ICES Journal of Marine Science, 2024, Vol. 82, Issue 5, fsae179
https://doi.org/10.1093/icesjms/fsae179
Received: 30 August 2024; **revised:** 20 November 2024; **accepted:** 22 November 2024
Original Article

Population structure of the swordfish, *Xiphias gladius*, across the Indian Ocean using next-generation sequencing

Thomas Chevrier [1,2,*], Dominique A. Cowart [2], Anne-Elise Nieblas [2], Grégory Charrier [3], Serge Bernard [4], Hugues Evano [1], Blandine Brisset [1], Jérémie Chanut [2], Sylvain Bonhommeau [1]

[1]IFREMER Délégation Océan Indien (DOI), Rue Jean Bertho, 97420 Le Port, La Réunion, France
[2]Company for Open Ocean Observations and Logging (COOOL), 97436 Saint-Leu, La Réunion, France
[3]Univ. Brest, CNRS, IRD, Ifremer, LEMAR, Rue Dumont d'Urville, 29280 Plouzané, France
[4]LIRMM-CNRS, université Montpellier-2, 161, rue Ada, 34392 Montpellier cedex 05, France
*Corresponding author. Company for Open Ocean Observations and Logging (COOOL) Saint-Leu, 97436 La Reunion. E-mail: t.chevrier.cooolresearch@gmail.com

Abstract

Swordfish (*Xiphias gladius*) is of significant economic importance as it is the second most exploited billfish in the Indian Ocean. While the Indian Ocean Tuna Commission (IOTC) considers swordfish to be a single panmictic population in the Indian Ocean, several studies have examined the potential for spatial variations within this highly migratory species with conflicting results, including emerging evidence that population structuring does indeed exist within swordfish. These findings therefore raise questions about the current guidelines for management adopted by the IOTC. In the present study, we address questions about the genetic structuring of swordfish in the Indian Ocean through the analysis of three datasets: (i) neutral SNPs, (ii) with, and (iii) only SNPs under potential selection identified from 1694 swordfish originating from 24 distinct locations across the Indian Ocean. A discriminant analysis of principal components showed the presence of two swordfish subpopulations in the Indian Ocean in the north and the south and was confirmed by admixture methods. This genetic differentiation may be explained by a chromosomal inversion, indicating that both populations could be demographically connected but remain differentiated by this structural variant.

Keywords: population structure; fisheries management; single nucleotide polymorphism; discriminant analysis of principal component

Introduction

The development of accurate stock assessments relies upon the definition of discrete biological units or populations, which may have different responses to harvest (Carvalho and Hauser 1994). Fish species can form demographically disconnected subpopulations that are differentiated by behavior and/or geographical distribution, which can result in adaptive genetic differences (Ciftci and Okumus 2002, Reiss et al. 2009). Consequently, identifying this genetic population structure is an essential prerequisite for effective and sustainable management of fisheries resources (Reiss et al. 2009, Neilson et al. 2013). Poor definition of management units can lead to depletion or even extinction of the most vulnerable or local subpopulations, resulting in a loss of genetic diversity and reducing the species ability to recover from the stress of commercial harvesting or adapt in the face of global change (Carvalho and Hauser 1994, Righi et al. 2020). Over the past decade, the declines of many fish stocks have been observed as a result of inadequate management measures (Reiss et al. 2009, Hilborn et al. 2020). This observation is particularly true for large pelagic fish, where population structuring is defined by low levels of intraspecific heterogeneity due to the absence of geographical barriers and the high mobility of the fish, which facilitates gene flow (Ward et al. 1994).

Swordfish, *Xiphias gladius*, is a highly migratory, mesopelagic, opportunistic predator, distributed in the temperate and tropical regions of the Atlantic, Pacific, and Indian Oceans, as well as various seas from 45°N to 45°S (Palko et al. 1981, Nakamura 1985, Hernandez-Garcia 1995). In the Indian Ocean, *X. gladius* has the largest commercial value among billfish fisheries and is one of the most exploited billfish species (23 917 t in 2021, i.e. the second billfish species after sailfish, 37 310 t in 2021) (IOTC 2022, Thoya et al. 2022). While not presently defined as overfished/overexploited, this species is subjected to increasing fishing pressure at the industrial and semi-industrial scale (IOTC 2022). As such, there is a continued interest in a more comprehensive understanding of this species' population dynamics to ensure more robust stock assessments.

The migratory nature, wide depth distribution, and broad temperature tolerance of *X. gladius* make the investigation of the population structure and dynamics of this fish particularly challenging. At the global scale, previous studies based on molecular markers of mitochondrial DNA (mtDNA) variants revealed one panmictic population in the North Pacific (Grijalva-Chon et al. 1994) and the presence of two divergent clades in the Pacific Ocean and in the Atlantic Ocean and Mediterranean Sea (Bremer et al. 1996, Rosel and Block 1996).

In the Indian Ocean, sequence variation at the mtDNA control region suggested the existence of two distinct population units in waters north of Madagascar and the Bay of Bengal, whereas the rest of the Indian Ocean was considered as a single population (Lu et al. 2006). By comparison, the

combination of mtDNA and microsatellite markers identified a population subdivision between the Indian Ocean, the Atlantic Ocean (waters off South Africa and Namibia), and the Pacific Ocean (Coral Sea) yet also supported the presence of a single panmictic swordfish population in the Indian Ocean (Muths et al. 2009, 2013). However, the hypothesis of population structure between the southwest and the northern Indian Ocean was proposed (Muths et al. 2009). Most recently, single nucleotide polymorphic (SNP) markers were used to identify subtle population structure of swordfish within the Indian Ocean, finding at least two genetically differentiated groups present north and south of the equator (Grewe et al. 2020). Overall, several studies, using a variety of genetic approaches, have confirmed restricted gene flow of swordfish between the Atlantic, Pacific, and Indian Oceans, as well as some differentiation between the northern and southern hemispheres; however, contrasting results remain (Kotoulas et al. 1995, 2007, Bremer et al. 1996, 2007, Reeb et al. 2000, Jean et al. 2006, Smith et al. 2015, Lu et al. 2016, Darnaude et al. 2020, Grewe et al. 2020). Therefore, to date, swordfish are considered to be a single panmictic population in the Indian Ocean by their management body.

In the present study, we investigate the population structure and dynamics of *X. gladius* by inspecting the genetic diversity of this species across the Indian Ocean to inform the stock assessments and help resolve previous conflicting findings. We generated three datasets to test for neutral and putatively adaptive differences among potential populations. We genotyped 2068 individuals, including juveniles and adults from 24 locations across the Indian Ocean targeting SNPs from genome-wide coverage, using the DarTseq methodology (Sansaloni et al. 2011).

Materials and methods

Sampling design

Samples were selected from the Ifremer IOSSS—ESPADON project (Bourjea et al. 2010). During the IOSSS project, 3127 *X. gladius* samples were collected between the years 2006 and 2012, and 2359 samples originating from 24 different sampling surveys (Supplementary Table S1) were identified as having enough tissue to enable DNA extraction, sequencing, and genotyping. We selected 2030 samples for genetic analyses, by prioritizing samples collected within the IOTC convention area (Fig. 1) during 2009 to 2011 (but also including samples from 2006 collected around the Glorioso Islands) (Supplementary Table S1). For each sampling site, particular attention was paid to keep balanced sample sizes between life stages (i.e. juvenile or adult) and between sexes. Samples were selected from as many sites as possible across the Indian Ocean to better assess the spatial distribution of genetic variability within this ocean basin (Fig. 1). In addition, 40 samples originating from outside of the IOTC convention area (eastern Australia) were included in the set, which were used to examine levels of genetic differentiation between the IOTC and Pacific areas. A total of 2068 samples were finally retained.

Sequencing

Sequencing, SNP discovery, and genotyping of all samples were performed by Diversity Arrays Technology (DArT PL, Canberra, Australia), using the DArTseq methodology (Sansaloni et al. 2011). DArTseq integrates DArT complexity reduction methods with next-generation sequencing platforms, allowing optimized sequencing of genomic representations for different organisms while enhancing the precision and adaptability of genomic assays across various species. For *X. gladius*, the PstI–SphI enzyme combination was chosen, and DNA samples were processed using a modified Kilian et al. (2012) protocol (supplementary information), which incorporates Illumina-compatible adaptors with unique barcodes for efficient sequencing. DArTseq has previously been successfully applied in population genetics studies of other large pelagic fishes such as southern bluefin tuna *Thunnus maccoyii* (Davies et al. 2020), yellowfin tuna *T. albacares* (Grewe et al. 2015, Anderson et al. 2019), Atlantic bluefin tuna *T. thynnus* (Bravington et al. 2022), and blue skate *Dipturus batis* (Delaval et al. 2023). The DArTSeq workflow was implemented on all 2068 samples. Resulting FASTQ files were processed using a DArT analytical pipeline first to filter poor-quality sequences. The filtering steps performed on the raw sequence dataset are detailed in the supplementary information. Sequences were then aligned to the *X. gladius* genome available on the National Center for Biotechnology Information (NCBI) (Accession: GCF_016859285.1) (Wu et al. 2021) to call SNP loci. Final reports were produced and contained the genotyped samples, identified SNPs, call rates, and the co-dominant status of each sample (Supplementary Table S2) (Anderson et al. 2019).

Sequencing data were produced for 2228 individuals, which included 2038 samples and 189 technical replicates (including one sample that was replicated four times instead of twice) that were successfully processed, and 30 individuals that failed sequencing (Supplementary Table S2). During the application of the workflow, eight of these failed individuals were identified as matching either a species distinct from *X. gladius* or interspecific hybrids; these non-swordfish samples were not analyzed further, which resulted in 2030 samples that underwent downstream analyses.

Computational processing

Data filtering

Data were represented in a table of SNP bases (A, T, C, and G) with two states for each individual at each locus as *X. gladius* is a diploid organism. Genotypes were coded as "0" for the homozygous reference allele, "1" for the homozygous alternative allele, and "2" for heterozygotes, the reference allele defined to be the most common. Missing data were coded as "NA" or "-." All analyses were performed using four packages: *dartR* (Gruber et al. 2018), *adegenet* (Jombart 2008), *kinference* (M. Bravington and S. Baylis, personal communication), and *gbasics* (M. Bravington and S. Baylis, personal communication).

Metadata for each locus were included in the SNP report, as shown in Supplementary Table S2. In addition, a metadata file was created containing information associated with each individual fish. This file contained an "id" column that listed individual sample names and a "pop" column that listed the cluster or population of origin of individuals without any spatial geo-referencing. The SNP report was imported into *R* with the individual metadata file using the gl.read.dart() function from *dartR*. SNP data coded in this way were held in a *genlight* object defined with *adegenet* (Jombart 2008).

A total of 86 409 bi-allelic SNPs was returned for the 2030 swordfish samples. SNPs were filtered based on the sequencing

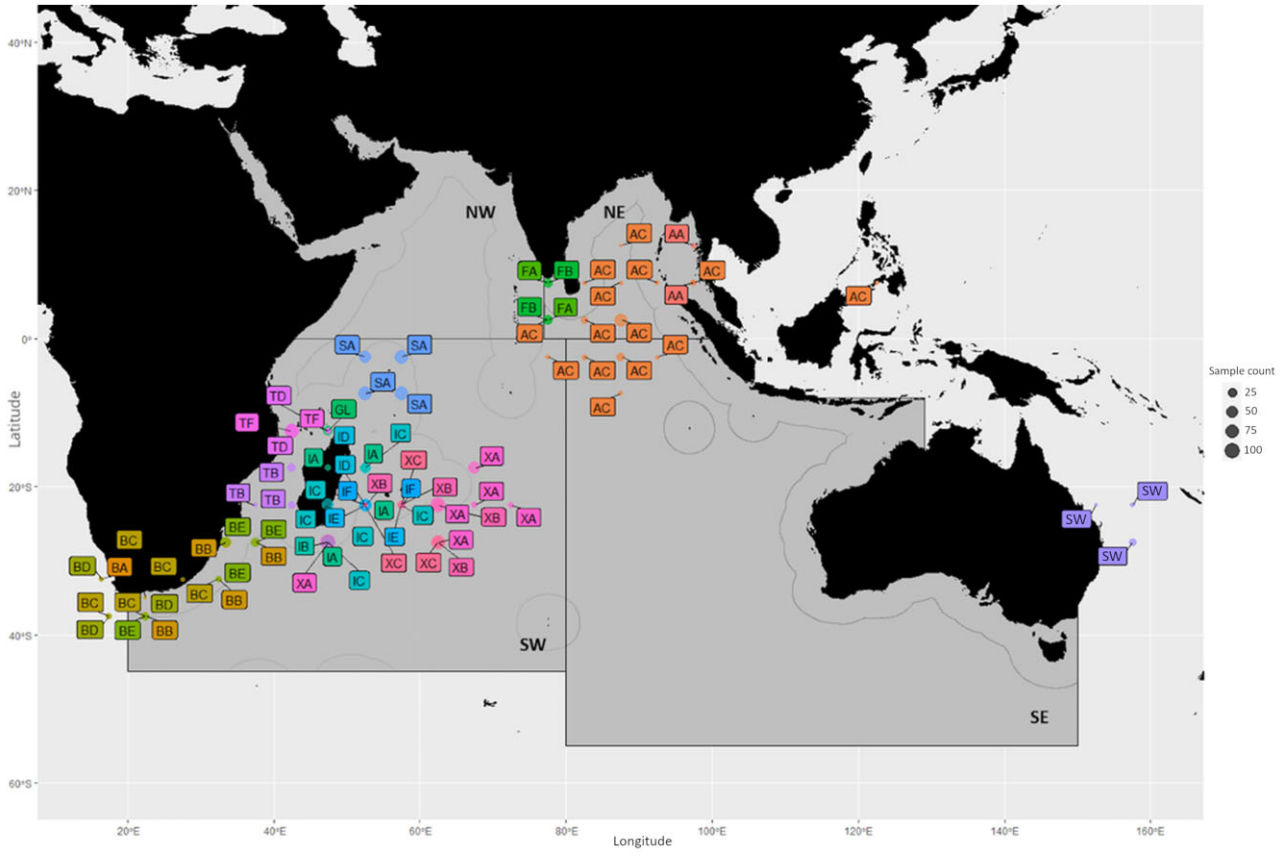

Figure 1. Sampling locations for each campaign in the different IOTC convention subareas (in gray): NW: northwest; NE: northeast; SW: southwest; and SE: southeast.

depth, reproducibility, call rate, minor allele frequency, linkage disequilibrium, Hardy–Weinberg disequilibrium, and signals of potential selection to keep only reliable variants for population structure analyses. Individuals were filtered based on their call rate. At this stage, samples were organized into the 24 sample areas and the conformance of loci to the Hardy–Weinberg proportions was tested in each group. The loci that did not conform to the Hardy–Weinberg proportion in at least six populations were removed. Sample areas containing <10 fish per site were deleted to avoid any bias linked to the sampling effort. The selection enabled us to obtain a reliable dataset for population structure from 22 locations with only unlinked SNPs. The SNP filtering method is detailed in Supplementary Table S3. Once the *genlight* object was filtered, it was converted into a *SPAgeno* object with *gbasics* R package, to be use with *kinference*. After these steps, data were again filtered for samples with atypical genotypes based on probability calculation using the *kinference R* package, which may indicate individuals which may have been contaminated or those whose DNA was too degraded, as well as technical replicates.

Three datasets were generated to investigate the population structure. For the three, all individuals were organized into the sampling sites, as we have no information on their spawning areas. In the ALL dataset, all SNPs obtained after the different filtering steps, including neutral and potentially under selection, were retained. The second dataset NEUT is based on ALL but without the SNPs which could be considered under selection. SNPs potentially under selection are kept in a third dataset OUTLIER.

Detection of potential adaptive loci

Loci potentially under selection were determined with *dartR* using *OutFLANK* (Whitlock and Lotterhos 2015). This method is used to detect SNPs under selection by comparing their allele frequencies to those expected under a neutral model, using the F_{ST} statistic to identify SNPs with significant genetic differentiation. The *OutFLANK* filter was applied using the *gl.outflank()* function, setting a minimum F_{ST} threshold of 0.05 for outlier detection. The SNP is identified as an outlier if its F_{ST} value was $\geq$0.05 and its q-value, which balances detection of true signals with the control of false positives, was $\leq$0.05.

Global F_{ST} values for each SNP across all sampling sites were estimated using the *–weir-fst-pop* option in *VCFtools*. A Manhattan plot was generated with the *qqman* package (Turner 2018) to visualize F_{ST} values for each SNP in relation to their chromosomal positions. Outlier markers, identified with *OutFlank*, were highlighted to indicate potential selection signals from specific chromosomes. The presence of chromosomal inversion was also checked using PLINK and VCFTools by calculating the determination coefficient r^2 between each pair of SNPs across each chromosome.

Population assignment

The number of distinct genetic clusters in the dataset was assessed without *a priori* information using the *K*-means clustering method implemented in *adegenet* (Jombart 2008). The find.clusters() function from *adegenet* was then used for the assignment of individuals to each of the identified clusters. As

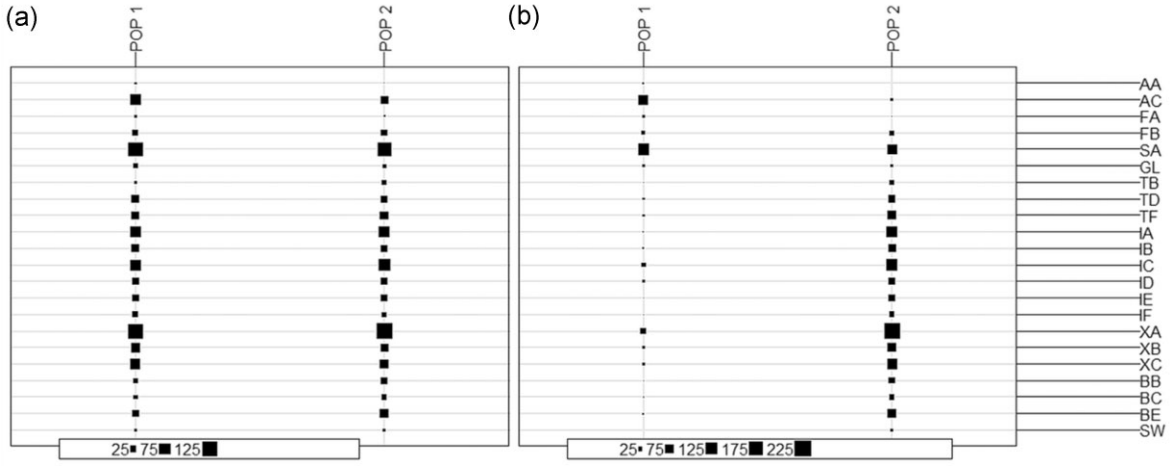

Figure 2. Sample assignments without preconceptions for $K = 2$: (a) the NEUT dataset with only neutral loci and (b) the ALL dataset with loci under potential selection.

advocated in previous studies (Fraley and Raftery 1998, Lee et al. 2009), the Bayesian information criterion (BIC) was used to assess the best-supported model, and therefore the number and nature of clusters. To evaluate the potential effect of selective forces on the genetic structure of swordfish populations, the clustering method was performed with the NEUT and ALL datasets. To complement the clustering approach, admixture analyses using *STRUCTURE* software were conducted on both datasets (NEUT and ALL) for $K = 2$–4. The length of the burn-in period was set to 5000, and the number of Markov Chain Monte Carlo (MCMC) repetitions after burn-in was set at 50 000.

Discriminant analysis of principal component

We used a discriminant analysis of principal components (DAPC, Jombart et al. 2010) on both NEUT and ALL data sets. DAPC is used for dimensionality reduction and classification in multivariate data analysis by combining the advantages of principal component analysis (PCA) and linear discriminant analysis (LDA). DAPC begins with a PCA to reduce the dimensionality of the data and extract the principal components, then LDA is applied to these principal components to maximize the separation between predefined groups or classes. DAPC is particularly useful when dealing with complex datasets with overlapping or closely related groups, and allows for effective discrimination between these groups. In addition to providing insights into the most discriminatory variables, aiding in classification and prediction tasks, DAPC has been widely applied in various fields for exploring population structure, identifying important variables, and classifying individuals into groups based on their multivariate characteristics. The number of PCs retained in each DAPC was decided using cross-validation to provide an objective optimization procedure to identify the "goldilocks point" in the trade-off between retaining too few or too many PCs in the model (Supplementary Fig. S2). In this case, we retained the number of PCs associated with the lowest root mean squared error (RMSE) as the "optimum" *n.pca* in the DAPC analysis.

Pairwise differentiation

Pairwise fixation index of genetic differentiation (F_{ST}) and *P*-values were estimated for the three datasets (ALL, NEUT, and OUTLIER) between the 22 different sampling sites. The method proposed by Wright (1949) and updated by Weir and Cockerham (1984) was used through the *gl.fst.pop* function from *dartR*, which is based on the implementation of the *StAMPP* package. In addition, the statistical significance of pairwise F_{ST} estimates was evaluated by calculating confidence intervals and *P*-values over 10 000 bootstraps.

Results

Based on the *K*-means clustering method, a single population was found ($K = 1$) when only neutral loci were considered (NEUT dataset; Supplementary Fig. S1A). However, two clusters were found when loci under potential selection were included (ALL dataset; Supplementary Fig. S1B). As the BIC values were similar, an assignment without preconceptions was made for $K = 2$–5 for both datasets. No notable patterns were observed for the $K = 3$–5 for both datasets. For $K = 2$, samples appeared to be randomly and homogeneously split between both clusters when considering only neutral loci (NEUT dataset; Fig. 2a), which seems to confirm that $K = 1$ in this configuration. In contrast, two groups emerged from the dataset that included SNPs with a selection signature (OUTLIER dataset; Fig. 2b). The first cluster (POP_1) was predominant in samples from the northeastern sector of the Indian Ocean (AC, AA, and FA, except FB), whereas the second cluster (POP_2) was largely dominant in all the samples from the southern sector of the Indian Ocean, except for SA (Seychelles) and GL (Glorioso Island) (Fig. 2b). The two clusters were found in roughly equal proportions in FB, SA, and GL.

Admixture analysis was conducted on the 22 remaining locations for $K = 2$–4. At $K = 2$, the admixture analysis based on the ALL dataset revealed two distinct genetic groups (Fig. 3). The northern samples (locations 1–3) were strongly associated with the second cluster (green), whereas the southern cluster (locations 7–21) predominantly associated with the red cluster (cluster 1). Equatorial locations (4–6) showed a mix of the two clusters, indicating a potential transitional zone. The east Australian population (subpopulation 22) aligned closely with the southern Indian Ocean areas, sharing a dominant red cluster. At $K = 3$, samples from northern areas remained strongly assigned to cluster 2, with minimal influence from the other clusters. The introduction of

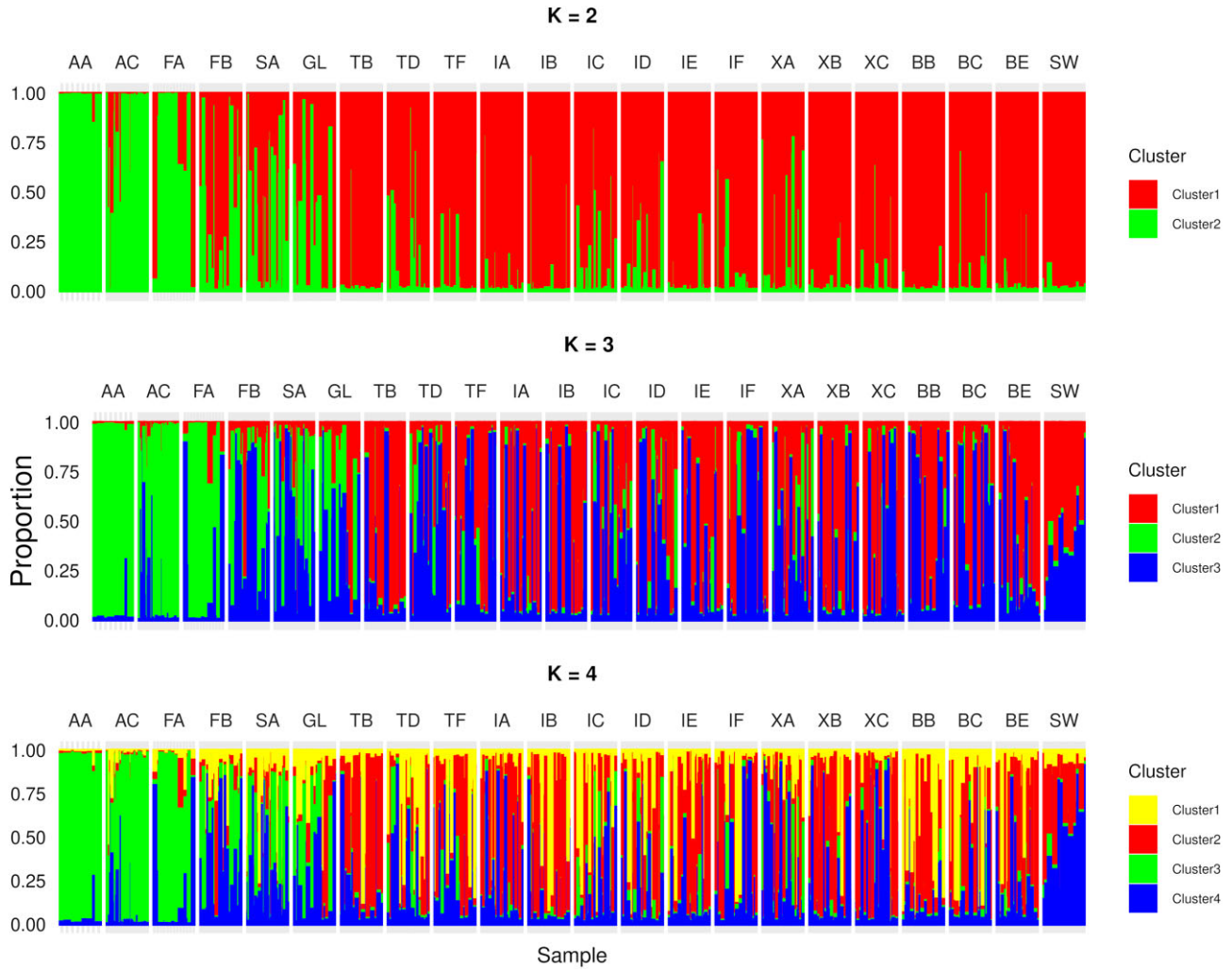

Figure 3. Individual bar plot of the probability of assignment to a particular *K* genetic group for $K = 2$–4 using the ALL dataset. Samples are latitudinally ordered from north (left) to south (right). SW sampled from the east coast of Australia are to the far right of the barplot.

cluster 3 revealed a more complex genetic structure in the equatorial and southern Indian Ocean regions, with individuals showing mixed ancestry from clusters 1, 2, and 3. The presence of blue indicated further sub-structuring or gene flow within this broad region. For $K = 4$, except for individuals from the north, which were always strongly assigned to the green cluster, $K = 4$ did not show any clear pattern for individuals from the central/equatorial zone and the south, which had a complex mixture of the four clusters. For the SW (East Australia) individuals, the majority were assigned to a mix of two clusters. Admixture analysis was also performed using the NEUT dataset for $K = 2$–4 (Fig. 4). No structure pattern was observed for $K = 2$–4, and all individuals were assigned equally to each of the clusters.

The DAPC and PCA (Supplementary Fig. S5) approaches for the ALL dataset confirmed the previous results of two distinct groups and an intermediate group (Fig. 5b). The first cluster was composed of samples taken in the northeastern Indian Ocean, including part of the Indian and Sri Lankan samples (AA, AC, and FA), which were discriminated on the first axis. The other group was composed of all the southern samples. However, between these two distinct clusters, an intermediate group appeared containing samples with locations close to the equator (SA, GL, and FB). Furthermore, the SW samples (eastern Australia) used as outgroups were mixed with samples from the southwestern Indian Ocean. DAPC results were quite similar between the NEUT and ALL datasets, with greater discrimination for the dataset including SNPs with potential selection (Fig. 5).

An F_{ST} matrix was developed for each dataset. The ALL dataset confirmed the observations made with all previous analyses (e.g. *K*-clustering, Fig. 2; admixture, Fig. 3; and the DAPC, Fig. 5b), with the largest F_{ST} values (blue) observed between northern (AC, AA, and FA) and southern samples. However, F_{ST} values were small, with a maximum of about 0.02 between AA and SW (Fig. 6a). For the NEUT dataset, all pairwise F_{ST} values were small (Fig. 6b) with a largest value of 0.007 between AA and SW, which confirmed the clustering method results ($K = 1$) (Supplementary Fig. S1A). The F_{ST} matrix based on only outlier SNPs was similar to that obtained with the ALL dataset, but the F_{ST} values were larger (Fig. 6c) with 0.363 between AA and SW. As with the other dataset, the largest F_{ST} values were found for the samples between the northeastern Indian Ocean (named FA-AC) and southern samples (Fig. 6).

As the samples included both adult and juvenile individuals, all of the analyses were also carried out on adult individuals during the breeding season (November to April) to ensure

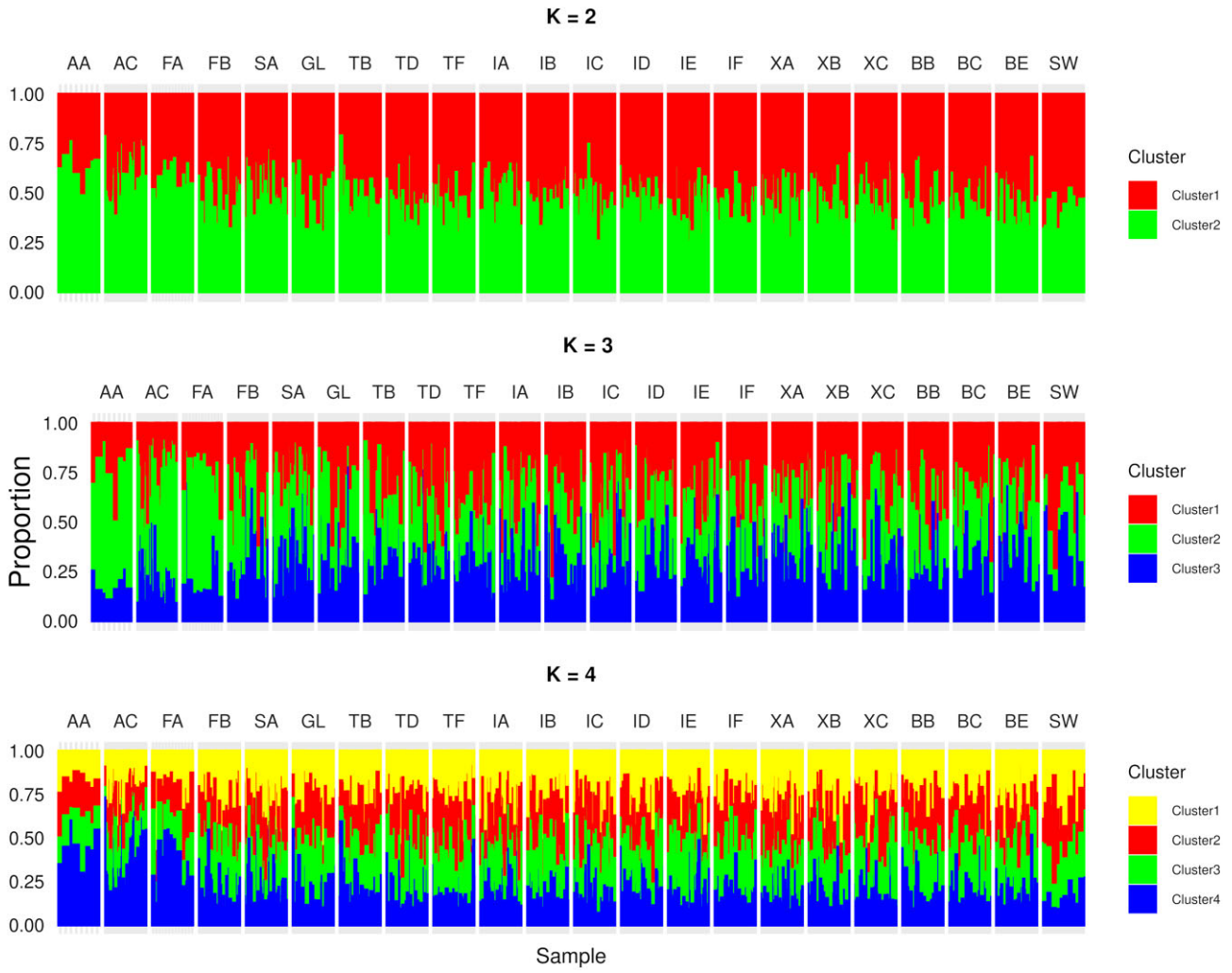

Figure 4. Individual bar plot of the probability of assignment to a particular *K* genetic group for $K = 2$–4 using the NEUT dataset. Samples are latitudinally ordered from north (left) to south (right). SW sampled from the east coast of Australia are to the far right of the barplot.

that no bias was observed due to the stage of maturity and the time of year (Nielsen et al. 2009). Results obtained with this subset led to the same conclusions (Supplementary Figs S3 and S4).

OutFLANK classified 70 loci as outliers, with ~75% (53 loci) located on the fifth chromosome (Fig. 7a). As indicated by their positions, these SNPs appear to be grouped on a portion of a chromosome (Fig. 7b). The linkage disequilibrium across chromosomes suggested the presence of chromosomal inversions on chromosome 5 (Fig. 7c). Nothing was observed on the other chromosomes (Supplementary Fig. S6).

Discussion

Genetic differentiation of swordfish between the north and south of the Indian ocean

A distinct genetic differentiation between swordfish populations in the northern and southern regions of the Indian Ocean was identified using a subset of putatively adaptive loci. In contrast, analyses of presumably neutral loci indicated a pattern consistent with panmixia across the entire Indian Ocean. These findings were corroborated by results from PCA, DAPC, genotypic clustering in STRUCTURE, and pairwise F_{ST} analyses. The enhanced resolution provided by putatively adaptive loci compared to neutral loci has been observed in previous RADseq studies on tuna (Grewe et al. 2015, Pecoraro et al. 2018, Vaux et al. 2021) and other marine fish species (Mamoozadeh et al. 2019, Longo et al. 2020). These findings underline the effectiveness of adaptive loci in detecting population structure in highly migratory marine species such as swordfish (Gagnaire et al. 2015, Gagnaire and Gaggiotti 2016, Vaux et al. 2021). Overall, our results suggest the existence of at least two population units in the Indian Ocean, one in the northeastern region (AA, AC, and FA), one in southern waters, and possibly a third unit north of Madagascar (SA, GL, and FB). These results are in agreement with Lu et *al.* (2006), who showed the existence of genetic differences between swordfish sampled in the Indian Ocean with two distinct population units in waters north of Madagascar (in our case, GL and SA) and the Bay of Bengal (here AC, AA, FA, and FB), whereas the rest of the Indian Ocean was considered as a single population. They obtained high levels of F_{ST} when samples from the northern Indian Ocean (IMadN and INor) were compared with samples from the rest of the Indian Ocean and the Pacific Ocean, respectively. These results are also in line with those

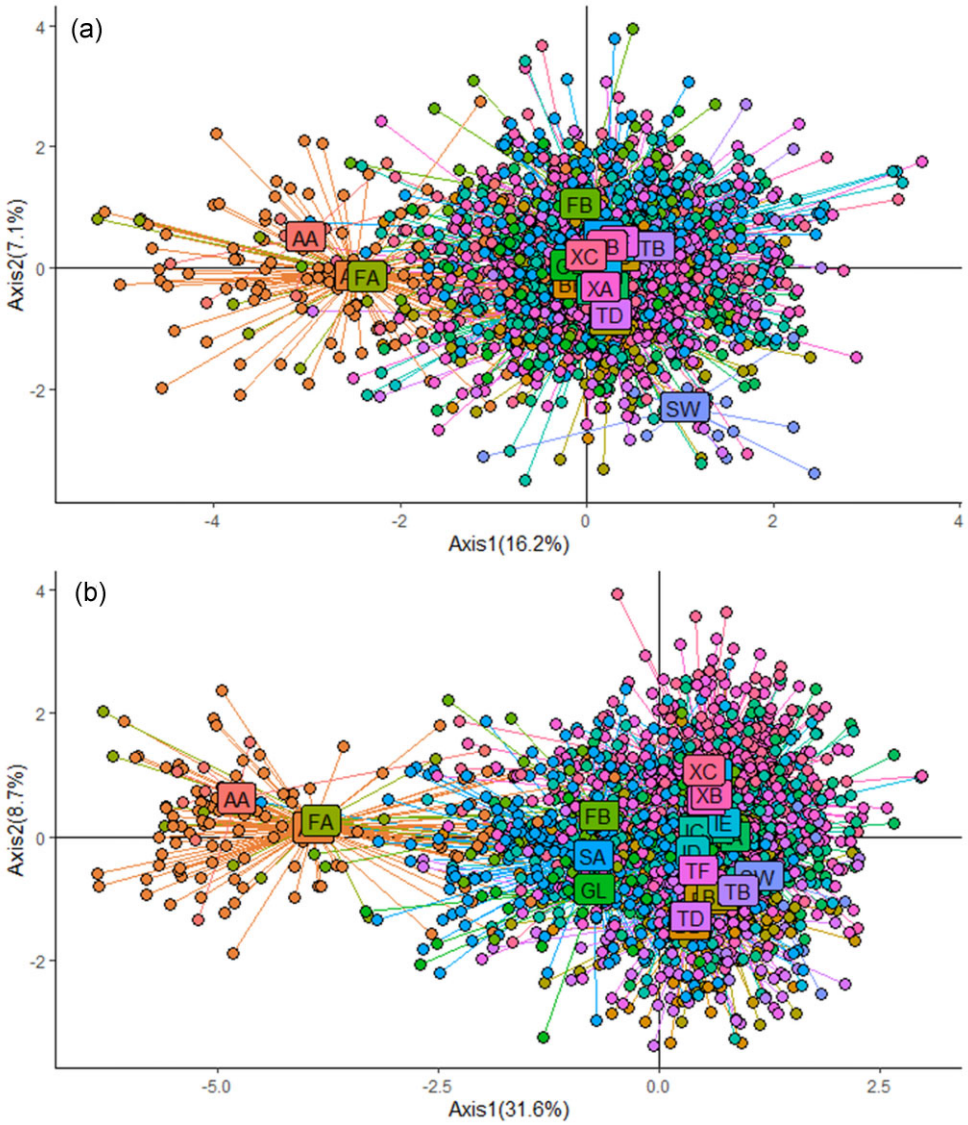

Figure 5. Representations of the DAPC according to the first two axes, with samples grouped by their sampling areas: (a) dataset NEUT and (b) dataset ALL.

obtained by Grewe et al. (2020), who highlighted two subpopulations, one in the north and the other in the south of the Indian Ocean.

Identification of a potential chromosomal inversion

The putatively adaptive loci were found to have a close spatial proximity, with the majority found on the fifth chromosome. This finding allows us to highlight a structural variant and, more precisely, a chromosomal inversion of ~10 Mb. Chromosomal inversions are structural rearrangements where a segment of a chromosome is reversed end to end. These inversions can have profound effects on meiosis, gamete viability, and, consequently, fertility. Their role in reproductive isolation and speciation has been a subject of considerable interest in evolutionary biology (Noor et al. 2001, Hoffmann and Rieseberg 2008, Mérot 2020). In this case, this inversion includes 70% of the potential underselected loci that we found (Fig. 7c). It seems to explain the segregation between individuals from the north and south clusters. It is unknown, however, if this chromosomal inversion impacts individual fertility or signals the beginning of a possible speciation event, as the current reference genome does not include the centromere for chromosome 5 to allow us to determine whether this inversion is pericentric or paracentric. Examining the potential biological implications of this inversion, including a detailed understanding of the biological functions of the genes involved could further enrich the interpretation of how these genetic variations may influence traits such as fitness, behavior, local adaptation, or reproductive isolation. While the biological functions of potentially underselected genes have been specified in Supplementary Table S3, it should be noted that these annotations are based on data from humans or zebrafish, with potential differences in functional context between species. Further, the reference genome utilized in this study originates from an individual sampled in the Pacific Ocean (the China Sea), and developing a reference genome from individuals native to the Indian Ocean could provide a relevant framework for more in-depth analyses. Additionally, future transcriptomic and proteomic studies are necessary to better understand the potential role of each gene.

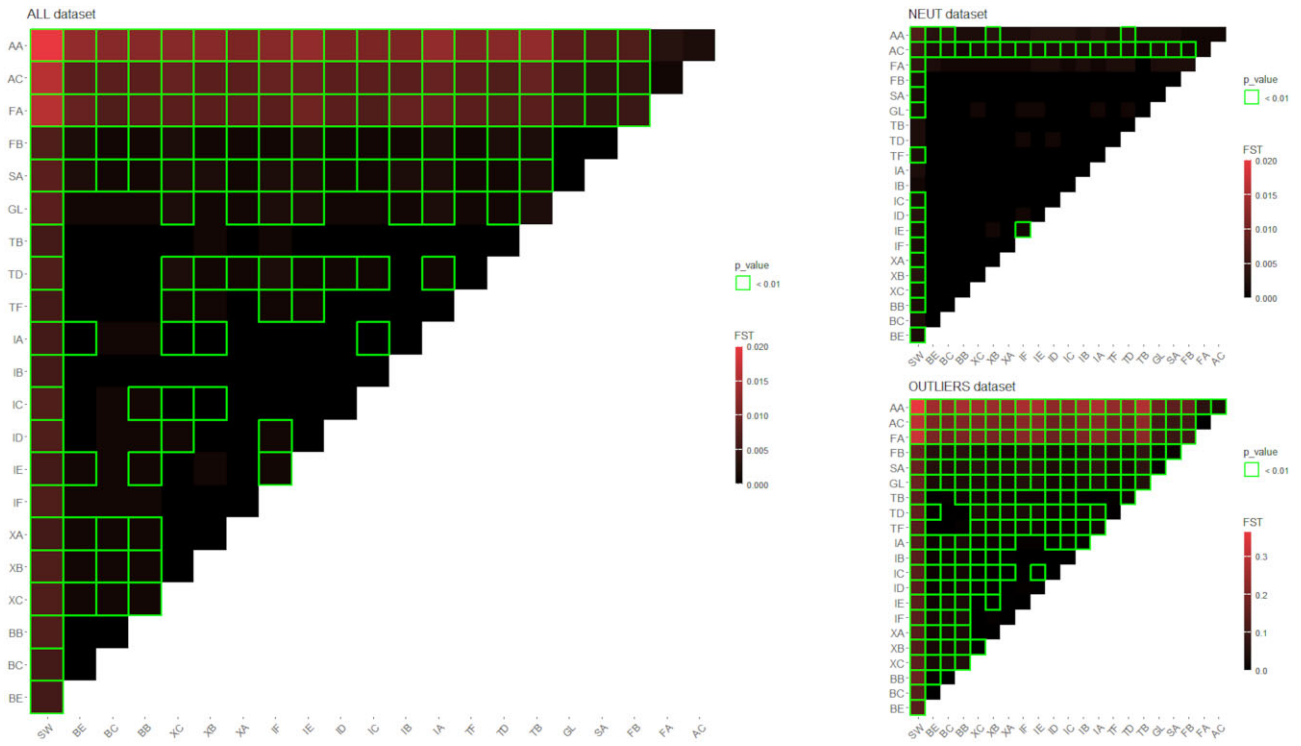

Figure 6. Heatmap representations of the pairwise fixation index (F_{ST}) between the different sampling areas: (a) dataset ALL, (b) dataset NEUT, and (c) dataset OUTLIERS.

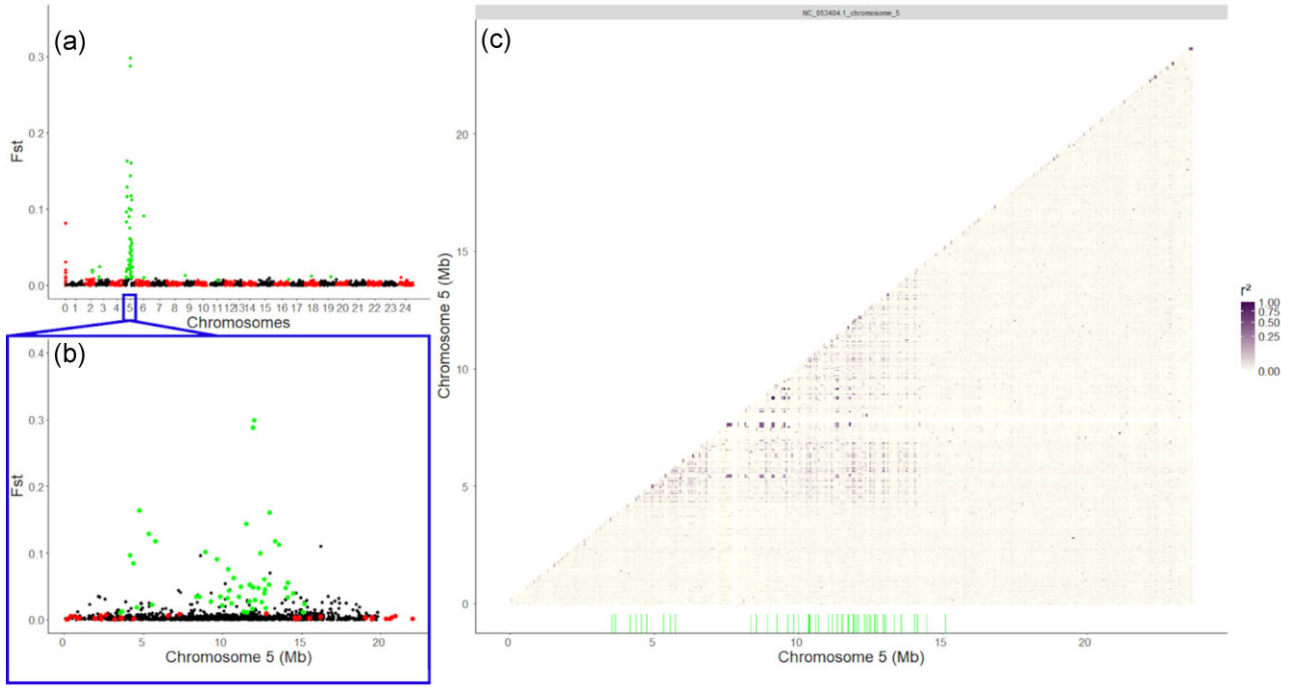

Figure 7. (a) Genome-wide distribution of F_{ST} values across all chromosomes. Each dot represents a SNP, with different shading: lighter shaded dots correspond to the SNPs under potential selection, whereas medium and dark shaded dots correspond to neutral SNPs. (b) Close-up view of chromosome 5, showing a region with elevated F_{ST} values. This zoomed-in section highlights a cluster of SNPs with higher genetic differentiation corresponding to SNPs under potential selection (lightly shaded dots) indicating a potential region of divergence between populations. (c) Linkage disequilibrium (LD) heatmap of chromosome 5, showing the pairwise LD (r^2 values) between SNPs. Darker colors (close to 1) represent higher levels of LD, suggesting regions of genomic similarity or potential structural variations, such as inversions. The lightly shaded ticks at the bottom mark the positions of the SNPs under potential selection.

Eastern Australian samples are more distinct from northern Indian Ocean samples than they are from southern samples

Samples from eastern Australia (SW), typically considered part of the Pacific swordfish population, were included in our analysis as an outgroup. Our findings align with those of Lu et al. (2006), who observed smaller F_{ST} values when comparing swordfish from the southern Indian Ocean with those from the Pacific, rather than with samples from the northern Indian Ocean. Interestingly, despite the significant F_{ST} values indicating genetic differentiation between SW samples and those from the southern Indian Ocean, other analyses, including DAPC, admixture, and *K*-clustering reveal a high degree of genetic similarity between these groups. This suggests that while there is measurable differentiation, the overall genetic structure between these populations remains closely aligned. This observation raises important concerns for fisheries management, as it suggests that current management practices may not accurately reflect the true biological population structure of swordfish—a mismatch often seen in exploited marine species (Reiss et al. 2009). The limited number of samples from eastern Australia, compared to the Indian Ocean, may introduce biases in clustering analyses, potentially masking subtle genetic differences. However, this sampling imbalance is unlikely to affect other analyses, such as pairwise F_{ST} comparisons corroborating the actual management plan for this species (Kalinowski 2005). To further explore the genetic structure between the Indian Ocean and eastern Australian swordfish populations, we plan to increase sampling efforts in eastern Australia, ensuring a more balanced and robust analysis.

Stock assessment models should consider two Indian Ocean stocks: north and south

Our results show that swordfish in the Indian Ocean do not appear to form a panmictic population. Although there is low genetic differentiation between the different regions, a north–south pattern is consistently observed, suggesting that there are two distinct demographic populations. The identification of distinct genetic stocks has important implications for the design of fisheries management strategies (Okumus and Çiftçi 2003). Finally, gathering data on the movement of individual swordfish obtained from satellite tags (Neilson et al. 2014) could complement the present genetic data and enhance our understanding of connectivity patterns that may shape the population structure between the northern and southern Indian Ocean, ultimately informing sound management strategies for these subpopulations.

Acknowledgments

We thank the team at DArT for their work and guidance during the course of this project. Furthermore, we are sincerely grateful to Laura Babet for her assistance during the sample preparation phase of the project. We are thankful to the partners of the IOSSS-Espadon project (European Fund for Fisheries #DAE4/20090164) for the sample collection over the Indian Ocean. Furthermore, we are sincerely grateful to Shane Baylis and Mark Bravington for sharing with us the *kinference* and *gbasics* packages: Kinference: pairwise kin-finding from genotypes and R package version 1.0.2, respectively. We would also like to thank Natalia Diaz Arce from AZTI for her time and advice.

Author contributions

All authors were thoroughly involved in each step of the workflow; however, S. Bonhommeau and S. Bernard were responsible for the conceptualization of this study, while S. Bonhommeau and S. Bernard were responsible for acquiring funding. A-E.N. and D.A.C. led the process of sample design and selection. T.C. led the sample handling process with the help of H.E., B.B. and J.C. T.C. performed the sequence data analyses with the help of G.C., while T.C. and D.A.C. produced the first draft of the manuscript. All authors provided interpretations of the data, as well as reviewed and approved the manuscript.

Supplementary data

Supplementary data is available at *ICES Journal of Marine Science* online.

Conflict of interest: None declare.

Funding

This work has been funded by the European Maritime and Fisheries Fund (EMFF, number PFEA400018DM0980002) for the project FLOPPED. The IOSSS—Espadon project has been co-funded by the European Fund for Fisheries #DAE4/20090164.

Data availability

For this manuscript, the datasets generated and analyzed can be obtained from the corresponding author upon reasonnable request.

References

Anderson G, Lal M, Hampton J *et al.* Close kin proximity in Yellowfin Tuna (*Thunnus albacares*) as a driver of population genetic structure in the tropical western and central Pacific Ocean. *Front Mar Sci* 2019;**6**:341. https://doi.org/10.3389/fmars.2019.00341

Bourjea J, Le Couls S, Muths D *et al.* IOSSS—ESPADON—Etude de la structure du stock d'espadon de l'Océan Indien: Synthèse de la première année. Période novembre 2008–décembre 2009. 2010. https://archimer.ifremer.fr/doc/00018/12914/

Bravington M, Grewe PM, McDowell JR *et al.* Low levels of sibship encourage use of larvae in western Atlantic bluefin tuna abundance estimation by close-kin mark-recapture. *Sci Rep* 2022;**12**:18606. https://doi.org/10.1038/s41598-022-20862-9

Bremer JRA, Mejuto J, Gómez-Márquez J *et al.* Genetic population structure of Atlantic swordfish: current status and future directions. *Col Vol Sci Pap ICCAT* 2007;**61**:107–18.

Bremer JRA, Mejuto J, Greig TW *et al.* Global population structure of the swordfish (*Xiphias gladius* L.) as revealed by analysis of the mitochondrial DNA control region. *J Exp Mar Biol Ecol* 1996;**197**:295–310. https://doi.org/10.1016/0022-0981(95)00164-6

Carvalho GR, Hauser L. Molecular genetics and the stock concept in fisheries.*Rev Fish Biol Fisheries* 1994;**4**:326–50.

Ciftci Y, Okumus I. Fish population genetics and applications of molecular markers to fisheries and aquaculture: i-basic principles of fish population genetics. *Turk J Fish Aquat Sci* 2002;**2**:145–55.

Darnaude AM, Labonne M, Petit C *et al.* Otolith microchemistry suggests probable population structuring in the Indian Ocean for the broadbill swordfish *Xiphias gladius*. *Indian Ocean Tuna Commission*. Victoria, Seychelles, 2020. https://archimer.ifremer.fr/doc/00652/76400/

Davies C, Bravington M, Eveson P *et al.* Next-generation close-kin mark recapture: Using SNPs to identify half- sibling pairs in Southern Bluefin Tuna and estimate abundance, mortality and selectivity . *CSIRO Fish Res Develop Corp*. 2020, https://www.frdc.com.au/project/2016-044

Delaval A, Bendall V, Hetherington SJ *et al.* Evaluating the suitability of close-kin mark-recapture as a demographic modelling tool for a critically endangered elasmobranch population. *Evol Appl* 2023;**16**:461–73. https://doi.org/10.1111/eva.13474

Fraley C, Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput J* 1998;**41**:578–88. https://doi.org/10.1093/comjnl/41.8.578

Gagnaire P-A, Gaggiotti OE. Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Curr Zool* 2016;**62**:603–16. https://doi.org/10.1093/cz/zow088

Gagnaire P, Broquet T, Aurelle D *et al.* Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol Appl* 2015;**8**:769–86. https://doi.org/10.1111/eva.12288

Grewe P, Feutry P, Foster S *et al.* Genetic population structure of sailfish, striped marlin, and swordfish in the Indian Ocean from the PSTBS-IO Project. *GT IOTC, Tasmania, Australia*. 2020. https://hal.science/hal-03351823v1

Grewe PM, Feutry P, Hill PL *et al.* Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Sci Rep* 2015;**5**:16916. https://doi.org/10.1038/srep16916

Grijalva-Chon JM, Numachi K, Sosa-Nishizaki O *et al.* Mitochondrial DNA analysis of North Pacific swordfish *Xiphias gladius* population structure. *Mar Ecol Prog Ser* 1994;**115**:15–9. https://doi.org/10.3354/meps115015

Gruber B, Unmack PJ, Berry OF *et al.* dartr: an r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour* 2018;**18**:691–9. https://doi.org/10.1111/1755-0998.12745

Hernandez-Garcia V. The diet of the swordfish *Xiphias gladius* Linnaeus, 1758, in the central east Atlantic, with emphasis on the role of cephalopods. *Fish Bull* 1995;**93**:403–11.

Hilborn R, Amoroso RO, Anderson CM *et al.* Effective fisheries management instrumental in improving fish stock status. *Proc Natl Acad Sci* 2020;**117**:2218–24. https://doi.org/10.1073/pnas.1909726116

Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst* 2008;**39**:21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

IOTC. Review of the statistical data available for Indian Ocean swordfish (1950-2020). *Indian Ocean Tuna Commission*. Victoria, Seychelles, 2022. https://iotc.org/documents/WPB/20/07e

Jean C, Bourjea J, Jouen E *et al.* Stock structure of the swordfish (*Xiphias gladius*) in the SouthWest Indian Ocean: a preliminary study. *Bull Mar Sci* 2006;**79**:521–6.

Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 2010;**11**:94. https://doi.org/10.1186/1471-2156-11-94

Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008;**24**:1403–5. https://doi.org/10.1093/bioinformatics/btn129

Kalinowski ST. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* 2005;**94**:33–6. https://doi.org/10.1038/sj.hdy.6800548

Kilian A, Wenzl P, Huttner E *et al.* Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol* 2012;**88**:67–89.

Kotoulas G, Magoulas A, Tsimenides N *et al.* Marked mitochondrial DNA differences between Mediterranean and Atlantic populations of the swordfish, *Xiphias gladius*. *Mol Ecol* 1995;**4**:473–82. https://doi.org/10.1111/j.1365-294X.1995.tb00241.x

Kotoulas G, Mejuto J, Antoniou A *et al.* Global genetic structure of swordfish (*Xiphias gladius*) as revealed by microsatellite DNA markers. *Collect Vol Sci Pap ICCAT* 2007;**61**:79–88.

Lee C, Abdool A, Huang C-H. PCA-based population structure inference with generic clustering algorithms. *BMC Bioinf* 2009;**10**:S73. https://doi.org/10.1186/1471-2105-10-S1-S73

Longo GC, Lam L, Basnett B *et al.* Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome. *Evol Appl* 2020;**13**:2536–54. https://doi.org/10.1111/eva.13037

Lu C-P, Chen C, Hui C-F *et al.* Population genetic structure of the swordfish, *Xiphias gladius* (Linnaeus, 1758), in the Indian Ocean and West Pacific inferred from the complete DNA sequence of the mitochondrial control region. *Zool Stud* 2006;**45**:269–79.

Lu C-P, Smith BL, Hinton MG *et al.* Bayesian analyses of Pacific swordfish (*Xiphias gladius* L.) genetic differentiation using multilocus single nucleotide polymorphism (SNP) data. *J Exp Mar Biol Ecol* 2016;**482**:1–17. https://doi.org/10.1016/j.jembe.2016.03.010

Mamoozadeh N, Graves J, McDowell JR. Genome-wide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (*Kajikia audax*), a broadly distributed and highly migratory pelagic species. *Evol Appl* 2019;**13**:677–98.

Mérot C. Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Mol Ecol* 2020;**29**:2513–6. https://doi.org/10.1111/mec.15500

Muths D, Couls SL, Evano H *et al.* Multi-genetic marker approach and spatio-temporal analysis suggest there is a single panmictic population of swordfish *Xiphias gladius* in the Indian Ocean. *PLoS One* 2013;**8**:e63558. https://doi.org/10.1371/journal.pone.0063558

Muths D, Grewe P, Jean C *et al.* Genetic population structure of the Swordfish (*Xiphias gladius*) in the southwest Indian Ocean: sex-biased differentiation, congruency between markers and its incidence in a way of stock assessment. *Fish Res* 2009;**97**:263–9. https://doi.org/10.1016/j.fishres.2009.03.004

Nakamura I. *Billfishes of the world: an annotated and illustrated catalogue of marlins, sailfishes, spearfishes, and swordfishes known to date. FAO Species Catalogue; FAO Fisheries Synopsis*. Vol. **5**. FAO species catalogue 1985, 125. https://www.fao.org/4/ac480e/ac480e10.pdf

Neilson J, Arocha F, Cass-Calay S *et al.* The recovery of Atlantic swordfish: the comparative roles of the regional fisheries management organization and species biology. *Rev Fish Sci* 2013;**21**:59–97. https://doi.org/10.1080/10641262.2012.754842

Neilson JD, Loefer J, Prince ED *et al.* Seasonal distributions and migrations of Northwest Atlantic swordfish: inferences from integration of pop-up satellite archival tagging studies. *PLoS One* 2014;**9**:e112736. https://doi.org/10.1371/journal.pone.0112736

Nielsen E, Wright P, Hemmer-Hansen J *et al.* Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals. *Mar Ecol Prog Ser* 2009;**376**:213–25. https://doi.org/10.3354/meps07798

Noor MAF, Grams KL, Bertucci LA *et al.* Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci* 2001;**98**:12084–8. https://doi.org/10.1073/pnas.221274498

Okumus I, Çiftçi Y. Fish population genetics and molecular markers: II- molecular markers and their applications in fisheries and aquaculture. *Turk J Fish Aquat Sci* 2003;**3**:51–79.

Palko BJ, Beardsley G, Richards WJ. Synopsis of the biology of the swordfish, Xiphias gladius Linnaeus. *FAO Fisheries Synopsis* 1981;**127**. https://www.fao.org/4/ap932e/ap932e.pdf

Pecoraro C, Babbucci M, Franch R *et al.* The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. *Sci Rep* 2018;**8**:13890. https://doi.org/10.1038/s41598-018-32331-3

Reeb CA, Arcangeli L, Block BA. Structure and migration corridors in Pacific populations of the Swordfish *Xiphias gladius*, as inferred through analyses of mitochondrial DNA. *Mar Biol* 2000;**136**:1123–31. https://doi.org/10.1007/s002270000291

Reiss H, Hoarau G, Dickey-Collas M *et al.* Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish Fish* 2009;**10**:361–95. https://doi.org/10.1111/j.1467-2979.2008.00324.x

Righi T, Splendiani A, Fioravanti T *et al.* Mediterranean swordfish (*Xiphias gladius* Linnaeus, 1758) population structure revealed by microsatellite DNA: genetic diversity masked by population mixing in shared areas. *PeerJ* 2020;**8**:e9518. https://doi.org/10.7717/peerj.9518

Rosel PE, Block BA. Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. *Mar Biol* 1996;**125**:11–22. https://doi.org/10.1007/BF00350756

Sansaloni C, Petroli C, Jaccoud D *et al.* Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of eucalyptus. *BMC Proc* 2011;**5**:P54. https://doi.org/10.1186/1753-6561-5-S7-P54

Smith BL, Lu C-P, García-Cortés B *et al.* Multilocus bayesian estimates of intra-oceanic genetic differentiation, connectivity, and admixture in Atlantic swordfish (*Xiphias gladius* L.). *PLoS One* 2015;**10**:e0127979. https://doi.org/10.1371/journal.pone.0127979

Thoya P, Kadagi NI, Wambiji N *et al.* Environmental controls of billfish species in the Indian Ocean and implications for their management and conservation. *Divers Distrib* 2022;**28**:1554–67. https://doi.org/10.1111/ddi.13525

Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw* 2018;**3**:731. https://doi.org/10.21105/joss.00731

Vaux F, Bohn S, Hyde JR *et al.* Adaptive markers distinguish North and South Pacific Albacore amid low population differentiation. *Evol Appl* 2021;**14**:1343–64. https://doi.org/10.1111/eva.13202

Ward RD, Woodwark M, Skibinski DOF. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *J Fish Biol* 1994;**44**:213–32. https://doi.org/10.1111/j.1095-8649.1994.tb01200.x

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984;**38**:1358–70.

Whitlock MC, Lotterhos KE. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F(ST). *Am Nat*. 2015 Oct;**186** Suppl 1:S24–36. https://doi.org/10.1086/682949

Wright S. The genetical structure of populations. *Ann Eugen* 1949;**15**:323–54. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Wu B, Feng C, Zhu C *et al.* The genomes of two billfishes provide insights into the evolution of endothermy in Teleosts. *Mol Biol Evol* 2021;**38**:2413–27. https://doi.org/10.1093/molbev/msab035

Handling Editor: W. Stewart Grant