Australia's National
Science Agency

Genetic population structure of sailfish, striped marlin, and swordfish in the Indian Ocean from the PSTBS-IO Project

Peter Grewe, Pierre Feutry, Scott Foster, Jorden Aulich, Matt Lansdell, Scott Cooper, Naomi Clear, Paige Eveson, Daniel Fernando, Audrey M. Darnaude, Natacha Nikolic, Zulkarnaen Fahmi, Francis Marsac, Jessica Farley, Campbell Davies

18 August 2020

Paper submitted to the 10th meeting of the Working Party on Billfish

Abstract

Swordfish, striped marlin and sailfish are the dominant billfish species caught in the Indian Ocean. These species are currently assessed and managed as single stocks. Evidence from population genetics studies to date has not provided evidence that a revision to this approach is warranted. Exploring and understanding the level of population heterogeneity is a priority for sustainable management of these fisheries. This paper presents results from a recent investigation of population structure of swordfish, striped marlin and sailfish using cutting-edge sequencing technology as part of a larger collaborative project "Population Structure of IOTC species and sharks of interest in the Indian Ocean (PSTBS-IO)". Sampling of striped marlin was challenging with only three individuals collected within the Indian Ocean, permitting only a rudimentary population comparison to a single outlier location in the Coral Sea, SW Pacific Ocean. Similarly, sailfish also proved difficult to obtain, however, sufficient samples were obtained to permit analysis from two Indian Ocean locations. No population structure was revealed for this comparison, which was consistent with sailfish representing a highly mixed Indian Ocean population. For both striped marlin and sailfish, sampling coverage was insufficient for the current study to unequivocally conclude whether there was evidence of sufficiently restricted gene flow to consider whether the Indian Ocean basin contains stocks of these species that are effectively isolated population(s) from the Atlantic and/or Pacific Oceans for fisheries management purposes. In contrast, swordfish samples permitted a broad spatial analysis that included 6 sampling locations within the Indian Ocean and an outlier sample location from the southwest Pacific in the Coral Sea. Differences in gene-frequencies of swordfish between the Indian and south-west Pacific Ocean, in the current study, and results from earlier studies from the south-east Atlantic Ocean support the assessment and management of swordfish within the Indian Ocean as an independent unit. Notwithstanding this conclusion, the potential for connectivity between the NE Indian and western Pacific Oceans requires further investigation. The indication of different northern and southern genetic groupings of swordfish, which may represent different reproductive populations, warrants targeted investigation via additional structured sampling and analysis that includes and expands the areas covered in this project and, ideally, includes sampling of spawning adults. For the three billfish species examined, insufficient sampling coverage was obtained for both sailfish and striped marlin to enable adequate assessment of population genetic structure. However, results for swordfish indicate consideration of management strategy refinement may be warranted, and highlights the need for targeted research to confirm the temporal stability of the results presented here, and in particular, further extending of sample coverage in the northern regions of the IO. Additional sampling both within and from locations outside the Indian Ocean should also be completed in order to provide a more comprehensive understanding of inter-ocean population boundaries for all three Indian Ocean billfish species.

Introduction

Indo Pacific sailfish, striped marlin, and swordfish are among commercially important billfish species managed within the Indian Ocean by the IOTC. Little investigation of stock structure exists for Indo-Pacific sailfish (*Istiophorus platypterus*) as they are not a target species for large fisheries

in the Indian ocean; however long line fisheries in Iran, Sri Lanka, India and Pakistan have reported significant bycatch of Indo-Pacific sailfish (Ganga et al., 2008). Previous work by Hoolihan et al., (2004) using mtDNA indicated that significant genetic differences exist between populations inside and outside of the Arabian Gulf; no differentiation between populations within the Indian ocean was found. Speare (1995) indicated that finer stock structure might exist within the east coast of Australia, however this work did not include the use of DNA markers for Indo-Pacific sailfish and relied on the study of parasite communities present on the sailfish. Striped marlin (*Kajikia audax*) are found throughout tropical and temperate waters in both the Indian and Pacific Oceans. While there have been reports of striped marlin from the South East Atlantic Ocean in waters near South Africa, reproduction appears to be solely confined to the Indian and Pacific Oceans. Examination of striped marlin population structure using SNP technology has shown that six genetically distinct populations could be genetically identified among sampling locations broadly representing the full species range (Mamoozadeh et al., 2020). Mamoozadeh (et al., 2020) provided the first examination of population structure of this species within the Indian Ocean and identified the presence of genetically differentiated western and eastern groups of striped marlin. Swordfish (*Xiphias gladius*) is one of the most widely distributed pelagic species found in both tropical and temperate waters of the Atlantic, Indian and Pacific Oceans. In contrast to its broad distribution and occurrence in diverse habitats, tagging data has suggested a high degree of residency and homing behaviour (Sedberry and Loefer, 2001). Restricted gene flow was confirmed by a number of population structure studies that have shown population differentiation of swordfish, particularly in the Atlantic and Pacific Oceans where there is clear evidence of stock structure between northern and southern hemispheres (Kotoulas et al., 1995; Bremmer et al., 1996; Reeb et al., 2000; Bremer et al., 2005; Smith et al., 2016). In particular, differentiation of mtDNA haplotypes may even indicate existence of a cryptic species of swordfish the Mediterranean Sea (Pappalardo et al., 2011). While the studies of Lu et al. (2006) and Bradman et al. (2009) indicated genetic heterogeneity was present for swordfish within in the Indian Ocean, others have only demonstrated inter-ocean differentiation (Jean et al., 2006; Muths et al., 2009; Muths et al., 2013; Smith et al., 2015). In some studies, examination of maternally inherited mitochondrial DNA demonstrated genetic heterogeneity while others examining bi-parental DNA microsatellite markers were unable to detect population structure within the same samples (Lu et al., 2006; Bradman et al., 2011; Muths et al., 20013; Muths et al., 2003). While nuclear DNA markers have been used to clearly detect population structure in both the Atlantic and Pacific Ocean, the limited the limited number of loci explored using DNA microsatellites have either lacked appropriate resolution to adequately examine genetic differentiation or authors have contended male mediated gene flow within the Indian Ocean has been sufficient to homogenize populations (Muths et al., 2013). The current study extends earlier work and further characterises population genetic structure of these three billfish species through examination of nuclear DNA variation at SNP markers among fish sampled from broadly distributed locations located in the Indian and Pacific Oceans.

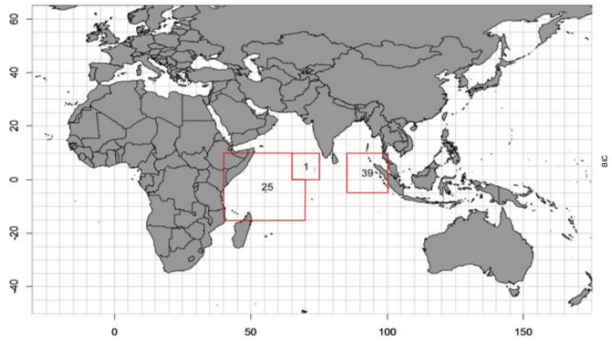
Methods

Sampling

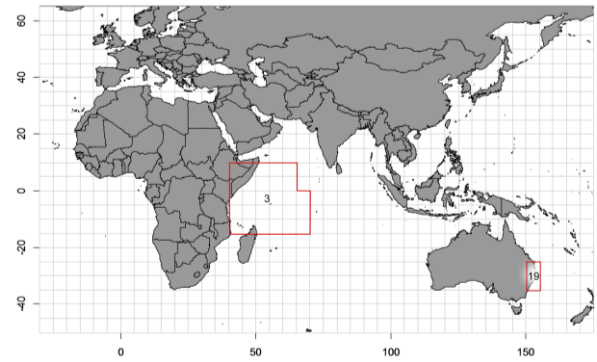
Tissues from individual fish were collected from several locations within the Indian Ocean between 2017 and 2019 (Fig. 1; Table 1) and preserved in RNA-Later following protocols developed for the project (Anon 2018). All samples were shipped to the CSIRO labs in Hobart Australia.

Figure 1. Sample location and numbers for sailfish (A), striped marlin (B), and swordfish (C).

(A)



(B)



(C)

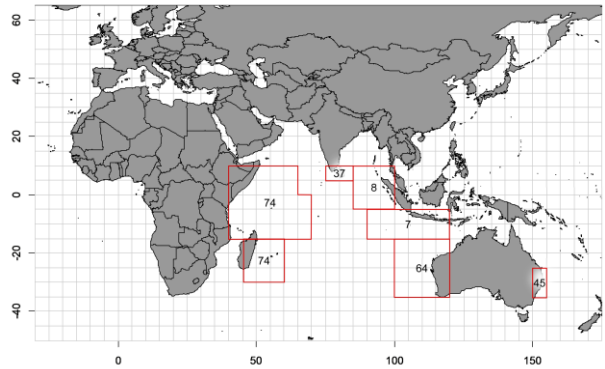


Table 1. Summary of sailfish (SFA), striped marlin (MLS) and swordfish (SWO) used in this study.

Species	No. tissue samples	No. selected for sequencing	No. analysed (post quality control)	Range of lengths (cm)
SFA	84	79	65	50-250
MLS	27	22	20	Length data not available
SWO	616	417	309	80-250

DArTSeq library preparation and sequencing

DNA extractions were prepared from approximately 15mg of tissue subsampled from individual biopsies. Samples were extracted on an Eppedorf EP motion 5057 liquid robotic handler using a modification of the QIAamp® 96 DNA QIAcube HT Kit (QIAGEN, Hilden, Germany). This extraction includes a lysis step in the presence of Proteinase K followed by bind-wash-elute QIAGEN technology. Low quality/degraded samples were re-extracted using the modified CTAB method following Grewe et al. (1993). DNA aliquots were shipped to Diversity Array Technologies (DArT) in Canberra where DNA complexity reduction and library construction was performed prior to sequencing that was used to generate genotype data for each individual.

The genotyping procedure followed the one used by Grewe et al. (2015). In brief, DNA sample libraries were created in digestion/ligation reactions using two restriction enzymes, *Pst*I and SphI. The PstI site was compatible with a forward adapter that included an Illumina flow cell attachment sequence and a sequencing primer sequence incorporating a "staggered", varying length barcode region. *Sph*I- generated a compatible overhang sequence that was ligated to a reverse adapter containing a flow cell attachment region and reverse priming sequence. Only "mixed fragments" (*Pst*I-*Sph*I) were effectively amplified by PCR. PCR conditions consisted of an initial denaturation at 94°C for 1 min followed by 30 cycles of 94°C for 20 sec, 58°C for 30 sec and 72°C for 45 sec, with a final extension step at 72°C for 7 min. After PCR, equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to cBot (Illumina) bridge PCR, followed by sequencing on an Illumina Hiseq2000 using 77 cycles. Sequence clustering and SNP calling was then performed by DArT using their proprietary DArT-Soft14 analytical software.

Quality control filtering of loci and individuals

A step wise process for data quality control using the package RADIATOR (Gosselin 2020) was performed to filter out both poor-quality DNA markers (SNP loci) and poor-quality individual samples (i.e. low DNA quality/quantity or DNA contamination). The following parameters were evaluated as part of the SNP quality control: reproducibility (proportion of repeatable genotype calls estimated via inclusion of technical replicates, i.e. some DNA samples are prepared and sequenced twice); call rate (proportion of samples genotyped); minor allele count (how many times the allele in lowest frequency was observed); coverage (how many times a locus was observed in average across all individuals); position of the SNP on the fragment being sequenced; linkage disequilibrium, multiple SNP on the same fragment are not independent of each other and we only retained the one with the highest minor allele counts (most informative); Hardy Weinberg equilibrium within each sampling location; missingness (proportion of times a loci was reported as NA); minor allele frequency per sampling region. For the quality control of individuals, the following parameters were evaluated: missingness (proportion of missing values); heterozygosity (proportion of locus exhibiting two different alleles); total coverage (total number of sequences obtained per individual); genetic distance (measure of how similar a pair of individuals are). Thresholds used during the filtering process foreach of these parameters are provided in appendix 1-3.

Statistical analysis

Groups were sought in the genetic data using the approach outlined in Foster et al. (2018) as implemented in the R package stockR (Foster 2018). This statistical method aims to find the groups of individuals within which the genetic profiles are more similar to those between groups. Foster et al. (2018) showed that an effective way to navigate the parameter space, to maximise the likelihood, is to use an initial grouping based on a K-means analysis of dimension-reduced data, via principle components. This implies that the analysis uses the method of Jombart et al. (2010) as starting values. However, while these are excellent starting values, for any individual data set there may still be a higher likelihood peak. We altered the choice of starting values very slightly to add confidence in the estimation process and the full set of starting values used are: (i) completely random starts. Each individual is assigned at random to one of the putative groups; (ii) random assignment of sampling regions. This attempts to use the extra information that regions are *likely*, but not necessarily, to contain a similar genetic profile. When there are fewer putative genetic groups in the analysis (K) than sampling regions, then sampling regions are randomly assigned amongst the groups. When there are more putative genetic groups than sampling regions, then the 'extra' sampling groups are made up of a random sample of individuals from all the sampling regions; (iii) K-means clustering based on PCA-rotated data (as in Foster et al.; 2018) but with a random number of principle components. The standard method calls for 100 components, which was thought to be an upper bound. The random number is taken from a minimum of 1 and a maximum of 100.

It is important to note that none of the methods associated with stockR incorporate spatial information, which makes assessing the number of likely groups difficult. To adjust for this deficiency, our primary approach to choosing the number of groups is pragmatic: we take the largest number of groups that maintain geographical coherency. Here geographical coherency is defined (albeit subjectively) as the case where many/most of the individuals within a sampling region share a similar genetic grouping. Additionally, we used a resampling method similar to cross-validation to give an empirical indication of performance for our grouping. Here we repeatedly resample the genetic data and see how well the groupings match those from the analysis of the full data. Whilst we label this a cross-validation, as it has many similarities, we note that it technically is not due to the fact that we don't observe the true groupings – we infer them from the full data set.

The groupings are displayed using an individual fish's probability of belonging to each genetic group. These probabilities are obtained using bootstrap methods (Foster et al. 2018), using 250 resamples. The affinity to a genetic group is measured by an individual's probability – high, or low, probability means that it is more, or less, likely to be part of that group, respectively.

Results and discussion

Indo-Pacific Sailfish – population genetics

Little investigation of stock structure exists for Indo-Pacific sailfish as a large targeted fishery does not exist in the Indian Ocean; however long-line fisheries in Iran, Sri Lanka, India and Pakistan have

reported significant bycatch of this species (Ganga et al., 2008). Previous work by Hoolihan et al., (2004) using mtDNA indicated that significant genetic differences exist between populations inside and outside of the Arabian Gulf; no differentiation between populations within the Indian ocean was found. Speare (1995) indicated that finer stock structure might exist within the east coast of Australia, however this work did not include the use of DNA markers for Indo-Pacific sailfish and relied on the study of parasite communities present on the sailfish.

In this project we genotyped 99 Indo-Pacific Sailfish samples (including technical replicates) for a total of 54,827 SNPs. Our Radiator quality control left 65 unique individuals and 49,030 SNPs available for downstream analysis (see appendix 1).

Our stockR analysis revealed no genetic structure between sampling locations (Seychelles and Lampulo) in the Indian ocean (Figure 2; Figure 9, Appendix 1). Further clarification of stock structure including samples from the Arabian Gulf and possibly the East Coast of Australia would be required in order to assess the appropriateness of a single stock assumption underpinning the current management arrangements for sailfish within the Indian Ocean.

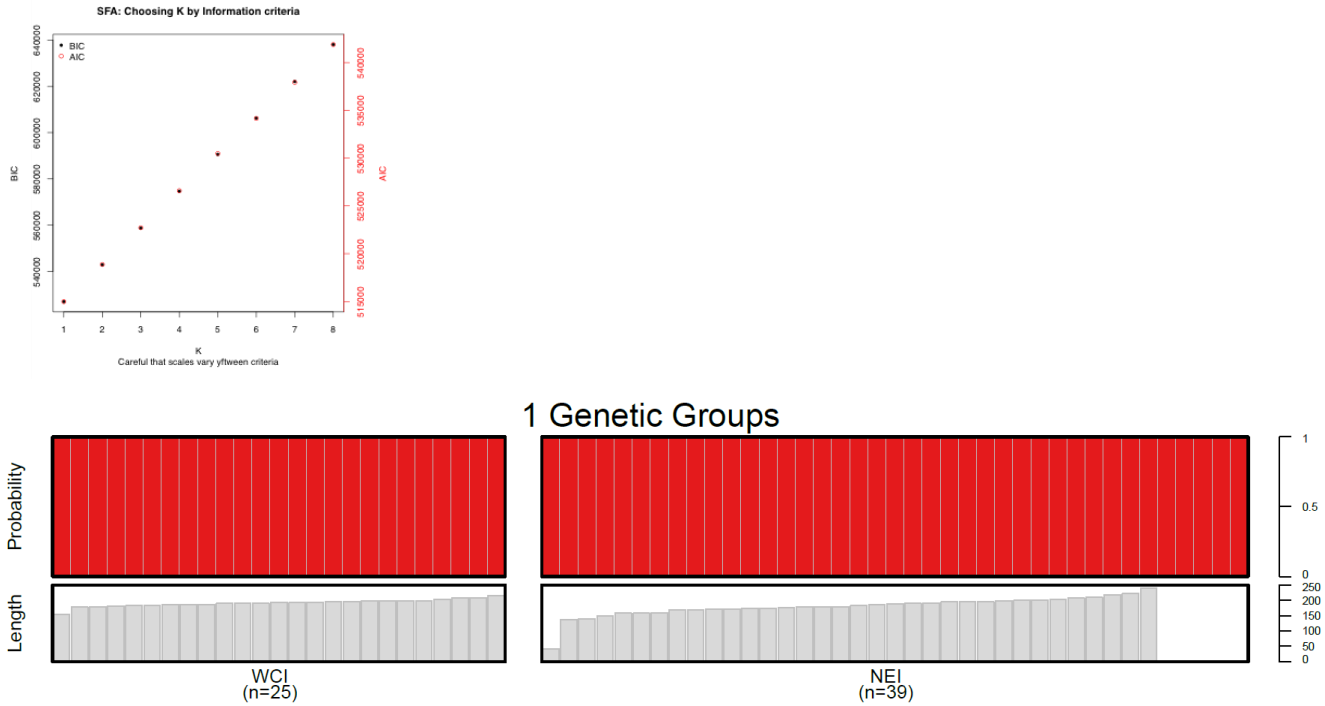


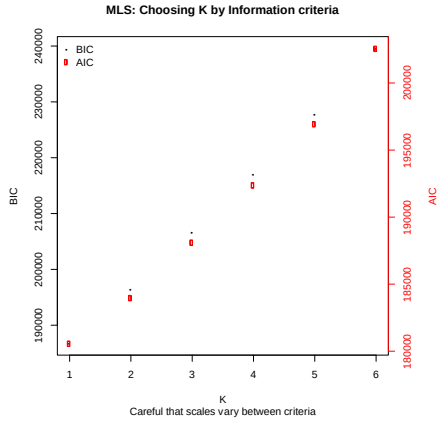
Figure 2. Top : Information criterion used to assess the likelihood of different numbers of genetic groups (k). Bottom: plot of K=1 genetic groups and numbers of fish assessed for population structure indicated in brackets.

Striped Marlin – population genetics

Striped marlin (*Kajikia audax*) are found throughout tropical and temperate waters in both the Indian and Pacific Oceans. While there have been reports of striped marlin from the South East Atlantic Ocean in waters near South Africa, reproduction appears to be solely confined to the Indian and Pacific Oceans. Examination of striped marlin population structure using SNP technology has shown that six genetically distinct populations could be identified among sampling locations broadly representing the full range of the species (Mamoozadeh et al., 2020).

Mamoozadeh (et al., 2020) provided the first examination of population structure of this species within the Indian Ocean and identified the presence of genetically differentiated western and eastern groups of striped marlin. Compared to this previous study our results, which were based on comparison of three western Indian Ocean fish against 17 from the Coral Sea (south western Pacific Ocean), were unable to demonstrate genetic differentiation between these two regions with reasonable statistical precision (i.e. greater than 95%). However, average individual heterozygosity levels were different between the west Indian Ocean samples compared to those from the Pacific (27% versus 32% respectively). Admittedly while this result is based on only three west Indian Ocean samples versus 17 from the Pacific, the differences are indicative of taking samples from two genetically distinct cohorts of fish (Figure 3). Further temporal and spatial comparison of population structure of striped marlin will require more focussed engagement and cooperation of local fisheries at broader sampling scales within the Indian Ocean in order to secure sufficient samples for more powerful analyses.

(A)



(B)

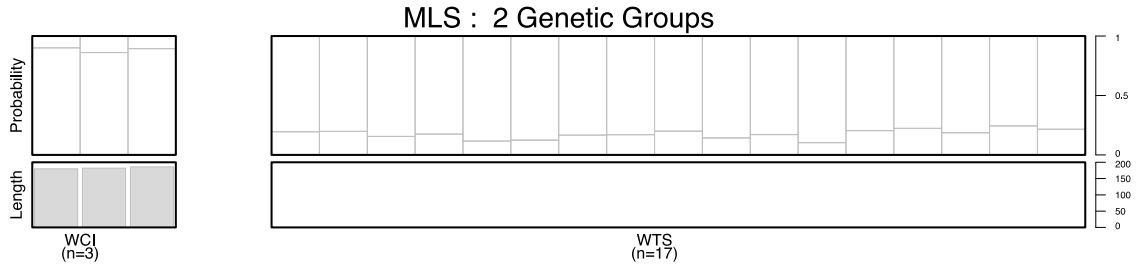


Figure 3. (A) Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. (B) Results of population structure analysis of DArTSeq using StockR for striped marlin for 2 genetic groups.

Swordfish (*Xiphias gladius*) – population genetics

The population structure of swordfish was examined through analysis of variation at single nucleotide polymorphic (SNP) loci from 6 sample locations within the Indian Ocean and one from the Coral Sea in the south-west Pacific Ocean (Figure 1). The sizes of sampled fish ranged from sub-adults (80-100cm) to large (>200cm) mature adults (Figure 10, Appendix 2). Following quality control filtering using Radiator (Gosselin 2019, https://github.com/thierrygosselin/radiator) the DArT-SEQ SNP sequencing data produced by Diversity Arrays Technology produced genotype information at 15,070 SNP loci using the filtering parameters listed in Appendix 3. The genotype data was further analysed using the program StockR to examine geographical distribution for each individual when modelled for K genetic groupings from 1 through 7 (Figure 10, Appendix 2). Examination of gene frequency results for values of K>1 suggest K=2, or possibly, K=3 based on geographical pattern of groupings (Figure 6; Figure 10, Appendix 2). In contrast, values calculated for the BIC and AIC, which was based on results of population differentiation analysis using the package stockR, indicated that K=1 as the preferred value for K (Figure 4). However, when modelled at K=2and K=3, there is appears to be a strong clustering of genotypic groups restricted to samples from the northernmost locations with the highest (almost fixed) frequency of individual members at the NCI location. This genetic grouping was also present, but at reduced frequency, in the two closest locations, WCI (40%) and NEI (50%). In southern sample locations this genetic group was rare to absent, for example, SWI (3%) and ECI (0%) and SEI (0%), (Figure 5; Figure 10, Appendix 2). Initial analysis of fish sampled from WTS (western Tasman Sea) failed to be assigned to a single specific group using assignment probabilities greater than 80%. However, a more focussed comparison of the eastern Indian locations (ECI, SEI) and the WTS clearly revealed partitioning of swordfish into two geographically partitioned groups of swordfish between these Indian and Pacific Ocean sample regions (Figure 6). The separation of WTS from Indian Ocean, as well as the close relationship among SWI, SEI, and ECI sample locations was consistent across StockR model results at higher values of K and, in particular, at K=5 (Figure 8; Figure 10, Appendix 2). In addition, analysis of gene flow (F_{st}) indicated restricted gene flow between fish sampled from WTS (western Tasman Sea) and those from NCI (Sri Lanka) with NCI also the most consistently divergent from other Indian Ocean sampling locations (Figure 7). Interestingly, the southwest Indian and the southeast Indian were the least genetically differentiated samples based on F_{st} (Figure 7), while the SWI and WCI samples included a substantially different mix of genetic groups, indicating they were sampled from genetically different groups of fish. When modelled at K=2 the major difference between SWI and WCI appears to be a large proportion (~40%) of a genotype in the WCI sample that is present at only 5% in SWI. Phenotypic sex data was only available for samples from SWI and WCI to examine potential sex bias in our analysis. Further scrutiny of the WCI population using the available sex data indicated that males represented 56% (28/50) of sampled fish and demonstrated that a biased sex ratio could not explain the low proportion observed for this genetic grouping for this location. Interestingly, this same genotypic group is present in the NCI (Sri Lanka) sample at a nearly fixed frequency of 95% (30/32 fish).

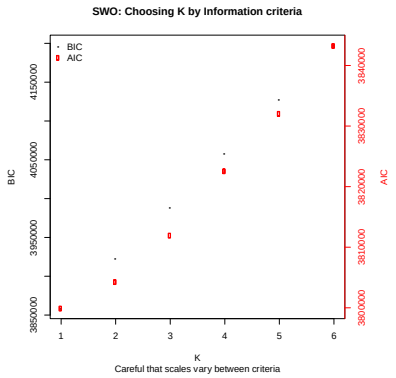


Figure 4. Plot of information criteria for swordfish from each tested model of from K=1 through K=6.

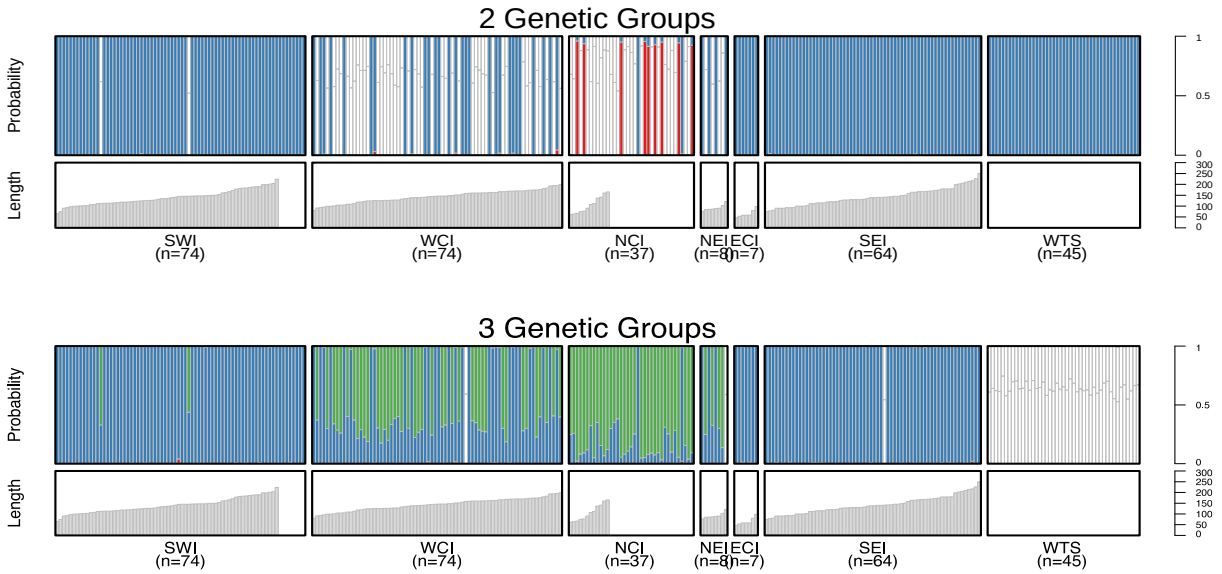


Figure 5. Individual bar plot probability of assignment to a particular K genetic group when modelled at K=2 and K=3 genetic groups. Bars lacking colour indicate uncertainty in assignment of individuals to a specific K group is below 90%. Fish lengths of individuals are plotted below each bar and sample size at location is in brackets.

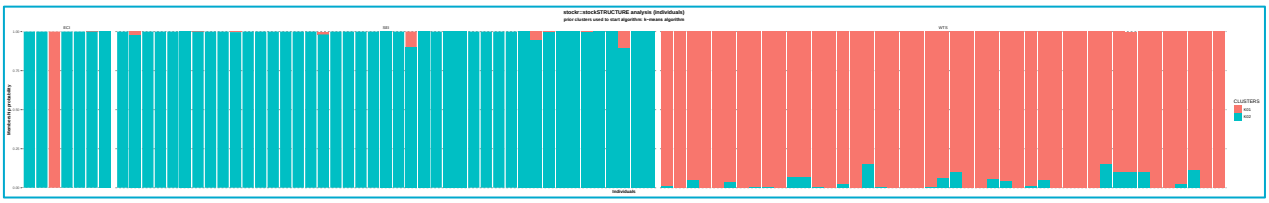


Figure 6. Results of focussed comparison of eastern Indian Ocean (ECI and SEI) and western Tasman Sea (Pacific Ocean) sampling locations (WTS) Individual bar plot probability of assignment to a particular K genetic group when modelled at K=2 (cyan and red). From left to right are ECI (east central Indian), SEI (south east Indian), WTS (west Tasman Sea).

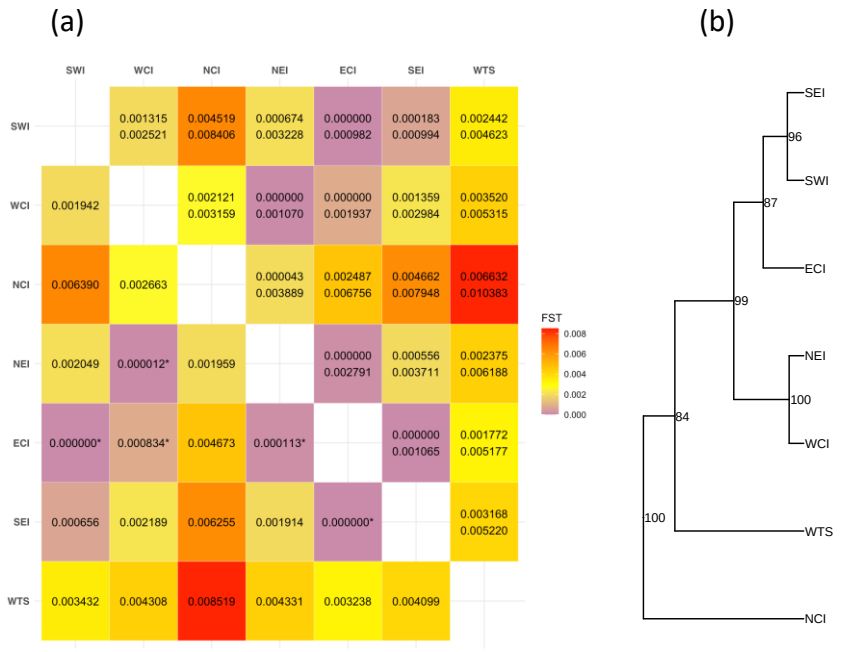
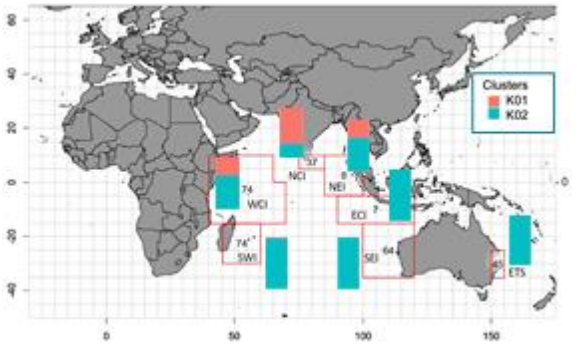
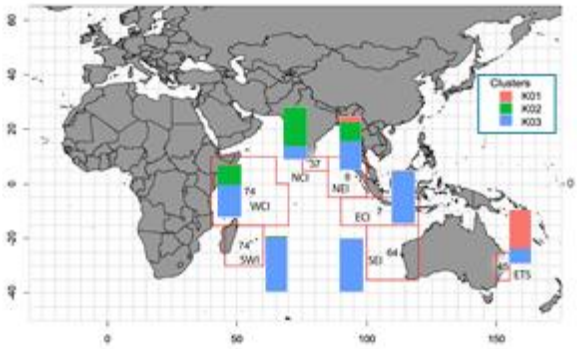


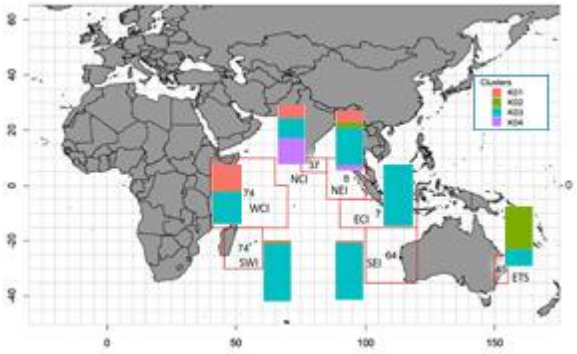
Figure 7 (a) Heat map of F_{st} values between pairs of sampling locations (below diagonal) and 95% confidence range (above diagonal). Colour levels as indicated by the scale bar to highlight relative F_{st} values calculated between pairs of sample locations with low (<0.003, purple to light purple), medium (0.003 to .005, yellow to mustard), and high d (>0.005, orange to red). (b) genetic distance phenogram to indicate relative relationships using F_{st} as a proxy for genetic distance. Numbers indicate bootstrap values calculated for each node.



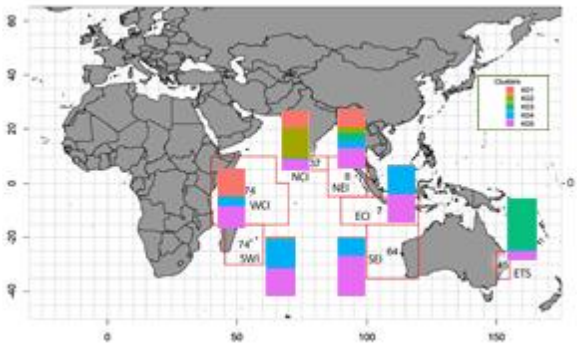
K=2



K = 3



K = 4



K = 5

Figure 8. Cumulative probabilities by location modelled at various levels of K genetic groups.

Summary

Previous studies and use of different genetic markers

Analysis of SNP genotypes has demonstrated presence of population heterogeneity of swordfish within the Indian Ocean. While population heterogeneity of Indian Ocean swordfish has previously been detected through examination of maternally transmitted mtDNA markers (e.g. Lu et al., 2006; Bradman et al., 2011), other studies examining nuclear DNA microsatellite markers have failed to demonstrate evidence of population structure in (Muths et al., 2009; Muths et al., 2013). The main defining difference among these earlier studies, which sampled at similar spatial and temporal scales to the current study, was examination of maternally versus nuclear transmitted genetic markers. This prompted the authors to suggest the discrepancy between results and subsequent failure to detect population structure was largely due to maternal philopatry and male mediated gene flow. An equally, and perhaps more parsimonious explanation, which can explain different observed levels of population differentiation revealed by SNP markers versus more rapidly evolving DNA markers, is that homoplasy resulting from elevated mutation rates in DNA microsatellite and mtDNA d-loop markers returns alleles to the population at a similar rate to that at which they are lost due to genetic drift subsequently blurring any signal of population structure. For example, a study of Indian Ocean swordfish identified genetic homoplasy through sequencing of two mitochondrial DNA fragments (ND2 gene versus d-loop section) and clearly demonstrated that erroneous conclusions could be drawn from studies relying on the rapidly evolving d-loop section of the mitochondrial genome which contains high levels of backward mutation of DNA markers (Bradman et al., 2011). Homoplasy within mtDNA d-loop and DNA microsatellites can lead to a mistaken case of “identity by state” rather than “identity by descent”, which has potentially led to false conclusions of panmixia in previous studies, and suggests cautious interpretation of studies where either marker type has failed to detect population structure. The advent of SNP marker technology has clearly overcome this disadvantage and provided a powerful platform to investigate fine scale resolution of population structure.

Population structure of Swordfish within the Indian Ocean

The analysis of SNP data suggests subtle population structure of swordfish within the Indian Ocean with at least two genetically differentiated groups present north and south of the equator. While swordfish are known to exhibit sex-specific north-south migrations, with females undertaking large north-south seasonal migrations, the sex ratios of the samples in this study indicate sexually dimorphic migration patterns do not explain the observed genetic pattern. This suggests a strong likelihood for the presence of separate northern and southern reproductively isolated stocks. Interestingly, there was no evidence of a longitudinal partitioning of swordfish between the SWI and SEI sample locations. These southern locations at the eastern and western extremes of the Indian Ocean were the least differentiated based on F_{st} (**Error! Reference source not found.**). Furthermore, the results from both the StockR and F_{st} analyses, the SWI and SEI sampled fish are also most genetically similar to those from the ECI location (**Error! Reference source not found.** and **Error! Reference source not found.**). While neither maturation state nor sex of the fish was recorded, the samples from this location were taken near the well-known spawning grounds in the Eastern Indian Ocean between late October through late December close to the time of peak

spawning activity in this area (Nishikawa et al., 1985). In contrast, SWI (Reunion Island), which is ~3000 nautical miles from SEI, showed clear genetic differentiation from the geographically close (~800 nm) WCI location (2% versus 40 % northern genotypic groups respectively at K=2 and K=3). These WCI samples were taken close to a putative spawning grounds near the Mozambique Channel and Seychelles, indicating the two locations may not be genetically related (Nishikawa et al., 1985; García-Cortés and Mejuto, 2003). This may indicate the WCI sample, which has only 40% of the genotypic group found at a nearly fixed frequency (95%) in the NCI (Sri Lanka) location, potentially represents either a mixed feeding aggregation or alternatively a third reproductively isolated population , (Figure 5; Figure 10 Appendix 2). Analysis of SNP data has demonstrated the potential for resolving stock structure and delivered evidence for both northern and southern stocks of swordfish in the Indian Ocean. Extension of this approach to include additional targeted sampling at specific locations and, ideally spawning adults, should distinguish whether fish from SWI caught near Reunion Island potentially originate from WCI and/or ECI spawning aggregations and further refine the population structure of swordfish within the Indian Ocean.

Global Connectivity between the Atlantic, Indian Ocean, and Pacific Oceans.

Gene flow from Atlantic Ocean and Pacific Ocean is an important consideration for sustainable management of Indian Ocean swordfish. Chow and Takeyama (2000) examined both mtDNA and nuclear markers and resolved four reproductively isolated swordfish breeding units located in north western Atlantic, Mediterranean Sea, tropical to South Atlantic, and Indo-Pacific. Phylogenetically, divergence of the Mediterranean origin swordfish has since prompted a suggestion they may even represent a cryptic species with no gene flow between either of the two north-south equatorially partitioned Atlantic Ocean stocks (Alvarado-Bremer et al., 1999; Smith et al., 2015). Lack of detectable heterogeneity among six Atlantic sample locations broadly covering the southern hemisphere has indicated very limited genetic divergence in this region consistent with the presence of a single panmictic south Atlantic breeding population (Smith et al., 2015). Furthermore, differentiation observed between samples from Namibia (south eastern Atlantic Ocean) compared to those from south western Indian Ocean are consistent with both tagging studies and previous genetic results confirming Atlantic and Indian Ocean swordfish represent two reproductively isolated populations (Chow and Takayama, 2000; Alvarado-Bremer et al., 2005; Kadagi et al., 2011; Muths et al., 2013). Clear genetic differentiation between eastern Indian (ECI and SEI) and south western Pacific (WTS) is also consistent with previous results from tagging and genetic studies (Stanley, 2006; Muths et al., 2013; Lu et al., 2016). Lu et al. (2016) further described intra-oceanic differentiation of populations in the Pacific Ocean and also demonstrated a lack of potential geneflow into the Indian Ocean that could have arisen through larval transport via the Indonesian throughflow described by Meyers et al., (1995). In summary, evidence from both the current and previous studies indicates lack of significant gene flow among the Atlantic, Indian and Pacific oceans that has promoted detectable inter-oceanic scale genetic differentiation and indicates the Indian Ocean should be considered a closed and reproductively isolated system.

Future Directions and Management implications for billfish within the Indian Ocean

Analyses of data from both the current and previous studies have indicated a significant lack of gene flow into the Indian Ocean from the Atlantic and Pacific populations of swordfish. While the Indian Ocean should be considered a closed genetic system with a minimum of two reproductively isolated populations further targeted spatial and temporal sampling is required to refine the

population structure of swordfish and identify the relationship between major spawning grounds, feeding areas and fishing grounds. In particular, identifying genetic markers of major spawning aggregations would be valuable for developing provenance markers to better inform sustainable swordfish management both within the Indian Ocean and globally.

Our study has advanced the understanding of population structure of billfish species within the Indian Ocean. From a stock assessment and management perspective, these results indicate that monitoring, population biology and management programs should be designed to support management of multiple stocks of swordfish in the Indian Ocean. While, we recognize this presents considerable challenges, given the data poor nature of many of the fisheries in the region, it is evident from the results of this basin-scale collaborative study, among others, that the necessary advances can be achieved.

Acknowledgements

We wish to thank all those who assisting with the sample collection and analysis, in particular,

- Hughes Evano, Jérôme Huet and Sylvain Bonhommeau, Ifremer, Estelle Crochelet (IRD), Loïc De Foulgoc, CapRun, for sampling on albacore and swordfish caught by the Reunion local longliners;
- Jenny and Ray Davies, Ocean Wild Tuna, for allowing us to sample from their catches; skippers Ray Davies and Michael West, and their crews, for collecting samples while at sea;
- The many vessel owners, skippers, observers, crew and processors who provided fish for sampling and access to facilities;
- The team at Diversity Array Technologies Pty Ltd for the high quality of the sequencing and genotyping services and the willingness to fast-track processing and data delivery;
- Partner institutions and IOCT Secretariat for the project support, cooperation and patience with complex contracting arrangements.

The project is supported by funding from CSIRO Oceans and Atmosphere, AZTI Tecnalia, Institut de Recherche pour le Développement (IRD), and Indonesia's Center for Fisheries Research (CFR) and financial assistance of the European Union (GCP/INT/233/EC –Population structure of IOTC species in the Indian Ocean). The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

Bibliography

Anon. (2018). Population Structure of IOTC species and sharks of interest in the Indian Ocean: Estimation with Next Generation Sequencing Technologies and Otolith Micro-chemistry Sampling Protocols. Report to the Indian Ocean Tuna Commission, May 2017 46p.

Foster, SD, P Feutry, PM Grewe, O Berry, FKC Hui, CR Davies. (2018). Reliably discriminating stock structure with genetic markers: Mixture models with robust and fast computation. Molecular ecology resources 18:1310-1325.

Fu, Dan (IOTC Secretariat) (2017). An age-, sex- and spatially-structured stock assessment of the Indian Ocean Swordfish fishery 1950-2015, using stock synthesis IOTC–2017–WPB15–20_Rev1.

Ganga, U., Pillai, N. G. K., & Elayathu, M. N. K. (2008). Billfish fishery along the Indian coast with special reference to the Indo-Pacific sailfish *Istiophorus platypterus* (Shaw and Nodder 1792). Journal of the Marine Biological Association of India 50(2), 166-171.

García-Cortés, B., Mejuto, J., (2003). Sex ratio patterns and gonadal indices of the swordfish (*Xiphias gladius*) caught by the Spanish surface longline fleet in the Indian Ocean. In: IOTC Proceedings no. 6 (2003), pp. 287–299.

Givens, G.., Hoeting, J. 2013. Computational Statistics Wiley.

Gosselin, T., Lamothe, M., Devloo-Delva, F., and Grewe, P. (2019). RADseq Data Exploration, Manipulation and Visualization using R. DOI:10.5281/zenodo.2595083.

Gosselin T, Lamothe M, Devloo-Delva F, Grewe P. (2020). radiator: RADseq Data Exploration, Manipulation and Visualization using R. https://thierrygosselin.github.io/radiator/. doi : 10.5281/zenodo.3687060

Grewe P, Grueger C, Aquadro CF, Berminghmam E. (1993). Mitochondrial DNA Variation among Lake Trout (*Salvelinus namaycush*) Strains Stocked into Lake Ontario. Can. J. Fish. Aquat. Sci. 50, 2397–2403.

Hoolihan, J.P., Premanandh, J., D'Aloia-Palmieri, M. et al. (2004). Intraspecific phylogeographic isolation of Arabian Gulf sailfish *Istiophorus platypterus* inferred from mitochondrial DNA. Marine Biology 145, 465–475.

Hui, F. K.C., Warton, D.I., Foster, S.D. 2015. Order selection in finite mixture models: complete or observed likelihood information criteria? Biometrika, 102: 724-730

Humphreys, R., Campana, S. & DeMartini, E. (2005). Otolith elemental fingerprints of juvenile Pacific swordfish *Xiphias gladius*. Journal of Fish Biology 66, 1660–1670.

Humphreys, R. L. & DeMartini, E. E. (2008). Application of Otolith-Based Methods to Distinguish Nursery Areas of Juvenile Swordfish. Final Report to the Hawaii Fisheries Disaster Relief Program Joint Institute for Marine and Atmospheric Research, 8p.

Jombart T., Devillard, S., Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC genetics 11:94.

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, et al. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. Methods Mol Bio 888:67–88.

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? Molecular Ecology, 19(23), 5320-5320. doi:DOI 10.1111/j.1365-294X.2010.04878.x

Mahé, K., H. Evano, T. Mille, D Muths & J Bourjea. (2016). Otolith shape as a valuable tool to evaluate the stock structure of swordfish *Xiphias gladius* in the Indian Ocean. African Journal of Marine Science 38(4): 457–464.

Meyers G., Bailey, R., and Worby, A. (1995). Geostrophic transport of Indonesian throughflow. Deep-Sea Research. 42: 1163-1174.

Poisson, F. and and C. Fauvel (2009). Reproductive dynamics of swordfish (*Xiphias gladius*) in the southwestern Indian Ocean (Reunion Island). Part 1: oocyte development, sexual maturity and spawning. *Aquatic Living Resources* 22(1), 45–58.

Poisson, F and and C. Fauvel. (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.

Smith BL, Lu C-P, García-Cortés B, Viñas J, Yeh S-Y, Alvarado Bremer JR (2015) Multilocus Bayesian Estimates of Intra-Oceanic Genetic Differentiation, Connectivity, and Admixture in Atlantic Swordfish (Xiphias gladius L.). PLoS ONE 10(6): e0127979. doi:10.1371/journal.pone.0127979

Speare, P. (1995). Parasites as biological tags for sailfish *Istiophorus platypterus* from east coast Australian waters. Marine ecology progress series. Oldendorf, 118(1), 43-50.

Stanley C (2006). Determining the nature and extent of swordfish movement and migration in the eastern and western AFZ through an industry-based tagging program. CSIRO. 24 p.

Varghese, Sijo P., Kandachamy Vijayakumaran, A. Anrose, Vaibhav D. Mhatre. (2013). Biological Aspects of Swordfish, *Xiphias gladius* Linnaeus, 1758, Caught During Tuna Longline Survey in the Indian Seas. Turkish Journal of Fisheries and Aquatic Sciences 13: 529-540.

Wang, Sheng-Ping, Chi-Hong Lin and Wei-Chuang Chiang. (2010). Age and growth analysis of swordfish (*Xiphias gladius*) in the Indian Ocean based on the specimens collected by Taiwanese observer program. IOTC-2010-WPB-08.

Appendix 1 Sailfish extended analysis.

Radiator filtering steps for Indo-Pacific Sailfish *Istiophorus platypterus*, including threshold values and the number of individuals, locus and markers at the start of each step.

Filters	VALUES	Individuals / Locus / Markers
Filter DArT reproducibility	0.95	99 / 40445 / 54827
Filter monomorphic markers		99 / 35396 / 47182
Filter markers in common		99 / 35396 / 47182
Filter individuals based on missingness	0.3	99 / 31313 / 42046
Filter monomorphic markers		96 / 31313 / 42046
Filter MAC	10	96 / 25017 / 31576
Filter coverage min / max	10 / 100	96 / 9512 / 10318
Filter genotyping	0.3	96 / 6020 / 6582
Filter SNPs position on the read	all	96 / 5797 / 6299
Filter markers snp number	4	96 / 5797 / 6299
Filter short ld	mac	96 / 5797 / 6299
detect mixed genomes	0 / 0.267	96 / 5797 / 5797
Filter monomorphic markers		80 / 5797 / 5797
detect duplicate genomes	0.75	80 / 5797 / 5797
Filter monomorphic markers		65 / 5797 / 5797

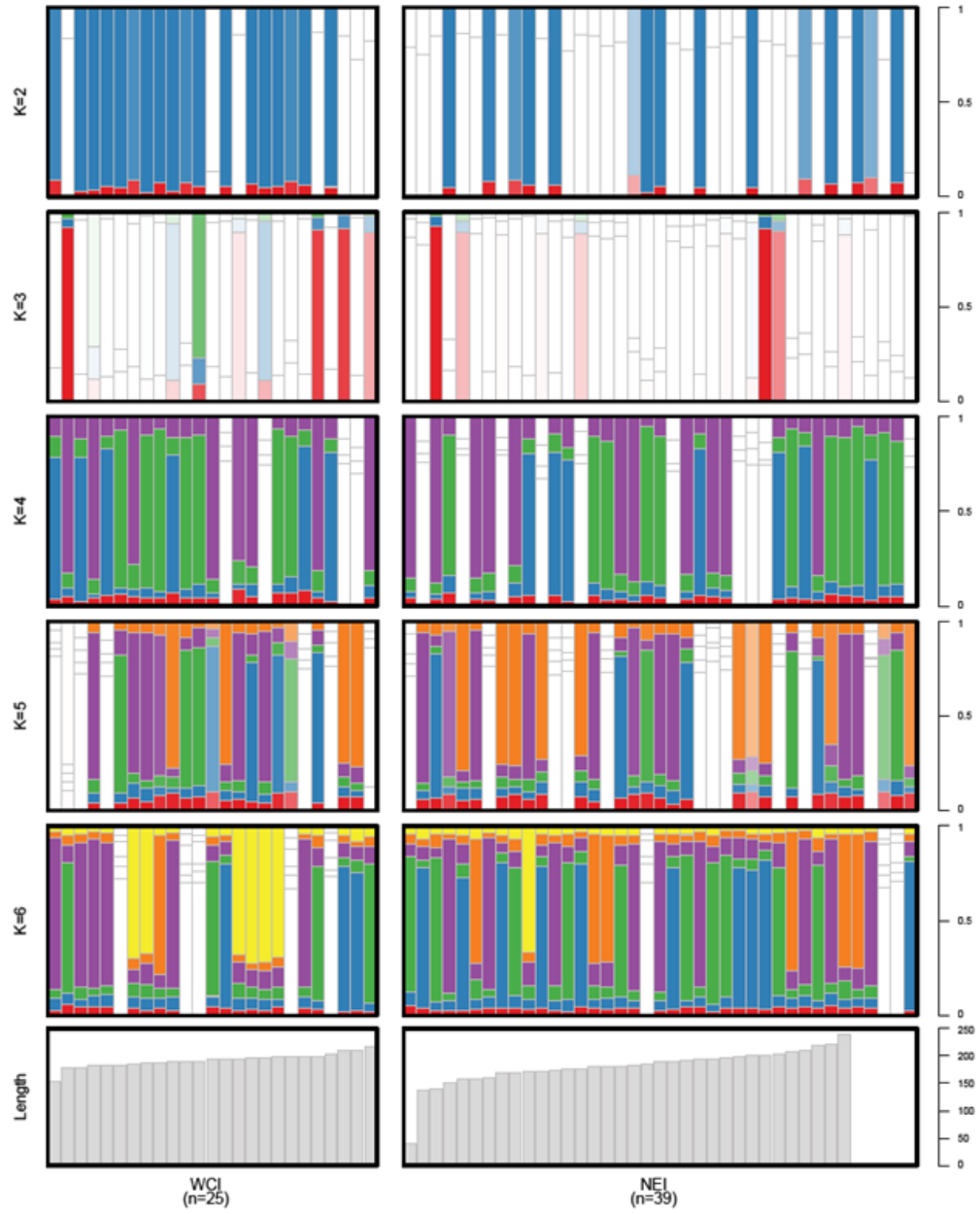


Figure 9. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for all Indo-Pacific sailfish tuna assuming 2-6 genetic groups.

Appendix 2 – Swordfish extended analysis.

Parameters used in the radiator QC filtering swordfish from 11 sampling sites (strata) using starting values for radiator of 802 individuals, 39,663 locus fragments, and 54,733 SNP markers.

Applied FILTER step	VALUES	Individuals / Locus / Markers
Filter DArT reproducibility	0.93	493 / 71021 / 126843
Filter monomorphic markers		493 / 71021 / 121665
Filter markers in common		493 / 68712 / 121665
Filter individuals based on missingness (with outlier stats or values)	0.25	492 / 68712 / 118575
Filter monomorphic markers		492 / 68710 / 118575
Filter MAC	8	492 / 49988 / 118557
Filter coverage min / max	12 / 145	492 / 25533 / 75959
Filter genotyping	0.8	492 / 25532 / 41831
Filter SNPs position on the read	all	492 / 25532 / 41828
Filter markers snp number	1	492 / 15084 / 41828
Filter short ld	not filtering	492 / 15084 / 15084
detect mixed genomes	0.070 to 0.110	358 / 15084 / 15084
Filter monomorphic markers		358 / 15073 / 15084
detect duplicate genomes	0.834	284 / 15073 / 15073
Filter monomorphic markers		284 / 15070 / 15073
Filter HWE	0.93	284 / 15070 / 15070

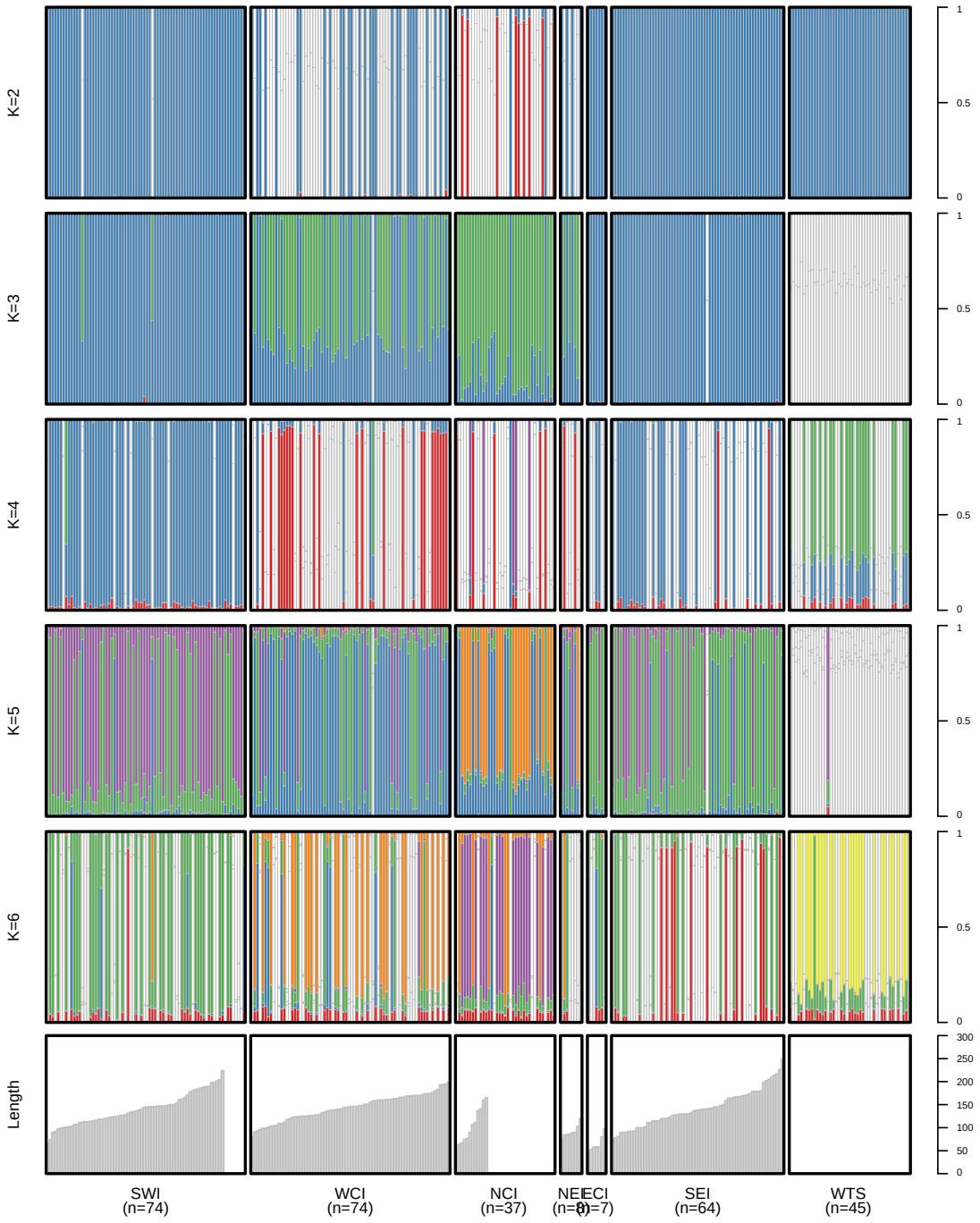


Figure 10. Membership probability barplots for each individual swordfish belonging to one of K genetic groupings. For each panel below there are K colours representing K genetic groupings. The probability of each individual is then plotted *proportionally* relative to assignment probability to each of the K groups with white profiles representing individuals where there is less than 95% certainty of belonging to a K genetic group.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiroenquiries@csiro.au
www.csiro.au

For further information

CSIRO Ocean and Atmosphere
Campbell Davies, Leader – International Fisheries
+61 3 **62325044**
campbell.davies@csiro.au
www.csiro.au