionally chose a set of SNPs that were lowly undifferentiated (or "neutral") and a few SNPs that were distinctive of sample 1b, to test whether this sample was actually unique as it behaved as an outlier in pilot analysis. The

neutral SNPs were randomly selected from the chromosomes underrepresented in the paired contrasts. We required these SNPs had a depth of coverage between 40x and 200x; were at least 10 bp away from nearby SNPs and indels; had an average allele frequency between 0.4 and 0.7; and had enough flanking sequence (± 120 bp) of good quality for primer design, which was visually evaluated with *IGV*. The final split of loci per region in the 100-SNP panel was: southern North Sea (n = 28), neutral loci (n = 24), north-south break (n = 13), 1b-western Ireland (n = 10), Alboran Sea (n = 13), southern Portugal (n = 4), 1a-western Ireland (n = 4), northern Africa (n = 4) (Figure S6).

A subset of 20 individuals each was selected from 8 of the 12 samples included in the Pool-Seq analyses (Table 2) for the SNP validation. Three or four individuals per sample were genotyped twice in order to test for genotyping errors. DNA extraction and SNP genotyping was undertaken by IdentiGEN, Dublin, Ireland using their proprietary IdentiSNP genotyping assay chemistry. The protocol utilises target specific primers and universal hydrolysis probes. Following the endpoint PCR reaction different genotypes are detected using a fluorescence reader.

Only individuals with >80% genotyping success and SNPs with >80% genotyping success were retained in the analyses. Deviations from Hardy–Weinberg equilibrium and linkage disequilibrium were tested with *Genepop* 4.2 – default settings (Rousset, 2008). *Microsatellite Analyzer* (*MSA*) 4.05 was used, under default settings, to calculate pairwise F_{ST} estimates (Dieringer & Schlötterer, 2003). In all cases with multiple tests, significance levels were adjusted using the sequential Bonferroni technique (Rice 1989). Discriminant Analysis of Principal Components (DPCA) and clustering analyses were performed in *R* using the *adegenet* package for the multivariate analysis of genetic markers (Jombart, 2008). It should be noted that sample sizes were small and therefore the results of the analyses presented in section 3.6 should be viewed as preliminary until further large-scale screening is undertaken. To illustrate the potential of the markers for individual assignment for stock identification, an exploratory assignment was also conducted in *GeneClass2* (Piry et al., 2004) and the *R* package *geneplot* (McMIllan & Fewster, 2017) with the Bayesian method of Rannala and Mountain (1997).

Table 2. The horse mackerel samples included in the SNP validation analyses

Stock	Area	Sample	Pool	Year	#individuals	# repeated
Western	West of Ireland	1a	1a	2016	20	4
Western	Southwest of Ireland	1b	1b	2016	20	4
North Sea	Southern North Sea	3	3	2016	20	4
North Sea	Southern North Sea	4b	4	2017	20	4
Southern	Northern Portugal	5a	5a	2016	20	4
Southern	Southern Portugal	5b	5b	2016	20	4
North African	Mauritania	7a	7	2016	4	0
North African	Mauritania	7b	7	2016	4	1
North African	Mauritania	7c	7	2016	8	1
North African	Mauritania	7e	7	2016	4	1
Western	Northern Spanish Shelf	8d	8	2016	20	3

3. Results

3.1 Sampling and DNA Isolation

A total of 33 collections comprising 716 individual fish were included in this study (Figure 4 and Table 1). Samples were aggregated into 12 pools based on spatial and temporal proximity, thus broadly representing most of the geographical range of the species in the northeast Atlantic and the western part of the Mediterranean Sea.

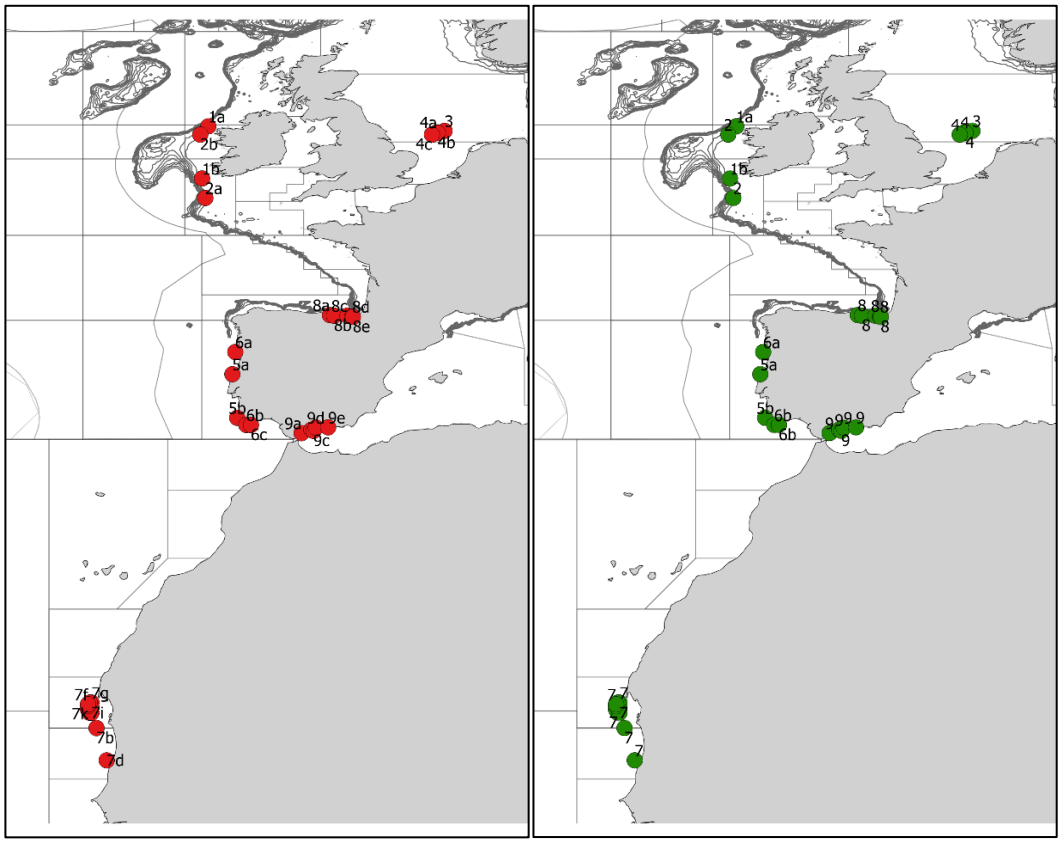


Figure 4. Sampling locations of the Atlantic horse mackerel included in this study. (Left) Sample batches collected at each location, (right) Pooled samples.

Four of the available samples corresponded to temporal replicates collected one year apart, which allowed us to examine the short-term stability of the genetic composition at these sites. Pool 2 was a mix of the replicates of the two samples collected in western Ireland (1a and 1b); pools 6a and 6b were temporal replicates of pools 5a and 5b from Northern and Southern Portugal, respectively; and pool 4 was the replicate of pool 3 from southern North Sea).

3.2 High-throughput sequencing, QC of raw reads, and read mapping

A total of 490-764 million high-quality reads were obtained for each pool. Mean read depth of coverage per pool ranged between 25.7x and 46.3x, mean mapping quality (MQ) was larger than 35 for all pools, and GC content was ~42% for most samples except for the African pool (46.6%) (Table S2).

A comparison of the mapping statistics of all pools showed that three of them (6a, 6b, 7) might be affected by technical artefacts. The two temporal replicates from Portugal (6a, 6b), which were extracted with Chelex and had a SPLAT library preparation, had a smaller mean coverage and shorter insert size (~245 bp vs. ~400-465 bp) than the other pools (Figure S2). The sample from Africa had a

flatter and wider coverage distribution, higher GC content, and higher missing rate (Figure S2) with respect to the other pools, which could be the result of certain degradation of the starting genetic material that was noticeable during DNA quantification. Given the difficulty to rule out the effect of technical biases from biological variation in these samples, they were excluded from some analyses.

3.3 Variant calling and filtering

From the three depth of coverage thresholds tested (Figure S3), we chose the range of 20x-300x because in a pilot analysis it provided a large number of SNPs and similar genetic patterns as the more stringently filtered sets. A total of ~12.8 million polymorphic biallelic SNPs passed all the quality filters and were used in the population analysis.

3.4 Population genetic structure

The large set of genetic variants here analysed indicated that overall, there are low levels of genetic differentiation among Atlantic horse mackerel populations distributed across the broad geographic area here represented (Figure 5) (global mean pool-F_{ST} = 0.007, pairwise pool-F_{ST} values ranged between 0.001 and 0.015). The genetic differences among populations constituted less than 1.5% of their entire genome.

The pairwise pool-F_{ST} values revealed a north-south genetic break along mid Portugal, distinguishing a “north” group comprising southern North Sea (3, 4), western Ireland (1a, 2), northern Spanish shelf (8) and northern Portugal (5a), from a “south” group including southern Portugal (5b), northern Africa (7), and the Alboran Sea (9) samples (Figure 5). These statistics also showed that the sample from the Alboran Sea (pool 9) was the most genetically distinct of all (pool-F_{ST} 0.01-0.015), followed by Southern Portugal (5b) and northern Africa (7), respectively (pool-F_{ST} 0.005-0.007). In contrast, the two samples collected one year apart from southern North Sea (pools 3 and 4) were the most genetically similar of all (pool-F_{ST} 0.001).

For the PCA we excluded samples 1b, 6a and 6b, as in a pilot analysis they appeared as outliers. The PCA agreed with the previous observations of a north-south break and it additionally revealed sub-structuring within the “north” and “south” groupings. The first two PCs show that the genetic differences among the samples within the “north” group (1a, 2, 3, 4, 5a, 8) are very small (all cluster together near the centre) with respect to the differences between the three samples in the “south” group (5b, 7, 9). PC1 shows that within the “south” group, genetic differences exist between the Alboran Sea (9), southern Portugal (5b) and northern Africa (7). PC2 indicates that differences also occur between northern Africa (7) and the Alboran Sea (9) and southern Portugal (5b). PC3 separates the “north” and “south” groups, being southern Portugal (5b) closer to the “north” group than northern Africa (7) and the Alboran Sea (9). PC4 distinguishes western Ireland (1a) and northern Portugal (5a) and also shows the high genetic similarity (tight clustering) between the two samples from the southern North Sea (3, 4).

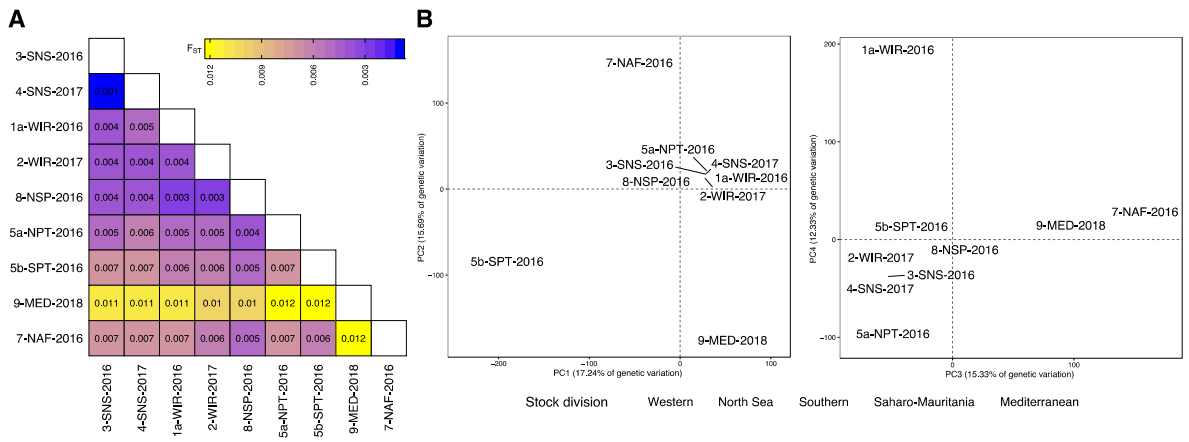


Figure 5. Population genetic structure of the 9 pool samples analysed. **A**. Pairwise pool-F_{ST} statistics, **B**. PCA of 9 pools; (left) PC1-2, (right) PC3-4.

3.5 Detection of loci putatively under selection

The genome-wide scans for the identification of candidate loci under selection revealed a number of genomic regions with elevated allele frequency differences for three contrasts: i) “north” vs. “south” groupings; ii) southern North Sea vs. others; and iii) Alboran Sea (9) vs. others.

The comparison between the “north” and “south” groups disclosed that a single large locus, likely corresponding to a chromosome structural variation (SV), underlies the north-south genetic break (Figure 6). This locus on chromosome 21 appears as a large block of SNPs with elevated allele frequency differences spanning 9.9 Mb. The large genomic size and abrupt change in allele frequencies (well-defined edges) at this locus are common characteristics of SVs with suppressed recombination (e.g. inversions). A further exploration of the allele frequency patterns of some of the most differentiated SNPs at this locus (dAF ≥ 0.72) showed that one allele occurs at high frequency among all northern samples and in the Alboran Sea; at intermediate frequencies in southern Portugal (Figure 6, inset box); and the alternative allele occurs at high frequency in northern Africa, the southernmost sample studied.

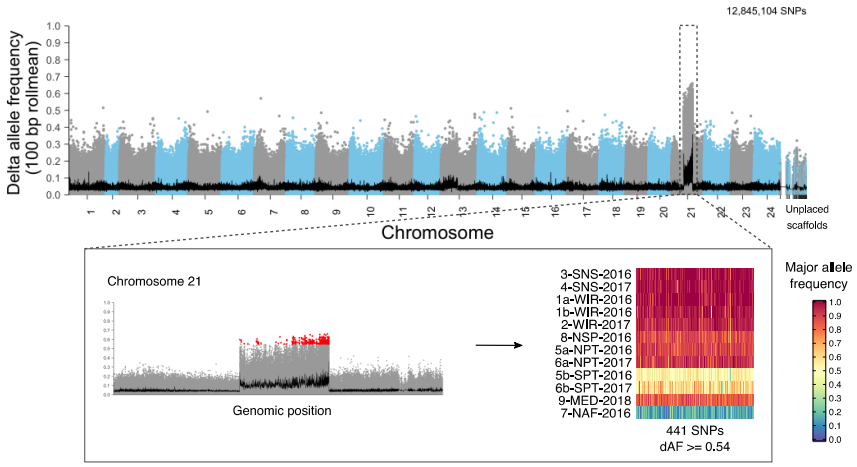


Figure 6. Manhattan plot representing the dAF of each SNPs along the genome for the north-south contrast. Each dot corresponds to a single SNP, the x-axis shows its genomic position, and the y-axis indicates its dAF frequency value for a given contrast. The line in black corresponds to the rolling mean of dAF calculated over 100 SNPs. The inset box shows a zoom-in of the putative chromosomal structural variant found in chromosome 21. The red dots correspond to the SNPs with a dAF ≥ 0.72. The heatmap plot at the right-hand side of the inset

shows the major allele frequencies of these top SNPs. In the heatmap plot, rows correspond to pool samples, and columns to SNP variants.

The comparison of the southern North Sea samples against all others disclosed that seven genomic regions distinguish this population. Two of these regions are located on chromosome 1, and the others are on chromosomes 4, 7, 11, 20, and 21 (Figure 7); they stand out as a "peak" or aggregate of SNPs with elevated differences in allele frequencies in respect to the neighbouring variants. Further examination of the allele frequencies of some of the most divergent SNPs at each locus show the large agreement in allele frequency patterns that exists between the two southern North Sea temporal replicates, and that they are distinctive of this population (Figure 7, inset boxes).

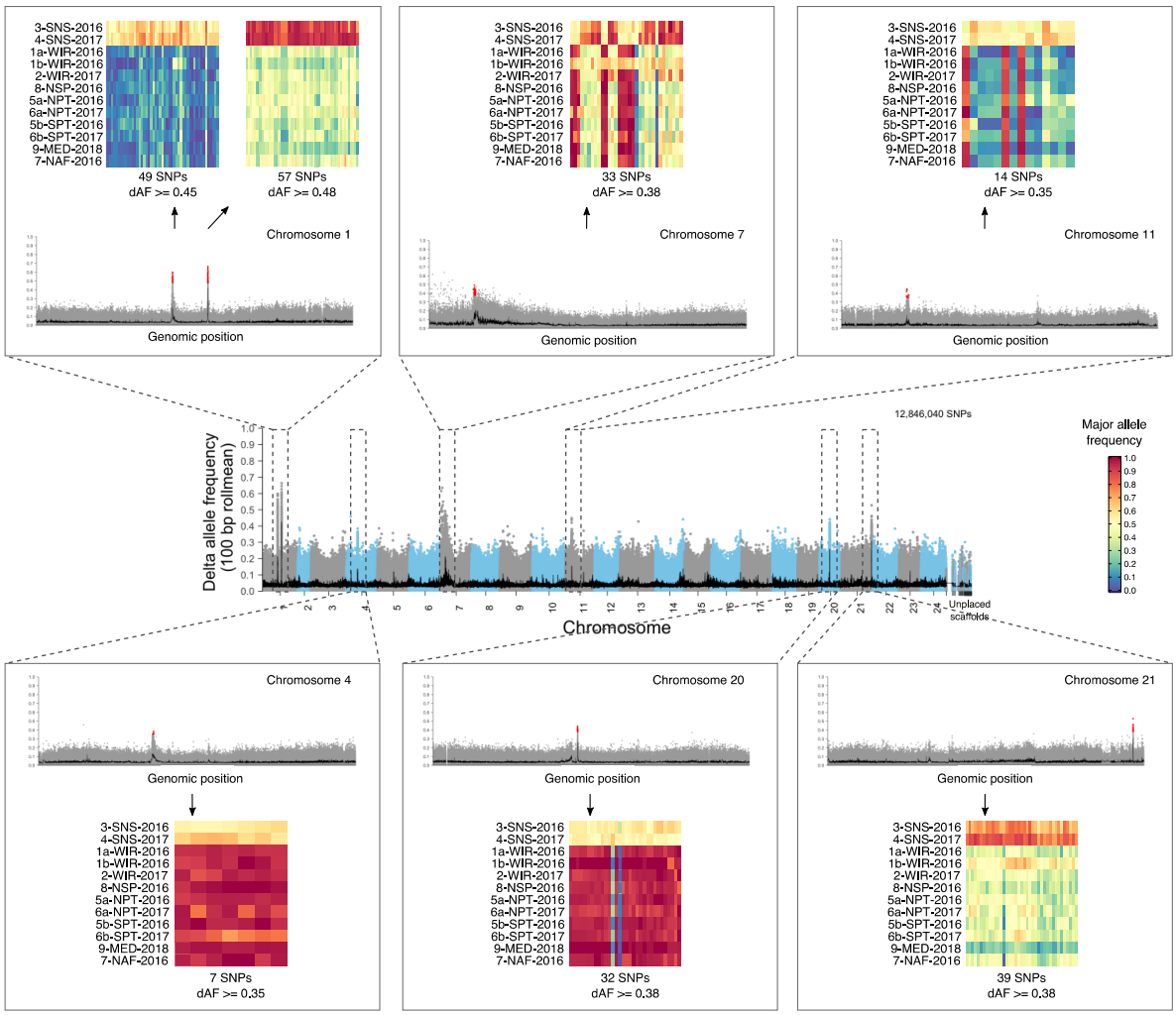


Figure 7. Manhattan plot of the dAF of each SNPs along the genome for the contrast distinguishing the southern North Sea samples. Each dot is a single SNP. The line in black corresponds to the rolling mean of dAF over 100 SNPs. The inset boxes show a zoom-in into the 7 genomic regions across chromosomes 1, 4, 7, 11, 20, and 21, characteristics of the North Sea samples. The red dots in the zoomed dAF profile of each chromosome correspond to the most highly differentiated SNPs per genomic region. The heatmap plot at the right-hand side of the inset shows the major allele frequencies of these top SNPs. In the heatmap plot, rows correspond to pool samples, and columns to SNP variants.

The contrast of the Alboran Sea sample against all others showed that two regions, one on chromosome 5 and another on chromosome 21, distinguish this sample from other samples (Figure 8). In this case the "peaks" of divergence were not as evident as in the other contrasts, for which it was necessary to focus more on the patterns shown by the rolling mean in dAF values. The

examination of allele frequencies of the most differentiated SNPs showed that the Alboran Sea sample had a characteristic allele frequency pattern.

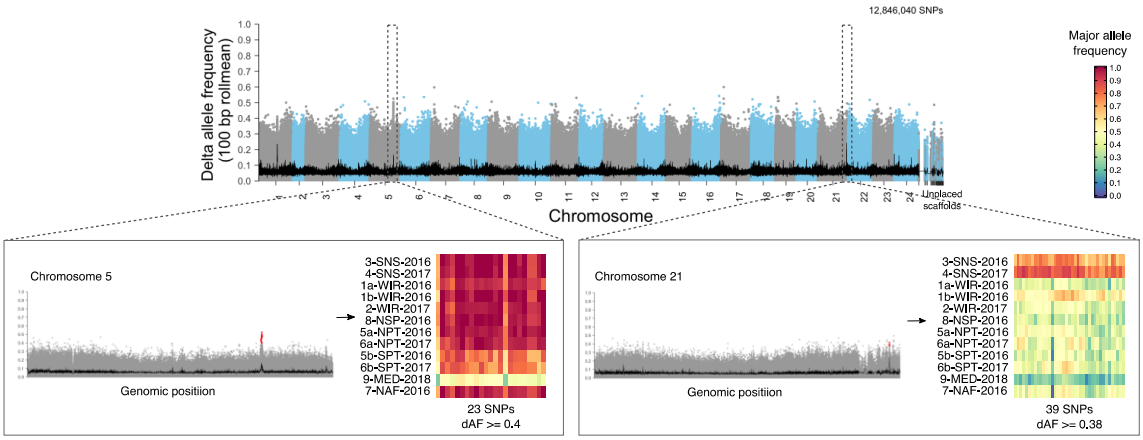


Figure 8. Manhattan plot of the dAF of each SNPs along the genome for the contrast distinguishing the Alboran Sea (from the western part of the Mediterranean Sea) sample. Each dot is a single SNP. The line in black corresponds to the rolling mean of dAF over 100 SNPs. The inset boxes show a zoom-in into the two genomic regions in chromosomes 5 and 21 showing high differentiation between the Alboran Sea sample and other samples. The red dots in the zoomed dAF profile of each chromosome correspond to the most highly differentiated SNPs per genomic region. The heatmap plot at the right-hand side of the inset shows the major allele frequencies of these top SNPs. In the heatmap plot, rows correspond to pool samples, and columns to SNP variants.

3.6 Individual validation of informative markers for stock assessment

The strong correlation between population allele frequencies obtained with individual genotyping and with Pool-Seq confirms the main genomic regions of divergence discovered with Pool-Seq (Figure S5). A total of 72 out of the 100 SNPs included in the panel had a genotyping success >80% (Table 3). Of these, six SNPs had indication of deviation from Hardy-Weinberg Equilibrium (HWE), two markers (12_3119866 and 17_972744) were not polymorphic and one had evident scoring errors (24_5252083). After removing these nine markers, the resulting dataset had 63 SNPs and 157 out of 160 individuals with a genotyping success >80%.

Table 3. Details of the 100 SNPs tested in the validation analyses. The SNPs highlighted in red did not reach the 80% genotyping success threshold or failed to amplify. The SNPs highlighted in orange deviated from HWE, were not polymorphic or had scoring errors and were removed from the analyses. '*LD*' indicates significant linkage disequilibrium between samples and '*Assumed*' indicates assumed LD based on chromosome position. * indicates SNPs that were included in the 17 SNP dataset.

SNP Name	>80% success	Chromosome	Position	Contrast	LD Group	Comment
1_17504018*	Yes	1	17504018	Southern North Sea	Assumed	
1_17506941	Yes	1	17506941	Southern North Sea	LD	
1_17510324	Yes	1	17510324	Southern North Sea	LD	
1_17517550	Yes	1	17517550	Southern North Sea	LD	
1_17521852	Yes	1	17521852	Southern North Sea	LD	
1_17523218	Yes	1	17523218	Southern North Sea	LD	
1_17525646	Yes	1	17525646	Southern North Sea	Assumed	
1_17558501	Yes	1	17558501	Southern North Sea	LD	
1_22046469	Yes	1	22046469	Southern North Sea	LD	
1_22046756	Yes	1	22046756	Southern North Sea	LD	
1_22047461	Yes	1	22047461	Southern North Sea	LD	
1_22049353	Yes	1	22049353	Southern North Sea	LD	
1_22053057*	Yes	1	22053057	Southern North Sea	LD	
1_22081696	No	1	22081696	Southern North Sea	Assumed	
3_2811572	No	3	2811572	Neutral markers		
3_18949602	No	3	18949602	Neutral markers		
3_18951336	Yes	3	18951336	Neutral markers		
3_33715024	No	3	33715024	Neutral markers		
4_13086614*	Yes	4	13086614	Southern North Sea	LD	
4_13088818	Yes	4	13088818	Southern North Sea	LD	
4_13098092	Yes	4	13098092	Southern North Sea	LD	
5_22983273	No	5	22983273	Western Ireland (1a)		
5_28197435	Yes	5	28197435	Med and/or S Portugal		
5_28205448	Yes	5	28205448	Med and/or S Portugal		
5_28240764	Yes	5	28240764	Med and/or S Portugal		
5_28240785	Yes	5	28240785	Med and/or S Portugal		
5_28241356*	Yes	5	28241356	Med and/or S Portugal		
5_28242757	No	5	28242757	Med and/or S Portugal		
5_28243095	Yes	5	28243095	Med and/or S Portugal		
5_28274875	No	5	28274875	Med and/or S Portugal		
6_18368752*	Yes	6	18368752	Neutral markers		
6_24275858	No	6	24275858	Neutral markers		
6_33295851*	Yes	6	33295851	Neutral markers		
7_5053296*	Yes	7	5053296	Southern North Sea		
7_5108289	Yes	7	5108289	Southern North Sea		
8_2410897	No	8	2410897	Neutral markers		
8_3426603*	Yes	8	3426603	Neutral markers		
11_6942036	Yes	11	6942036	Southern North Sea		Out of HWE in 2 pops
12_3119866	Yes	12	3119866	Neutral markers		Not polymorphic
12_10994158	No	12	10994158	Neutral markers		
12_27660258	Yes	12	27660258	Neutral markers		Out of HWE in 3 pops
13_4844455	No	13	4844455	Western Ireland (1b)		
13_4874422	Yes	13	4874422	Western Ireland (1b)	LD	
13_4874692	Yes	13	4874692	Western Ireland (1b)	LD	
13_4874725	Yes	13	4874725	Western Ireland (1b)	LD	
13_5015377*	Yes	13	5015377	Western Ireland (1b)		
13_5092546	Yes	13	5092546	Western Ireland (1b)		
16_22440492	No	16	22440492	Africa		
17_955542	No	17	955542	Western Ireland (1b)		
17_955717	Yes	17	955717	Western Ireland (1b)		Out of HWE in 1 pop
17_961283	No	17	961283	Western Ireland (1b)		
17_972744	Yes	17	972744	Western Ireland (1b)		Not polymorphic
18_4093892*	Yes	18	4093892	Africa		
19_4188265	No	19	4188265	Neutral markers		
19_4189387	No	19	4189387	Neutral markers		
19_4194438	No	19	4194438	Neutral markers		
19_13550308	No	19	13550308	Neutral markers		
20_11636865	Yes	20	11636865	Southern North Sea	LD	
20_11638825*	Yes	20	11638825	Southern North Sea	LD	
20_11640406	Yes	20	11640406	Southern North Sea	LD	
20_11643211	Yes	20	11643211	Southern North Sea	LD	
20_11644062	Yes	20	11644062	Southern North Sea	LD	
20_11647497	Yes	20	11647497	Southern North Sea	LD	
20_11647537	Yes	20	11647537	Southern North Sea	LD	
20_11649644	Yes	20	11649644	Southern North Sea	LD	
21_13901383	Yes	21	13901383	North-South pattern		
21_15195721	Yes	21	15195721	Southern Portugal		
21_15619806*	Yes	21	15619806	North-South pattern		
21_16093398	Yes	21	16093398	North-South pattern		
21_18106603	Yes	21	18106603	North-South pattern		
21_19507025	Yes	21	19507025	Southern Portugal		Out of HWE in 1 pop
21_20477335	Yes	21	20477335	North-South pattern		

Table 3. *Continuation*.

SNP Name	>80% success	Chromosome	Position	Contrast	LD Group	Comment
21_20646321	Yes	21	20646321	North-South pattern	LD	
21_20838721	Yes	21	20838721	North-South pattern	LD	
21_21340446	Yes	21	21340446	North-South pattern	LD	
21_21591928	Yes	21	21591928	North-South pattern		
21_21801450	Yes	21	21801450	North-South pattern		
21_22552517	Yes	21	22552517	North-South pattern		
21_23412586*	Yes	21	23412586	North-South pattern	LD	
21_23420067	Yes	21	23420067	North-South pattern	LD	
21_34276436	No	21	34276436	Southern Portugal		
21_34279224	No	21	34279224	Southern Portugal		
21_34570675	Yes	21	34570675	Med and/or S Portugal	LD	
21_34571601	No	21	34571601	Med and/or S Portugal		
21_34571721	Yes	21	34571721	Med and/or S Portugal	LD	
21_34573582*	Yes	21	34573582	Med and/or S Portugal	LD	
21_34578009	No	21	34578009	Med and/or S Portugal		
22_253248	No	22	253248	Africa		
22_29332559	Yes	22	29332559	Western Ireland (1a)		Out of HWE in 5 pops
22_29369048*	Yes	22	29369048	Western Ireland (1a)		
22_29400293	Yes	22	29400293	Western Ireland (1a)		
24_2630784	No	24	2630784	Neutral markers		
24_2631095	No	24	2631095	Neutral markers		
24_3769194	No	24	3769194	Neutral markers		
24_5252083	Yes	24	5252083	Africa		Scoring error
24_5255627	No	24	5255627	Neutral markers		
24_10305770*	Yes	24	10305770	Neutral markers		
24_10306442	Yes	24	10306442	Neutral markers		Out of HWE in 1 pop
24_14507474	No	24	14507474	Neutral markers		
24_19228299*	Yes	24	19228299	Neutral markers		

As expected, analyses of linkage disequilibrium (LD) indicated significant linkage between a number of SNPs located in close proximity on the same chromosomes (Table 3). Though LD was not statistically significant in some cases (e.g. SNPs on chromosome 5), these were considered to be linked due to the closeness of the SNPs. In order to identify the most informative SNPs for discriminating the samples, the F_{ST} per locus was analysed by marker and by population (Figure 9). The most informative SNP (highest average F_{ST}) per linkage group was retained, yielding a 17 SNP dataset comprising 155 out of 160 individuals with a genotyping success >80%. Further analyses were conducted with both the 63_SNP and the 17_SNP datasets (individual genotypes in each SNP set are shown in Figure S7).

There was no significant genetic differentiation between the North Sea temporal replicates or between the two west of Ireland samples (Table 4). There was also no significant genetic differentiation between the northern Spanish shelf sample, the northern Portugal sample and the two west of Ireland samples (Table 4). Discriminant Analysis of Principal Components (DAPC) and clustering analyses of the 63_SNP and 17_SNP datasets indicated the same pattern as the F_{ST} analyses with the North Sea temporal replicates clustering together, the west of Ireland, northern Spanish shelf and northern Portugal samples clustering together and the southern Portugal and northern African samples forming two separate clusters (Figure 10). Due to the lack of genetic differentiation, the two North Sea samples were combined into one sample and the two west of Ireland samples were combined into one sample for further analyses.

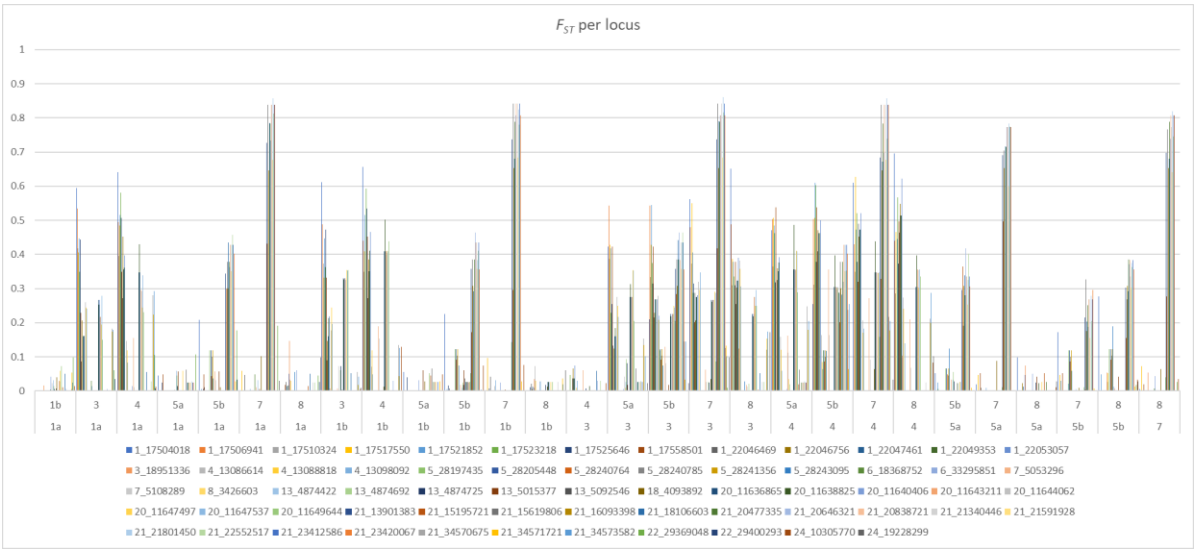


Figure 9. The pairwise F_{ST} per locus for the 63_SNP dataset

Table 4. Pairwise multi-locus F_{ST} (above the diagonal) and associated *P*-values (below the diagonal) for the 63_SNP dataset (top panel) 17_SNP dataset (bottom panel). *P*-values highlighted in red were still significant after sequential Bonferroni correction.

	1a	1b	3	4	5a	5b	7	8
1a		0.004	0.195	0.260	-0.004	0.135	0.361	0.004
1b	0.28		0.198	0.265	-0.006	0.124	0.352	-0.003
3	0.00	0.00		-0.006	0.180	0.243	0.417	0.218
4	0.00	0.00	0.60		0.241	0.287	0.446	0.286
5a	0.61	0.71	0.00	0.00		0.101	0.323	-0.002
5b	0.00	0.00	0.00	0.00	0.00		0.080	0.111
7	0.00	0.00	0.00	0.00	0.00	0.00		0.334
8	0.29	0.54	0.00	0.00	0.51	0.00	0.00	

	1a	1b	3	4	5a	5b	7	8
1a		0.016	0.138	0.196	0.004	0.088	0.241	0.009
1b	0.09		0.121	0.190	0.003	0.075	0.221	-0.002
3	0.00	0.00		-0.002	0.102	0.137	0.297	0.168
4	0.00	0.00	0.53		0.154	0.187	0.340	0.233
5a	0.32	0.35	0.00	0.00		0.033	0.183	0.006
5b	0.00	0.00	0.00	0.00	0.01		0.055	0.068
7	0.00	0.00	0.00	0.00	0.00	0.00		0.209
8	0.17	0.52	0.00	0.00	0.26	0.00	0.00	

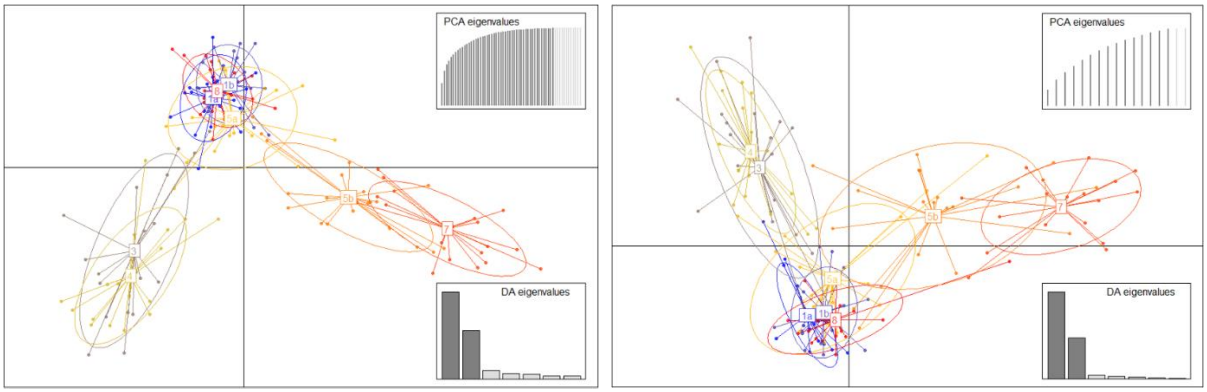


Figure 10. Discriminant Analysis of Principal Components of the 63_SNP dataset (left panel) and the 17_SNP dataset (right panel).

Membership probability plots of the two datasets also indicated the close affinity between the west of Ireland samples and the northern Spanish shelf and northern Portugal samples. A degree of mixing or admixture is evident in a small number of individuals (3-4) in the North Sea sample that have a high probability of originating from the western group. Similarly, the southern Portugal sample had a number of outliers which appear to originate from the western group (n=3) or from the African group (n=2).

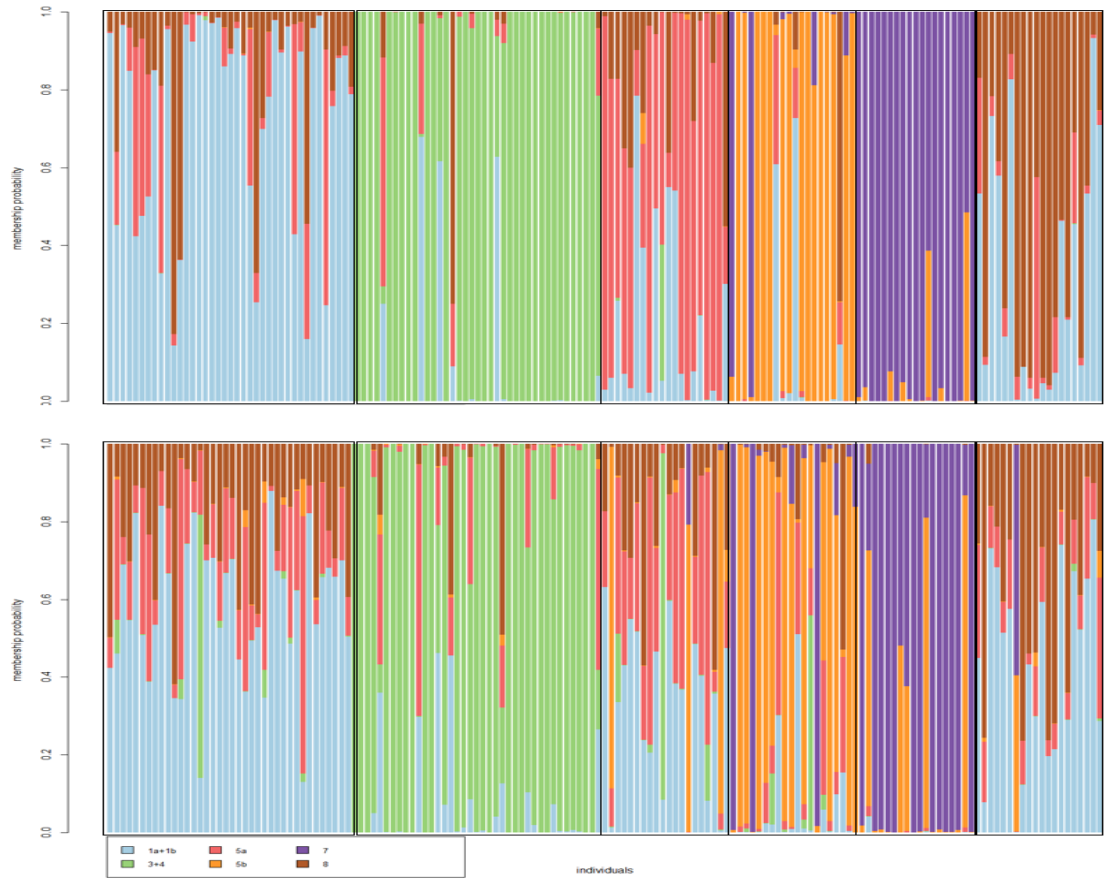


Figure 11. Membership probability plots the 63_SNP dataset (top panel) and the 17_SNP dataset (bottom panel). Samples 1a and 1b are combined into one sample and samples 3 and 4 are combined into one sample. Samples are delineated by the black boxes.

An exploratory assignment was conducted for illustration purposes using a combined 1a, 1b, 8 sample to represent what is currently considered to be the Western Stock and a combined 3, 4 sample to represent the North Sea. Only the 17_SNP dataset was used in order to avoid the violation of the assumption of independent markers, which is a prerequisite of the Rannala and Mountain approach. *Geneplot* indicated a self-assignment rate of 93% and *geneclass2* a self-assignment rate of 95%, indicating significant power to discriminate between mixed samples from these areas.

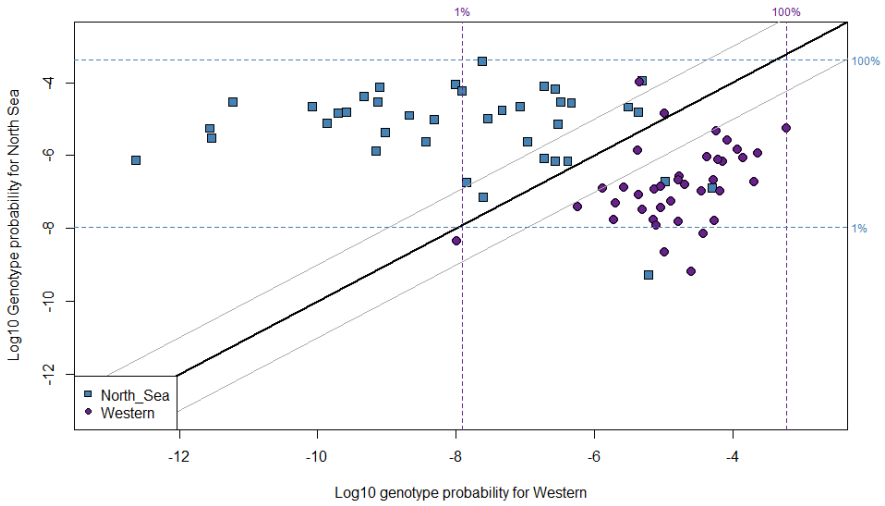


Figure 12. Plot generated with *genePlot* based on the 17_SNP dataset of the Western and North Sea stock samples. Each point represents an individual. The horizontal axis shows the posterior log-probability of obtaining each individual's genotype from the Western stock; the vertical axis shows the same, but with respect to the North Sea stock. The thick diagonal line shows equal probability with respect to Western and the North Sea. The vertical dashed lines shows the 0% and 100% percentile lines, that is, the minimum and maximum log-genotype probability, for the Western stock; the horizontal lines show the 0% and 100% percentile lines for the North Sea population.

4. Discussion

This study represents the largest and most comprehensive genetic assessment of the Atlantic horse mackerel to date. The combination of extensive geographic sampling and analysis of a large number of SNP markers derived from whole-genome sequencing, provided a powerful dataset that allowed us to discover, for the first time, genomic regions supporting population subdivision within the species. The genetic differences largely separate five groups: i) southern North Sea, ii) western Ireland - northern Spanish shelf - northern Portugal, iii) southern Portugal, iv) Alboran Sea/Mediterranean, and v) northern Africa. With the exception of the Southern stock, these genetic-based subdivisions are in agreement with the main horse mackerel stocks proposed by the HOMSIR project using morphometry, parasites, and life history traits (Abaunza et al., 2008). Our genetic data suggest that the samples from the southern stock in Portuguese waters do not come from a single biological population. The samples from northern Portugal appear to be genetically closer to the Western stock, while samples from southern Portugal form their own group. Further wide scale sampling is required to confirm these findings and assess the spatial and temporal trends in mixing between these areas. We additionally demonstrated that 63 of the most genetically differentiated SNP markers tag the genetic subdivisions and, thus, could be used as a genetic tool to inform the appropriate level of data collation for fisheries stock assessment. In fact, using a reduced panel of 17 markers, we demonstrated that it is possible to differentiate between individuals collected in the North Sea and Western stocks with a potential accuracy up to 95%.

Population structuring detected at loci putatively under selection

Genetic analysis of horse mackerel revealed that populations distributed across the broad geographic area spanning from the North Sea to northern Africa (Figure 5) differ by less than 1.5% of their DNA (Global mean pool-F_{ST} = 0.007, pairwise pool-F_{ST} values ranged between 0.001 and 0.015). This result indicates that gene flow occurs across the distribution range of the species. The observed genetic differences, despite representing a small fraction of the genome, are highly significant as they correspond to outlier SNPs putatively under selection and support population structuring within the species. A pattern of low genome-wide differentiation at neutral loci and high differentiation at adaptive loci is becoming a relatively common observation among various highly dispersive marine species inhabiting heterogeneous environments [e.g. Atlantic cod (Clucas et al., 2019); Atlantic herring (Lamichhaney et al., 2017)]. Many of these species, including the horse mackerel (Abaunza et al., 2008; Bozano et al., 2015; Cimmaruta et al., 2008; Farrell & Carlsson, 2018; Healey et al., 2020), were previously assumed to be panmictic, largely because prior genetic techniques did not provide enough genomic resolution. New genomic sequencing techniques enable the thorough examination of the genetic variation of non-model species and are revealing unprecedented levels of structuring, as we accomplished here for the horse mackerel. The large population sizes and high dispersal and gene flow presumed to be characteristic of numerous marine species may explain the low levels of genome-wide structuring observed, as the role of genetic drift in population structuring becomes negligible in these circumstances. The presence of well-defined parts of the genome showing high differentiation, so called “genomic islands of divergence or speciation” are generally associated with ecological adaptation or reproductive isolation (Seehausen et al., 2014; Turner et al., 2005). Theory predicts that when genetic variants are advantageous in a local environment, natural selection would favour their frequency in the local population (Yeaman & Whitlock, 2011). Thus, when different populations are locally adapted to heterogenous environments, it would be expected to see large differences in allele frequencies between them. This scenario goes in line with the fact that the horse mackerel exhibits a broad spatial distribution encompassing heterogeneous environments, for which, populations should be exposed to diverse selective pressures that can promote genetic differentiation, and thus, local adaptation.

Indeed, we hypothesize that the large chromosomal structural variant (9.9 Mb) underlying the cryptic north-south genetic break discovered here for the horse mackerel along mid Portugal, is associated with differential responses of populations to contrasting environmental conditions. Interestingly, a similar genetic pattern has also been observed in the boarfish (*Capros aper*) (Farrell et al., 2016), a pelagic fish with overlapping distribution and similar life-history characteristics in the northeast Atlantic. This suggests that a major biogeographic barrier may exist in Portugal waters, which could be leading to differentiation of biota inhabiting this area.

The structural variant exhibits high frequency of homozygotes for one allele among populations from the "north" (southern North Sea, west of Ireland, northern Spanish shelf, northern Portugal) and the Alboran Sea; heterozygotes are predominant in southern Portugal; and homozygotes for the alternative allele are in high frequency in the "south", at coastal areas near Mauritania, northern Africa. These contrasting allele frequency patterns are in concordance with differences in sea water conditions at the local spawning peak in each area. For example, oceanographic data collected in previous horse mackerel egg surveys (ICES, 2019) suggest that reproduction along the west of Ireland and the northern Spanish shelf may occur at temperatures around 12.5-14°C. Similarly, reproduction at the northern coast of Portugal may occur at sea water temperatures around 12.5° and also at lower salinities associated with freshwater discharge from rivers. In contrast, reproduction at the southern coast of Portugal may happen at warmer sea water temperatures around 17° and higher salinity with to respect to the northern coast of Portugal (ICES, 2019).

Out of the 12 samples included in this study, the sample from the Alboran Sea, at the western part of the Mediterranean Sea, was the most genetically distinct of all. This result may be explained by the ecological (Coll et al., 2010; Emig & Geistdoerfer, 2004) and geological (Garcia-Castellanos et al., 2009) differences existing between the Mediterranean Sea and the Atlantic Ocean. Moreover, the genetic data supports the consideration of the Mediterranean Sea as a separate stock, as proposed by the HOMSIR project based on morphometry, otoliths, and life history traits (Abaunza et al., 2008). The genetic distinctiveness of the Alboran Sea sample suggests that it likely constitutes a separate population, although its genetic closeness with the sample from southern Portugal indicates that gene flow may occur between these two areas. This observation is also in agreement with data collected in the HOMSIR project, indicating the mixed nature of the Alboran Sea populations (Abaunza et al., 2008).

Our genetic analysis provides evidence that the North Sea stock represents a distinct population. As many as 7 specific genomic regions distinguished the southern North Sea samples. The allele frequency patterns at these genomic regions were nearly identical between the 1-year temporal replicates, which also showed the smallest genome-wide differentiation of the 12 samples analysed (pool-F_{ST} 0.001). The North Sea samples were the northeastern most samples included in this study. Thus, we hypothesize that the observed genetic differentiation may be associated with local adaptation to colder sea water conditions experienced during spawning or at early life-history stages. We expect that further gene annotation of the novel horse mackerel genome, will help understand the putative role of these genomic regions in the differentiation of the North Sea stock. Regardless, a subset of the top outlier SNPs distinguishing the North Sea samples could be used for conservation and management purposes, as these genetic markers could help elucidate the extent of mixing between the Western and North Sea stocks along the English Channel and in ICES area 4a in the northern North Sea.

The samples from the Western stock, west of Ireland and the northern Spanish shelf, and the northern part of the Southern stock, northern Portugal, appear to form a genetically close group. This result

lends support to the inclusion of the Spanish shelf in the Western stock as proposed by the HOMSIR project, and also points to the need of an extended genetic study along the Spanish shelf and northern Portugal to determine whether the southern boundary of the Western stock should be extended.

Individual genotyping confirms Pool-Seq findings and constitute an informative SNP panel

The individual genotype data for the subset of samples corroborate the main results of the Pool-Seq analyses (Figure S4). The same pattern of sample clustering was observed with temporally stable samples in the North Sea that were distinct from all others. The two samples collected west of Ireland did not display any significant genetic differentiation between themselves or the northern Spanish Shelf sample. The northern Portuguese sample was also closely affiliated with these western samples and could not be robustly separated based on the reduced marker panels. The southern Portuguese samples formed a separate cluster, however there was evidence of mixing between this and the northern Portuguese group. As expected, the outlier group consisting of the African samples was significantly differentiated to all other samples but most closely related to the most geographically close sample in southern Portugal. Whilst these results should be treated with caution, as the sample sizes were small and temporal stability was not tested in all populations, they do prove the potential for using the reduced marker panels to investigate the population structure of horse mackerel on a larger scale.

Limitations and recommendations

While this study made important contributions to our understanding of the population structuring of the horse mackerel, we acknowledge there is room for improvement and emphasize the importance of follow-up studies. Firstly, the sampling, conducted over three consecutive years and three spawning seasons, while it covered a large area of the distribution of the species, is spatially and temporally limited. A more extensive spatial sampling within each stock area could, for instance, help identify the boundaries between the Western and Southern stocks, and between the Western and North Sea stocks. Repeated genetic monitoring (e.g. every one or two years) are necessary to assess the long-term stability of genetic sub-divisions. The Mediterranean Sea was a notable exclusion, as only a single sample from the Alboran Sea was studied. Whilst analysis of this sample indicates limited connectivity with the adjacent southern Portuguese samples, it does not enable any further conclusions the be drawn regarding population structure within the Mediterranean Sea. Secondly, whilst every effort was made to collect spawning fish from each putative stock this proved to be difficult in some areas and as such the best available alternative samples were included. Future sampling efforts should focus both on the collection of spawning baseline samples from each of the putative populations and also the collection of potentially mixed samples outside of the spawning season. Lastly, while the Pool-Seq approach is a powerful method to perform genome scans, it is sensitive to poor DNA sample quality, and variation in laboratory procedures such as pooling and library preparation. Thus, high quality DNA and standard laboratory procedures among samples are highly recommended to minimize technical biases.

5. Acknowledgements

The authors would like to thank Martin Pastoors, Maria Manuel Angélico, Finlay Burns, Gersom Costas, Cindy Van Damme, Cristina Nunes, Brendan O'Hea, Ciaran O'Donnell, Michael O'Malley and all scientists and crew on the survey and commercial vessels involved in sampling. We thank to the SNP&SEQ Technology Platform and to UPPMAX in Uppsala for the provided service in high-throughput sequencing and computational infrastructure under the SNIC projects 2020/5-36 and 2020/16-14. These facilities are part of the National Genomics Infrastructure (NGI) of Sweden and Science for Life

Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

6. References

Abaunza, P., Gordo, L., Karlou-Riga, C., Murta, A., Eltink, A.T.G.W., Santamaria, M.T.G., Zimmermann, C., Hammer, C., Lucio, P., Iversen, S.A., Molloy, J., and Gallo, E. 2003. Growth and Reproduction of Horse Mackerel, *Trachurus trachurus* (Carangidae). Fish Biology and Fisheries, 13:27–61.

Abaunza, P., Murta, A. G., Campbell, N., Cimmaruta, R., Comesaña, A. S., Dahle, G., García Santamaría, M. T., Gordo, L. S., Iversen, S. A., MacKenzie, K., Magoulas, A., Mattiucci, S., Molloy, J., Nascetti, G., Pinto, A. L., Quinta, R., Ramos, P., Sanjuan, A., Santos, A. T., … 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. https://doi.org/10.1016/j.fishres.2007.09.022

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

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

Borges, M.F. and Gordo, L.S. 1991. Spatial distribution by season and some biological parameters of horse mackerel (*Trachurus trachurus*) in the Portuguese continental waters (Division IXa). ICES Document CM 1991/H: 54. 16 pp.

Bozano, Mariani, Barratt, Sacchi, Boufana, & Coscia. (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. https://doi.org/10.3318/bioe.2015.20

Broad Institute. (2018). *Picard tools*. http://broadinstitute.github.io/picard/

Brunel, T., Farrell, E.D., Kotterman, M., Kwadijk, C., Verkempynck, R., Chen, C and 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 57 pp.

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. https://doi.org/10.1016/j.fishres.2007.09.030

Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications*, *12*(10), 1971–1987. https://doi.org/10.1111/eva.12861

Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., Ballesteros, E., Bianchi, C. N., Corbera, J., Dailianis, T., Danovaro, R., Estrada, M., Froglia, C., Galil, B. S., Gasol, J. M., Gertwagen, R., Gil, J., Guilhaumon, F., Kesner-Reyes, K., … Voultsiadou, E. (2010). The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. *PLoS ONE*, *5*(8), e11842. https://doi.org/10.1371/journal.pone.0011842

Dieringer, D. & Schlötterer, C. 2003 Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Resources, 3: 167–169.

Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, *36*(16), e105–e105. https://doi.org/10.1093/nar/gkn425

Dransfeld, L., Dwane, O., Molloy, J., Gallagher, S., and Reid, D. G. 2005. Estimation of mackerel (*Scomber scombrus* L., 1758) and horse mackerel (*Trachurus trachurus* L., 1758) daily egg production outside the standard ICES survey area. ICES Journal of Marine Science, 62: 1705–1710.

Emig, C. C., & Geistdoerfer, P. (2004). The Mediterranean deep-sea fauna: historical evolution, bathymetric variations and geographical changes. *Carnets de Géologie (Notebooks on Geology)*, *Articles*. https://doi.org/10.4267/2042/3230

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.

https://doi.org/10.1093/bioinformatics/btw354
Farrell, E. D., & Carlsson, J. (2018). *Genetic stock Identification of Northeast Atlantic Horse mackerel, Trachurus trachurus*. A report prepared for the members of the Northern Pelagic Working Group. 40pp.
Farrell, E. 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. https://doi.org/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. https://doi.org/10.1371/journal.pone.0048588
Garcia-Castellanos, D., Estrada, F., Jiménez-Munt, I., Gorini, C., Fernàndez, M., Vergés, J., & De Vicente, R. (2009). Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature*, *462*(7274), 778–781. https://doi.org/10.1038/nature08555
Gordo, L.S., Costa, A., Abaunza, P., Lucio, P., Eltink, A.T.G.W. and Figueiredo, I. 2008. Determinate versus indeterminate fecundity in horse mackerel. Fisheries Research, 89: 181–185.
Healey, A. J. E., Farthing, M. W., Nunoo, F. K. E., Potts, W. M., Sauer, W. H. H., Skujina, I., King, N., Becquevort, S., Shaw, P. W., & 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. https://doi.org/10.1111/jfb.14276
Heessen, H.J.L., Daan, N. and Ellis JR. 2015. Fish atlas of the Celtic Sea, North Sea, and Baltic Sea. Heessen HJL, Daan N, Ellis JR, editors. Wageningen Academic Publishers; 2015. 572 p.
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. https://doi.org/10.1534/genetics.118.300900
ICES. 1991. Report of the Horse Mackerel (Scad) Otolith Reading Workshop. Lisbon, 21–27 November 1990. ICES C.M. 1991/H:59, 59 pp.
ICES. 1995. Report of the Working Group on the assessment of mackerel, horse mackerel, sardine and anchovy. 21 June-1 July 1994, ICES Headquarters, Copenhagen. ICES CM 1995/Assess:2: Part 1: 165pp.
ICES. 1998. Report of the Working Group on the assessment of mackerel, horse mackerel, sardine and anchovy. 9-18 September 1997, ICES Headquarters, Copenhagen. ICES CM 1998/Assess:6: Part 1: 176pp.
ICES. 2015. Report of the Workshop on Maturity Staging of Mackerel and Horse Mackerel (WKMSMAC2), 28 September–2 October 2015, Lisbon, Portugal. ICES CM 2015/SSGIEOM:17. 93 pp.
ICES. 2017. Final Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys. WGMEGS Report 2017 24-28 April 2017. Vigo, Spain. ICES CM 2017/SSGIEOM:18. 134 pp.
ICES. (2019). Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS). In *ICES Scientific Reports: Vol. 1:66*. https://doi.org/http://doi.org/10.17895/ices.pub.5605
Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24: 1403-1405.
Kasapidis, P. and Magoulas, A. 2008. Development and application of microsatellite markers to address the population structure of the horse mackerel *Trachurus trachurus*. Fisheries Research, 89: 132–135.
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. https://doi.org/10.1534/genetics.110.123059
Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., Singh, R., Ruzzante, D. E., & 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. https://doi.org/10.1073/pnas.1617728114

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Macer, C.T. 1974. The reproductive biology of the horse mackerel *Trachurus trachurus* (L.) in the North Sea and English Channel. Journal of Fish Biology, 6: 415-438.

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 University. 12pp.

McMillan, L.F. and Fewster, R.M. 2017. Visualizations for Genetic Assignment Analyses using the Saddlepoint Approximation Method. Biometrics 73, 1029–1041.

Ndjaula, H.O.N., Hansen, T., Kruger-Johnsen, M. and Kjesbu, O. S. 2009. Oocyte development in captive Atlantic horse mackerel *Trachurus trachurus*. ICES Journal of Marine Science, 66: 623–630.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), btv566. https://doi.org/10.1093/bioinformatics/btv566

Piry, S., Alapetite, A., Cornuet, J-M., Paetkau, D., Baudouin, L., Estoup, A. 2004. geneclass2: a software for genetic assignment and first-generation migrant detection. Journal of Heredity, 95: 536–539.

R Core Development Team. (2020). *R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing*.

Rannala, B. and Mountain, J. L. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. USA 94: 9197-9221.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution, 43: 223–225.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rousset, F. 2008 Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources, 8: 103-106.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic Islands of Speciation in Anopheles gambiae. *PLoS Biology*, *3*(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Van Damme, C.J.G., Thorsen, A., Fonn, M., Alvarez, P., Garabana, D., O'Hea, B., Perez, J.R. et al. 2014. Fecundity regulation in horse mackerel. ICES Journal of Marine Science, 71: 546–558.

Vartia, S., Collins, P.C., Cross, T.F., FitzGerald, R.D., Gauthier, D.T., McGinnity, P., Mirimin, L. & Carlsson J. 2014. Multiplexing with three-primer PCR for rapid and economical microsatellite validation. Hereditas, 151: 43–54.

Vartia, S., Villanueva, J.L., ., Finarelli, J., Farrell, E.D., Hughes, G., Carlsson, J.E.L., Collins, P.C., Gauthier, D.T., McGinnity, P., Cross, T.F., FitzGerald, R.D., Mirimin, L., Cotter, P. & Carlsson, J. 2016. A

novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. Royal Society Open Science, 3 DOI: 10.1098/rsos.150565.

Waldron, M.E. and Kerstan, M. 2001. Age validation in horse mackerel (*Trachurus trachurus*) otoliths. ICES Journal of Marine Science, 58: 806-813.

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

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

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

7. Annex

Table S1. The international maturity scale for horse mackerel, *Trachurus trachurus.*

Stage	Name	Female	Male
1	Immature	Ovaries small. Ovaries wine red and clear, torpedo shaped.	Testes small, when fresh pale flattened and transparent. When frozen it may be opaque.
2	Developing	Ovaries occupying 1/4 to almost filling body cavity. Opaque eggs visible in ovaries giving pale pink to yellow to orange coloration. Largest oocytes may have oil globules.	Gonads occupying 1/4 to to almost filing body cavity. Testes off-white to creamy white., milt not running. When frozen testes can be bleuish.
3	Spawning	Ovaries characterized by externally visible hyaline oocytes no matter how few or how early the stage of hydration. Ovary size variable from full to < 1/4 of body cavity. Ovaries can be bloodshot.	Testes from filling to < 1/4 of body cavity, milt freely running. Testes can be shrivelled (wrinkled and contracted) at anus. When frozen there might be a change of structure and the testes needs a little pushing before running.
4	Regressing Regenerating	Ovaries occupying 1/4 or less of body cavity. Ovaries reddish and often murky (dark and gloomy) in appearance, sometimes with a scattering or patch of opaque eggs. The empty ovaries will ripple when pushed together.	Ovaries occupying 1/4 or less of body cavity. Testes opaque with brownish tint and no trace of milt. When frozen testes can be bleuish ore purple.
5	Omitted spawning	No evidence of omitted spawning	No evidence of omitted spawning
6	Abnormal	No evidence of abnormal ovaries	No evidence of abnormal testes

Table S2. Read mapping summary statistics of the Pool-Seq data of 12 horse mackerel samples included in this study. Abbreviations: W: Western, SW: Southwestern, S: South, N: North, MQ: Mapping quality, cov.: coverage.

Area	Sample	Total reads	% reads aligned	%GC	Median insert size	Mean MQ	Median cov.	Mean cov.
W Ireland	1a-WIR-2016	496686692	99.0	42.4	405	39.05	83	30.7
SW Ireland	1b-WIR-2016	594538427	99.1	42.2	416	38.97	99	35.2
SW Ireland	2-WIR-2017	573044377	99.0	42.4	465	38.95	96	35.5
S North Sea	3-SNS-2016	724017069	99.1	42.3	416	39	122	45.1
S North Sea	4-SNS-2017	764658923	99.1	42.3	419	38.97	128	46.3
N Portugal	5a-NPT-2016	571274302	99.2	42.4	404	38.9	95	35.2
S Portugal	5b-SPT-2016	494209199	99.1	42.9	426	39.13	83	29.0
N Portugal	6a-NPT-2017	490808045	98.1	41.8	*248*	39.32	75	26.1
S Portugal	6b-SPT-2017	514732597	99.2	42.3	*245*	39.12	79	27.5
Africa Mauritania	7-NAF-2016	714009211	98.5	*46.6*	425	38.49	91	25.7
N Spanish Shelf	8-NSP-2016	720020789	98.9	43.3	438	38.96	122	41.0
Mediterranean - Alboran Sea	9-MED-2018	671149600	98.8	42.5	422	35.13	112	41.5

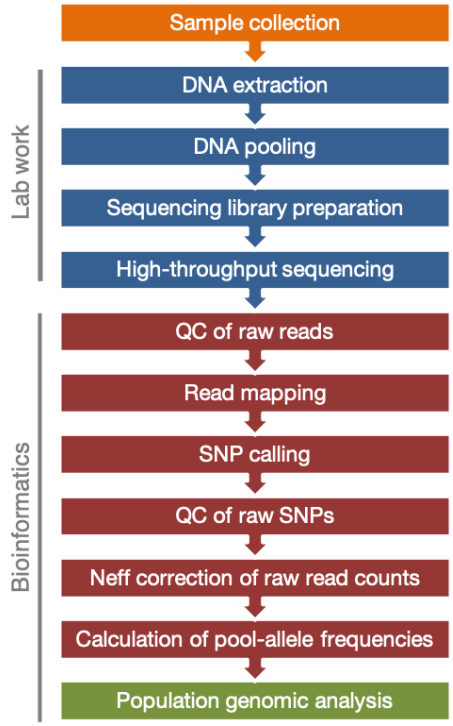


Figure S1. Schematic summary of steps followed for data generation.

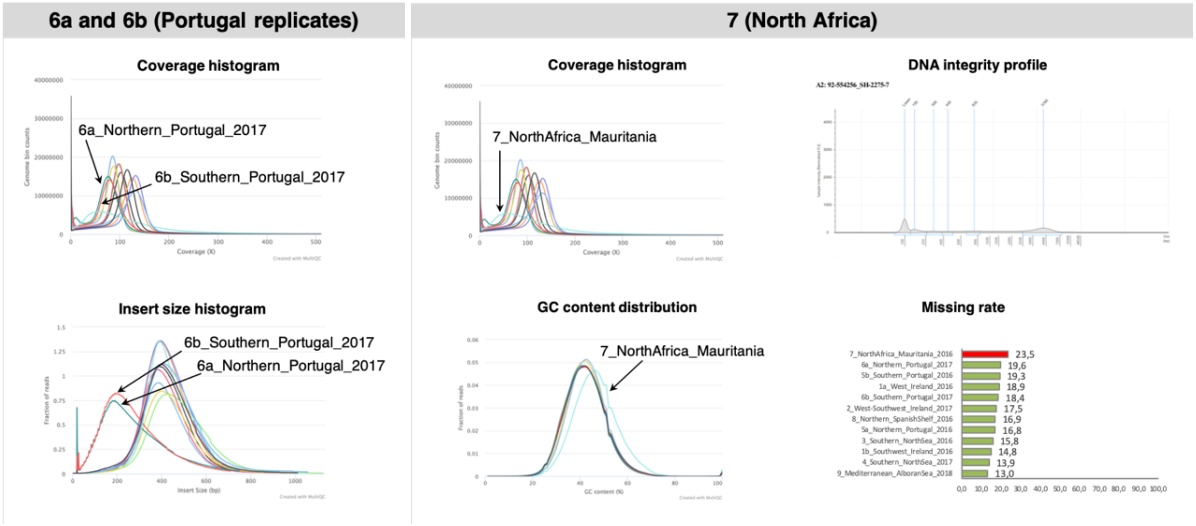


Figure S2. Read mapping statistics supporting that samples 6a, 6b, 7 were likely affected by technical artefacts. Plots obtained with *MultiQC*. (Left) Coverage and insert size distribution plots for the 12 samples, denoting the lines corresponding to samples 6a and 6b. (Right) Left, coverage and GC content distribution for all 12 samples, sample 7 is highlighted. Right, DNA integrity profile for the African sample and comparison of missing rate percentage for all 12 samples, the African sample is denoted in red.

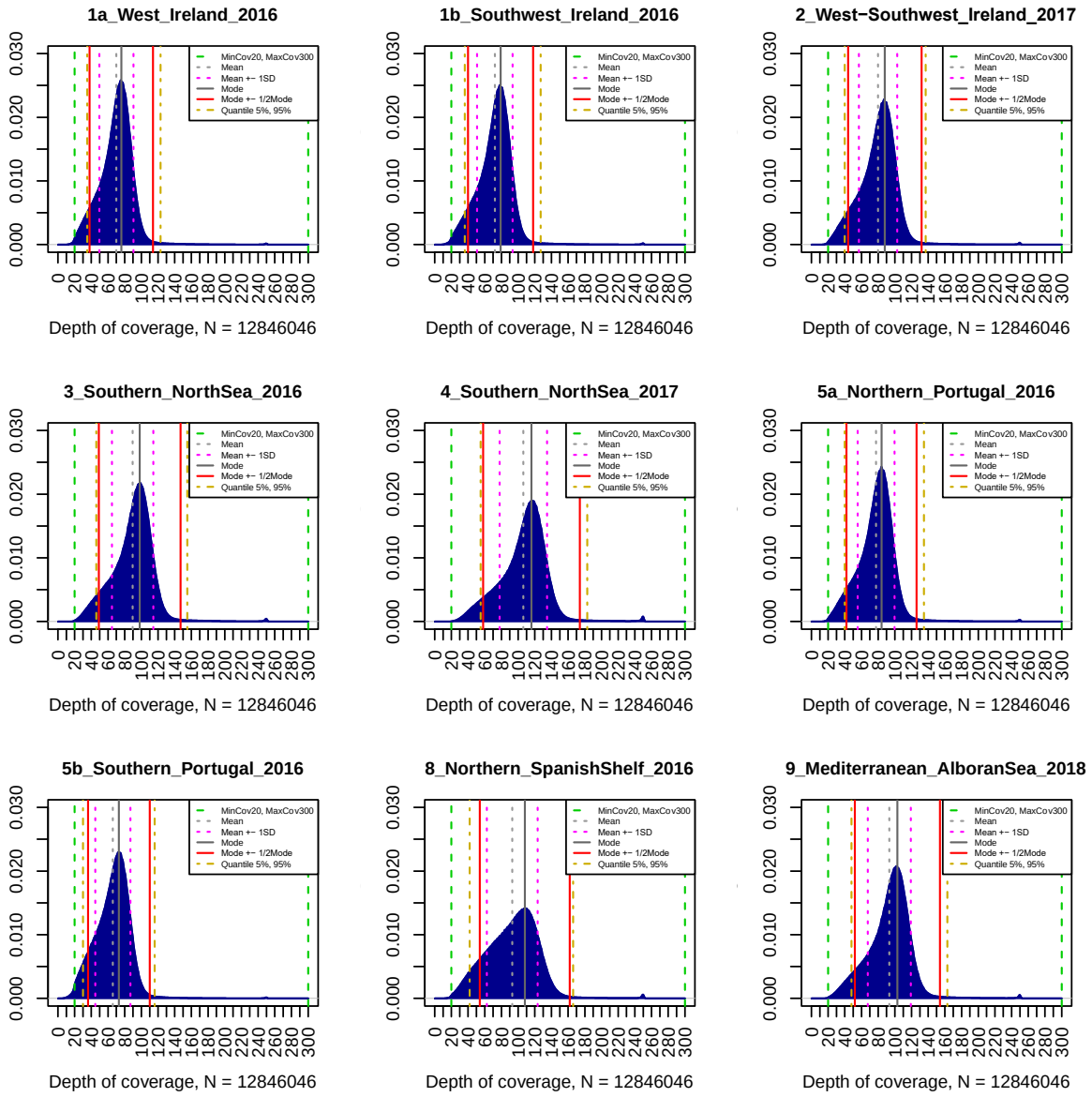


Figure S3. Depth of coverage distribution of 9 horse mackerel pools based on the SNPs that passed quality filters (~12 million). The different vertical lines correspond to the various lower and upper depth of coverage cut-off values examined.

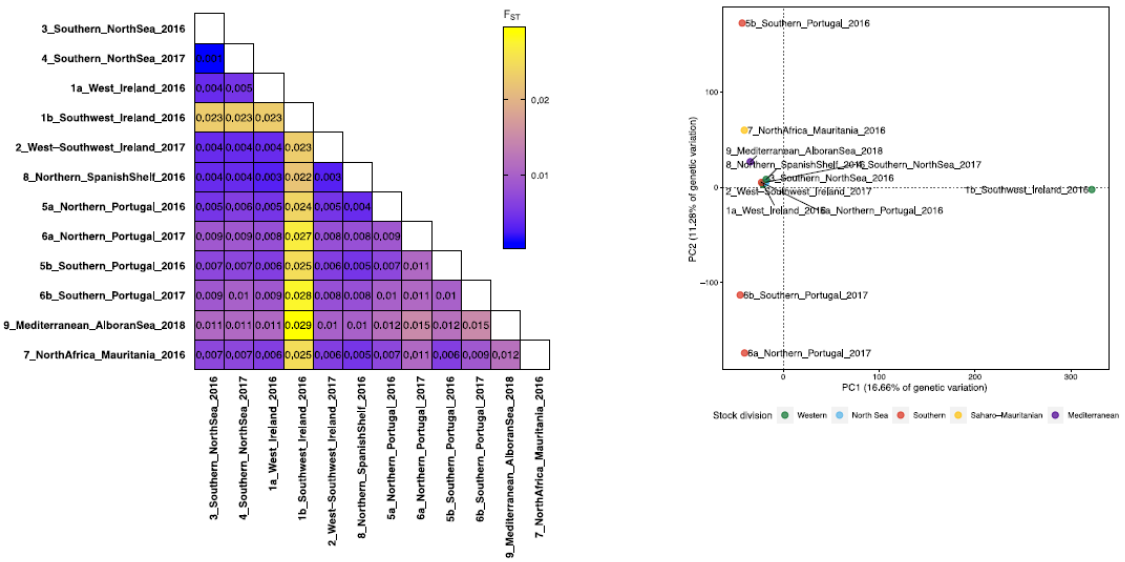


Figure S4. Exploratory population structure analysis for the 12 pools of the horse mackerel showing that samples 1b, 6a, and 6b correspond to outlier samples. (Left) Pairwise F_{ST}. (Right) PCA.

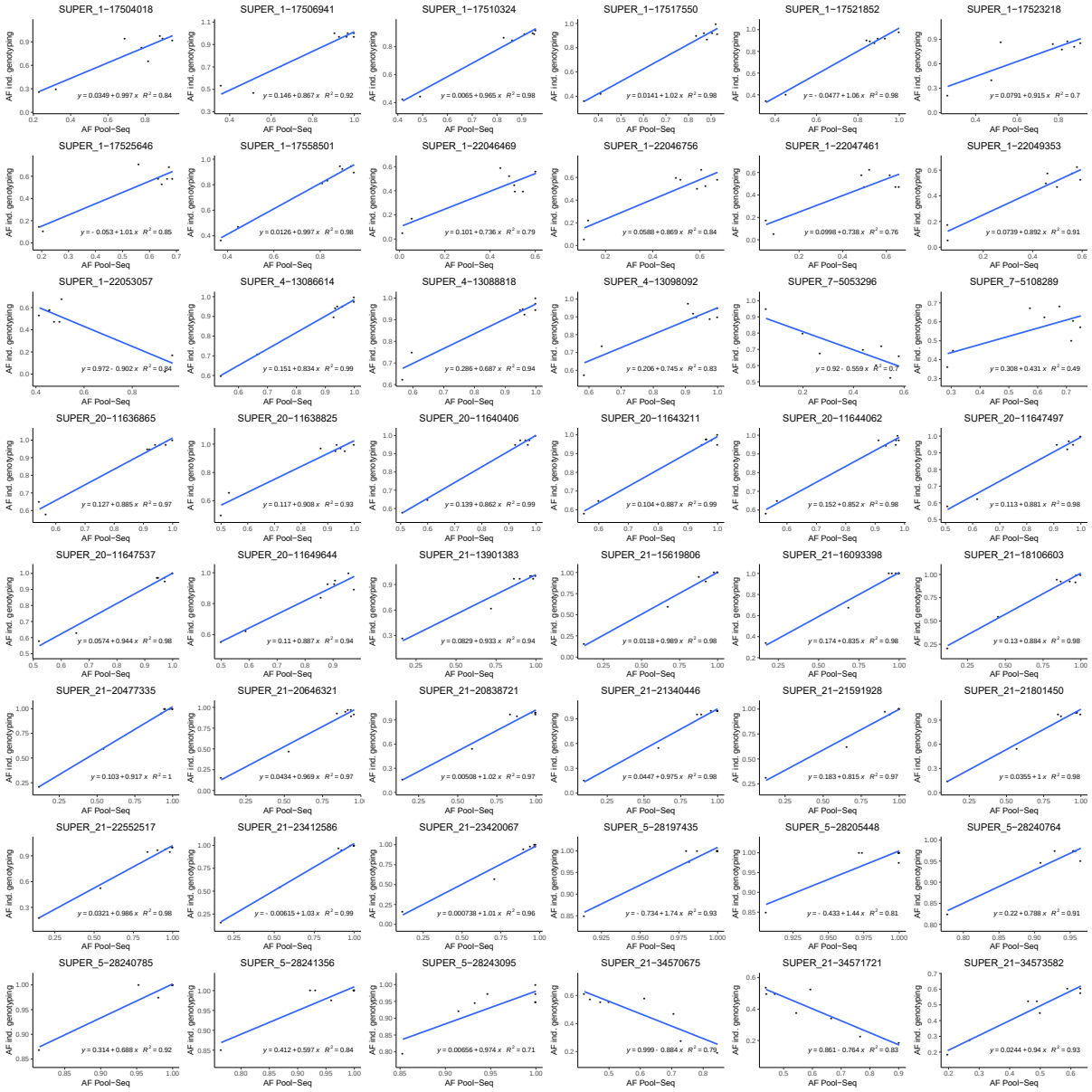


Figure S5. Comparison of population allele frequencies obtained with Pool-Seq and individual genotyping for the 48 SNPs putatively under selection.

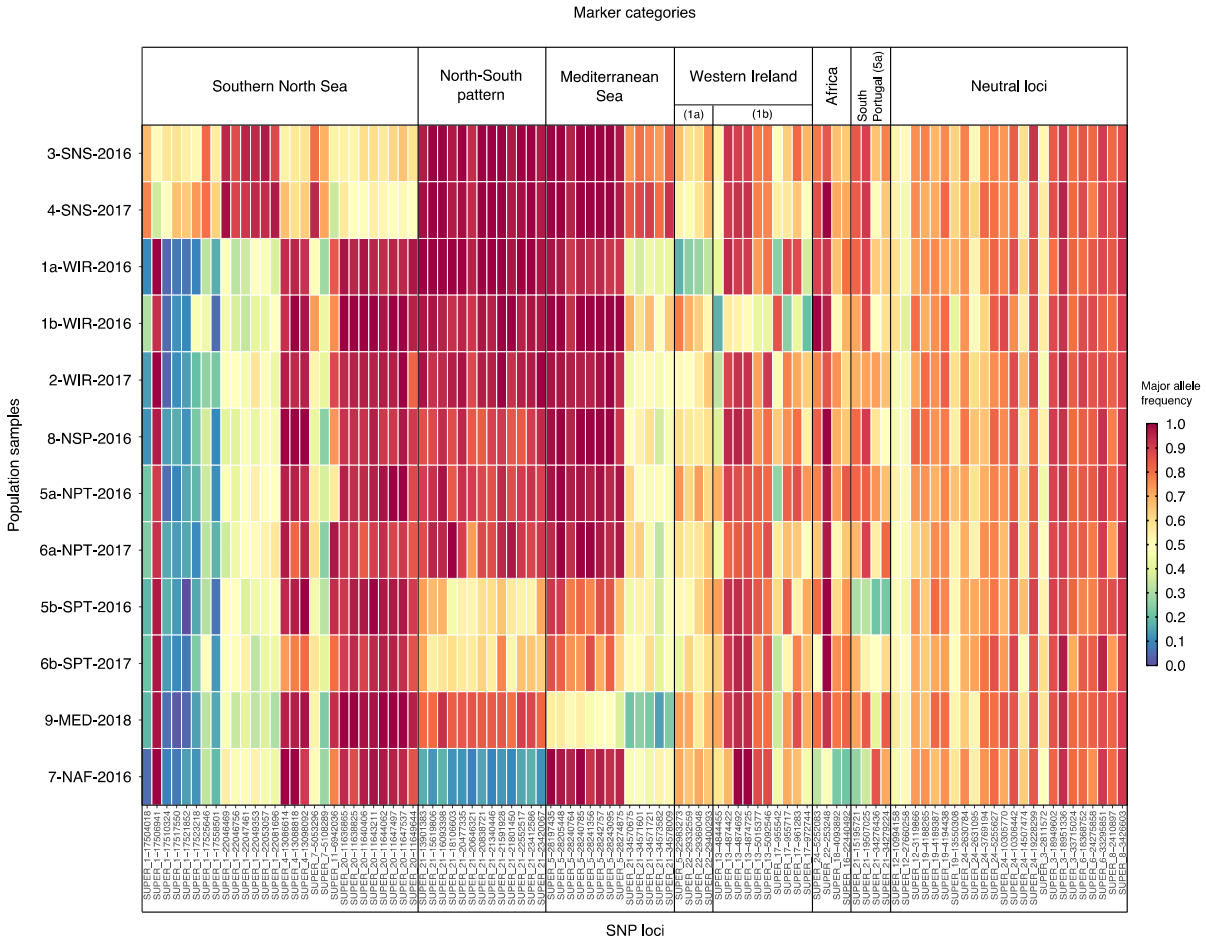


Figure S6. Heatmap plot representing the population allele frequencies of the 100 genetic markers included in the SNP panel. Rows correspond to samples and columns to SNP loci.

Figure S7. Heatmap plot representing the genotype of 157 individuals screened in 63 of the most informative SNPs for the horse mackerel. Squares in blue highlight the genotypes distinguishing the southern North Sea and the north-south genetic break.