Australia's National
Science Agency

Population Structure of IOTC species and sharks of interest in the Indian Ocean:

Estimation with Next Generation Sequencing Technologies and Otolith Micro-chemistry

Study of population structure of IOTC species and sharks of interest in the Indian Ocean using genetics and microchemistry: an update on progress and preliminary results.

Campbell Davies, Francis Marsac, Hilario Murua, Igaratza Fraile, Zulkarnaen Fahmi, Jessica Farley, Peter Grewe, Craig Proctor, Naomi Clear, Matt Lansdell, Jorden Aulich, Pierre Feutry, Scott Cooper, Scott Foster, Naiara Rodríguez-Ezpeleta, Iraide Artetxe-Arrate, Natacha Nikolic, Iñigo Krug, Iñaki Mendibil and Agostino Leone, Maylis Labonne, Audrey Darnaude, Sophie Arnaud-Haond, Wudianto, Toni Ruchimat, Fayakun Satria, Pratiwi Lestari, Muhammad Taufik, Asep Priatna, Achmad Zamroni.

30 November 2019

Working paper submitted to the IOTC 22nd Scientific Committee Meeting, Karachi, Pakistan.

Citation

Davies, C.R., Marsac, F, Murua, H., Fraile, I., Fahmi, Z., Farley, J., Grewe, P., Proctor, C., Clear, N., Lansdell, M., Aulich, J., Feutry, P., Cooper, S., Foster, S., Rodríguez-Ezpeleta, N., Artetxe-Arrate, I., Krug, Mendibil, I., Agostino, L., Labonne, M., Nikolic, N., Darnaude, A., Arnaud-Haond, S., Wudianto, Ruchimat, T., Satria, F., Lestari, P., Taufik, M., Priatna, A., and Zamroni, A.

Copyright

Important disclaimer

Contents

Figures

Tables

Acknowledgments

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.

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

- Umair Shahid, Syed Meesum Reza Kazmi, Saeed ul Islam, Mariam Tariq, Sidra Zafar, Jaweria Zaidi, WWF Pakistan;
- Hamid Badar Usmani, Kashifa Zehra and Mohammad Wasim Khan, Marine Fisheries Department, Government of Pakistan;
- Daniel Fernando, Blue Resource Trust, Sri Lanka;
- Denham Parker, Malcolm Francis, Charlene de Silva, Department of Agriculture, Forestry and Fisheries, South Africa, and Stewart Norman, CAPFISH, South Africa;
- Anais Medieu, Marianne Pernak, Pascal Bach, IRD for tissue sampling and otoliths extraction on fish caught by the Seychelles local longliners;
- Hughes Evano, Jérôme Huet and Sylvain Bonhommeau, Ifremer for otoliths extraction on swordfish caught by the Reunion local longliners;
- Cecile Petit for microchemistry analyses at University of Montpellier, and Cathy Liautard-Haag for genetic analyses at Ifremer;
- Mohamed Ahusan, Ahmed Haikal and Shafiya Naeem, Marine Research Centre, Ministry of Fisheries, Marine Resources and Agriculture, Maldives;
- Shiham Adam, the Ministry of Fisheries, Marine Resources and Agriculture, Maldives;
- Seychelles Fishing Authority
- Grant Johnson, Northern Territory Department of Primary Industry and Fisheries, Australia and Rik Buckworth, Seasense consulting, Darwin, Northern Territory;
- Joanne Langstreth, John Cavallaro, Queensland Department of Primary Industry and Fisheries, Australia;
- Ruth Sharples for her contributions to compiling the SOP and preparation and coordination of sampling kits;
- The many vessel owners, skippers, observers, crew and processors who provided fish for sampling and access to facilities;
- Those who collected and curated samples from previous projects that have been made available to this project (e.g. Blueshark);
- 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.

Executive summary

In 2017, CSIRO (Australia) in collaboration with AZTI Tecnalia (Spain), IRD (France) and RITF (Indonesia) commenced a 3-year collaborative project on tuna, billfish and sharks of the Indian Ocean. The aim of the project was to describe the population structure and connectivity of a range of tuna and tuna-like species within the Indian Ocean (and adjacent Pacific and Atlantic waters as appropriate), as well as some of the key shark species that interact with Indian Ocean Tuna Commission (IOTC) fisheries. Genetics and microchemical analysis of hard parts (otoliths, shark vertebrae) were used as complimentary techniques in investigating population structure. The project also aimed to develop and extend research networks among partners and contribute to technical capacity building in participating coastal states, where possible.

The project is now nearing completion with 2 years of sampling completed, almost 6,000 samples collected of which 3,635 have been genotyped and 689 processed for otolith microchemistry. Initial data analyses have been completed for the majority of the species for genetics. A summary of the preliminary results from these analyses are presented in this working paper. Micro-chemistry analyses are complete for swordfish, albacore, skipjack tuna and yellowfin tuna and the preliminary results from these are also included here. The microchemistry analyses for the neritic species and bigeye tuna are in progress and will be reported on in the near future.

The sample collection and processing phases of the project have provided a sound foundation for many of the species, but improvements/additions are needed for others before substantive interpretations and conclusions about population structure will be able to be made. Very preliminary results suggest structure for some species while not providing evidence of structure for others. Further detailed analyses are required before any substantive conclusions can be drawn. We invite initial comment and input from the Scientific Committee on the work completed to date and the very preliminary results presented here. There will be further opportunity to review the more substantive results and initial interpretations in the draft final report (due 20 Dec 2019) which will be followed by the final report (due 31 March 2020) and subsequent papers to the relevant working groups in 2020.

1 Introduction

1.1 Background

There are at least 10 tuna and tuna-like species, 5 billfish species, and 7 shark species of substantial commercial and food security value in the Indian Ocean (IO). All of these species are assumed to be highly migratory, and straddle multiple coastal EEZs and international waters, necessitating a multi-national effort for effective fisheries management. The Indian Ocean Tuna Commission (IOTC) is responsible for the management of these species (with the exception of southern bluefin tuna). Some of these species have been assessed with modern, data-intensive, integrated population modelling techniques in recent years (yellowfin, skipjack, bigeye and albacore tunas), while many of the neritic tuna species have never been formally assessed. Attempts have been made to quantify movement within the IO for yellowfin, skipjack and bigeye tunas primarily on the basis of tag displacements observed in the Regional Tuna Tagging Programme (RTTP-IO). Unfortunately, constraints to the RTTP-IO release design and low tag reporting rates for longline and artisanal fleets has meant that movements to/from areas outside of the Western equatorial region are difficult to quantify, even for these tagged species. All assessments to date have assumed a single panmictic spawning population within the Indian Ocean. However, there have been studies suggesting that there may be distinct population structure at a much smaller scale than the IO basin (e.g. for yellowfin: Dammannagoda et al. 2008, Swaraj et al. 2013; skipjack: Dammannagoda et al. 2011, Menezes et al. 2012; and bigeye: Nugraha et al. 2011). Similarly, analyses of tagging data in the Indian Ocean and elsewhere (e.g. western Pacific) have suggested that movement/mixing rates may not be consistent with the large spatial regions that are typically assumed in tuna assessments. If the scientific stock assessment advice is based on invalid assumptions, management may not achieve stated objectives related to conservation and optimal economic use of the resource. Specifically, if populations are distinct (or mixing rates are very low within a panmictic population), some populations (or sub-regions) could be locally over-exploited and management measures might be directed toward the wrong populations.

1.2 Need

There was a clear need to underpin assessment and management advice with a basic understanding of population structure and connectivity among populations within the Indian Ocean (and potentially with adjacent populations in the Atlantic and Pacific Oceans). This is of particular importance for small island and developing states with short range fishing fleets. Responsible management is equally in the interest of the distant water fishing nations, and addressing these fundamental concerns will assist with the attainment of sustainable product endorsements for all fisheries, regardless of whether the populations are revealed to be well-mixed or fragmented.

1.3 Approach

1.3.1 Next Generation Sequencing and Single Nucleotide Polymorphisms (SNPs)

High throughput, next generation sequencing (NGS) technologies represent a cost effective option for revealing population structure through examination of Single Nucleotide Polymorphisms (SNPs). SNP markers also lend themselves easily to routine and inexpensive (AU$10-$20 per fish) screening methodologies. Recently developed NGS approaches (e.g. Restriction site Associated DNA markers or RADtags) represent a major advancement over classical techniques (i.e. based on allozymes, DNA microsatellites, and mitochondrial DNA). These latter approaches are more labour intensive in terms of ability to screen the quantity of suitable markers necessary to discriminate structure present in large marine fish populations. Furthermore, limited screening of classical marker loci is less sensitive and results can potentially be misleading when no stock structure is revealed due to limited genetic resolution.

1.3.2 Otolith microchemistry

Genetic markers have been widely used with success in identifying population structure of marine fish. However, genetic methods can struggle to resolve regional demographic and life history patterns over time scales relevant to population dynamics and operational fisheries management. Complimentary methods, such as otolith microchemistry, can provide insights into population structuring over time scales relevant for individual movements (Proctor 2019). Chemical markers in otoliths have significant potential for determining natal origin and population connectivity of tuna species (Rooker et al. 2007; Proctor et al. 2019). A combination of trace-element chemistry and carbon and oxygen stable isotopes (δ13C and δ18O) were used to investigate population structure of the non-shark study species.

Table 1. Study species and summary of responsibility for species and analyses method (genetics, otolith micro-chemistry among partners. * Note that the original project proposal included exploration of the utility of microchemistry on shark vertebrae for population structure analysis, this was not pursued due to the logistic difficulties of obtaining shark vertebrae across the study area.

Species	Genetic		Otolith chemistry	
	Lead partner	Method	Lead partner	Method
Longtail tuna	CSIRO	DArTSeq	CSIRO	Trace element
Kawakawa	CSIRO-CFR	DArTSeq	CSIRO-CFR	Trace element
Narrow-barred Spanish mackerel	CSIRO	DArTSeq	CSIRO	Trace element
Skipjack tuna	AZTI	RAD-Seq	AZTI	Trace element & stable isotope
Albacore	IRD	DArTSeq	IRD	Trace element
Yellowfin tuna	CSIRO	DArTSeq	AZTI	Trace element & stable isotope
Bigeye tuna	AZTI	DArTSeq	CSIRO	Trace element
Striped marlin	CSIRO	DArTSeq	-	-
Indo-Pacific sailfish	CSIRO	DArTSeq	-	-
Swordfish	CSIRO	DArTSeq	IRD	Trace element
Blue shark	IRD	DArTSeq	-	*
Scalloped hammerhead	CSIRO	DArTSeq	-	*

2 Sampling design and Biological Sample Collection

2.1 Sampling design

The project included the 3 species of neritic tunas, 3 tropical tuna, 1 temperate tuna, 3 billfish and 2 shark species listed in Table 1. Sampling locations, or regions, for the first round of sampling aimed to include the approximate extremes (four locations) of the known range of each species in the Indian Ocean and from one, or more, 'outlier' locations in the Pacific and/or Atlantic Oceans for species that appear to form a continuum across oceans (e.g., albacore, swordfish and blue shark). The intent of this staged design was to obtain sufficient samples of each species, from each location, to enable statistically robust analyses of the level of variation in genetic markers and otolith/vertebrae chemistry within and between sample locations and provide a basis for refining the sampling design with "intermediate" locations for the second round of sampling; in the case there was evidence of population structure from the analysis of samples from round 1. In practice, it proved too challenging to coordinate and synchronise the logistics of sampling, distribution of samples and analyses over such a large geographic range (see results). The sampling regions for all species are shown in Figure 1. The subset of locations for each species and method (genetics/microchemistry are provided in the summary results sections (sections 3 and 4).

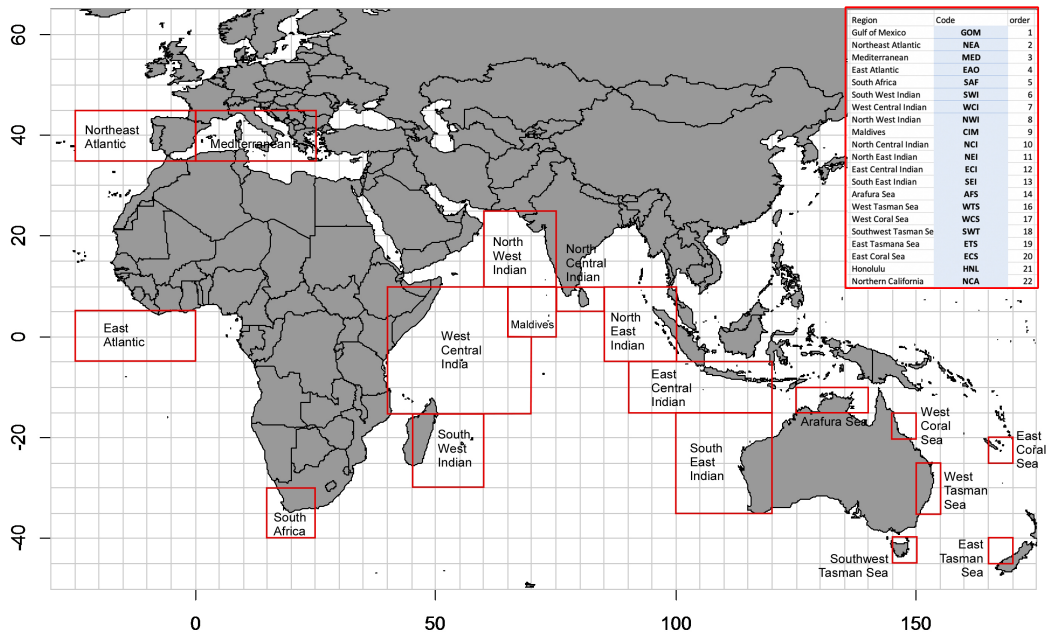

Figure 1. Distribution of sampling locations for both rounds of sampling across Indian Ocean and outlier locations in the Pacific and Atlantic Oceans. Note: this includes locations for active sampling as well as locations for samples provided from earlier studies (see text for details).

The sampling strategy for each species was based on the collective knowledge, of the project team and collaborators, of species distribution, fisheries characteristics and landing locations, likely environmental barriers to connectivity, sampling objectives and interested parties who may be able to source and collect samples. Ideally, samples were to be collected from spawning adults on the spawning grounds, as opposed to sampling fish transiting through or foraging in the area and breeding elsewhere. However, on review of the available information and the expert knowledge of partners and collaborators on distribution, timing and location of spawning and the logistics of access to spawning fish via fisheries, it was concluded it was not logistically feasible to only focus sampling on spawning adults for all species. Hence, the final design focussed on spawning adults and/or young-of-the-year (YOY). The rationale for the latter is that YOY are less likely to have moved far from their natal spawning grounds.

The size ranges of fish for each species targeted are provided in Table 2. Proposed maximum number of tissue samples (for genetics), otoliths and vertebrae to be collected in 2017 and 2018 from throughout the Indian Ocean. The minimum number of samples for each species to be collected in each area was 50, with a maximum number of 100 to allow for sub-sampling and prioritising of different size fish, quality of samples, or regions for final genetic or microchemistry analyses (Table 3). In addition, the protocol stipulated that no more than 10 samples from a species should be taken from one 'batch', so as to increase the representativeness of final sample, and minimise the possibility that a large proportion of the sample from any one location could have come from the same school.

Table 2. Proposed maximum number of tissue samples (for genetics), otoliths and vertebrae to be collected in 2017 and 2018 from throughout the Indian Ocean

Species	Young of the year (YOY)	Adults	Muscle tissue	Otolith	Vertebrae
Longtail tuna	<30 cm?	>50 cm Mar-May & Jul-Dec	100	100	
Kawakawa	<25 cm	East >40 cm	100	100	
Narrow-barred Spanish mackerel	<40 cm	>100 cm	100	100	
Skipjack tuna	<35 cm	>50 cm	100	100	
Albacore	<40 cm	>95 cm	100	100	
Yellowfin tuna	<45 cm	>120 cm	100	100	
Bigeye tuna	<45 cm	>120 cm	100	100	
Striped marlin	<80 cm?	>195-200 cm	100	100	
Indo-Pacific sailfish	<80 cm		100	100	
Swordfish	<80 cm?	Females 190 cm OFL, Males: 110 cm OFL	100	100	
Blue shark	<60 cm?	>280 cm	100		100
Scalloped hammerhead	<60 cm	>210 cm	100		100

Table 3. Summary of sampling design and number of fish by ocean and species in round 1 (top table) and round 2 (bottom table) of sampling, included proposed samples sizes for genetic and microchemistry analyses. Note that the range of some species does not extend into the Atlantic Ocean.

Species	IO priority sites	Intermediate sites	Pacific	Atlantic	Total	Min (50) samples	Max (100) samples	No. analysed genetics (50/site)	No. analysed otolith chem (20 x 4 sites)**	No. analysed vertebrae chem (20 x 4 site)
Longtail tuna	4		1		5	250	500	250	80	
Kawakawa	4		1		5	250	500	250	80	
Narrow-barred Spanish mackerel	4		1	?	5	250	500	250	80	
Skipjack tuna	4		1	1	6	300	600	300	80	
Albacore	4		1	1	6	300	600	300	80	
Swordfish	4		1	1	6	300	600	300	80	
Blue shark	4		1	1	6	300	600	300		80
Scalloped hammerhead	4		1		5	250	500	250		80
Yellowfin tuna	4		1	1	6	300	600	300	80	
Bigeye tuna	4		1	1	6	300	600	300	80	
Striped marlin	4		1	?	5	250	500	250		
Indo-Pacific sailfish	4		1		5	250	500	250		
Total					66	3300	6600	3300	640	160

Species	IO priority sites	Intermediate sites	Pacific	Atlantic	Total	Min (50) samples	Max (100) samples	No. analysed genetics (50/site)	No. analysed oto chem (20 x 6 site)**	No. analysed vertebrae chem (20 x 4 site)
Longtail tuna	4	2	1		7	350	700	350	120	
Kawakawa	4	2	1		7	350	700	350	120	
Narrow-barred Spanish mackerel	4	2	1	?	7	350	700	350	120	
Skipjack tuna	4	2	1	1	8	400	800	400	120	
Albacore	4	2	1	1	8	400	800	400	120	
Swordfish	4	2	1	1	8	400	800	400	120	
Blue shark	4	2	1	1	8	400	800	400		180
Scalloped hammerhead	4	2	1		7	350	700	350		180
Yellowfin tuna	4	2	1	1	8	400	800	400	120	
Bigeye tuna	4	2	1	1	8	400	800	400	120	
Striped marlin	4	2	1	?	7	350	700	350		
Indo-Pacific sailfish	4	2	1		7	350	700	350		
Total					90	4500	9000	4500	960	360

2.2 Sampling protocol

A Standard Operating Procedure (SOP) (Anon 2018) was developed to provide a standardized set of sample collection techniques for use by trained observers at sea and for port sampling by project staff and participating coastal states. This included standardised collection kits (sampling equipment, vials, data sheets, data entry templates etc) that were assembled by CSIRO and provided to sampling teams (see SOP for details).

2.3 Sample collection

2.3.1 Port Sampling

Detailed sampling plans for port sampling, including direct involvement of a number of coastal states, were developed by the project partners and in-country collaborators (see Table 1 and Table 4). The final plans were reviewed by the Project Leadership Team and managed by senior member from each partner:

- IRD: South-Western coastal states
- AZTI: Central-West and NW purse-seine
- CSIRO: Central, North-West coastal states, North-east coastal states and South-east

Dedicated port sampling exercises were conducted in:

- Australia – CSIRO: Mooloolaba, Perth, Great Australian Bight, Tasmania.
- Indonesia – CFR and CSIRO: Lampulo (Aceh), Palabuhanratu (West Java)
- Maldives –MRC of the Maldives and CSIRO
- Reunion - IRD
- Seychelles – IRD, SFA and AZTI

2.3.2 Sample collection and distribution

The tissues and otolith samples collected in the field as part of the project were distributed to the lead partner for the particular species, with the exception of tissue samples for the albacore, blue shark and bigeye tuna. In this case, as the genetic sequencing for these species was being done using the DArT facility in Canberra, the tissues were sent to CSIRO and the extraction, sequencing and data management handled by CSIRO. The final sequencing data from DArT was then provided to IRD (albacore, blue shark) and AZTI (bigeye tuna) for analysis, following some initial checking and QC at CSIRO.

2.3.3 Samples obtained from other sources

In addition to the new samples collected directly as part of project, project partners made available samples from existing holdings. These included:

- Blue shark samples from , France and Department of Agriculture, Forestry and Fisheries, South Africa and a previous project (SELPAL).
- Samples from a range of tuna and billfish species and blue shark from Research Institute for Tuna Fisheries (RITF) in Bali, Indonesia.

2.4 Summary of samples available for analysis

A summary of the number of tissue and otolith samples collected, or made available, for the project and the equivalent for the samples processed for genetics and microchemistry are provided in Table 4). The spatial distribution of samples collected and processed for genetics is shown for each species and sampling round in Appendix 1.

As noted above, the intent had been to complete the sampling in two "rounds". The first to cover the extent of the range and the second to provide temporal replication of the first round and to also sample at some additional locations "intermediate" to those sampled in round 1 in the cases where there was evidence of population structure from the analyses of round 1 samples. In practice, this was not possible and the schedule and logistics of the project required that some locations were still being sampled to complete "round 1" at the same time that the second round of sampling need to commence in other locations. Hence, the extent to which the final data set will allow for temporal variation to be examined explicitly will vary among species and locations.

Table 4. Summary of total number of tissue and otolith samples collected, or made available, and selected for genetic and microchemistry analyses across both rounds of sampling. * Note the CITES listed status of Scalloped Hammerhead Shark (SPL) meant that it has not been possible to date to transport the samples between countries for sequencing.

Species	No. tissue samples	No. selected for genetics	No. otolith samples	No. selected for microchemistry
LOT	316	298	161	70
KAW	546	362	309	104
COM	256	210	173	86
SKJ	940	385	531	81
YFT	1206	664	868	99
BET	717	521	434	100
ALB	415	288	185	79
SWO	616	417	313	70
MLS	27	22	1	-
SFA	84	79	35	-
BSH	544	364	-	-
SPL	~100*	-	-	-
Total	5767	3635	3010	689

3 Sequencing and Genetic analysis of population structure

3.1 Reduced representation library preparation and sequencing

As listed in Table 1, two forms of reduced representation library approaches for SNP discovery and genotyping were used: DArTSeq and RAD-seq. DArTseq™ genotyping is a set of proprietary methods developed by Diversity Arrays Technologies (DArT) that represents a set of complexity reduction methods coupled with sequencing on and Illumina HiSeq platform (Kilian, et al. 2012; Courtois, et al. 2013; Von Mark, et al. 2013; Raman, et al. 2014). Therefore, DArTseq™ represents a new implementation of sequencing of complexity reduced representations (Altshuler, et al. 2000) and more recent applications of this concept on the next generation sequencing platforms (Baird, et al. 2008; Elshire, et al. 2011). RAD-seq was developed by Baird et al. (2008) and has been widely used for assessing population structure of non-model organisms. Here, all species were run using DArTSeq except skipjack, which was analysed through RAD-seq. For methods comparison purposes, some skipjack samples were processed through DArTSeq and some bigeye samples, through RAD-seq.

3.1.1 DArTSeq library preparation and sequencing

DNA from all species (except skipjack) were extracted in the CSIRO O&A lab facility and shipped to Diversity Arrays Technologies (DArT) in Canberra for 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.

DNA sample libraries were created in digestion/ligation reactions using a 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. The sequencing (single read) was run for 77 cycles.

3.1.2 RADSeq

RAD-seq libraries were prepared at AZTI. Genomic DNA was extracted from about 20 mg of muscle tissue using the Wizard® Genomic DNA Purification kit (Promega, WI, USA) following manufacturer's instructions for "Isolating Genomic DNA from Tissue Culture Cells and Animal Tissue". Extracted DNA was suspended in Milli-Q water and concentration was determined with the Quant-iT dsDNA HS assay kit using a Qubit® 2.0 Fluorometer (Life Technologies). DNA integrity was assessed by electrophoresis, migrating about 100 ng of GelRed™-stained DNA on an agarose 1.0% gel and assigning values 1, 2 o 3 depending if they are poor, medium or high quality.

Restriction-site-associated DNA libraries were prepared following the methods of Etter et al. (2011). Briefly, starting DNA (ranging from 250 to 600ng, depending on integrity) was digested with the SbfI restriction enzyme and ligated to modified Illumina P1 adapters containing 5bp unique barcodes. Pools of 32 individuals were sheared using the Covaris® M220 Focused-ultrasonicator™ Instrument (Life Technologies) and size selected to 300-500 bp by cutting agarose migrated DNA. After Illumina P2 adaptor ligation, each library was amplified using 14 PCR cycles. Each pool was sent for paired-end sequenced (100 bp) on one third of a Illumina HiSeq2000 lane. Skipjack and bigeye FASTQ files were provided to CSIRO to be used for methods comparison analyses.

3.2 Post-processing of DArT Sequencing data for neritics, albacore tuna, yellowfin tuna, billfish, and sharks

DNA genotype data was generated from sequencing runs completed at DArT using a proprietary DArTseq analytical pipeline (DArT-Soft14 version) for all species (albacore, blue shark, kawakawa, Indo-Pacific sailfish, longtail tuna, narrow-barred Spanish mackerel, scalloped hammerhead, striped marlin, swordfish, yellowfin tuna), except for the populations of bigeye and skipjack analysed at AZTI. In the primary DArT-Soft14 pipeline, the FASTQ files were first processed to filter away poor-quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence. In that way the assignments of the sequences to specific samples carried in the "barcode split" step was very reliable. Processed genotype data from the DArTSoft14 pipeline was transmitted to CSIRO for further processing. In addition, DArT and RAD-seq sequence data processing for bigeye and skipjack tuna was performed following the approaches optimized at AZTI. Raw FASTQ files were also provided to IRD (albacore and blueshark) and AZTI (bigeye tuna) for downstream processing.

3.2.1 Species identification

Field identification of albacore, bigeye, longtail, and yellowfin tuna species were genetically confirmed following restriction digestion of a mitochondrial PCR amplicon (PCR-RFLP) as described by Chow and Inoue (1993) with further modifications described by Takayama et al. (2001). Size specific banding patterns representing restriction-fragment-length-polymorphisms (RFLPs) for all species were resolved on 1.2% agarose gels using standard lab practices.

3.2.2 Quality control filtering of Loci generated by DArT sequencing

A step wise process for data quality control using the package RADIATOR (Gosselin 2017) was performed to filter out both poor quality DNA markers (SNP loci) and bad DNA quality individual samples. Filtering of SNP loci included examination of marker reproducibility, removal of monomorphic markers, filtering on minor allele counts to remove poor quality loci exhibiting sequencing artefacts, minimum (poorly amplified markers) and maximum read depth to identify and remove paralogous loci (e.g. repetitive DNA), the number of SNPs present in a sequencing fragment, and whether loci comply with assumption of within population Hardy Weinberg equilibrium (Andrews 2010).

3.2.3 Quality control filtering of Individuals

DNA quality of individual samples were filtered at three key steps: 1. missing high proportion of genotype calls (indicative of individuals with poor quality DNA); 2. Higher than average heterozygosity (identifies and removes individuals with DNA cross contamination); 3.removal of highly similar/duplicate genotypes (removes technical and accidental tissue replicates).

3.2.4 QC of individuals for low quality DNA or cross-contamination based on Heterozygosity

We used significant deviations from average heterozygosity as a proxy for DNA cross-contamination and very poor quality DNA extracts. On average individuals within a population will have the same level of heterozygosity as each other. However, if the heterozygosity observed for the DNA profile of an individual deviates from this average then this likely reflects sample cross contamination – introduced at the point of sampling in the field, during handling or subsampling in the lab – and often is the symptom of poor tissue sampling skills or inadequate cleaning protocols (e.g. not cleaning the knife or scalpel blade in between samples, not cleaning hands when handling multiple samples). Conversely, samples with lower than average heterozygosity are likely an indication of poor DNA quality that results in a homozygous excess as a result of introduced artefactual sequencing bias. An important step in assessing the quality of samples is therefore to identify samples that are either too homozygous or too heterozygous compared to the average observed level of heterozygosity observed in a population. To do this, the level of genome-wide mean heterozygosity is calculated. For the current study, individual samples with a mean heterozygosity above and below statistical threshold values of higher and lower confidence limits were identified and filtered out of datasets for further quality control.

3.2.5 QC of duplicate and related genotypes

Related individuals were identified using the “show duplicates” filter in RADIATOR. Genetic similarity was used to identify potentially related and duplicate individuals that show higher levels of genetic similarity and by extension show lower levels of genetic distance between them relative to average genetic distance between unrelated pairs. In essence, non-related individuals should have genotypes that are dissimilar (because they have no common relatives to derive their genes from). However, when cross-contamination or technical mishaps occur (e.g. labelling two samples collected from the same individual as different animals), samples with similar or almost identical

genotypes can occur among individuals sampled from a population. Care needs to be taken in examining individuals with similar genotypes to determine if values of genetic distance are reflective of relatedness or the result of human error. The sequencing process also includes a number of technical replicates (these are duplicate DNA aliquots sequenced from selected individuals representing 5-10% of the total individuals for each plate) that are run by DArT as internal controls to assess marker quality (e.g. reproducibility of SNP calls at a locus). The individual exhibiting the lowest call rates from each pair was removed from the final dataset.

3.3 Processing of RAD FASTQ (skipjack) and DArT FASTQ (bigeye) data

3.3.1 Raw read pre-processing and quality control

Generated RAD-tags were pre-processed with the *process_radtags* module of Stacks 2.4 (Catchen et al. 2013). Sequences were demultiplexed based on unique barcodes and only those that contained the restriction enzyme cut site and whose overall average Phred (quality) score was higher than 20 were included. PCR duplicates were removed using *clone_filter.*

3.3.2 Genotype table generation for bigeye and skipjack

Generated cleaned tags for skipjack and bigeye were analysed with Stacks 2.4 (Catchen et al. 2013), calling the specific stacks modules from two custom scripts (assemble.sh and genotypeTables.sh). Bigeye reads, processed through DArTSeq, where further checked using *process_radtags* (to check for presence of restriction enzyme site, to remove adaptor sequences and to truncate all reads to the same length). Putative orthologous tags per individual were assembled using *ustacks* with a minimum of 3 identical reads to create a stack and a maximum of 4 mismatches allowed between stacks. A catalogue of RAD loci was built using *cstacks* with a maximum of 6 mis-matches between sample tags and matches to the catalogue of individual samples were searched using *sstacks* and transposed using *tsv2bam*. Using only samples with a minimum of 40,000 RAD-tags, the module *gstacks* was used to assembly paired-end reads into contigs, merging them to the single-end locus and identifying and genotyping SNPs. SNP selection and genotype table building were performed using a custom script (*genotypeTable.sh*). First, using populations we selected the SNPs found in tags present in at least 90% of the individuals and exported into PLINK (Purcell et al. 2007). This was used to select individuals with genotyping rate larger than 0.75 and SNPs with a genotyping rate larger than 0.95, and a minimum allele frequency larger than 0.05.

3.4 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 fish within which the genetic profiles are more similar to those between groups. The method is a ‘soft classification’ method as it assigns fish to groups on a probabilistic basis, rather than with a ‘hard’ decision rule.

Information about the number of groups that the data support is obtained using two sources. The first are information criteria (AIC and BIC, see Miller 2002), which are obtained parametrically from the model and the model's likelihood (how well the model fits the data). The second source is using a resampling method similar to cross-validation. The resampling method gives an empirical indication of performance. Here we repeatedly resample (only 20 times in this initial analysis) 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 only 20 resamples in this initial analysis. The affinity to a genetic group is measured by a fish's probability – high, or low, probability means that it is more, or less, likely to be part of that group, respectively.

It is important to note that the sampling regions are not used in this analysis. The only information entered are the genetic data themselves. This means that the analysis does not intentionally seek spatially consistent grouping, but if there is a real spatial signal then this should show in any case.

3.5 Results and discussion

3.5.1 Status of the preliminary results and opportunity for further input

The timing of completion of genotyping and QC of these data means there has been very limited time for analysis and interpretation prior to Scientific Committee meeting. Hence, the results presented are very preliminary and the data will be subject to further diagnostics, additional analysis and interpretation. They should be considered as indicative only. We make no conclusions about what the results may mean in terms of population structure or the potential implications for assessment and management, as this would be premature and needs to await the results of more detailed and comprehensive analyses and the input of the Scientific Committee and Working Groups.

In light of the above, we invite and encourage feedback from the Scientific Committee on these preliminary results and specific suggestions on additional work (e.g. analyses or initial input on interpretation) that can be used by the project team in refining these analyses for the draft final report due with the IOTC Secretariat on 20 December 2019.

The draft final report will include more complete, but still not final, analyses for each species and provide a more substantive opportunity for review and input from the SC and WPs. This input will be used by the project team in a further round of refinement of analyses, interpretation and synthesis between Feb-March 2020, which will be reported in the final report to the IOTC Secretariat before 31 March 2020.

3.5.2 Presentation of genetics results

We have attempted to present the substance of the results in a clear and straight forward fashion. For each species we include: i) a map with the distribution of the samples that were processed for the species using the method (e.g. genetics or microchemistry); a summary of the fit of the population model (where one was used), and; a representation of the distribution of samples (individual fish) among locations for the most likely number of groups, or "*k*".

In most cases for the genetics results, the summary text provides the number of samples actually used in the preliminary population structure analysis presented. This can differ to the number processed as the data for some individuals may have been excluded as part of the final quality control tests, prior to the population structure analysis.

The "Choosing "*k*" by Information Criterion", summarises the results of the model fits for population structure and is one way of identifying the most likely number of genetic groups, or "k". In these plots the lowest value indicates the most likely, or preferred number of groups. If the most likely *k* is equal to one, then the plot will look like the results for Indo-Pacific sailfish, shown below in Figure 2. If the most likely number of groups is greater than one, then there will be a dip, or U-shape pattern in the plot with the bottom of the "U" being the most likely number of groups, as shown in the example for narrow barred Spanish mackerel in Figure 2 below.

The bottom plot illustrates the probability of membership of groups among the sample locations. In most cases this is shown for the most likely number of groups. In cases where the

most likely number of groups is one, then the plot is omitted, as it is uninformative: all locations and individuals are shown as the one colour. In the case of swordfish, we have included two versions of this group membership plot, as even though the model fit had a preference for a single group, prior knowledge and the results for k = 2 or 3 indicate the potential for structure consistent with that suggested by earlier studies.

The other form of analysis used to examine the potential for differences among sampling locations is Principle Components Analysis, or PCA, as a form of multivariate statistical test. This has been used for the genetics for skipjack and for the microchemistry for skipjack, yellowfin, albacore, and swordfish.

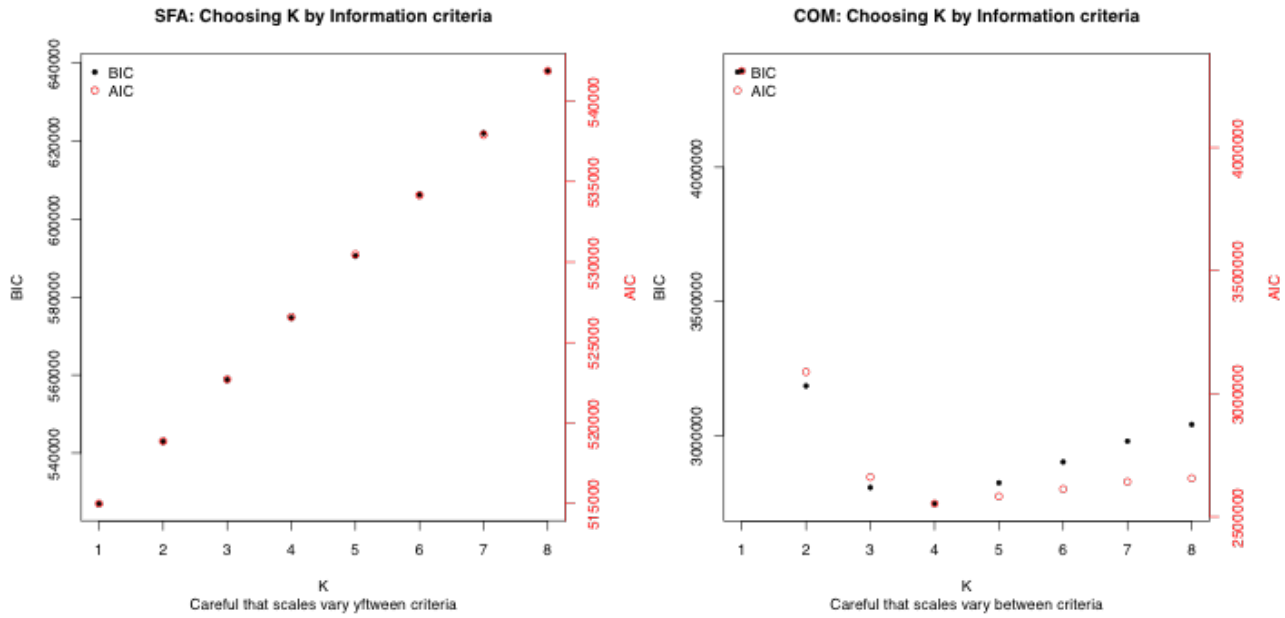

Figure 2. The Information Criterion plots summarise the results of model fits for the most likely number of genetic groups from the distribution of SNP data in the sample. The AIC and BIC are two forms of statistic used to summarise how well the model fits the data with the lower the value the more likely the *k*. In these examples, the result on the left for Indo-Pacific sailfish indicates that k = 1, or a single group is most consistent with the data, while the example on the right for narrow-barred Spanish mackerel, with the bottom of the "U" at 4 for both AIC and BIC, indicates that four groups is most likely number of genetic groupings in the data.

3.5.3 Neritic tuna

Longtail tuna (*Thunnus tonggol*)

- The sample coverage for longtail tuna covers a reasonable proportion of the Indian Ocean range. Unfortunately, it was not possible to source samples from the NW and W parts of the range;
- A total of 316 samples from 3 Indian Ocean sampling regions were collected and samples from a Pacific outlier are being sought. A total of 298 samples were sequenced using DArTSeq and included in the preliminary analysis of population structure (Figure 3);
- The preliminary population analysis based on StockR indicates a strong preference for two or three genetic groupings within the Indian Ocean with 3 groups being most likely based on the model fits (Figure 3).

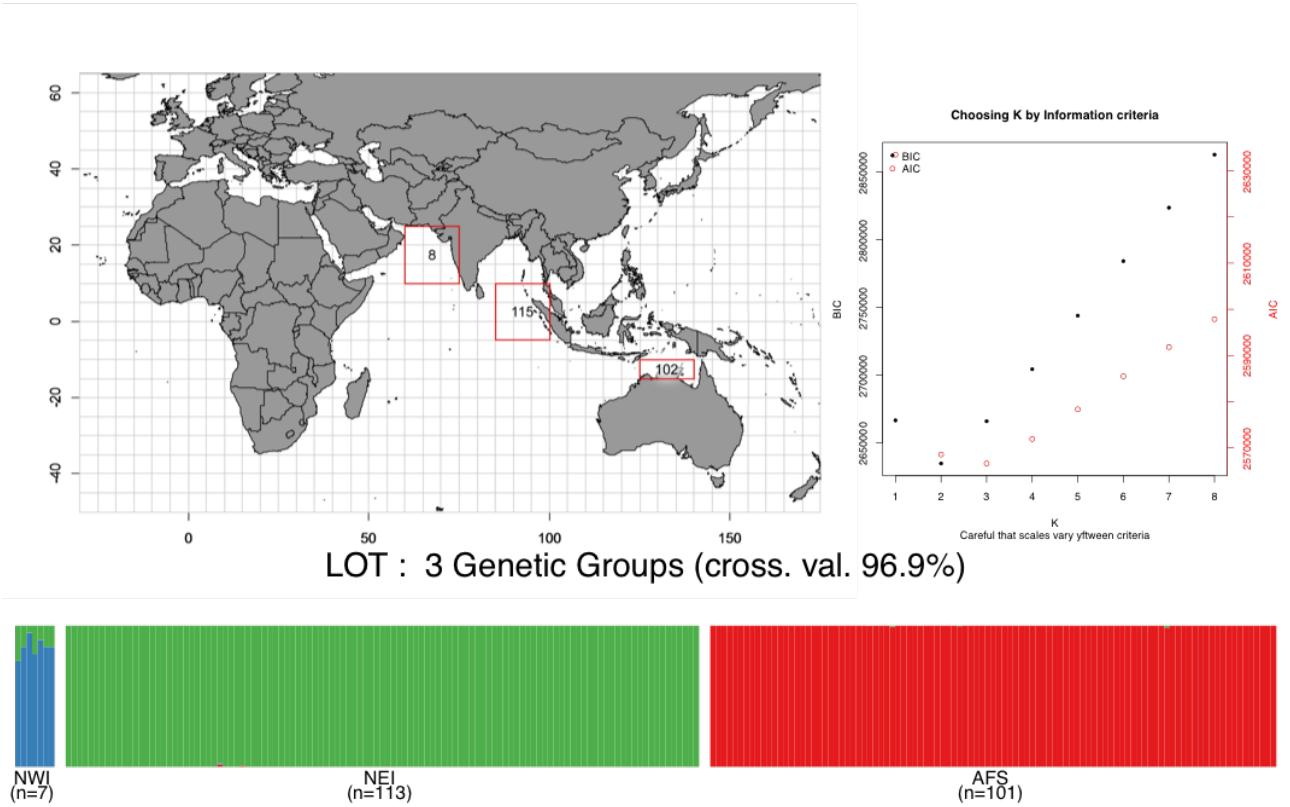

Figure 3. Top Left: Distribution of samples (N=316) of Longtail tuna (*Thunnus tonggol*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: very preliminary results of population structure analysis of DArTSeq using StockR for longtail tuna for 3 genetic groups.

Kawakawa (*Euthynnus affinis*)

- The sample coverage for kawakawa is generally very good across much of the Indian Ocean range of the species. Additional samples from the central-west and south-west Indian Ocean would complete the coverage of the range;
- A total of 546 samples from 7 Indian Ocean sampling regions were collected. A total of 362 were sequenced using DArTSeq and included in the preliminary analysis of population structure (Figure 4);
- The preliminary population analysis based on StockR indicates a preference for 2 genetic groupings within the Indian Ocean based on the model fits (Figure 4).

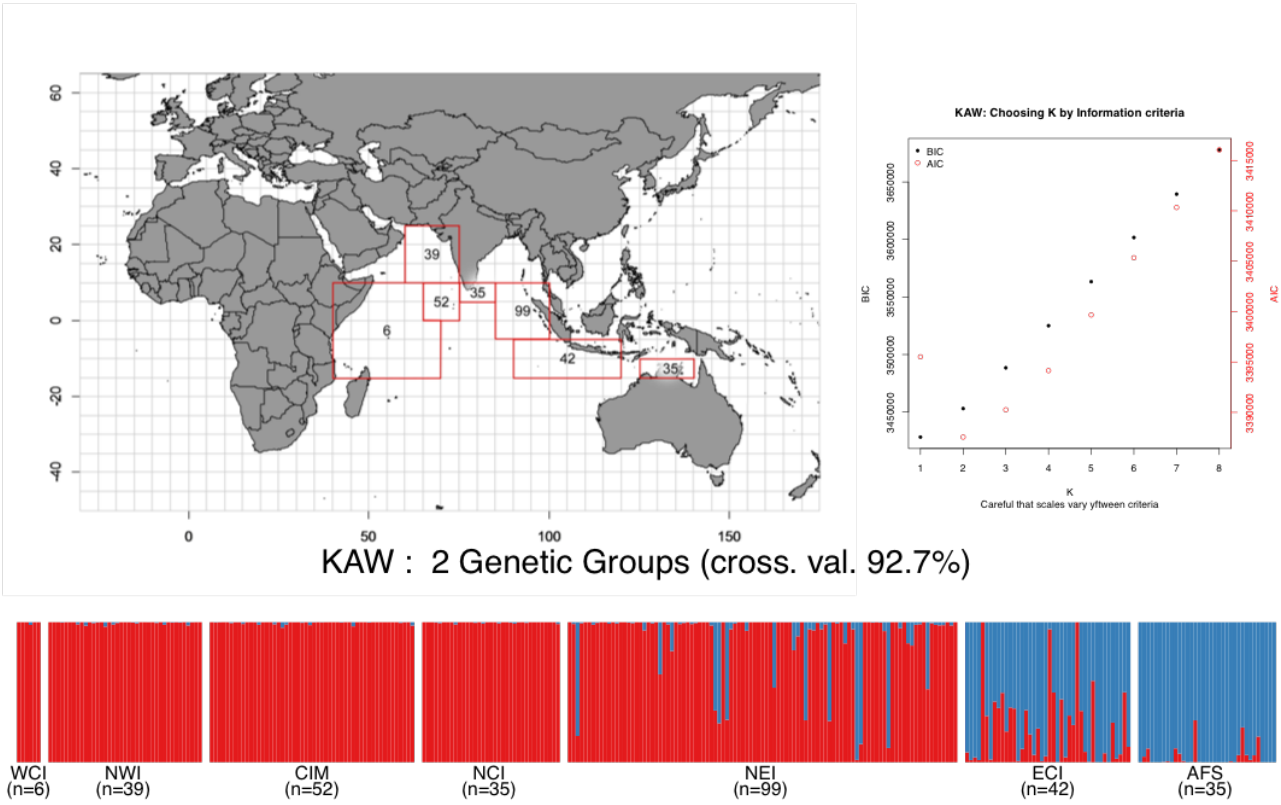

Figure 4. Top Left: Distribution of samples (N=362) of kawakawa (*Euthynnus affinis*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: very preliminary results of population structure analysis of DArTSeq using StockR for kawakawa suggesting for 2 genetic groups within the Indian Ocean.

Narrow-barred Spanish mackerel (*Scomberomorus commerson*)

- The sample coverage for narrow-barred Spanish mackerel is good for the eastern and northern regions of the Indian Ocean range of the species and for the Pacific outlier location. Samples have been collected in the SWI, but unfortunately have not yet been sequenced and were not available for inclusion in this analysis;
- A total of 256 samples from 4 Indian Ocean sampling regions and one Pacific outlier location were collected. A total of 210 were sequenced using DArTSeq and included in the preliminary analysis of population structure (Figure 5);
- The preliminary population analysis based on StockR indicates a strong preference for four genetic groupings within the Indian Ocean based on the model fits (Figure 5);
- Based on this preliminary analyses, these correspond quite closely to 3 regions within the Indian Ocean (NWI+NCI, NEI and ECI) and one consisting of the AFS and WCS.

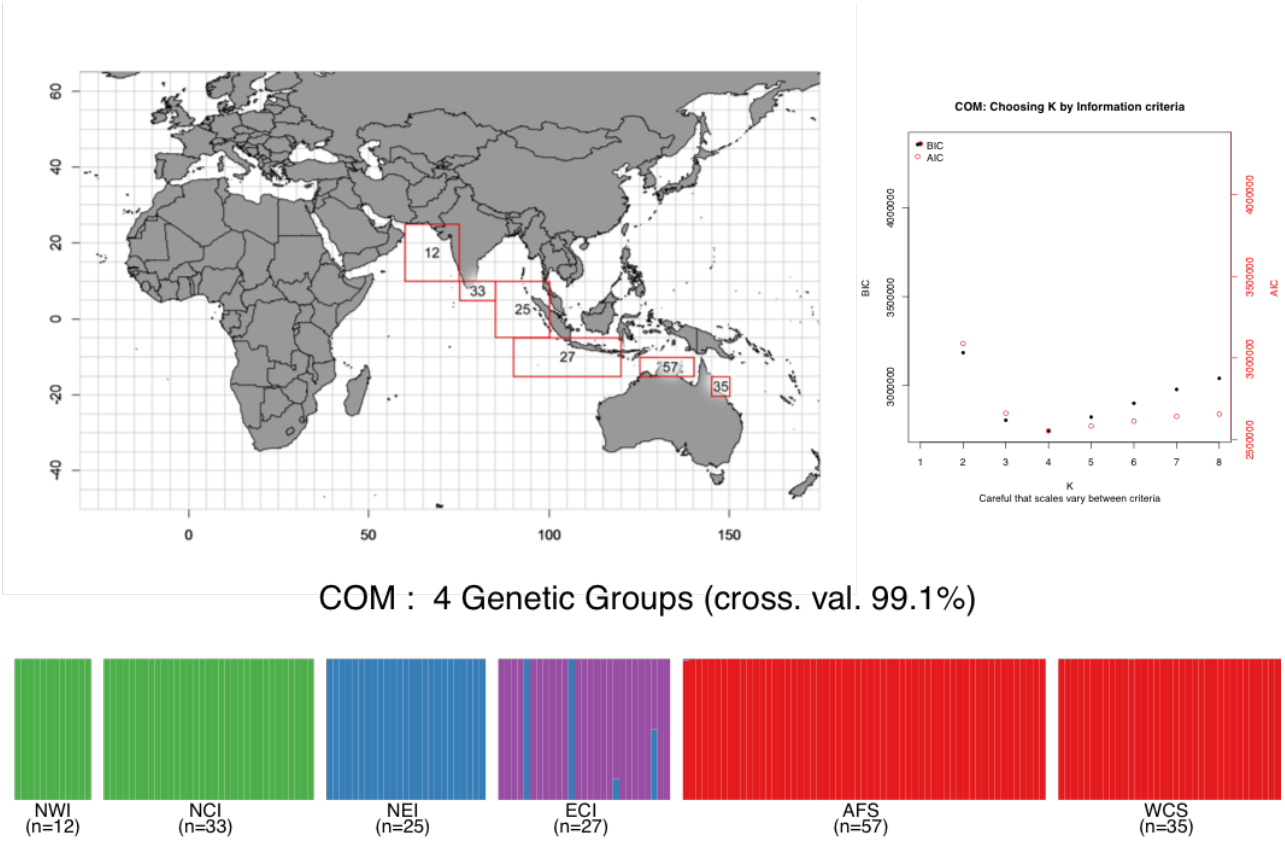

Figure 5. Top Left: Distribution of samples (N=316) of narrow-barred Spanish mackerel (*Scomberomorus commerson*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: very preliminary results of population structure analysis of DArTSeq using StockR for narrow-barred Spanish mackerel for 4 genetic groups.

3.5.4 Tropical tunas

Skipjack tuna (*Katsuwonus pelamis*)

- The sample coverage for skipjack was generally very good across the range with the exception of the north west;
- A total of 940 samples from 9 Indian Ocean sampling regions and two outlier locations (east Atlantic and south-west Pacific) were collected. A total of 385 samples were sequenced using RADSeq and included in the preliminary analysis of population structure (Figure 6);
- Very preliminary population analysis, based on 261 individuals using PCA, does not provide evidence of genetic differentiation among locations within the Indian Ocean (Figure 6).

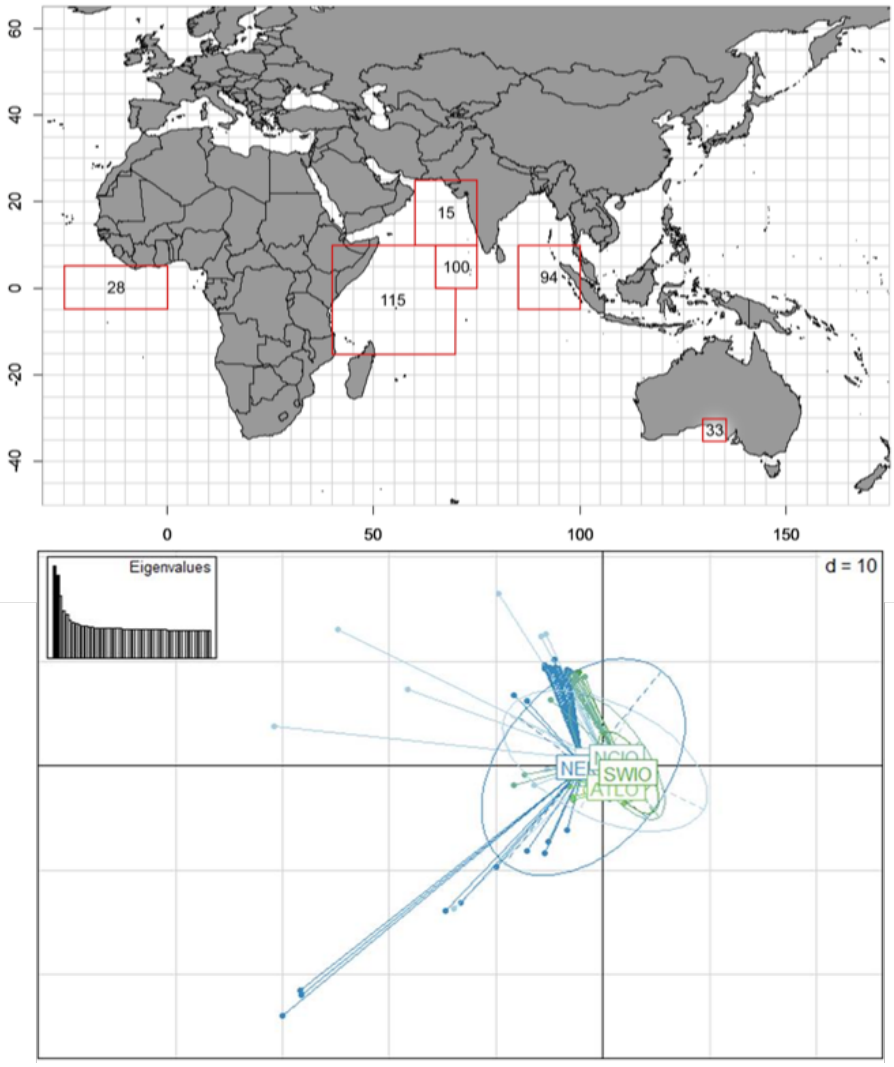

Figure 6. Top Left: Distribution of samples (N=385) of skipjack (*Katsuwonus pelamis*) for both rounds of sampling sequenced using RADSeq by sampling region for PSTBS-IO project. Bottom: very preliminary results of population structure analysis of RADSeq using PCA for 261 skipjack after filering. ARLO - Atlantic Ocean, SWIO – south-west Indian Ocean, NWIO – north-west Indian Ocean, NCIO – north-central Indian Ocean, NEIO – north-east Indian Ocean.

Yellowfin tuna (*Thunnus albacares*)

- The sample coverage for yellowfin was generally very good. As for many species the samples consist of a mix of YoY fish and mature adults;
- A total of 1206 samples from 9 Indian Ocean sampling regions and two outlier locations (east Atlantic and south-west Pacific) were collected;
- A total of 664 samples were sequenced using DArTSeq and included in the preliminary analysis of population structure (Figure 7);
- Very preliminary population analysis, based on StockR, suggests that 2 genetic groupings within the Indian Ocean are more likely than 1, with the preference for 1 and 3 groups being similar (Figure 7).

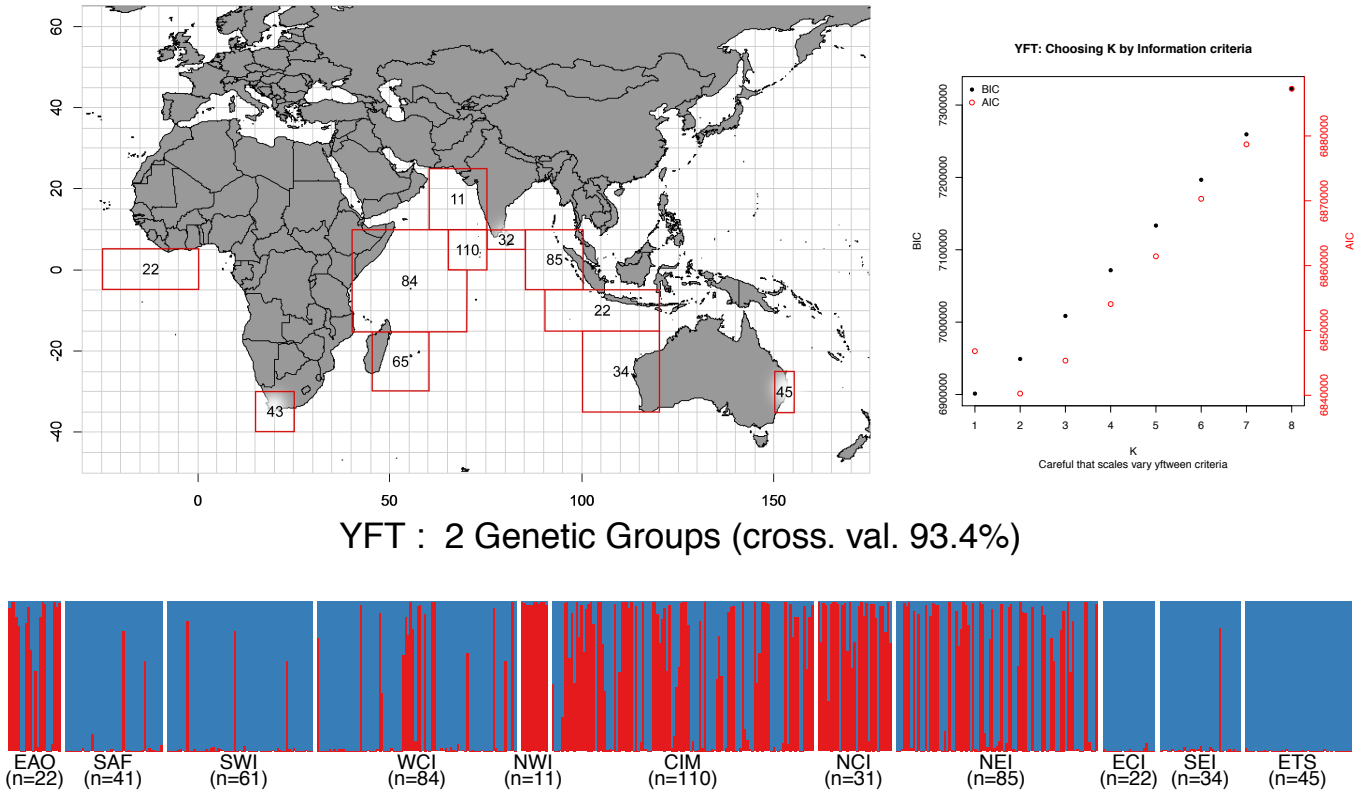

Figure 7. Top Left: Distribution of samples (N= 664) of yellowfin tuna (*Thunnus albacares*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k). Bottom: very preliminary results of population structure analysis of DArTSeq using StockR tuna for 2 genetic groups.

Bigeye tuna (*Thunnus obesus*)

- The sample coverage for bigeye tuna was generally good, although small sample size was obtained from the north central and middle Indian Ocean;
- A total of 717 samples from 9 Indian Ocean sampling regions and two outlier locations (east Atlantic and southwest) were collected;
- A total of 717 samples were sequenced using DArTSeq and 701 included in the preliminary analysis of population structure.
- After filtration for missing values per loci and individuals, mf 0.05, HWE, 1 SNP per tag, 701 individuals and 5132 loci were kept
- Very preliminary analyses show differentiation between Atlantic and Indian Ocean locations. Future analyses will focus on the possible structure within the Indian Ocean

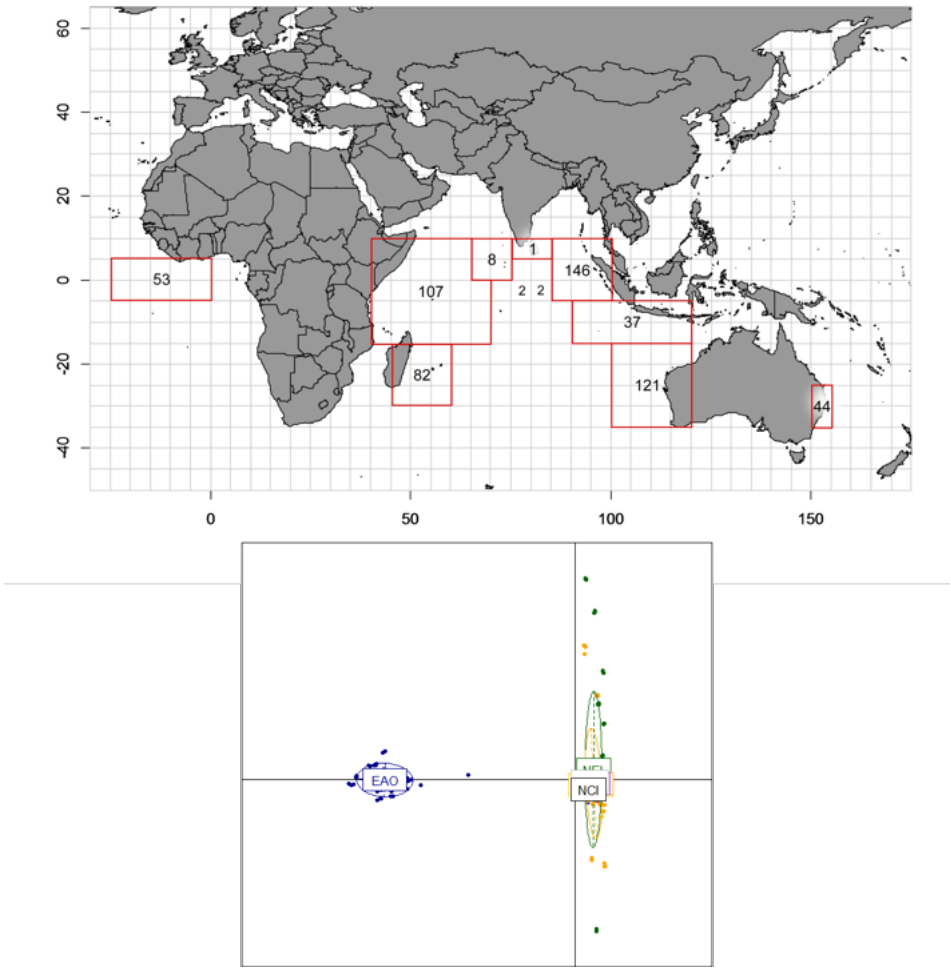

Figure 8. Top : Distribution of samples (N=717) of bigeye (*Thunnus obesus*) for both rounds of sampling sequenced using RAD-Seq by sampling region for PSTBS-IO project. Bottom: very preliminary results of population structure analysis of RAD-Seq showing a PCA plot of the 701 samples and 6553 SNP of bigeye tuna.

3.5.5 Temperate tunas

Albacore (*Thunnus alalunga*)

- The sample coverage for albacore for the western Indian Ocean and western Pacific was good. However, the coverage in the eastern Indian Ocean was poor due to difficulties in accessing samples;
- A total of 415 samples from 3 Indian Ocean sampling regions and two Pacific outlier locations (east Tasmania and west Tasman Sea, in the Pacific) and one Atlantic location were collected (Figure 9
- A total of 288 samples were sequenced using DArTSeq and included in the preliminary analysis of population structure (**Error! Reference source not found.**);
- Very preliminary population analysis based on DAPC and FST using 5,826 SNPs and 269 individuals suggests 1 genetic grouping within the Indian Ocean and 1 in southwest Pacific. Significant heterogeneity was detected between-ocean comparisons (Atlantic North, Atlantic Southeast, Indian, and Pacific Southwest oceans). (Figure 9).

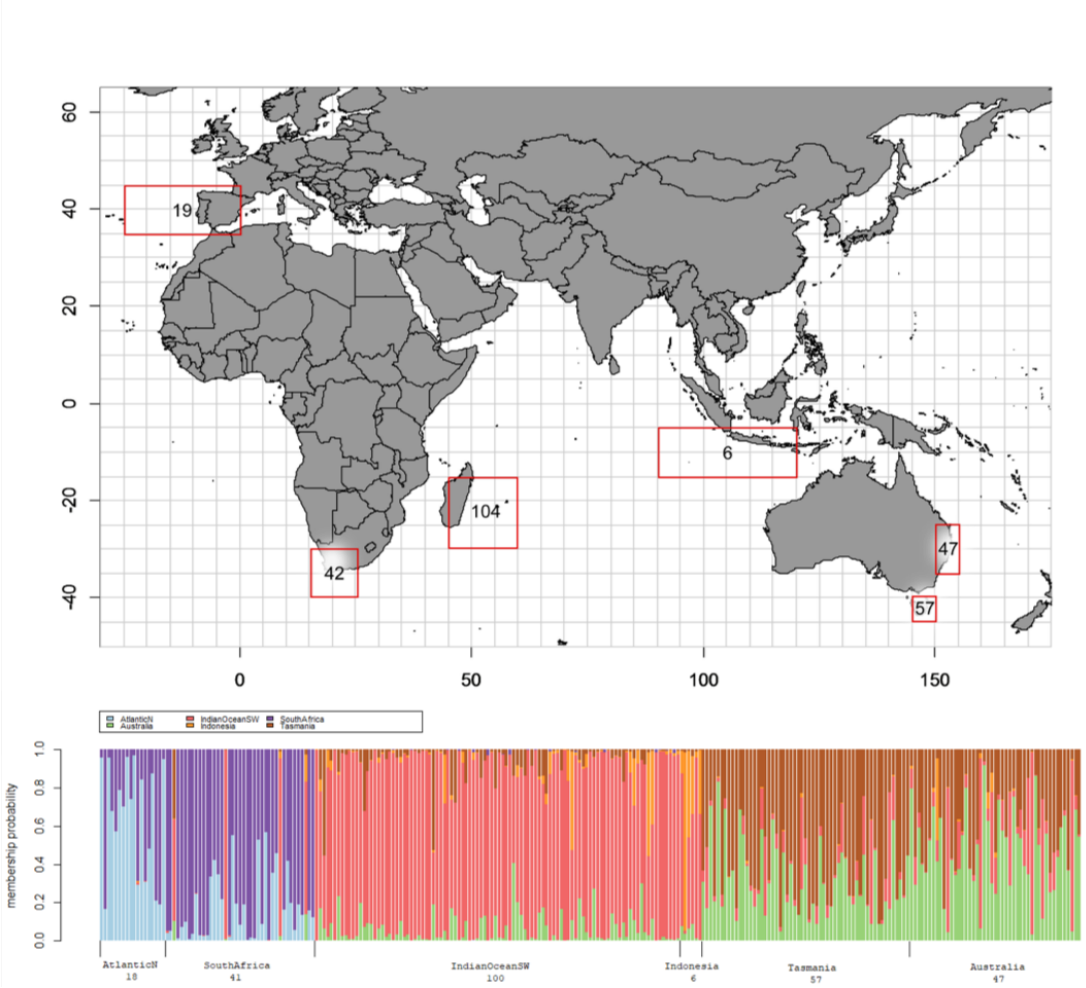

Figure 9. Top Left: Distribution of samples (N= 301) of albacore tuna (*Thunnus alalunga*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Bottom: very preliminary results of population structure analysis of DArTSeq using DAPC for albacore tuna over 5,826 SNPs and 269 individuals. Each colour represents a location.

3.5.6 Billfish

Swordfish (*Xiphias gladius*)

- The sample coverage for swordfish across the Indian Ocean was good with one outlier location;
- A total of 616 samples from 6 Indian Ocean sampling regions and one outlier location in the western coral sea;
- A total of 417 samples were sequenced using DArTSeq and included in the preliminary analysis of population structure (**Error! Reference source not found.**);
- Very preliminary population analysis, based on StockR, suggests a single genetic grouping across all sample locations, although analyses assuming 2 or 3 groupings suggest a more complex structure may be present (**Error! Reference source not found.**).

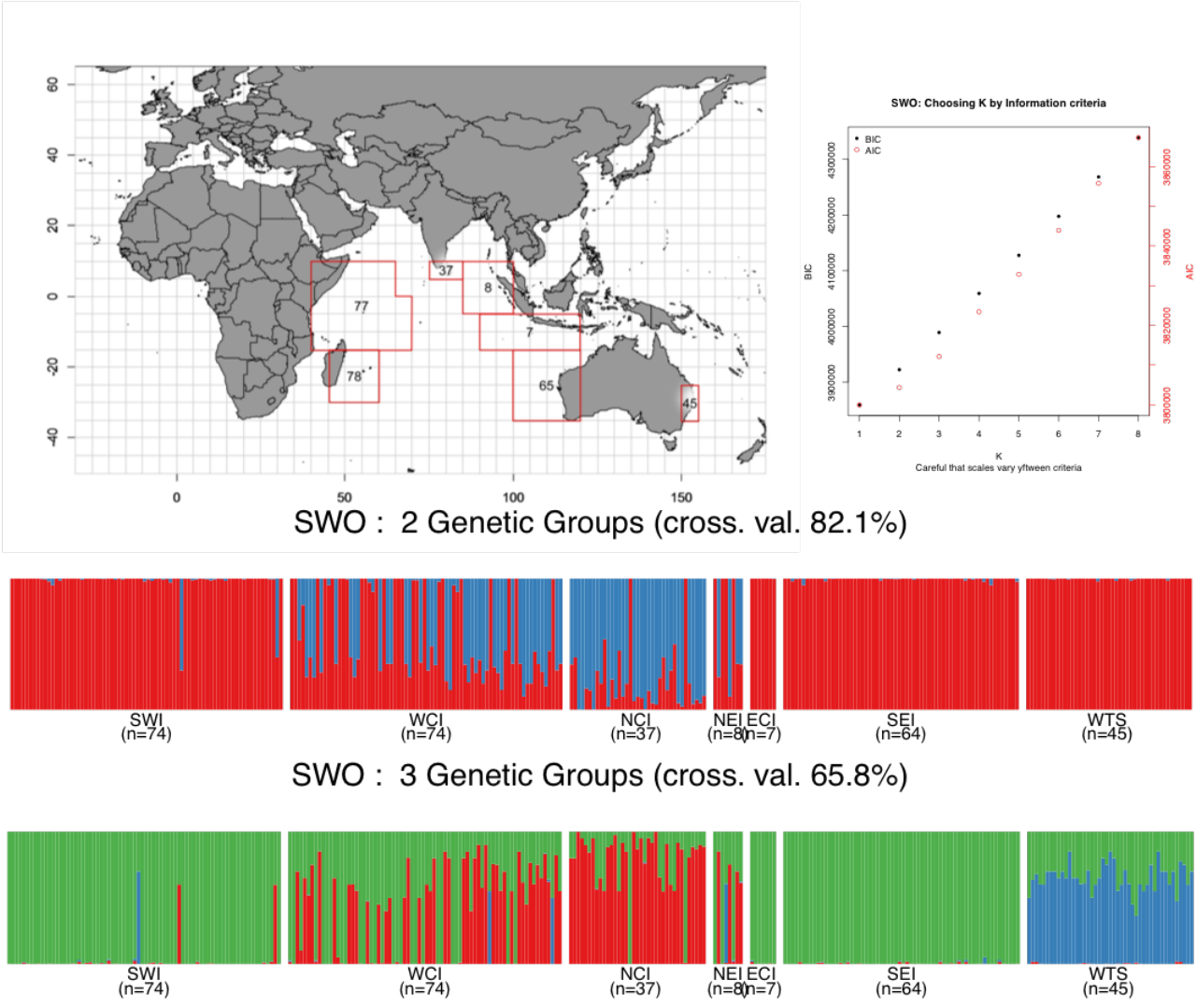

Figure 10. Top Left: Distribution of samples (N= 417) of swordfish (*Xiphias gladius*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k). Bottom: very preliminary results of population structure analysis of DArTSeq using StockR for swordfish assuming 2 and 3 genetic groups, noting the model fits suggest 1 genetic grouping is most likely.

Striped marlin (*Tetrapturus audax*)

- The sample coverage for striped marlin was poor with a total of 3 samples from the Seychelles and 19 samples for the western Coral Sea;
- A total of 22 samples were sequenced using DArTSeq. These will be included in a wider analysis striped marlin as part of the final report (**Error! Reference source not found.**).

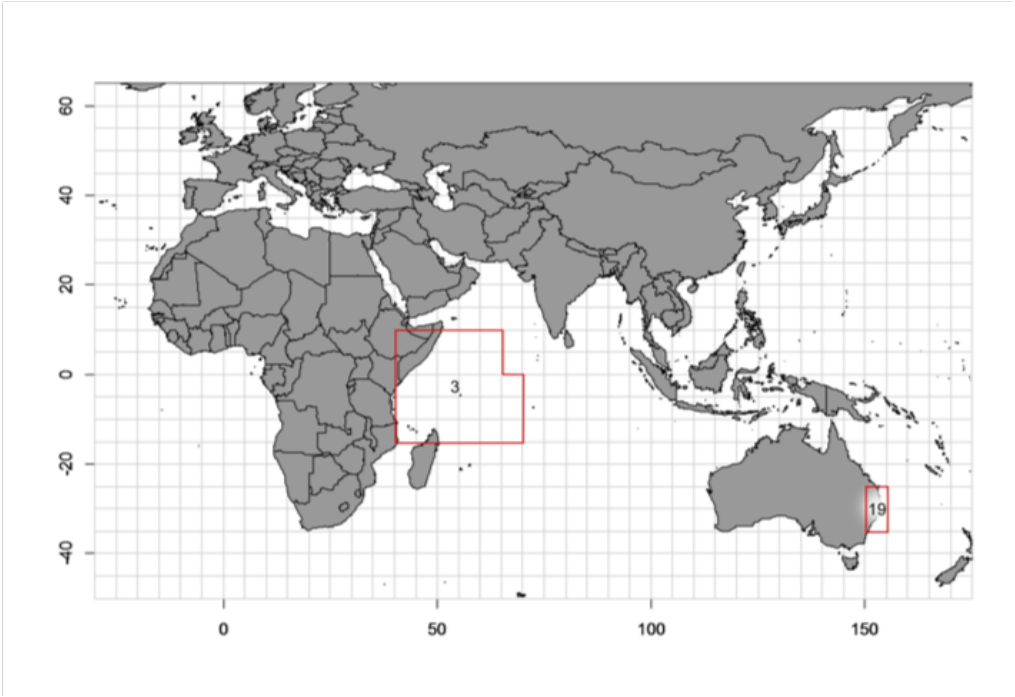

Figure 11. Distribution of samples (N= 22) of striped marlin (*Tetrapturus audax*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project.

Indo-Pacific sailfish (*Istiophorus platypterus*)

- A total of 84 Indo-Pacific sailfish were sampled at 3 locations across the Indian Ocean, with sufficient samples for analysis from the Seychelles in the west and Lampulo in the east;
- A total of 79 samples were sequenced using DArTSeq and included in the preliminary analysis of population structure (**Error! Reference source not found.**);
- Very preliminary population analysis, based on StockR, suggests a single genetic grouping across all sample locations in the Indian Ocean (**Error! Reference source not found.**).

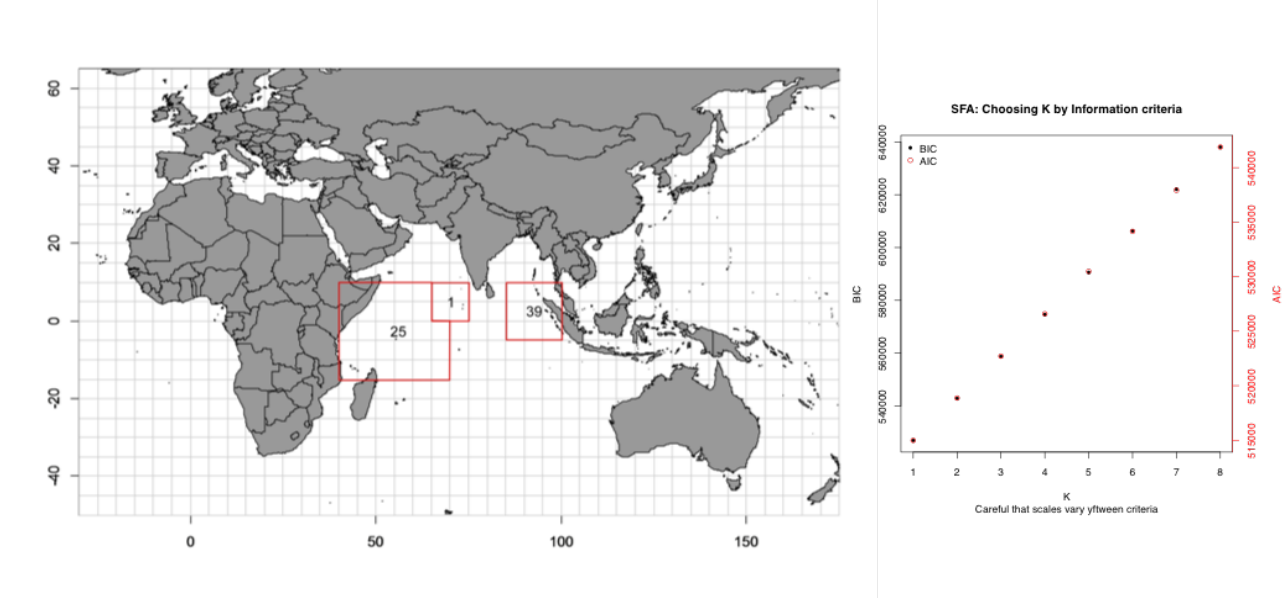

Figure 12. Top Left: Distribution of samples (N= 417) of Indo-Pacific sailfish (*Istiophorus platypterus*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k).

3.5.7 Sharks

Blue shark (*Prionace glauca*)

- The sample coverage for blue shark across the Indian Ocean from direct sampling was difficult. Hence this was supplemented with existing sample collections from Indian Ocean and a number of outlier locations;
- A total of 544 samples from 5 Indian Ocean sampling regions and 6 outlier locations in the Atlantic, Mediterranean and Pacific;
- A total of 376 samples were sequenced using DArTSeq and included in the preliminary analysis of population structure (**Error! Reference source not found.**);
- Very preliminary population analysis, based on DAPC, PCoA, and FST 16,466 SNPs and 348 individuals, revealed very low values of differentiation between areas. Nevertheless significant index was detected between-ocean comparisons (ex. FST) but not within (ex. FST not significant between Indian ocean samples). (**Error! Reference source not found.**).

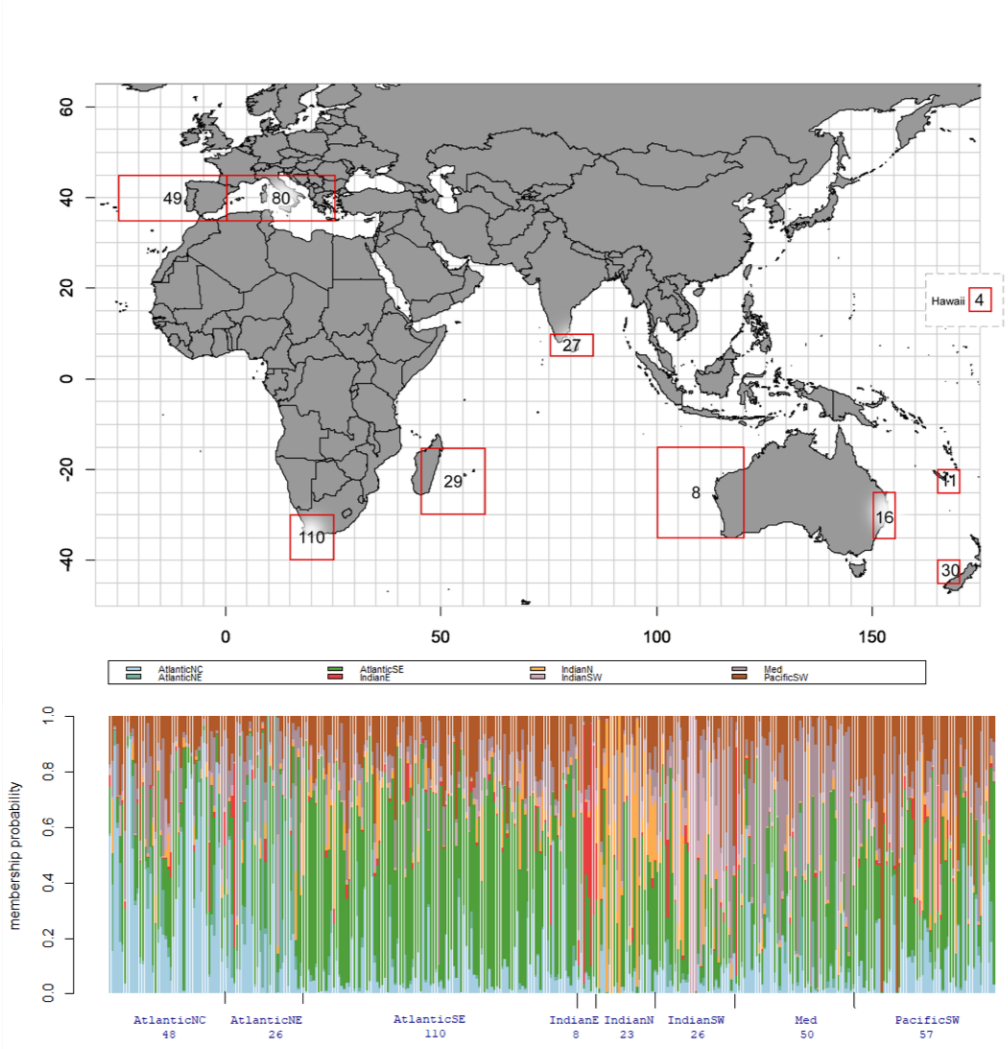

Figure 13. Top Left: Distribution of samples (N=364) of blue shark (*Prionace glauca*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k). Bottom: very preliminary results of population structure analysis of DArTSeq using DAPC for blue shark over 16,466 SNPs and 348 individuals. Each colour represents a location.

4 Microchemistry analysis of population structure

4.1 Sample selection for analysis

Representative individuals from each of the sampling locations and species were chosen from the total sample pool collected for the project (Table 4).

4.2 Methods

Sagittal otoliths were selected by fish length. One of the paired otoliths was analysed for trace element chemistry and where possible, the second otolith was analysed for stable isotopes, δ13C and δ18O. Transverse sections were cut from the otoliths and then polished until the primordium (or core) and growth rings became visible. The areas of particular interest along the section were the primordium, where otolith material reflects the environment of the spawning grounds, the near-core material, deposited when juvenile fish were in their nursery grounds, and the otolith edge, reflecting the water mass in which fish were caught.

The partners analysed otoliths at different facilities. The use of different facilities and their reference samples means that the comparisons among locations for individual species will be consistent, but the potential for differences in the calibrations between different facilities mean that cross-species comparisons can only be done between species processed at the same facility.

In general, trace elements were measured using laser ablation-ICPMS and were performed either as a continuous transect from primordium to margin or by spot analysis. For stable isotope analyses, material from each otolith was isolated using a high resolution MicroMill system consisting of a microscope and imaging system. Carbon and oxygen stable isotopes of the otolith material were analysed on a mass spectrometer coupled to automated carbonate preparation device.
Discriminant functional analyses were carried out on both the stable isotope and element data to examine similarity and differences in the otolith composition among groups of fish at different life stages and locations. Other statistical methods, such as principal component analyses and analyses of variance, may also be applied.

4.3 Statistical analysis

When age cannot be determined, age groups are determined during the otolith's treatment as a function of the length of the ICPMS transect, which is also proportional to the size of the otolith. Principal Component Analyses are performed to detect which trace-element is characterizing each age group and area. Finally parametric tests are used to determine if trace-element compositions at each age group and areas are significantly different.

4.4 Results and discussion

4.4.1 Neritic tuna

Sampling is complete for each of the neritic tuna species and the subset of otoliths to be included in the microchemistry analysis have been selected, cleaned and are being processed. A delay in the preparation process (sectioning) for LA-ICPMS analysis means that the results are not available for these species (or bigeye tuna) for this report. The processing will be completed over the coming month and included in the final synthesis and in the Final Report and publications for the project. Distributions of samples available for otolith microchemistry for these species are shown in the individual species maps below (Figure 14; Figure 15; Figure 16; Figure 19).

Longtail tuna (*Thunnus tonggol*)

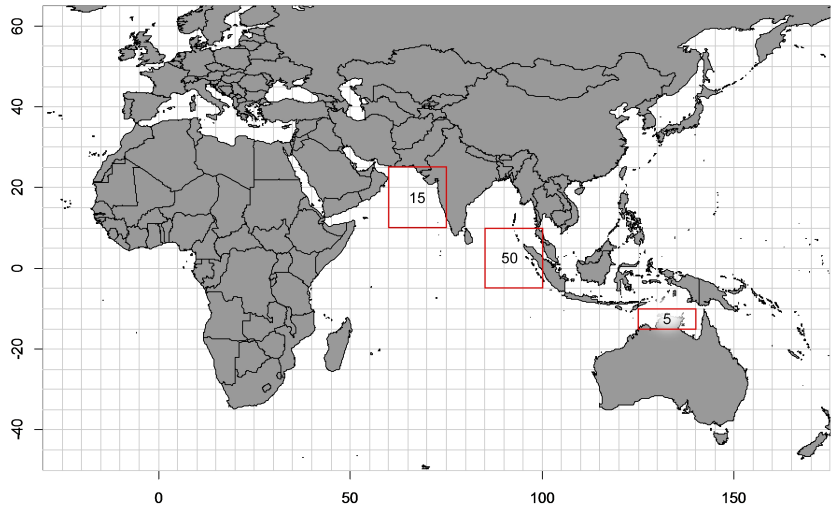

Figure 14. Distribution of samples (N=70) of longtail tuna (*Thunnus tonggol*) for multi-elemental micro-chemistry analysis.

Kawakawa (*Euthynnus affinis*)

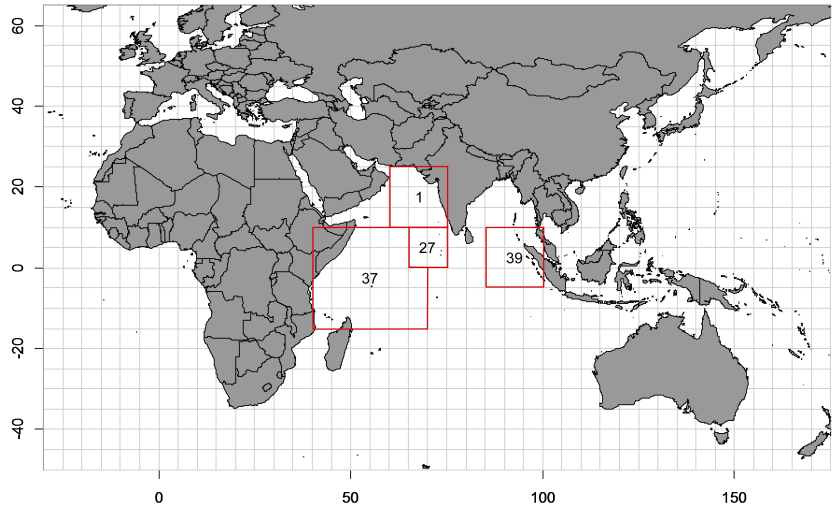

Figure 15. Distribution of samples (N=104) of Kawakawa (*Euthynnus affinis*) for multi-elemental micro-chemistry analysis.

Narrow-barred Spanish mackerel (*Scomberomorus commerson*)

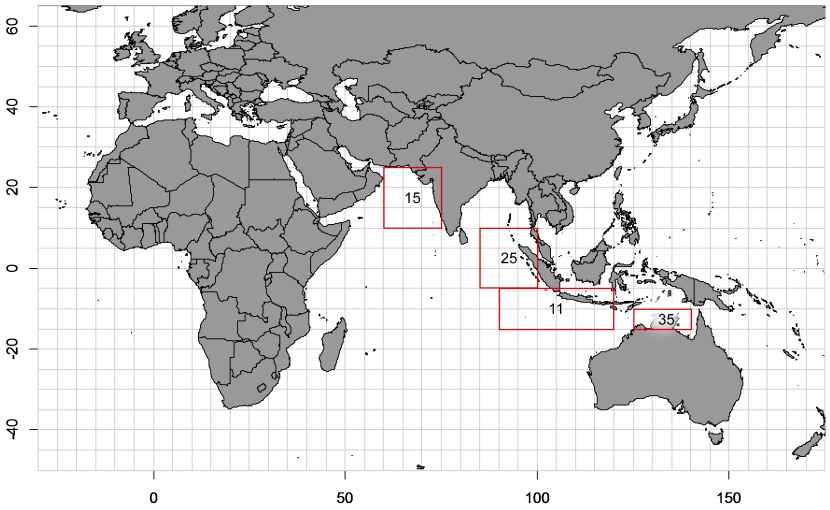

Figure 16. Distribution of samples (N=86) of Narrow-barred Spanish mackerel (*Scomberomorus commerson*) for both rounds of sampling for multi-elemental micro-chemistry.

4.4.2 Tropical tunas

Skipjack tuna (*Katsuwonus pelamis*)

- A total of 531 otoliths were collected for skipjack from locations within the Indian Ocean and 75 were processed and included in the preliminary microchemistry analysis for population structure (Figure 17);
- The trace elements that provided maximum discrimination accuracy (^{138}Ba, ^{88}Sr, ^{25}Mg) were selected for analysis;
- Significant differences were found between NEIO and SWIO (PERMANOVA, P= 0.01), but not between the other locations included in the analysis (Figure 17).

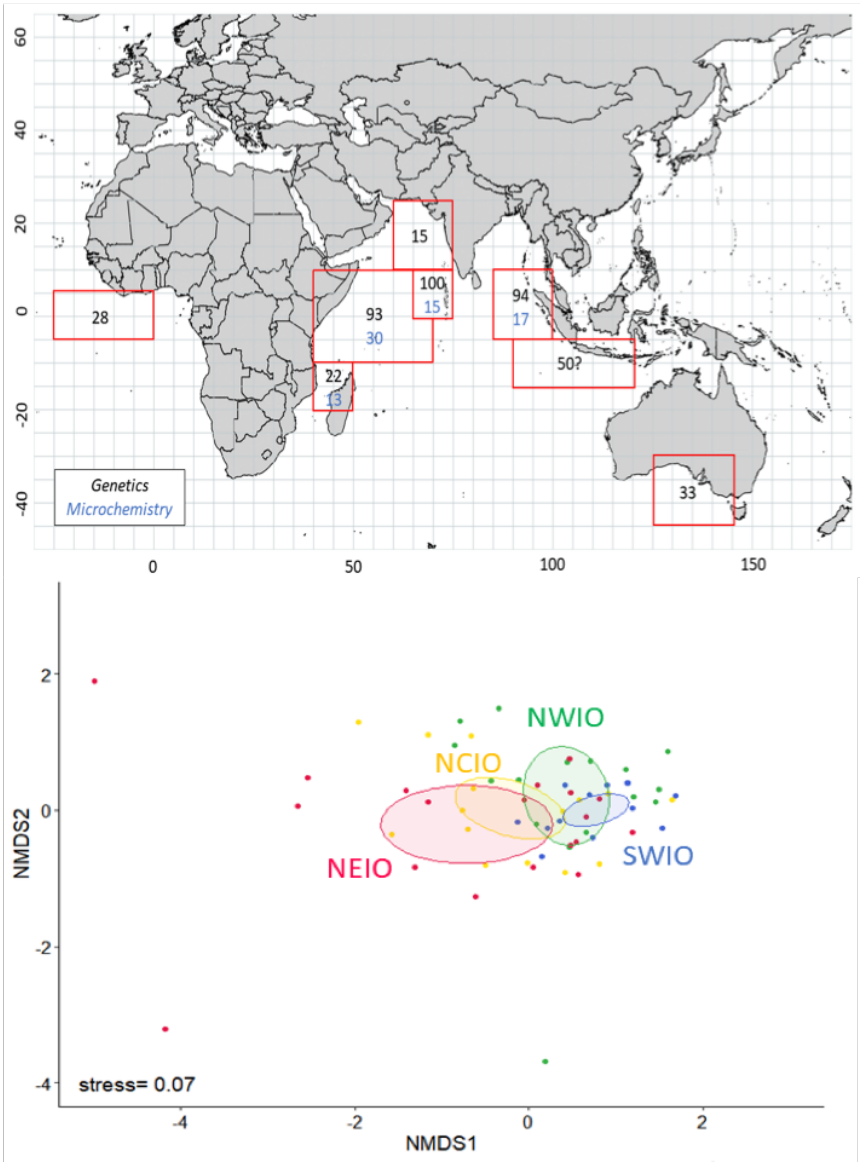

Figure 17. Top: Distribution of otolith (blue) and tissues (black) samples for skipjack tuna. Bottom: NMDS plot for multi-elemental micro-chemistry (^{138}Ba, ^{88}Sr, ^{25}Mg) analysis of YoY skipjack tuna. Ellipses represent 95% confidence interval around group centroids. SWIO – south-west Indian Ocean, NWIO – north-west Indian Ocean, NCIO – north-central Indian Ocean, NEIO – north-east Indian Ocean.

Yellowfin tuna (*Thunnus albacares*)

- A total of 868 otoliths were collected for yellowfin tuna from locations within the Indian Ocean and 99 were processed and included in the preliminary microchemistry analysis for population structure (Figure 18);
- The trace elements and stable isotopes which provided maximal discrimination accuracy (55Mn, 88Sr and δ 18O) were selected for analysis;
- Significant differences were found (PERMANOVA, P=0.01) between the two western groups (NWIO and SWIO) and the central and eastern groups (NCIO and NEIO) (Figure 18).

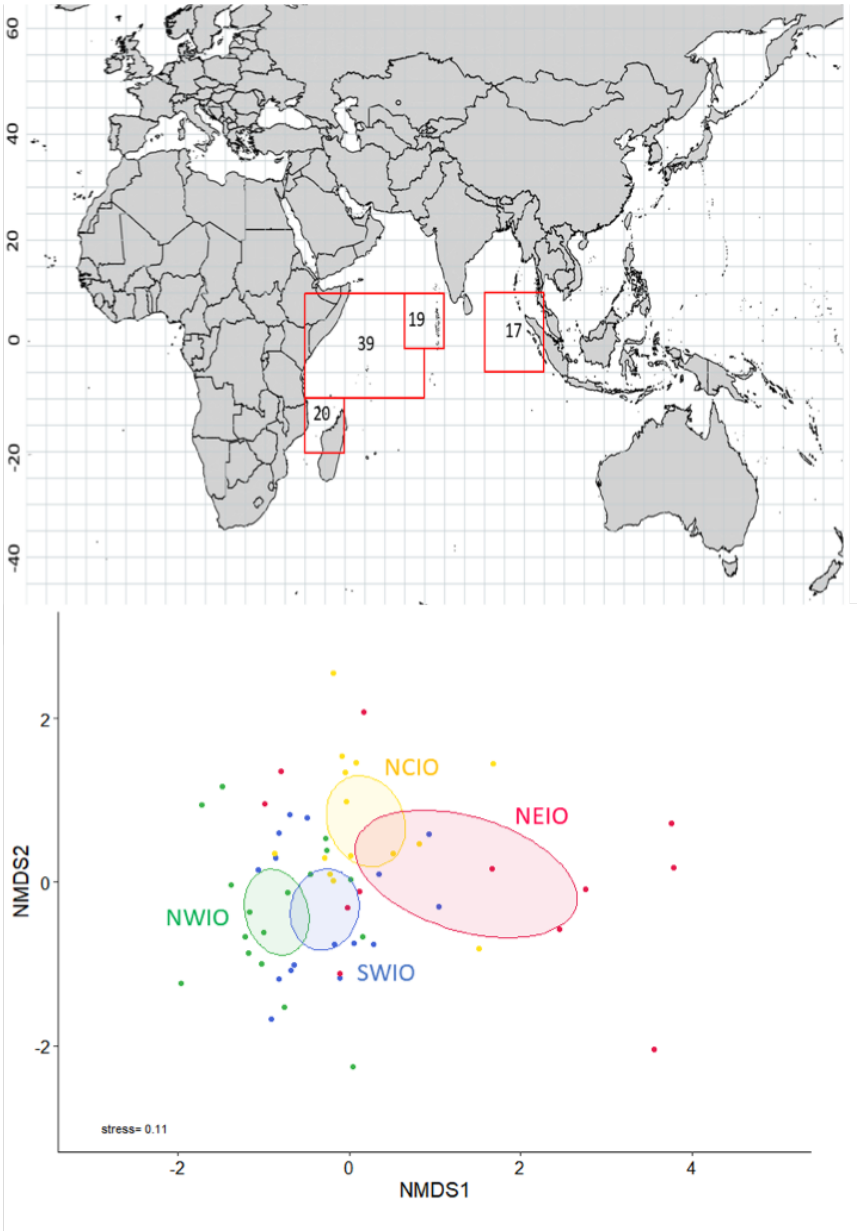

Figure 18. Top: Distribution of samples (N=99) of yellowfin tuna (*Thunnus albacares*) for multi-elemental micro-chemistry analysis. Bottom: NMDS plot for multi-elemental micro-chemistry (55Mn, 88Sr and δ 18O) analysis of YoY yellowfin tuna. Ellipses represent 95% confidence interval around group centroids. SWIO – south-west Indian Ocean, NWIO – north-west Indian Ocean, NCIO – north-central Indian Ocean, NEIO – north-east Indian Ocean.

Bigeye tuna (*Thunnus obesus*)

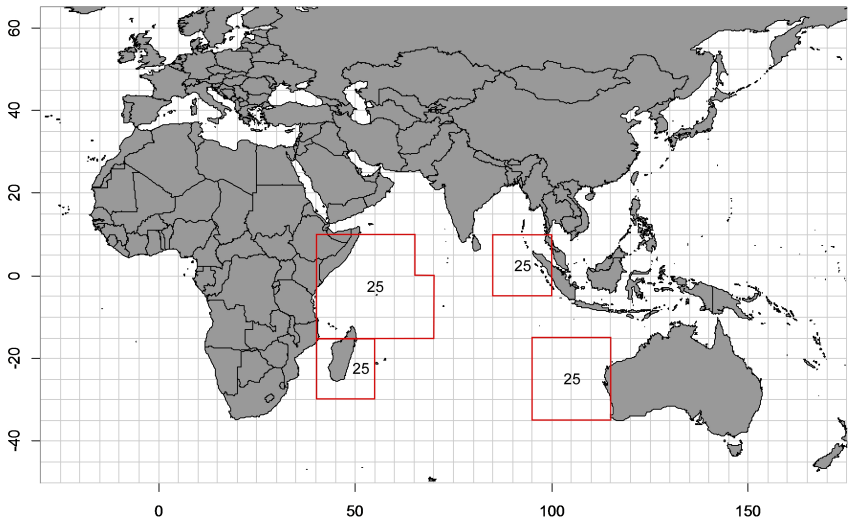

Figure 19. Distribution of samples (N=100) of bigeye tuna (*Thunnus obesus*) for both rounds of sampling for multi-elemental micro-chemistry analysis.

4.4.3 Temperate tunas

Albacore (*Thunnus alalunga*)

- Samples were received from Reunion 2018 and 2019 and from South Africa 2018 (n=49). Samples from Tasmania 2019 (N=20) were received in November 2019 and will be analysed soon (Figure 20);
- PCA analysis were performed using B, Mg, P Cu, Zn, Sr, Ba chemical elements;
- Preliminary results for edge analysis indicate a stable chemical signature for Reunion among years and different from S. Africa samples. PCA analysis for nucleus signature indicate a similar signature for Reunion 2018 and South Africa 2018 (Figure 20).

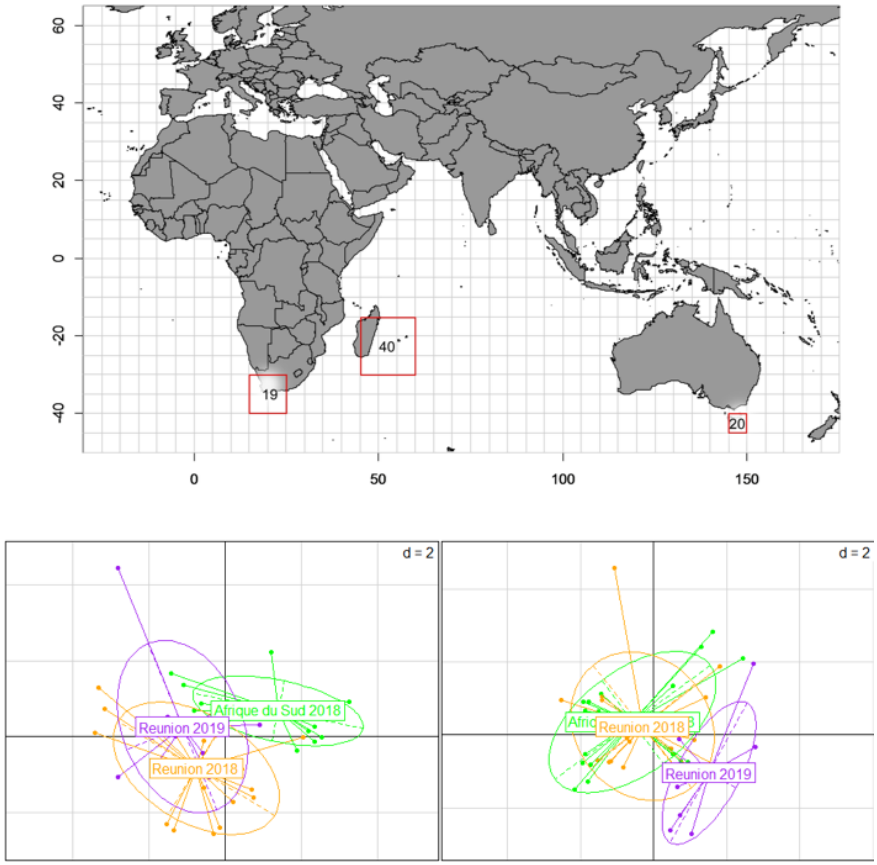

Figure 20. Top: Distribution of samples (N=79) of albacore tuna (*Thunnus alalunga*) for multi-elemental micro-chemistry analysis. Bottom: Results of PCA for elemental micro-chemistry for B, Mg, P Cu, Zn, Sr, Ba from transect from core to margin for albacore tuna from Reunion (2018, 2019) and South Africa (2018 only).

4.4.4 Billfish

Swordfish (*Xiphias gladius*)

- Samples were received from Australia 2018, Reunion 2018, Seychelles 2018 and Reunion 2019 (Figure 21);
- Preliminary results using PCA (B, Mg, P, Zn, Sr, Ba) indicate a large overlap of the edge signature;
- We observed similar nucleus signatures for all fish from Western I. O. (Seychelles and Reunion) and different from those of Eastern I.O. fish (Australia) (Figure 21).

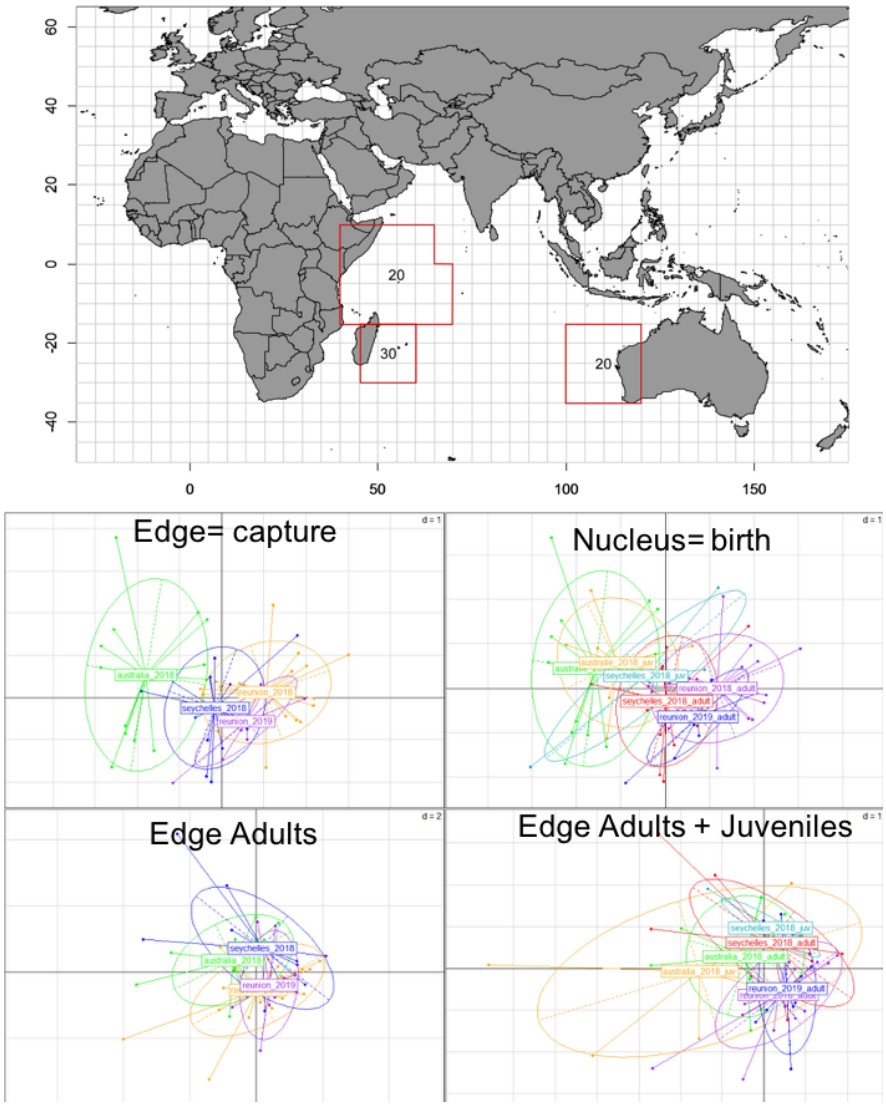

Figure 21. Top: Distribution of samples (N=70) of swordfish (*Xiphias gladius*) for for multi-elemental micro-chemistry analysis. Bottom: Results of PCA for elemental otolith microchemistry for swordfish for B, Mg, P, Zn, Sr, Ba for: Middle – Nucleus and Edge, and Bottom: Edge for adults compared to Edge for juveniles.

5 Summary

The project has now completed two years of sampling, accumulating 5,767 tissue samples and 3010 otoliths. The spatial coverage of sampling had been reasonable for most species, including the three neritic species, the three tropical species, swordfish and blue shark. There are gaps in coverage for albacore in the south and north-east. and greater coverage in these regions and in the far north and north-west would improve coverage of the full distribution for most species. In general, it was difficult to obtain the target sample sizes for the other two billfish. The CITES listing of scalloped hammerhead sharks shortly after the approval of the EoI for the project created administrative issues for the collection and transport of samples during the first round of sampling. Further collection of tissue samples for this species ceased following completion of the first round of sampling. The archived samples, DNA collection, and results of the genetics and otolith microchemistry analyses provide a sound foundation for population structure analyses for the project and for further research and development beyond the current project.

The genetics processing is completed for the samples from both rounds. There are potentially additional samples that could be analysed, as per the intent of the original adaptive sampling design. However, it is unlikely that there will be sufficient time to complete the laboratory and data analysis prior to the 31 March 2020 deadline for the current project. The otolith microchemistry is complete for some species, however, there has only been time to complete the most preliminary analyses. In the case of the neritic species and bigeye tuna, unforeseen delays in the otolith preparation process means that the results for these species are not yet available. They will be completed over the coming months and included in the final report for the project.

The preliminary results from the population structure analyses presented indicate the likely presence of population structure for some species, and particularly for the neritic species. For other species, such as skipjack tuna, blue shark and Indo-Pacific sailfish, the analyses completed to date do not suggest the presence of substantial genetic structure. Similarly, the results of the available microchemistry analyses indicate differences among locations for some species (e.g. yellowfin tuna), but not others (e.g. skipjack tuna).

The project team are seeking initial input from the Scientific Committee on the results of the sampling program presented, as well as on the preliminary population structure analyses. The reporting schedule for the project includes a draft final report at the end of this calendar year (20 December 2019), which will consist of an updated version of this paper, including updated analysis and interpretation. This report will be made available to members of the Scientific Committee and the relevant Working Parties to provide the opportunity for more in depth review and to provide the project team with input prior to a synthesis workshop in February 2020 and the drafting of the final report (due 31 March 2020). We recommend that this process of feedback to the project team be coordinated through the SC and WP chairs.

6 Literature Cited

Altshuler D, Pollara VJ, Cowles CR, van Etten WJ, Baldwin J, et al. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407: 513–516. doi: 10.1038/35035083.

Anonymous (2018). Sampling protocol for population structure study. Population Structure of IOTC species and sharks of interest in the Indian Ocean: Estimation with Next Generation Sequencing Technologies and Otolith Microchemistry. Progress report to the IOTC Secretariat. July 2018.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376. doi: 10.1371/journal.pone.0003376.

Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, *et al.* (2013). Genome-wide association mapping of root traits in a japonica rice panel. *Plos One* 8, e78037.Cruz VMV, Kilian A, Dierig DA. (2013). Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New OilseedCrop Lesquerella and Related Species. PLoS ONE 8(5): e64062. doi:10.1371/journal.pone.0064062

Dammannagoda ST, Hurwood DA, Mather PB. (2008). Evidence for fine geographical scale heterogeneity in gene frequencies in yellowfin tuna (*Thunnus albacares*) from the north Indian Ocean around Sri Lanka. Fish. Res. 90: 147-157

Dammannagoda ST, Hurwood DA, Mather PB. (2011). Genetic analysis reveals two stocks of skipjack tuna (*Katsuwonus pelamis*) in the northwestern Indian Ocean. Can J Fish Aquat Sci. 68(2):210–23

Elshire R, Glaubitz J, Sun Q, Poland J, Kawamoto K, Buckler E, Mitchell S. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. Plos One 6, e19379

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

Foster SD. (2018) stockR: Identifying Stocks in Genetic Data.

Grewe PM, Krueger CC, Aquadro CF, Bermingham E, Kincaid HL, May B. (1993). Mitochondrial DNA variation among lake trout strains stocked into Lake Ontario. Can J Fish aquat Sciences 50: 2397±2403

Kilian A, Wenzl P, Huttner E, Carling J, Xia L. *et al.* (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. In F. Pompanon & A. Bonin (Eds.), Data production and analysis in population genomics (pp. 67–89). Totowa, NJ: Humana Press.

Menezes MR, Kumar G, Kunal SP. (2012). Population genetic structure of skipjack tuna *Katsuwonus pelamis* from the Indian coast using sequence analysis of the mitochondrial DNA D-loop region. J Fish Biol. 80(6):2198–212

Miller A. (2002). Subset Selection in Regression. Chapman & Hall/CRC.

Nugraha B, Novianto D, Barata A. (2011). Keragaman genetik ikan tuna matabesar (*Thunnus obesus*) di Samudera Hindia. [Genetic diversity of bigeye tuna (*Thunnus obesus*) in Indian Ocean]. Ind. Fish. Res. J. Vol.17 (4), 285 - 292 (In Bahasa Indonesia, with English abstract).

Proctor C. H., Lester R. J. G., Clear N. P., Grewe P. M., Moore B. R., Eveson J. P., Lestari P., Wujdi A., Taufik M., Wudianto, Lansdell M. J., Hill P. L., Dietz C., Thompson J. M., Cutmore S. C., Foster S. D, Gosselin T. and Davies C. R. (2019). Population structure of yellowfin tuna (*Thunnus albacares*) and bigeye tuna (*T. obesus*) in the Indonesian region. Final Report as output of ACIAR Project FIS/2009/059. Australian Centre for International Agricultural Research, Canberra. 139 pp. Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N. (2014) Genome-wide delineation of natural variation for pod shatter resistance in Brassica napus. *Plos One* 9, e101673.

Rooker JR, Alvarado Bremer JR, Block BA, Dewar H et al. (2007). Life history and stock structure of Atlantic bluefin tuna (*Thunnus thyunnus*). Rev Fish Sci 15:265–310

Swaraj PK, Kumar G, Menezes MR, Meena RM. (2013). Mitochondrial DNA analysis reveals three stocks of yellowfin tuna *Thunnus albacares* (Bonnaterre, 1788) in Indian waters. Conserv. Genet 14: 205-213.

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

Oceans and Atmosphere
Dr Campbell Davies
+61 3 6232 5222
Campbell.davies@csiro.au
csiro.au