Population structural dynamics 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], Jérôme Bourjea, Sylvain Bonhommeau[1*]

[1] *IFREMER Délégation Océan Indien, Rue Jean Bertho, 97822 Le Port, La Réunion, France*

[2]*Company for Open Ocean Observations and Logging (COOOL), Saint-Leu, La Réunion, France*

[3]*Univ. Brest, CNRS, IRD, Ifremer, LEMAR, Rue Dumont d'Urville, 29280 Plouzané, France*

[4]*LIRMM-CNRS, rue Ada, 34000 Montpellier, France*

[*]Corresponding authors: thomas.chevrier@ifremer.fr, sylvain.bonhommeau@ifremer.fr

Abstract

The swordfish (*Xiphias gladius*) is of special economic importance in the Indian Ocean. At present, the Indian Ocean Tuna Commission (IOTC) considers the swordfish to be a single panmictic population in the Indian Ocean. Over the last few years, however, several population studies have contested this tenet, and through the implementation of different approaches, have provided conflicting results, including emerging evidence that population structuring exists within this species. Namely, Muths et al (2013) based their examination on the ND2 region of the mitochondrial locus and failed to identify multiple distinct populations within the Indian Ocean, while Grewe et al (2020) applied Single Nucleotide Polymorphic loci (SNPs) to highlight two subpopulations on either side of the equator. These past studies provided ambiguous results

regarding the structure of this species and consequently, called into question the guidelines for management to be adopted for the IOTC. In the present study, we have investigated the structure of the swordfish population in the Indian Ocean using a large number of samples spread over the Indian Ocean. We used a dataset of 1 990 SNPs loci for 1 694 swordfish from 23 distinct locations across the Indian Ocean, generating one of the most extensive datasets for the Indian Ocean swordfish to date. We implemented the Discriminant Analysis of Principal Components (DAPC) statistical technique, which is advantageous for analyzing overlapping or closely related groups. Our findings revealed a low genetic differentiation among swordfish in the Indian Ocean, rather than multiple genetically distinct stocks, further supporting that only one population should be considered for future stock assessment measures. The analysis shows the impact of gene under selection on the estimates of genetic differentiations, i.e. a much lower FST value was found when removing these specific genes.

Keywords: population structure, fisheries management, Single Nucleotide Polymorphism, Discriminant Analysis of Principal Component

Running title: Population dynamics of the Indian Ocean swordfish

Introduction

Fish stock assessment is a critical component of fisheries management worldwide and involves evaluating the abundance and health of fish populations towards developing sound management for promoting sustainable harvesting measures. The development of accurate stock assessments relies upon the definition of discrete biological population units which may have different responses to harvest (Carvalho and Hauser 1994). Fish species can form genetically structured sub-populations that are differentiated by their behavior and geographical distribution, which can result in genetic differences which are adaptive (Çiftci and Okumuş, 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). Poor spatial delimitation of management zones can lead to depletion or even extinction of the most vulnerable or local sub-populations, resulting in a loss of genetic diversity and reducing the species ability to recover from the stress of commercial harvesting and adapt in the face of global change (Carvalho and Hauser 1994; Righi et al. 2020). Over the past decade, the decline of many fish stocks has 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 their high mobility, which facilitates gene flow (Ward, Woodwark & Skibinski, 1994).

The swordfish (*Xiphias gladius*) is a highly migratory, mesopelagic, opportunistic predator, distributed in the temperate and tropical regions of the Atlantic, Pacific and Indian Oceans, and as well as various seas from 45 °N to 45 °S (Nakamura 1985; Hernandez-Garcia 1995). *X. gladius* has the largest commercial value among billfish fisheries and is one of the most

exploited billfish species in the Indian Ocean (23 917 t in 2021, i.e. the second billfish species after sailfish, 37310 t in 2021; IOTC 2015, 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 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* has made this species the center of multiple studies examining its putative population structure. Several molecular genetic tools were applied to *X. gladius* samples primarily to investigate the existence of discrete stocks in the world's oceans. One of the earliest studies examining swordfish was conducted by Grijalva-Chon et al. 1994, who identified one panmictic population in the North Pacific through the use of restriction fragment length polymorphism (RFLP) analysis of mitochondrial DNA (mtDNA). During the same period, the global population structure of *X. gladius* was addressed by Bremer et al. 1996 who examined DNA sequence variation contained in the mtDNA control region of 247 individuals from the Mediterranean sea, Atlantic and Pacific oceans. Two divergent clades were identified, one in the Pacific and the other was composed of individuals from the Atlantic and the Mediterranean sea (Bremer et al 1996). Similar results were obtained by Rosel and Block 1996, this time employing a 300 base pair segment of the 5' end of mtDNA control region. Pujolar et al. 2002 specifically evaluated the population structure of swordfish in the Mediterranean Sea through allozyme analyses, finding a lack of temporal or spatial heterogeneity, which supported the hypothesis of a single population in the Mediterranean Sea and adjacent waters of the Atlantic Ocean. The International Commission for the Conservation of Atlantic Tunas (ICCAT), however, has considered the North

Atlantic and the Mediterranean swordfish populations as two separate fish stocks and consequently recommend independent management actions for these two regions (Viñas et al. 2007; ICCAT 2002, 2021). While many studies agreed to establish the Strait of Gibraltar as the principal region restricting gene flow between the North Atlantic and the Mediterranean, Viñas et al. 2010 rejected the hypothesis of panmixia in the Mediterranean Sea.

Lu et al. 2006 used the mtDNA control region to screen 175 individuals obtained from across the Pacific and Indian Oceans, detecting distinct populations in the waters north of Madagascar and the Bay of Bengal while the rest of the Indian Ocean was considered as a single population. Later, Muths et al. (2009) concentrated their assessments on the southwest Indian ocean (the islands of La Réunion, Glorieuses, Seychelles, and southern Madagascar) through the use of a 517-bp fragment of the mtDNA control region and eleven microsatellite markers in 337 individuals. They recovered a weak differentiation between *X. gladius* caught from these areas, suggesting that the southwest Indian Ocean contains a single panmictic population. A follow-up study by Muths et al. (2013) amplified the mtDNA ND2 gene and as 19 microsatellite markers in 2231 *X. gladius* samples from 12 sites across the Indian Ocean, as well as from the waters off of South Africa and Namibia (Atlantic ocean) and the Coral Sea (Pacific Ocean), finding no evidence of multiple distinct populations within the Indian Ocean. Most recently, Grewe et al. 2020 suggested subtle population structure of swordfish within the Indian Ocean, with at least two genetically differentiated groups present North and South of the equator, as recognized through the use of single nucleotide polymorphism (SNPs) markers genotyped. Overall, several studies confirmed a pattern of gene flow restriction of swordfish to the Atlantic, Pacific and Indian oceans, as well as a structure between the northern and southern hemisphere through the use of a variety of genetic approaches; however, there remains contrasting results (Kotoulas et al.

1995; Bremer et al. 1996; Reeb et al. 2000; Bremer et al. 2007; Smith et al. 2015, Lu et al. 2016, Grewe et al. 2020). Therefore to date, swordfish are considered to be a single panmictic population in the Indian Ocean.

In the present study, we clarify population dynamics of this species as a basis for accurate stock assessments and resolve previous conflicting findings by inspecting whether there is genetic structuring of *X. gladius* across the Indian Ocean through the use of powerful assessment tools. We generated an large dataset of 1 990 Single Nucleotide Polymorphic loci (SNPs) for large-scale automated detection across the genome and ease of modeling mutational dynamics, which generally leads to improved estimates of fine scale differences within and among differing populations when compared to other marker types (Brookes 1999, Morin et al. 2004). Loci are genotyped in a dataset of 2068 female and male juvenile and adult individuals obtained from 23 locations across the Indian Ocean. We use one analytic approach to inferer the genetic cluster to each individual: the Discriminant analysis of principal components (Jombart et al. 2010).

Materials and Methods

Sampling design

Samples are selected from the Ifremer IOSSS - ESPADON project (Bourjea et al. 2010). During the IOSSS project, 3127 *X. gladius* samples have been collected over 2006-2007, 2009-2012, of which 2359 samples were identified as having enough tissue to enable DNA extraction, sequencing and genotyping. We selected 2000 samples for genetic analyses, by prioritizing IOTC area origins, the year of recovery, the individual life stage (i.e. juvenile or adult), sex and geographic origin within the IOTC area (Figure 1). We narrowed the sample set to only samples obtained from within the IOTC convention area, with the exception of 40

samples which were selected from outside of the IOTC convention area (Eastern Australia) to identify contrast in the data. Next, we selected samples primarily obtained during the years 2009 and 2010, and eventually included samples from the year 2011. They were selected as spread as possible in the Indian Ocean area. Since we aimed at doing a parallel CKMR study, samples have also been selected to have a balanced juvenile/adult ratio. A total of 2068 samples were retained (Fig. 1 and Table 1).

Sample submission and processing

Diversity Arrays Technology (DArT PL) based in Canberra, Australia, was chosen to perform SNP genotyping of all selected samples through the use of DarTseq (Sansaloni et al. 2011), a proprietary, cost effective procedure for generating high quality SNP loci datasets of non-model species. DArTseq has previously been used to successfully implement population studies of other large pelagic fishes such as southern bluefin tuna (Bravington et al. 2017), *Thunnus albacares* (Anderson et al. 2019), *Atlantic bluefin tuna* (McDowell et al. 2022) and *Dipturus batis* (Delaval et al. 2023). The DArTSeq workflow was implemented on all 2068 samples selected. Resulting FASTQ files were processed using a proprietary bioinformatic pipeline, beginning with quality control steps that included filtering poor-quality sequences. Sequences were then aligned to the *X. gladius* genome available on the National Center for Biotechnology Information (NCBI) (accession GCF_016859285.1) to call SNP loci, followed by additional quality filtering steps. Final reports were produced and contained genotyped samples, identified SNPs, call rates and co-dominant status for each sample (Supplementary table 2) (Anderson et al. 2019).

Computational processing

Data filtering

Data are 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. It was computationally convenient to code the data as "0" for the homozygous reference allele, "1" for the homozygous SNP, and "2" for heterozygotes. The reference allele is often arbitrarily taken to be the most common allele. "NA" or "-" indicated that the SNP could not be scored. All multivariate analyses were done using four CRAN packages: *dartR* (Gruber et al. 2018); *adegenet* (Jombart et al. 2008); *kinference* (Bravington et al., in prep) and *gbasics* (Bravington et al., in prep)

To import the data into R, a metadata file was created containing information associated with each individual fish. This file contained an "id" column which listed the sample name of each individual and a "pop" column which listed the cluster or population to which an individual belongs. The metadata for each locus were included in the report and an explanation for each locus metadata is shown in Supplementary table 2. Using the gl.read.dart() function from the *dartR* package, the report can be imported into CRAN R with the individual metadata file. SNP data coded in this way are held in a genlight object that is defined in the R package *adegenet*. In general, raw data contains several tens of thousands loci or more; however, not all loci are informative for population structure analyses. In the present study, it was found that less than 2000 SNPs should be retained. SNPs and individuals are filtered based on the sequencing depth, the reproducibility, loci call frequency as well as for individuals, minor allele frequency (MAF), SNPs linkage (linkage disequilibrium) and SNPs under selection. The selection enables us to obtain a reliable dataset for population structure with only unlinked SNPs. The SNP filtering method is described in supplemental table 1. Once the genlight object was filtered, it was

converted into a SPAgeno object with gbasics, which allowed the analyses with *kinference*. After these steps, data are again filtered for samples with atypical genotypes, individuals who may have been contaminated, as well as the deletion of technical replicates, but also those whose DNA is too degraded. All thresholds used for the different filters are reported in (Supplementary table 1).

Sequencing data was produced for 2228 individuals, which included 2038 samples and 189 technical replicates (including one sample that was replicated four times instead of twice) which were successfully processed, meaning that 30 individuals failed sequencing (Supplementary table 1). During the application of the workflow, eight of these 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 true swordfish samples that underwent downstream analyses. A total of 86,409 binary SNPs were returned for 2030 swordfish sampled from around the Indian Ocean. Then, once all the filters have been applied to the raw data, it resulted in a dataset with 1 990 SNPs for 1 694 individuals (Supplementary table 1).

Assignment

As swordfish group priors are unknown, the K-means clustering of principal components was used to identify groups of individuals. The find.clusters() function from *adegenet* (Jombart et al. 2008) was then used for the assignment. As advocated in previous studies (Fraley et al. 1998; Lee et al. 2009), Bayesian Information Criterion (BIC) was retained to assess the best supported model, and therefore the number and nature of clusters. Alternatively, to determine if the sampling location can be a good proxy of the population structure, another dataset was built assigning the IOSSS code to each fish. The clustering method was performed for two sets of loci

selections for comparison purposes, i.e including or not the loci under selection.

Discriminant Analysis of Principal Component (DAPC)

We used a Discriminant Analysis of Principal Components (DAPC, Jombart et al. 2010) on the final data file (report) provided by DArT. 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 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, such as what has been seen with *X. gladius*, which 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.

Distance matrix

The fixation index (F_{st}) was calculated using Nei equations between all the different sampling sites with *hierfstat* package.

Results

The clustering method shows that the population structure is most likely a single population as the Bayesian Information criteria (BIC) is the lowest for a number of clusters of 1 for the dataset without loci under selection (Fig.2b) whereas the lowest value of BIC is reach for 2 clusters if loci under selection are keeping (Fig. 2a).

The DACP approach shows 3 distinct groups. The first includes only Australian samples fished on the east coast (SW), used as the out group and considered as Pacific fish. A second group is made up from samples taken in the northern Indian Ocean, including part of the Indian and Sri Lankan samples (AA; AC and FA). Finally, the last group with all the remaining samples. Results between datasets including genes under selection or not are similar with a higher discrimination for the dataset when selection is not accounted for.

Two Fst matrices were done one for each dataset (with or without using gene under selection) as for DAPC. The first one (with gene under selection, Fig. 4a) confirmed the observations made with DAPC (Fig 3a) where the highest Fst value (blue) were observed between northern samples (AC, AA and FA) with all other samples and the same for Eastern Australia. However, Fst values are still to be low, the highest one ~ 0.003 is reached between AC and SW which suggest a gene flow between these clusters.

For the dataset without loci under selection, all pairwise Fst values are very low (< 1e -3), which seems to confirm the clustering method results (K=1), i.e. very weak population structure. The highest Fst values are found for the samples between the Northern and North-Eastern Indian Ocean (named FA-AC), Australia (SW), and the South Eastern Indian Ocean.

Discussion

Our findings revealed a low genetic differentiation among swordfish populations in the Indian Ocean. The analysis showed no distinct haplotype clusters, indicating the absence of multiple genetically distinct stocks (Figure 3). This lack of genetic structuring suggests a high gene flow among individuals, though a distinction could possibly be made between the North-Eastern, South-Western and the South-Eastern Indian Ocean sectors. The analysis of the population structure using a DAPC analysis shows that only one population should be considered for further stock assessment modeling, including for estimations of abundance through DNA-based approaches such as Close-Kin Mark-Recapture (CKMR) (Bravington et al. 2016). While a distinction could be made between the North-Eastern Indian Ocean, South-East Australia and South-West Indian Ocean, F_{ST} values are very low, indicating the near absence of population structure for this species (Figure 4).

The observed genetic diversity within the different samples suggests that local breeding and recruitment are not occurring and the population is largely panmictic. This indicates that swordfish populations in the Indian Ocean are mixed and are not likely influenced by local factors such as spawning grounds and oceanographic conditions. These results are in agreement with previous studies which analyzed the population structure of swordfish and found very low population structure in the Indian Ocean through the use of differing genetic markers (Muths et al. 2009; Muths et al. 2013) and particularly SNPs (Grewe et al. 2020).

Our results are similarly compared to Grewe et al. 2020, who implemented >15 000 SNP loci in < 300 individuals from 6 locations across the Indian Ocean. Grewe et al. 2020 found a subtle population structure of swordfish within the Indian Ocean with at least two genetically

differentiated groups present north and south of the equator. These results are similar to what we found in the dataset with genes under selection. For the second dataset which does not include the gene under selection, the results are distinct with an important reduction of the north-south differentiation which reflects a single panmictic population.

The identification of distinct genetic stocks has important implications for fisheries management (Okumuş and Çiftci 2003) and the design of CKMR studies (Trenkel et al. 2022). It is important to note that while our study incorporates a considerable number of samples spread across the Indian Ocean, it has certain limitations.

Acknowledgements

We would like to 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.

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, S. Bernard, A-E.N. and D.A.C. 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 bulk of the sequence data analyses with the help of G.C.,

while D.A.C. and T.C. produced the first draft of the manuscript. All authors provided interpretations of the data, as well as reviewed and approved the manuscript.

Funding

The analyses were founded by EU-FEAMP

Conflict of Interest

The authors have no conflicts of interest to declare.

Data Availability

Sequence data was deposited onto. Information for each sample is available at…

References

Anderson, G., Lal, M., Hampton, J., Smith, N., & Rico, C. (2019). Close kin proximity in yellowfin tuna (Thunnus albacares) as a driver of population genetic structure in the tropical western and central Pacific Ocean. Frontiers in Marine Science, 6, 341. https://doi.org/10.3389/fmars.2019.00341

Bourjea, J., Le Couls, S., Muths, D., & Evano, H. (2010). IOSSS-ESPADON-Etude de la structure du stock d'espadon de l'Océan Indien: Synthèse de la première année. Période novembre 2008-décembre 2009.https://archimer.ifremer.fr/doc/00018/12914/

Bravington, M. V., Eveson, J. P., Grewe, P. M., & Davies, C. R. (2017). SBT Close-Kin Mark-Recapture with Parent-Offspring and Half-Sibling Pairs: update on genotyping, kin-finding and model development. ESC, 1708, 12. CCSBT-ESC/1709/12

Bremer, J. R. A., Mejuto, J., Greig, T. W., & Ely, B. (1996). Global population structure of the swordfish (*Xiphias gladius* L.) as revealed by analysis of the mitochondrial DNA control region. Journal of Experimental Marine Biology and Ecology, 197(2), 295-310.https://doi.org/10.1016/0022-0981(95)00164-6

Bremer, J. A., Mejuto, J., Gómez-Márquez, J., Pla-Zanuy, C., Viñas, J., Marques, C., ... & Greig10, T. W. (2007). Genetic population structure of Atlantic swordfish: current status and future directions. Collective Volume of Scientific Papers, ICCAT, 61, 107-118.

Brookes, A. J. (1999). The essence of SNPs. Gene, 234(2), 177-186. https://doi.org/10.1016/S0378-1119(99)00219-X

Carvalho, G. R., & Hauser, L. (1994). Molecular genetics and the stock concept in fisheries. Reviews in fish biology and fisheries, 4, 326-350. https://doi.org/10.1007/BF00042908

Çiftci, Y., & Okumuş, İ. (2002). Fish population genetics and applications of molecular markers to fisheries and aquaculture: I-Basic principles of fish population genetics. Turkish Journal of Fisheries and Aquatic Sciences, 2(2).

Delaval, A., Bendall, V., Hetherington, S. J., Skaug, H. J., Frost, M., Jones, C. S., & Noble, L. R. (2023). Evaluating the suitability of close‐kin mark‐recapture as a demographic modelling tool for a critically endangered elasmobranch population. Evolutionary Applications, 16(2), 461-473. https://doi.org/10.1111/eva.13474

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. The computer journal, 41(8), 578-588. https://doi.org/10.1093/comjnl/41.8.578

Grijalva-Chon, J. M., Numachi, K., Sosa-Nishizaki, O., & De la Rosa-Vélez, J. (1994). Mitochondrial DNA analysis of North Pacific swordfish *Xiphias gladius* population structure. Marine Ecology Progress Series, 15-19. https://www.jstor.org/stable/24849727

Grewe, P., Feutry, P., Foster, S., Aulich, J., Lansdell, M., Cooper, S., ... Davies, C. (2020). Genetic population structure of sailfish, striped marlin, and swordfish in the Indian Ocean from the PSTBS-IO Project (Vol. 19, No. 10). IOTC-2020-WPB18-09

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular ecology resources, 18(3), 691-699. https://doi.org/10.1111/1755-0998.12745

Hernandez-Garcia, V. (1995). *Xiphias gladius* Linnaeus. Fishery Bulletin, 93(2), 403.

Hilborn, R., Amoroso, R. O., Anderson, C. M., Baum, J. K., Branch, T. A., Costello, C., ... & Ye, Y. (2020). Effective fisheries management instrumental in improving fish stock status. Proceedings of the National Academy of Sciences, 117(4), 2218-2224. https://doi.org/10.1073/pnas.1909726116

ICCAT Recommendation, establishing a swordfish statistical document program (2002). https://www.iccat.int/Documents/Recs/compendiopdf-e/2001-22-e.pdf

ICCAT swordfish statistical document (2021). https://media.fisheries.noaa.gov/dam-migration/iccat-swo-statistical-document-11-30-2021.pdf

IOTC (2015). Report of the 18th Session of the IOTC Working Party on Billfish. IOTC–2019–WPB18– R [E], https://archimer.ifremer.fr/doc/00710/82207/87022.pdf

IOTC (2022). Review of the statistical data available for Indian ocean swordfish. IOTC-2022-WPB20-07. https://iotc.org/documents/WPB/20/07e

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics, 11(1), 1-15. https://doi.org/10.1186/1471-2156-11-94

Kotoulas, G., Magoulas, A., Tsimenides, N., & Zouros, E. (1995). Marked mitochondrial DNA differences between Mediterranean and Atlantic populations of the swordfish, *Xiphias gladius*. Molecular Ecology, 4(4), 473-482. https://doi.org/10.1111/j.1365-294X.1995.tb00241.x

Lee, C., Abdool, A., & Huang, C. H. (2009). PCA-based population structure inference with generic clustering algorithms. BMC bioinformatics, 10(1), 1-13. https://doi.org/10.1186/1471-2105-10-S1-S73

Lu, CP, Chen, CA, Hui, CF, Tzeng, TD, & Yeh, SY (2006). 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. Zoological Studies. 45.

Lu, C. P., Smith, B. L., Hinton, M. G., & Bremer, J. R. A. (2016). Bayesian analyses of Pacific swordfish (*Xiphias gladius* L.) genetic differentiation using multilocus single nucleotide polymorphism (SNP) data. Journal of Experimental Marine Biology and Ecology, 482, 1-17. https://doi.org/10.1016/j.jembe.2016.03.010

McDowell, J. R., Bravington, M., Grewe, P. M., Lauretta, M., Walter III, J. F., Baylis, S. M., ... & Porch, C. (2022). Low levels of sibship encourage use of larvae in western Atlantic bluefin tuna abundance estimation by close-kin mark-recapture. Scientific reports, 12(1), 18606. https://doi.org/10.1038/s41598-022-20862-9

Morin, P. A., Luikart, G., Wayne, R. K., & SNP Workshop Group. (2004). SNPs in ecology, evolution and conservation. Trends in ecology & evolution, 19(4), 208-216. https://doi.org/10.1016/j.tree.2004.01.009

Muths, D., Grewe, P., Jean, C., & Bourjea, J. (2009). 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. Fisheries Research, 97(3), 263-269. https://doi.org/10.1016/j.fishres.2009.03.004

Muths, D., Le Couls, S., Evano, H., Grewe, P., & Bourjea, J. (2013). 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, 8(5), e63558.https://doi.org/10.1371/journal.pone.0063558

Nakamura, I. (1985). FAO Species Catalogue: vol. 5. Billfishes of the World. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and swordfishes known to date (No. 5).

Nei, M. (1987) Molecular Evolutionary Genetics. Columbia University Press

Okumuş, İ., & Çiftci, Y. (2003). Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. Turkish Journal of Fisheries and Aquatic Sciences, 3(1). https://dergipark.org.tr/en/pub/trjfas-ayrildi/issue/13292/160651

Poisson, F., & Fauvel, C. (2009). Reproductive dynamics of swordfish (*Xiphias gladius*) in the southwestern Indian Ocean (Reunion Island). Part 2: fecundity and spawning pattern. Aquatic Living Resources, 22(1), 59-68. doi:10.1051/alr/2009012

Pujolar, J. M., Roldan, M. I., & Pla, C. (2002). A genetic assessment of the population structure of swordfish (*Xiphias gladius*) in the Mediterranean Sea. Journal of Experimental Marine Biology and Ecology, 276(1-2), 19-29. https://doi.org/10.1016/S0022-0981(02)00245-9

Reeb, C., Arcangeli, L. & Block, B. Structure and migration corridors in Pacific populations of the Swordfish *Xiphias gladius*, as inferred through analyses of mitochondrial DNA. Marine Biology 136, 1123–1131 (2000). https://doi.org/10.1007/s002270000291

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

Righi, T., Splendiani, A., Fioravanti, T., Petetta, A., Candelma, M., Gioacchini, G., ... & Barucchi, V. C. (2020). Mediterranean swordfish (*Xiphias gladius* Linnaeus, 1758) population structure revealed by microsatellite DNA: genetic diversity masked by population mixing in shared areas. PeerJ, 8, e9518. https://doi.org/10.7717/peerj.9518

Rosel, P.E., Block, B.A. Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius* . Marine Biology 125, 11–22 (1996). https://doi.org/10.1007/BF00350756

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011, December). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. In BMC proceedings (Vol. 5, No. 7, pp. 1-2). BioMed Central. https://doi.org/10.1186/1753-6561-5-S7-P54

Smith, B. L., Lu, C. P., García-Cortés, B., Viñas, J., Yeh, S. Y., & Alvarado Bremer, J. R. (2015). Multilocus Bayesian estimates of intra-oceanic genetic differentiation, connectivity, and admixture in Atlantic swordfish (Xiphias gladius L.). PLoS One, 10(6), e0127979. https://doi.org/10.1371/journal.pone.0127979

Thoya, P., Kadagi, N. I., Wambiji, N., Williams, S. M., Pepperell, J., Möllmann, C., ... & Maina, J. (2022). Environmental controls of billfish species in the Indian Ocean and implications for their management and conservation. Diversity and Distributions, 28(8), 1554-1567. https://doi.org/10.1111/ddi.13525

Trenkel, V. M., Charrier, G., Lorance, P., & Bravington, M. V. (2022). Close-kin mark–recapture abundance estimation: practical insights and lessons learned. ICES Journal of Marine Science, 79(2), 413-422. https://doi.org/10.1093/icesjms/fsac002

Viñas, J., Bremer, J. A., Mejuto, J., de la Serna, J. M., García-Cortés, B., & Pla, C. (2007). Swordfish genetic population structure in the North Atlantic and Mediterranean. ICCAT Coll Vol Sci Pap, 61, 99-106.

Viñas, J., Pérez-Serra, A., Vidal, O., Alvarado Bremer, J. R., & Pla, C. (2010). Genetic differentiation between eastern and western Mediterranean swordfish revealed by phylogeographic analysis of the mitochondrial DNA control region. ICES Journal of Marine Science, 67(6), 1222-1229. https://doi.org/10.1093/icesjms/fsq031

Ward, R. D., Woodwark, M., & Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of fish biology, 44(2), 213-232. https://doi.org/10.1111/j.1095-8649.1994.tb01200.x

Tables

Table 1 :

Location	IOSSS code	Year	Number of fish
Australia	SW	2008 - 2009	16
Glorieuse Island	GL	2006	27
Madagascar	IA	2009 - 2010	122
	IB	2009	69
	IC	2009	142
	ID	2009	58
	XA	2009	248
Mayotte	TD	2009	66
	TF	2010	82
Mozambique channel	TB	2010	27
Reunion Island	IE	2010	57
	IF	2010	36
	XB	2010	90
	XC	2010	108
Seychelles	SA	2010 - 2011	88
	SB	2009	71
	TE	2011	55
Sri Lanka	FA	2009 - 2010	13
	FB	2010	48
South Africa	BB	2009	46
	BC	2009 - 2010	35
	BE	2009 - 2010	76
Thailande	AA	2010 - 2011	40
	AC	2009 - 2010	74

Figures

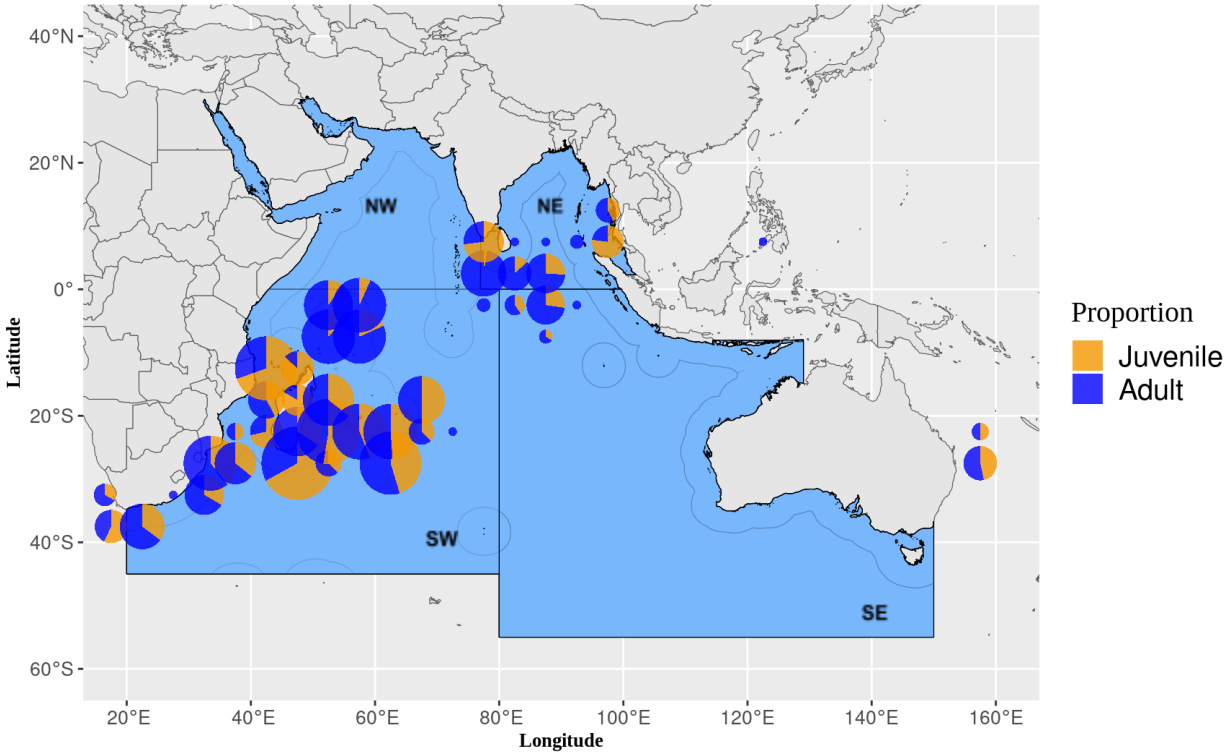

Figure 1: Sample location and log proportion of juvenile (orange) vs adult (blue) for each campaign in the different IOTC areas (NW : North-West ; NE : North East ; SW ; South-West ; SE : South-East)

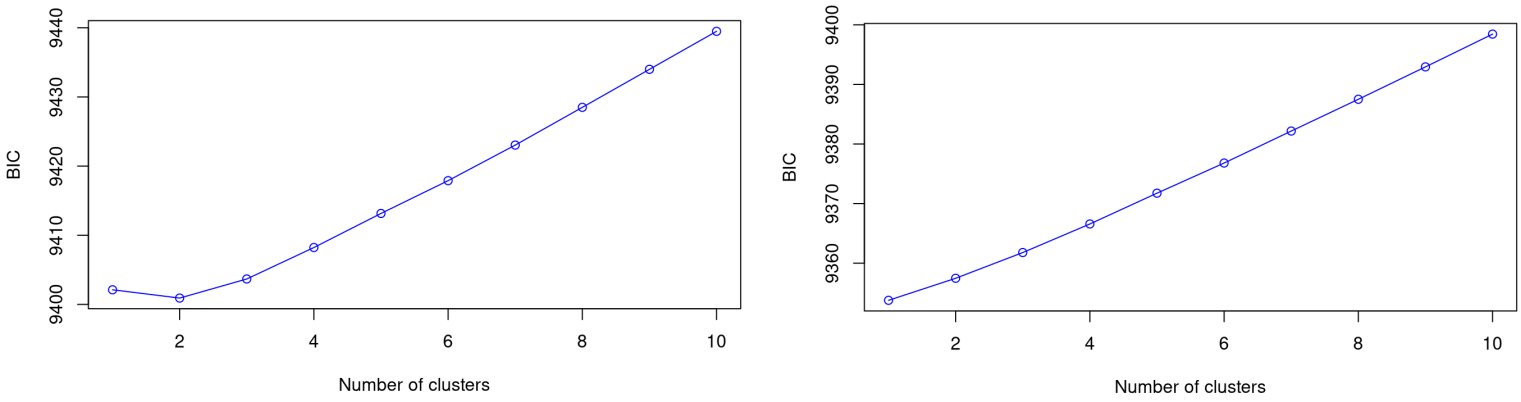

Figure 2: Goodness of the fit (Bayesian Information Criteria) for different numbers of populations from the clustering method. (a) Dataset with loci under selection (b) Dataset without loci under selection

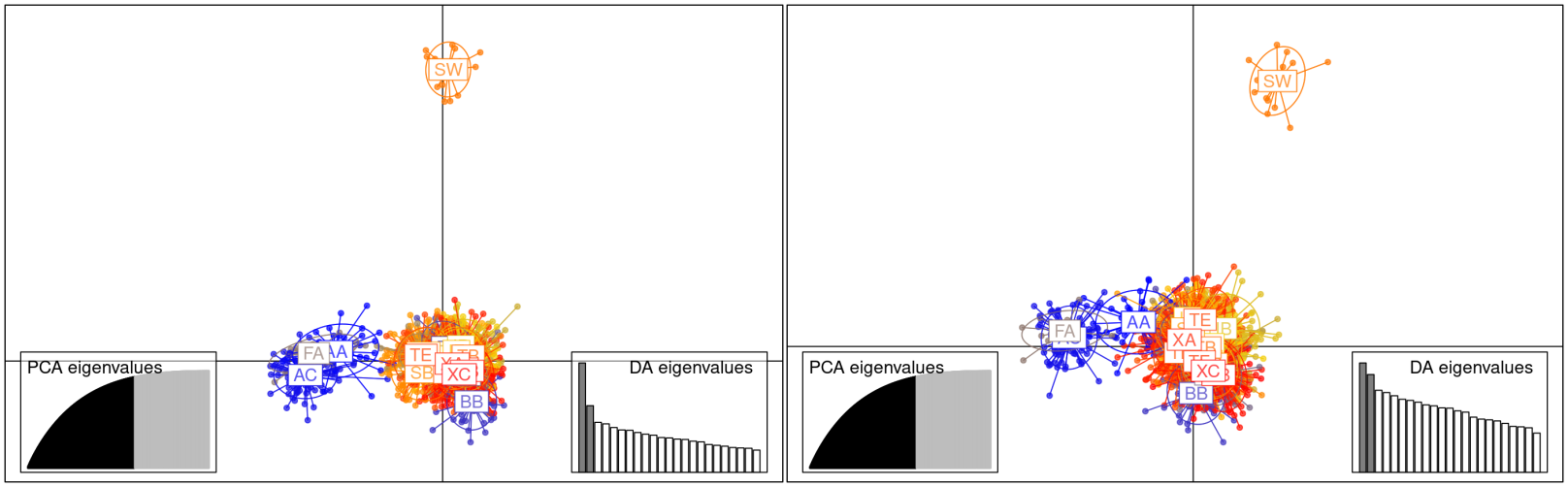

Figure 3: Discriminant Analysis of Principal Components (DAPC) for the first 2 axes showing how genetically close are the different samples, grouped by their sampling areas. 90% of the PCA eigenvalues and the first 2 discriminant analysis eigenvalues are used. (a) For dataset with loci under selection (b) without loci under selection.

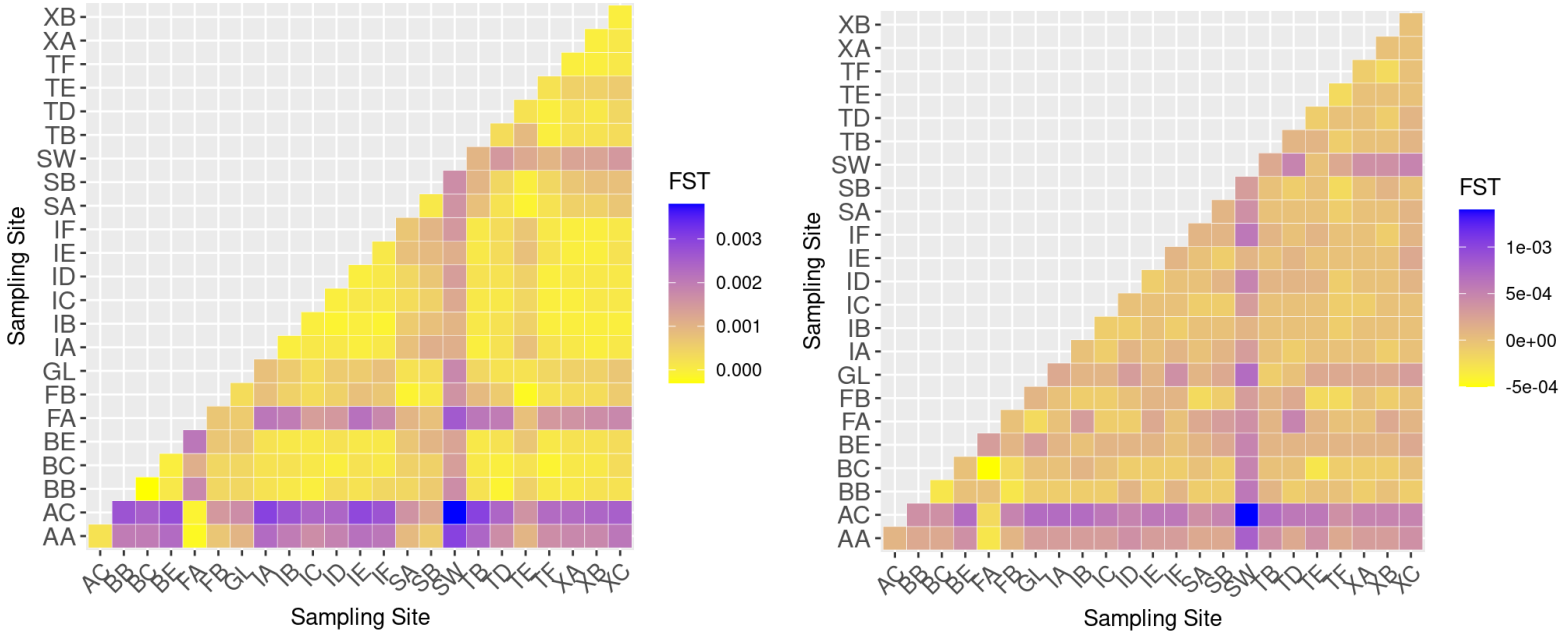

Figure 4: Pairwise fixation index (F_{ST}) between the different sampling areas (a) for dataset with loci under selection (b) without loci under selection

Supplementary data

Supplementary table 1

Parameters	Threshold	Loci filtered	Individuals filtered
Total samples submitted : 2068			
Total samples submitted : 2068			2068
Technical replicates	/	/	189
Total sample files remaining	/	/	2227
Samples not sequenced	/	/	30
Non-swordfish samples	/	/	8
Final swordfish samples/files for downstream processing : 2030			
Population structure bioinformatic process			
Linkage disequilibrium (dartR)	>1	41 830	/
Sequencing depth (*dartR*)	Between 20 and 100	25 845	/
Reproducibility (*dartR*)	0.95	46	/
Call Rate by locus (*dartR*)	0.99	7 313	/
Call Rate by individual (*dartR*)	0.95	/	2
Minor Allele Frequency (MAF) (*dartR*)	0.05	9 332	/
DNA quality and contamination (*kinference*)	/	/	327
Delete IOSSS group with not enough individuals	< 10 fish	/	7
Monomorphic loci	/	0	/
Under selected loci (outflank)	0.05	53	/
Final dataset for population structure : 1 694 individuals and 1 990 SNPs			

Supplementary table 2

Information	Meaning
SNP	Mutational change and its position in the sequence tag referenced from zero
SNP Position	Position (zero is position 1) in the sequence tag of the defined SNP variant base
Trimmed Sequence	The sequence containing the SNP or SNPs (the sequence tag) trimmed of adaptors
Call Rate	Proportion of samples for which the genotype call is nonmissing (that is not "-")
OneRatioRef	Proportion of samples for which the genotype is 0
OneRatioSNP	Proportion of samples for which the genotype is 2
FreqHomRef	Proportion of samples homozygous for Reference allele
FreqHomSNP	Proportion of samples homozygous for the Alternate (SNP) allele
FreqHets	Proportion of samples which score as heterozygous that is scored as 1
PICRef	Polymorphism information content (PIC) for the reference allele
PICSnp	Polymorphism information content (PIC) for the SNP
AvgPIC	Average of the polymorphism information content (PIC) of the Reference and SNP alleles
AvgCountRef	Sum of the tag read count for all samples, divided by the number of samples with non-zero tag read count, for the Reference allele row
AvgCountSnp	Sum of the tag read count for all samples, divided by the number of samples with non-zero tag read count, for the Alternate (SNP) allele row
RepAvg	Proportion of technical replicate assay pairs for which the marker score is consistent