



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: Final Report to IOTC**

Campbell Davies, Francis Marsac, Hilario Murua, Igaratza Fraile, Zulkarnaen Fahmi, Jessica Farley, Peter Grewe, Craig Proctor, Naomi Clear, Paige Eveson, 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, Florian Devloo-Delva, Clément Rougeux, Denham Parker, Natalia Diaz-Arce, Wudianto, Toni Ruchimat, Fayakun Satria, Pratiwi Lestari, Muhammad Taufik, Asep Priatna, Achmad Zamroni.

8 May 2020



Citation. Study of population structure of IOTC species and sharks of interest in the Indian Ocean using genetics and microchemistry: 2020 Final Report to IOTC.

Davies, C.R., Marsac, F., Murua, H., Fraile, I., Fahmi, Z., Farley, J., Grewe, P., Proctor, C., Clear, N., Eveson, P., 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., Devloo-Delva, F., Rougeux, C., Parker, D., Diaz-Arce, N., Wudianto, Ruchimat, T., Satria, F., Lestari, P., Taufik, M., Priatna, A., and Zamroni, A.

#### Copyright

© Commonwealth Scientific and Industrial Research Organisation 2020. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

#### Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

CSIRO is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document please contact [csiroenquiries@csiro.au](mailto:csiroenquiries@csiro.au).

# Contents

Acknowledgments.....	xi
Acronyms	xiii
Executive summary .....	xvi
1 Introduction .....	1
1.1 Background.....	1
1.2 Approach .....	1
2 Sampling design and Biological Sample Collection.....	6
2.1 Sampling design.....	6
2.2 Sampling protocol.....	9
2.3 Sample collection .....	9
2.4 Summary of samples available for analysis.....	10
3 Overview of regional oceanography.....	11
3.1 Annual cycle.....	11
3.2 Time series by sampling area, 2017-2019 .....	15
4 Genetic sequencing and otolith micro-chemistry methods.....	21
4.1 Reduced representation library preparation and sequencing.....	21
4.2 Post-processing of DArT Sequencing data for neritics, albacore tuna, yellowfin tuna, billfish, and blue shark.....	22
4.3 Processing of RAD FASTQ (skipjack) and DArT FASTQ (bigeye) data .....	24
4.4 Analysis of SNP data for population structure .....	25
4.5 Microchemistry methods .....	26
5 Results and discussion .....	30
5.1 Sample collection .....	30
5.2 Status of results and opportunity for further input .....	30
5.3 Presentation of genetic population structure results .....	31
5.4 Neritic tuna .....	33
5.5 Tropical tunas .....	48
5.6 Temperate tunas .....	65
5.7 Billfish .....	74
5.8 Sharks .....	82

6	Summary and conclusions .....	84
6.1	Sampling .....	84
6.2	Population structure.....	85
6.3	Extension and communication of results .....	88
7	Literature Cited .....	89
8	Appendix 1: Detailed species results .....	101
8.1	Neritic Tuna .....	101
8.2	Tropical tuna.....	137
8.3	Temperate tuna.....	189
8.4	Billfish .....	205
8.5	Sharks .....	223
9	Appendix 2: Project outputs .....	227
10	Appendix 3: Starting Values for Genetic Analyses.....	228



# Figures

Figure 1. Illustration of the “population grey zone”. On the left, the relationship between the index of differentiation  $F_{st}$  and the raw number of migrants ( $N_e m$ ) exchanged between two populations, illustrating the very quick drop in  $F_{st}$  values for levels of exchange clearly too small to ensure demographic connectivity. On the right, results of simulations for populations with effective population sizes of 100,000 and exchange of one migrant only per generation, illustrating the very large number of generation necessary to capture with a high level of probability (left axis, red line) the signature of population differentiation based on  $F_{st}$  estimates (right axis, blue line with 95% confidence intervals in dotted lines). Source: Bailleul et al., 2018. 3

Figure 2. 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)..... 6

Figure 3. Surface circulation during the northeast or winter (left) and southwest or summer (right) monsoons, with some choke point transport numbers ( $S_v = 106 \text{ m}^3 \text{ s}^{-1}$ ). Legend : South Equatorial Current (SEC), South Equatorial Counter-Current (SECC), Northeast/Southeast Madagascar Current (NEMC/SEMC), East African Coast Current (EACC), Somali Current (SC), Southern Gyre (SG) and Great Whirl (GW) and associated upwelling wedges, Socotra Eddy (SE), Ras al Hadd Jet (RHJ) and upwelling wedges off Oman, West Indian Coast Current (WICC), Laccadive High and Low (LH and LL), Sri Lanka dome (SD), East Indian Coast Current (EICC), Southwest and Northeast Monsoon Current (SMC and NMC), South Java Current (JC) and Leeuwin Current (LC). Reprinted from Schott & McCreary (2001)..... 12

Figure 4. Sea surface temperature average for January, April, July and October. White contours for isotherms 15°, 20°, 25° and 28°C. Maps generated from outputs of the Global Ocean Physical reanalysis product at 1/12° resolution, EU Copernicus Marine Environment Monitoring Service. .... 13

Figure 5. Surface salinity average for January, April, July and October. The white contour denotes the isohaline 35 g.kg<sup>-1</sup>. Maps generated from outputs of the Global Ocean Physical reanalysis product at 1/12° resolution, EU Copernicus Marine Environment Monitoring Service..... 14

Figure 6. Average dissolved oxygen (DO) concentration at two depth levels, 100 m (left column) and 300 m (right column) for January and July. Maps generated from outputs of the Global Ocean Biogeo-chemistry Hindcast product at 1/4° resolution, EU Copernicus Marine Environment Monitoring Service..... 15

Figure 7a. Monthly time series for (left column) dissolved oxygen at 100 m (oxy100) and 200 m (oxy200) and (right column) surface temperature (sst) and salinity (sss) in the South African sampling sites of the project..... 17

Figure 8b. Same as above, for the two western Indian Ocean sampling sites ..... 17

Figure 9c. Same as above, for the three Northern Indian Ocean sampling sites ..... 18

Figure 10d. Same as above, for the three eastern sampling sites..... 19

Figure 11e. Same as above, for the south-eastern sampling sites ..... 20

Figure 12. Summary of sampling performance for tissue (left) and otolith (right) collection relative to planned sample size. ....	30
Figure 13. 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. These results should be interpreted subjectively with the geographical distribution of the groups. It seems that the information criteria may give a lower bound for the number of genetic groups.....	32
Figure 14. Assignment probabilities for Kawakawa into $K=3$ genetic groups. Different colours represent different groups. Uncertainty is displayed as 'whiteness' -- those fish where the estimated group assignment probabilities is large are faded progressively to white. The bottom plots give the length distributions of the individuals sampled at that location. In this plot, we suggest that $K=3$ is overfitted and that a smaller $K$ is probable. This is due to the fact that the 'blue' genetic group is present in non-trivial proportions throughout many sampling locations. ....	32
Figure 15. Top Left: Number of samples of longtail tuna ( <i>Thunnus tonggol</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups ( $k$ ), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for longtail tuna for 3 genetic groups.....	33
Figure 16. Map showing the number of longtail otoliths analysed for each of three sampling locations, referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI) and Arafura Sea (Arafura), and the size range of fish at each location. ....	34
Figure 17. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the longtail otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. ....	35
Figure 18. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the longtail otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. ....	36
Figure 19. Top Left: Number of samples of kawakawa ( <i>Euthynnus affinis</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups ( $k$ ), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for kawakawa suggesting for 2 genetic groups within the Indian Ocean. ....	38
Figure 20. Map showing the number of kawakawa otoliths analysed for each of four sampling locations, referred to as North-West Indian Ocean (NWI), Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI), and the size range of fish at each location.	39

Figure 21. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the kawakawa otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	41
Figure 22. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the kawakawa otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	42
Figure 23. Top Left: Number of samples of narrow-barred Spanish mackerel sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Middle: results of first round of population structure analysis of DArTSeq using StockR for narrow-barred Spanish mackerel for 4 genetic groups. Bottom: results of second round of population structure analysis of DArTSeq using StockR for each potential cryptic species for 4 genetic groups. ....	43
Figure 24. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for the AFS and WCS subset assuming 2 genetic groups. ....	44
Figure 25. Map showing the number of Spanish mackerel otoliths analysed for each of the sampling locations, referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI), Eastern Central Indian Ocean (ECI) and Arafura Sea (Arafura), and the size range of fish at each location.....	44
Figure 26. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the Spanish mackerel otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	46
Figure 27. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the Spanish mackerel otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	47
Figure 28. Top: Number of samples of skipjack ( <i>Katsuwonus pelamis</i> ) sequenced using RAD-Seq and included in the analysis by sampling region. Bottom: Principal component analysis results using the filtered datasets. Left plots are identical to right plots but removing labels for clarity. ....	48
Figure 29. Map showing the number of skipjack otoliths analysed for each of three sampling locations, referred to as Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI), and the size range of fish at each location. ....	49
Figure 30. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the skipjack otolith core signatures for the 2018 and 2019 samples respectively. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	50

Figure 31. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the skipjack tuna otolith edge signatures. Individuals are coded by their sampling location and year. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	51
Figure 32. Top Left: Number of samples of yellowfin tuna ( <i>Thunnus albacares</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (K). Bottom: Individual bar plot probabilities 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 to a specific K group is below 80%. Fish lengths of individuals are plotted below each bar and sample size at location is in brackets. ....	54
Figure 33. Map showing the number of yellowfin tuna otoliths analysed for each of six sampling locations, referred to as South Africa (SAF), South-West Indian Ocean (SWI) , Western Central Indian Ocean (WCI), Maldives (CIM), North-East Indian Ocean (NEI), and south east Indian Ocean (GAB); and the size range of fish at each location. ....	55
Figure 34. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the yellowfin tuna otolith core signatures for the 2018 and 2019 samples respectively. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	56
Figure 35. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the adult yellowfin tuna otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. SAF= South Africa, SWI= South West Indian Ocean and SEI= South East Indian Ocean). (Note SEI is shown as “GAB” in this figure). ....	57
Figure 36. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the yellowfin tuna otolith edge signatures. Individuals are coded by their sampling location and year. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	58
Figure 37. Top: Distribution of samples of bigeye ( <i>Thunnus obesus</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for bigeye tuna for 3 genetic groups .....	59
Figure 38. Map showing the number of bigeye otoliths analysed for each of four sampling locations, referred to as Western Central Indian Ocean (WCI), North-East Indian Ocean (NEI), South-West Indian Ocean (SWI) and South-East Indian Ocean (SEI); and the size range of fish at each location. ....	60
Figure 39. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the bigeye otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. ....	62

Figure 40. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the bigeye otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. .... 63

Figure 41. Map of albacore samples genotyped with number of individuals for each of the 5 main sampling regions: Northeast Atlantic (Atlantic-N), Southeast Atlantic (Atlantic-SE or South-Africa), Southwest Indian Ocean (IndianOcean-SW), East Indian Ocean (IndianOcean-E or Indonesia), and Southwest Pacific (Pacific-SW)..... 66

Figure 42. Results of population structure analysis of DArTSeq using StockR (A), AssignPop (B), STRUCTURE and  $F_{st}$  (C), and DAPC and Evolutionary history from observed joint allele frequency spectrum (JAFS) presented the fittest model (D) for albacore samples over 20,220 SNPs (neutrals and with outliers) and 224 individuals. Each color represents an area. Bottom right: Information criterion used to assess one likelihood of different numbers of genetic groups (k), lower indicating more likely.  $N_{anc}$ : ancient population size. N: population size.  $m_{12}$ ,  $m_{21}$ : migration rate from the population 1 to the population 2 and vice versa.  $me_{12}$ ,  $me_{21}$ : reduced effective migration rates for loci influenced by selection when semi-permeability was assumed.  $T_{sc}$ : duration of the secondary contact.  $T_s$ : time from the initial split to present (Append Figure 69). .... 67

Figure 43. Map showing the number of albacore otoliths analysed for each of the three sampling locations, referred to as south west Indian Ocean (SWI), South Africa (SAf) and south west Tasman Sea (SWT); and the size range of fish at each location. .... 69

Figure 44. (a) Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the multi-elemental (B, P, Cu, Zn, Sr, Ba) signatures of the otolith cores of the 80 albacore analysed. Individuals are coded by their sampling location -- South Africa north (SA-N); South Africa south (SA-S), South West Indian Ocean (SWI) and Southwest Tasman Sea (SWTS). Their size on the graph is proportional to the quality of their representation in this plane. For the variables, the length of the arrow reflects the % of contribution to the total inertia. (b) Shows the same results as (a) but the colour coding for SWI is now broken down into the two sampling periods, Feb 2018 and May 2018. .... 71

Figure 45. Projection of individuals on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr) signatures of the otolith core of the 80 albacore analysed. Colours on the graph represent the spawning origins (1 to 3) for each fish and symbols represent its final sampling area. .... 72

Figure 46. Top Left: Number of samples of swordfish (*Xiphias gladius*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k). Bottom: 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 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. .... 75

Figure 47. Comparison of eastern Indian Ocean (ECI and SEI) and western Pacific (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). .... 75

Figure 48. Map showing the number of swordfish otoliths analysed for each of the three sampling locations, referred to as south west Indian Ocean (SWI), west central Indian Ocean (WCI) and south east Indian Ocean (SEI); and the size range of fish at each location. ....	76
Figure 49. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the swordfish otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.....	77
Figure 50. Projection of individuals on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr, Ba) signatures of the otolith core of the 70 swordfish analysed. Colours on the graph represent the spawning origins (1 to 3) for each fish and symbols represent its final sampling area.....	78
Figure 51. Top Left: Number of samples of striped marlin ( <i>Tetrapturus audax</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for striped marlin for 2 genetic groups.....	80
Figure 52. Top Left: Number of samples of Indo-Pacific sailfish ( <i>Istiophorus platypterus</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k). ....	81
Figure 53. Top Left: Number of samples of blue shark ( <i>Prionace glauca</i> ) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), indicating in fact no more likely group than a single 1. Bottom: results of population structure analysis of DArTSeq using StockR for blue shark for 2 genetic groups. ....	83

## Tables

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. ....	5
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 .....	7
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. ....	8
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.....	10
Table 5. Summary of sampling areas with boundaries used for the time series .....	16
Table 6. Details of LA-ICP-MS analysis of otolith sections from 8 species conducted by the 3 partners.....	27
Table 7. Papers to be submitted to IOTC meetings .....	31
Table 8. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.....	35
Table 9. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.....	40
Table 10. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.....	45
Table 11. Number, sampling period and estimated ages of fish for each of the three sampling locations WCI, CIM and ECI.....	49
Table 12. Number, sampling period and estimated ages of fish for each of the six sampling locations SAF, SWI, WCI, CIM, NEI and SEI.....	56
Table 13. The sampling period for bigeye otoliths analysed, and the estimated ages for bigeye from each of the sampling locations.....	61
Table 14. Top table: Pairwise $F_{st}$ below diagonal with p-value (<0.05) in bold based on 1,000 bootstrap of albacore samples over 20,220 SNPs (including outliers) and 224 individuals with Arlequin software. Bottom table: Pairwise $F_{st}$ below diagonal with p-value (<0.05) in bold based on 1,000 bootstrap of albacore samples over 20,038 SNPs (excluding 182 outliers) and 224 individuals with Arlequin software.....	68
Table 15. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.....	69
Table 16. Percent of individuals assigned to each spawning origin (SpO 1 to 3) in the albacore analysed (total N of fish tested = 80, FL = 55-116 cm) for each of the 3 sampling locations: South Africa (SA), South-Western (SWI) and Southwest Tasmanian Sea (SWTS).....	72
Table 17. Number, sampling period, size range and estimated ages of fish for each of the three sampling locations SWI, WCI and SEI.....	76
Table 18. Percent of individuals assigned to each spawning origin (SpO 1 to 3) in the sub-adult and adult swordfish analysed (total N of fish tested = 70, FL = 80-226 cm) for each of the 3 areas investigated in the Indian Ocean: South-Est (SEI), South-Western (SWI) and Central-Western (CWI).....	78
Table 19. Pairwise $F_{ST}$ values and their level of significance after correction with q-value (* $p < 0.01$ , ** $p < 0.001$ , *** $p < 0.0001$ ).....	83

# 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 and analysis, 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;
- Malcolm Francis, Charlene de Silva, Department of Agriculture, Forestry and Fisheries, South Africa, and Stewart Norman, CAPFISH, South Africa, Africa for sampling on blue shark and albacore;
- 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, Estelle Crochelet (IRD), Loïc De Foulgoc, CapRun, for sampling on albacore and 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;
- Leonid Danyushevsky and CODES staff at the University of Tasmania for providing technical advice on sample preparation and expertise in LA-ICP-MS analysis and subsequent data processing and evaluation;
- Kyne Krusic-Golub and Admir Sutrovic, Fish Ageing Services, for their expertise and adaptability in preparing otoliths from a suite of species, under time pressure;
- Mohamed Ahusan, Ahmed Haikal and Shafiya Naeem, Marine Research Centre, Ministry of Fisheries, Marine Resources and Agriculture, Maldives;
- Reunimer and Enez (particularly Hubert Chenede and Frédéric Payet) for facilitating access to work plans for sampling in good condition, Reunion Island;
- 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;
- Fook Lee, Universal Seafoods, for allowing us to sample during factory processing;
- Morgan Le Guernic, Rudy Levian, Mathias Hoarau, Samuel Kazambo, Sébastien Laffont, Rudy Levian, Leopold Corbrejaud of the domestic longline fleet for their continued support of the “PSTBS” project;
- Reunimer and Enez (particularly Hubert Chenede and Frédéric Payet) for facilitating access to work plans for sampling in good condition;
- Estelle Crochelet and Loïc Le Foulgoc for their participation in the sampling;
- Shiham Adam, the Ministry of Fisheries, Marine Resources and Agriculture, Maldives;



- Seychelles Fishing Authority;
- Skippers (Morgan Le Guernic, Rudy Levian, Mathias Hoarau, Samuel Kazambo, Sébastien Laffont, Rudy Levian, Leopold Corbrejaud) of the domestic longline fleet, Reunion Island;
- 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;
- Arthur Georges (University of Canberra) for his advices on dartR package and Thierry Gosselin on filter\_rad;
- Ruth Sharples for her contributions to compiling the SOP and preparation and coordination of sampling kits;
- Thor Carter for sampling and otolith preparation;
- 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. Blue shark);
- 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.

# Acronyms

AFS	Arafura Sea
AIC	Akaike Information Criterion
ALB	Albacore
ANOVA	One-way Analysis of Variance
ARLO	Atlantic Ocean
ARA	Arafura Sea
AS	Arabian Sea
AZTI	AZTI Technalia (Spanish technological research center)
BET	Bigeye Tuna
BIC	Bayesian Information Criterion
BoB	Bay of Bengal
BSH	Blue Shark
CFR	Indonesia's Center for Fisheries Research
	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CITES	
CMEMS	Copernicus Marine Environment Monitoring Service
CODES	Centre for Ore Deposits and Earth Sciences University of Tasmania
COM	Spanish Mackerel
CPCs	Cooperating non-Contracting Parties
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CTAB	Cetyl Trimethyl Ammonium Bromide DNA extraction method
DAPC	Discriminant Analysis of Principal Components
DArT	Diversity Arrays Technologies
DArTSeq	Diversity Arrays Technologies Sequencing method
DNA	Deoxyribonucleic Acid
DO	Dissolved Oxygen
EACC	East African Coast Current
ECI	Eastern-Central Indian Ocean
EEZ	Exclusive Economic Zone
EICC	East Indian Coastal Current
FASTQ	Standard format for storing high-throughput sequencing data
FL	Fork Length
$F_{st}$	Fixation index (measure of population differentiation due to genetic structure)
GW	Great Whirl
HSD	Honestly Significant Difference
HWE	Hardy-Weinberg Equilibrium
	Institut français de recherche pour l'exploitation de la mer (French oceanographic institution)
IFREMER	
IO	Indian Ocean
IOTC	Indian Ocean Tuna Commission
	Institut de Recherche pour le Développement (French science and technology research organisation)
IRD	
JC	Java Current
KAW	Kawakawa

KW	Kruskal-Wallis
LC	Leeuwin Current
LH	Laccadive High
LL	Laccadive Low
LJFL	Lower Jaw Fork Length
LOD	Limit of Detection
LOT	Longtail Tuna
MLS	Striped Marlin
NCI/NCIO	North-Central Indian Ocean
NEI/NEIO	North-East Indian Ocean
NGS	Next Generation Sequencing
NMC	Northeast Monsoon Current
NMDS	non-metric multidimensional scaling
NMEC	North East Madagascar Current
N-SAF	North - South Africa
NWI/NWIO	North-West Indian Ocean
O&A	Oceans and Atmosphere (CSIRO Business Unit)
OFL	Orbital Fork Length
OMZ	Oxygen Minimum Zones
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PERMANOVA	Permutational Multivariate Analysis of Variance
PSTBS-IO	Population Structure of Tuna, Billfish and Sharks-Indian Ocean
QC	Quality Control
RAD	Restriction-site Associated DNA Sequencing
RADIATOR	R package for RADseq data exploration, manipulation and visualization
RAD-Seq	Restriction-site Associated DNA Sequencing
RFLPs	Restriction Fragment Length Polymorphisms
RHJ	Ras al Hadd Jet
RITF	Research Institute for Tuna Fisheries
RTTP-IO	Regional Tuna Tagging Programme - Indian Ocean
S-SAF	South - South Africa
SAN	South Africa North (Saldana Bay)
SAS	South Africa South (Hout Bay)
SC	Scientific Committee
SC	Somali Current
SD	Sri Lanka Dome
SE	Socotra Eddy
SEC	South Equatorial Current
SECC	South Equator Counter Current
SEI/SEIO	South East Indian Ocean
SELPAL	Animal tracking project ( <a href="https://www.argos-system.org/the-selpal-project">https://www.argos-system.org/the-selpal-project</a> )
SEMC	South East Madagascar Current
SFA	Indo-Pacific Sailfish
SG	Southern Gyre
SKJ	Skipjack
SMC	Southwest Monsoon Current
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure

SPL	Scalloped Hammerhead Shark
S-SAF	South - South Africa
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SWI/SWIO	South-West Indian Ocean
SWO	Swordfish
SWTS	South West Tasman Sea
WCI/WCIO	Western Central Indian Ocean
WCS	West Coral Sea
WICC	West India Coastal Current
WIO	Western Indian Ocean
YFT	Yellowfin Tuna
YOY	Young-of-the-Year

# Executive summary

In 2017, CSIRO (Australia) in collaboration with AZTI Tecnalia (Spain), IRD (France) and CFR (Indonesia) commenced a 3-year collaborative project on population structure of tuna, billfish and sharks of the Indian Ocean funded by the European Union and the consortium partners. The aim of the project was to describe the population structure and connectivity of priority tuna and tuna-like species within the Indian Ocean (and adjacent Pacific and Atlantic waters as appropriate), as well as blue and scalloped hammerhead sharks. Genetic analyses of new and archived tissue samples was proposed as the primary method, complimented by microchemical analysis of otoliths and shark vertebrae. Unfortunately, the latter (shark vertebrae) were not completed due to logistic and administrative issues associated with their collection. The project also aimed to develop and extend research networks among partners and contribute to technical capacity building in participating coastal states.

Sampling was completed between late 2017 and early 2019 with a total of 5,767 tissue samples and 3,635 otoliths collected or made available to the project from partner archives, representing ~73% and 188% of the total sample sizes planned for collection for tissue and otoliths, respectively. Of these 3,635 tissue samples have been genotyped and 689 processed and analysed for otolith microchemistry following sample selection and quality control protocols.

The final data coverage for each species across their range within the Indian Ocean varied among the study species and between genetics and microchemistry methods. For genetics very good-good sample coverage was achieved for the six neritic and tropical tuna species and swordfish; while the coverage for albacore, the two other billfish species and blue shark limited the power of analyses to examine population structure within the Indian Ocean. The restrictions associated with international transports of samples of scalloped hammerhead following CITES listing precluded useful coverage of this species in the project.

The sample coverage for otolith microchemistry was often less complete for each species than for genetics due to the additional logistic difficulty in obtaining otoliths, relative to tissue samples, particularly in the case of larger, more valuable adults. Good coverage was achieved for kawakawa and Spanish mackerel and the three tropical tunas, whereas the lack of otolith samples from the south-east Indian Ocean for albacore limited the scope of inferences that could be made based on microchemistry for this species.

The project has provided a sound foundation for exploring hypotheses for population structure for the majority of the study species. For some species, such as albacore, improvements in sample coverage, or additional samples from locations with low sample sizes, are required before substantive interpretations and conclusions about population structure across the Indian Ocean can be made. It will be important, therefore, that the samples and data collected through this project are appropriately archived and curated to ensure they are available for use in future studies. The project partners will work with the IOTC secretariat and Scientific Committee to finalise arrangements for archiving, access and management of the samples and the data arising from the

project, so that this foundation data set is available to build the understanding of population structure of these species into the future.

There are a number of important considerations in considering the population structure results of the project. Firstly, while the project team has taken great care in conducting the analysis and interpretation, the detailed results are yet to be peer reviewed outside the project team. The project team will submit working papers to IOTC Working Party meetings through 2020 and, in parallel, prepare and submit scientific papers for the broader peer review literature. Hence, it is important the results presented here be considered preliminary until these standard scientific reviews have been completed. Second, in the case of the population genetic analysis, failure to detect population structure with a particular set of samples and method, does not prove absence of population structure, which may be important from an assessment or monitoring perspective. It is possible that more comprehensive sampling than was possible in this project and the application of new methods may provide additional information on the structure and connectivity of these populations, which is not evident in the current results. In this context, the results of this project should be seen as working hypotheses on population structure in the Indian Ocean and connectivity with adjacent oceans.

With the exception of blue shark, striped marlin and Indo-Pacific sailfish, the results are consistent with the Indian Ocean being considered a closed unit for stock assessment and management purposes. In the case of blue shark there is there is potential for some level of gene flow with the south-west Pacific. In the case of the two billfish species, the sampling coverage was not sufficient to resolve this question unequivocally.

There was genetic differentiation between samples from locations north and south of the equator for the three neritic species, skipjack and yellowfin tuna, but not bigeye and swordfish. This north-south genetic differentiation was not evident for blue shark, where there was reasonable sample coverage, nor for albacore, striped marlin or Indo-Pacific sailfish for which the sample coverage was insufficient.

There was strong genetic differentiation between the north-west Indian Ocean location from other locations for skipjack, longtail and yellowfin tuna and Spanish mackerel. In the case of yellowfin tuna, this NWI “signature” was also evident in samples from other locations north of the equator. Given the strength of this result and the fact it was restricted to one sampling location, confirming the strength of this genetic differentiation, for these and other species not sampled in the current project, should be a high priority for immediate future research.

The strongest evidence for population structure was observed for Spanish mackerel, with four genetic groups, and longtail tuna, with three. In both species the genetic partitioning corresponded closely with the sample locations. The microchemistry results for these species were generally consistent with the pattern observed in the genetics, however, cannot be unambiguously attributed to locations due to differences in the timing of the sample collection and age of the fish in the samples.

The level of genetic differentiation evident for longtail tuna and Spanish mackerel was not apparent for kawakawa. There was evidence of genetic structure between locations north and south of the equator, but not the finer scale structure evident for the other two neritic species. This is despite the fact that the sample coverage for genetics for kawakawa was the best of the neritic species. The results for long-tail tuna and Spanish mackerel is strong evidence for separate stocks within the

sampling range covered by this project. They provide a compelling case for future studies to: i) extend the sampling coverage to the full range of these species within the Indian Ocean; ii) increase sample sizes and temporal coverage (i.e. establish sampling capacity for multi-year program); and, iii) extend to the other identified priority neritic species (bullet and frigate tuna and Indo-Pacific king mackerel) to provide comprehensive description of population structure of important neritic species for stock assessment and management purposes.

The project has delivered a substantial contribution to the information base on population structure of neritic and pelagic tunas, billfish and blue shark in the Indian Ocean. It has also engaged a wide range of researchers, fisheries scientists (and fishing community), beyond the project team, in the design and implementation of the project. The next stage is to engage more directly with the wider IOTC scientific community on the interpretation of the results for population structure and their implications for how research, monitoring, stock assessment and management should be conducted in the future. The project team and collaborators will submit working papers for consideration at the appropriate working party meetings through 2020, with synthesis papers presented to the Scientific Committee at the end of 2020. In parallel to this engagement with the working parties, peer review papers for the wider scientific literature will be prepared and published, consistent with the publication protocols for the project.

# 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-lateral cooperation 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 integrated population modelling techniques in recent years (yellowfin, skipjack, bigeye and albacore tunas), while only half of the neritic tuna species have been formally assessed. All assessments to date have assumed a single panmictic spawning population within the Indian Ocean. 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. There also had been studies using a variety of genetic approaches indicating that there may be 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 spatially structured tuna assessments.

If the scientific stock assessment advice is based on invalid assumptions about stock structure and connectivity, there is the potential that management may not achieve the 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/components of the stock.

## 1.2 Approach

### 1.2.1 Population genetics

There have been high expectations of the potential for population genetics and its application to wildlife and conservation management (Frankham, 1995), including fisheries (Carvalho & Hauser, 1994; Ward, 2000), since advances in molecular biology in the early 1990s allowed empirical studies of wild populations. The ultimate aim of stocks identification is to define biological entities and the degree of connectivity between them and, in doing so, improve the understanding of population(s) inter(in)dependency and changes in the harvested resource (Abaunza et al., 2008) or conservation of threatened species. Population genetics was expected to provide information on stock discrimination, population expansions, bottlenecks or gene flow which would allow assessment and



management units to be defined and estimation of the impact of population interactions with environmental factors and/or human activities, such as fishing.

The first two decades of use of the population genetics toolbox (markers and theoretical models) has led to several significant achievements. In fact, genetic research at the population level, aimed at the identification of regional/stock origin, has demonstrated that many marine organisms are separated into more or less genetically distinct populations, which allows for genetic traceability (for a review, see Reiss, Hoarau, Dickey-Collas, & Wolff, 2009). Nevertheless, the use of population genetics as an indirect method to track patterns of migration and improve conservation and/or management was based on the expectation they would provide data reflecting demographic independence or inter-dependence of populations, or stocks, of interest. Unfortunately, several obstacles emerged that led to many inconclusive results, as the success of early studies depended strongly on the life history traits of the species being studied and suitability of the approaches being used. Among the life history traits that pose challenges for population genetics are those species characterised by large population sizes and high dispersal potential and/or migratory movements (Hedgcock, Barber, & Edmands, 2007; Waples, 1998), which are common for many of the species targeted by large commercial fisheries, such as tuna, billfish and sharks.

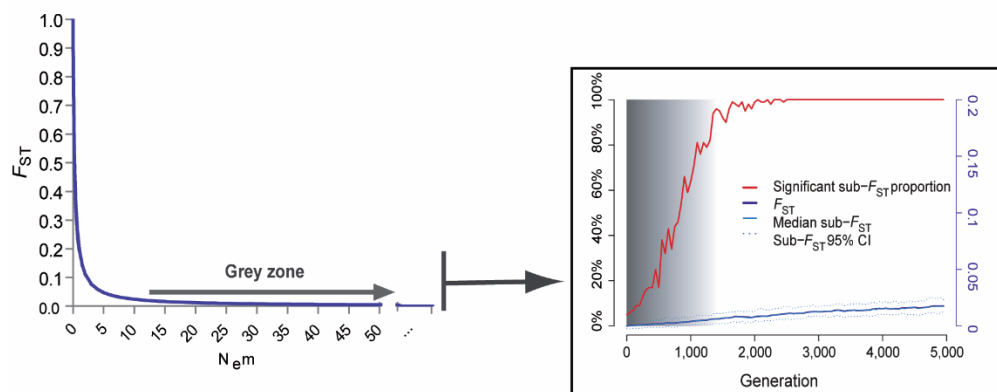
Management and conservation require the knowledge of contemporary connectivity between populations, which may occur via migration or larval dispersal. Classical genetic markers and models, however, reflect the integration of connectivity over many generations. In addition, the level of demographic dependence between populations depends mostly on the *rate* of migration  $m$ , whereas indices of population differentiation, such as  $F_{st}$ , result from the raw *number* of migrants exchanged across generations ( $N_e m$ , with  $N_e$  the effective population size). In other words, under the classical hypothesis underlying population genetic models<sup>1</sup>, the detection of genetic differentiation *does* reveal demographic independence of several populations, but the opposite is not true. That is, if a study does not detect genetic differentiation this does not necessarily equate to there being no population structure in reality. In fact, the lack of detection of differentiation can be the result of a large number of demographic scenarios (Lowe & Allendorf, 2010) spanning from the existence of a single demographic entity to that of several completely independent, yet cryptic, populations (Waples & Gaggiotti, 2006).

This is particularly relevant in the context of the current study. Most studies of large migratory species using classical genetic markers showed no population differentiation, and when results did demonstrate differentiation, they were often extremely weak (Allendorf et al., 2012; Waples 1998). An emblematic example of such situation is the blue shark, the most widely distributed vertebrate species with a global distribution except in polar regions (Compagno, 1984; Nakano & Seki, 2003). The first study aimed at unravelling genetic structure from the Mediterranean-Atlantic to the Pacific, based on a small number of microsatellites and one mitochondrial marker (*cytb*) showed mostly non-significant pattern of differentiation (Bailleul et al., 2018), with the exception of faint hints of differentiation for Mediterranean samples. Although the power of the set of markers and samples in hand was supposed to be high enough to detect even moderate levels of differentiation, this result raised the general question as to the ability of these methods to detect differentiation in large, widespread, highly migratory populations, such as those characterizing many pelagic and

---

<sup>1</sup> No selection and migration-drift equilibrium - i.e. no large stochastic changes in the size of population and pattern of migrant exchange during the last generations.

demersal fishes, depending on the time elapsed since the lowering or cease of migrant exchange among regions of their distribution. Simulations were performed to clarify expectations as to the evolution of  $F_{st}$  values under different scenarios of population size ( $N_e$  spanning from 10.000 to 1 million) and levels of migration ( $N_e m$  spanning from 0.1 to 10 per generation). The results of the simulations revealed what was called a “grey zone effect of population differentiation” (Bailleul et al., 2018), as an analogy to the grey zone of species differentiation proposed by de Queiroz (De Queiroz, 2007) to describe the time during which lineage sorting remains incomplete and the diagnostic of species is difficult to specify on the basis of the genetic data alone.



**Figure 1.** Illustration of the “population grey zone”. On the left, the relationship between the index of differentiation  $F_{st}$  and the raw number of migrants ( $N_e m$ ) exchanged between two populations, illustrating the very quick drop in  $F_{st}$  values for levels of exchange clearly too small to ensure demographic connectivity. On the right, results of simulations for populations with effective population sizes of 100,000 and exchange of one migrant only per generation, illustrating the very large number of generation necessary to capture with a high level of probability (left axis, red line) the signature of population differentiation based on  $F_{st}$  estimates (right axis, blue line with 95% confidence intervals in dotted lines). Source: Bailleul et al., 2018.

The simulations clearly illustrated (Figure 1) the large number of generations (200 to 1600 depending on  $N_e$  and  $m$ ) following the separation of populations necessary to allow detection of genetic differentiation with a probability of more than 0.95. Similar results were obtained for the detection of population differentiation following even drastic bottlenecks (160 to 2200 generations after a reduction of 99 or 90% of the population size).

These simulations indicate clearly one of the main reasons why classical population genetics did not fulfil all the initial promises when it came to applications to marine species with high dispersal potential at larval and/or adult stages (Hedgecock et al., 2007; Waples, 1998). The development of Next Generation Sequencing has demonstrated potential for improving on these classical population genetics approaches. The general availability and relatively low cost of high density genome sequencing is expected to provide finer grain information on patterns of genetic differentiation (but see Waples, 2015) and allow improved inference of parentage or kinship through coalescent analyses and to expand analyses based on linkage disequilibrium (Hellberg, 2009). Large genome scan analysis may thus help resolve the different demographic scenarios, through increase power allowing  $F_{st} < 0.002$  to be detected, and the access to many outlier loci that have proven to help detecting those levels of genetic differentiation that, although weak, reveal

strong demographic independence (Bierne, Welch, Loire, Bonhomme, & David, 2011; Gagnaire et al., 2015).

### **1.2.2 Otolith microchemistry**

Genetic markers have been widely used with success in identifying population structure of marine fish (Gagnaire et al. 2015, Reiss et al. 2009). 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 et al. 2019, Walther et al. 2019).

Both environmental and physiological factors influence the elemental composition of otoliths. Hence, even in the relatively homogenous environment some pelagic species inhabit, otolith microchemistry can discriminate separated groups of fish within a species, and previous studies have used otolith microchemistry to determine natal origin and population connectivity of some tuna and billfish species (Artetxe-Arrate et al. 2019, Proctor et al. 2019, Rooker et al. 2007, 2016).

The principal method proposed for identifying stock structure in this study will be Genotyping-By-Sequencing: that is, using frequency distributions of SNPs obtained by sequencing short segments of DNA. For a subset of priority species, where genetic analyses do not indicate population structure, analysis of trace element concentrations in otoliths will be used. So, the aim of the otolith microchemistry component of this study is to provide an alternative indicator of stock structure and/or finer scale differences that may not be apparent from the genetics methods.

Otoliths of young-of-the-year and adults captured in separate regions, including eastern Indian Ocean, central Indian Ocean, western Indian Ocean and the south-western limit of the Indian and Atlantic Oceans, were collected. Cross-sections of the otoliths were analysed for trace elements, using laser ablation-ICPMS, thus identifying an otolith elemental fingerprint for each fish. The areas of particular interest in the otoliths are the core, 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, deposited at the time just before the fish were captured and reflecting the water mass in which they were caught.

### **1.2.3 Multiple methods**

Abaunza et al. (Abaunza, Murta, & Stransky, 2014) proposed ways to integrate information from different methods within a single study even if we often need to integrate information from across many disparate studies. Presently, uncertainties remain as to the most appropriate way to standardize different types of data (e.g., genotypes, morphological traits, chemical signatures, parasitic fauna, etc.) and analyze such multivariate matrices. In addition to recommendations for definition of management units, the interdisciplinary analysis can also identify research recommendations, including refinement of fishery and resource monitoring approaches and the optimal sampling design for confirmatory analysis and possibly stock composition analysis for mixed-stock situations (Cadurin, Kerr, & Mariani, 2014).

The difficulties inherent to the study of connectivity have prompted researchers to use several approaches and types of data depending on distinct underlying assumptions and models for its

estimation. Genetic methods typically estimate rates of gene flow (Hellberg, Burton, Neigel, & Palumbi, 2002) while micro-chemical fingerprinting and stable isotopes are used to assign individuals to source populations or feeding areas (Thorrold, Zacherl, & Levin, 2007). It is also possible to make inferences about connectivity using biophysical models of ocean circulation and larval transport (Werner, Cowen, & Paris, 2007). Despite these advances, difficulties persist and it has become clear that we need to develop statistical methods that can combine different sources of information in a single analysis (Thorrold et al., 2007).

Next generation sequencing methods (DArTSeq and RAD-Seq) were used as the primary method complemented with otolith and vertebrae microchemistry. The distribution of species and methods among partners is provided in Table 1.

**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 Figure 2. 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 timing and logistics of sampling, distribution of samples and analyses over such a large geographic range and for the number of species (see results). The sampling regions for all species are shown in Figure 2. The subset of locations for each species and method (genetics/microchemistry) are provided in the summary results sections (section 5).

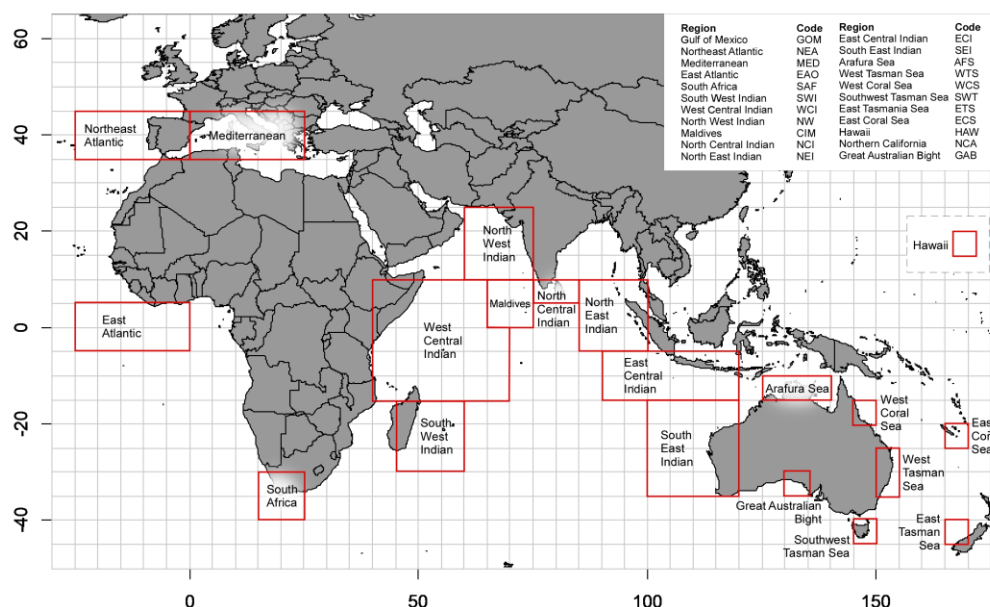


Figure 2. Distribution of sampling locations for both rounds of sampling across Indian Ocean and outlier locations in the Pacific and Atlantic Oceans. Note: this includes locations for active sampling as well as locations for samples provided from earlier studies (see text for details).

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 that were potentially 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 than larger, older individuals.

The size ranges targeted for each species are provided in Table 2. The minimum number of samples to be collected for each species 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 catch occasion/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 (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
- CFR: North-east, East-Central

Dedicated port sampling exercises were conducted in:

- Australia – CSIRO: Mooloolaba, Perth, Great Australian Bight, Tasmania, Northern Territory, Queensland.
- 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 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, Australia, the tissues were sent to CSIRO and the extraction, sequencing and data management handled by CSIRO at their Marine Laboratories in Hobart. The final sequencing data from DArT was then provided to IRD (albacore, blue shark) and AZTI (bigeye tuna) for analysis, following 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 collections. These included:

- Blue shark samples from IRD and IFREMER, 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), CFR, 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 and micro-chemistry is shown for each species in the results section for the respective species.

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. 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 in order to complete the sampling within the project schedule. Hence, the extent to which the final data set will allow for temporal variation to be examined explicitly varies 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	3610	3010	689

## 3 Overview of regional oceanography

The peculiarity of the Indian Ocean is to be landlocked in its northern part. This induces a wind forcing pattern over the ocean basin which is unlike the pattern observed in other oceans in their equatorial region. The wind annual cycle is characterized by a seasonal reversal of the surface winds (monsoons) over the northern hemisphere and the low latitudes of the southern hemisphere (Ramage, 1969) whereas the southeast trade wind regime prevails throughout the year south of 10°S from Madagascar to Indonesia, and south of 20°S in the Mozambique Channel. During the winter monsoon (December to March), winds blow from the northeast (turning to northwest south of the Equator). During the summer monsoon (June to September) winds blow from the southeast (trade-winds northward extension) turning to the southwest north of the Equator). April-May and October-November are the inter-monsoons with eastward winds blowing over the Equator.

In this section, we describe the annual cycle for several abiotic factors characterizing the habitat of tunas and billfishes by means of maps and time series for each of the sampled areas around the Indian Ocean.

### 3.1 Annual cycle

The data used to compute the maps presented in this section are the reanalyses produced by the European Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu/services-portfolio/access-to-products/>). Specifically, two products were used:

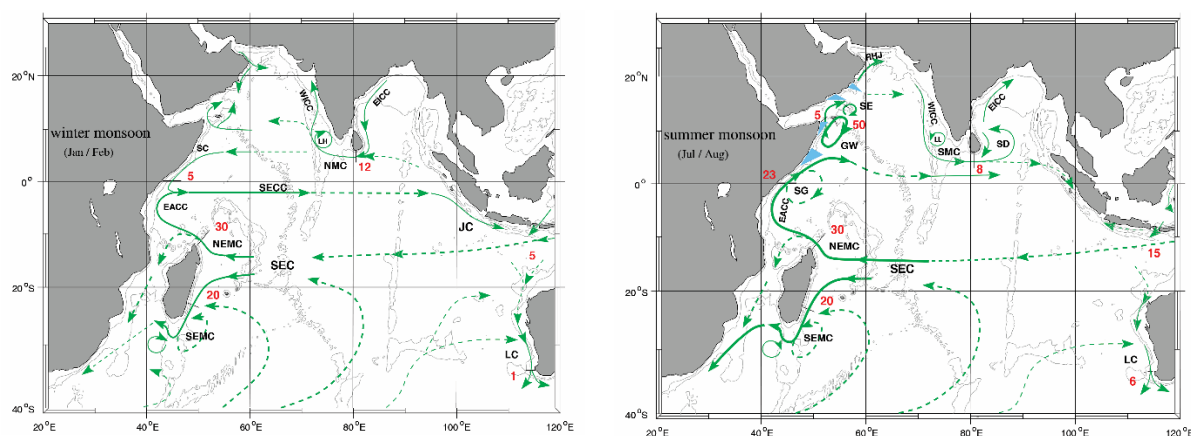
- the GLORYS 12V1 product, a global ocean eddy resolving (1/12° horizontal resolution –about 8 km- and 50 vertical levels) reanalysis covering the altimetry era 1993-2018 (Product identifier: GLOBAL\_REANALYSIS\_PHY\_001\_030). Monthly sea surface temperature and surface salinity were supplied by this product.
- the biogeochemical hindcast for global ocean (1/4° horizontal resolution – about 25 km- and 75 vertical levels) for 1993-2018 (Product identifier: GLOBAL\_REANALYSIS\_BIO\_001\_029). The dissolved oxygen concentration at 100 m and 300 m were supplied by this product.

We computed 20-year monthly climatological means over the period 1995-2014. Four months were selected to represent the seasonal variability driven by the monsoons for temperature and salinity. These were January, April, July and October to illustrate the winter monsoon, first inter-monsoon, summer monsoon and second inter-monsoon, respectively. Dissolved oxygen maps were only for January and July due to the lesser signal at depth. Maps were generated using Golden Software Surfer© v.17.

#### 3.1.1 Surface circulation

The surface circulation responds to the annual wind pattern (Figure 3). During the winter monsoon, a prominent feature south of the Equator is the development of a counter-current (SECC) flowing eastward supplied from the encounter at 2°S-4°S of the northward-flowing East African Coast

Current (EACC) with the southward-flowing Somali Current (SC). Along Indonesia, the Java Current (JC) extends the SECC flow towards the southeast. North of the Equator, the Northeast Monsoon Current (NMC) supplies the Arabian Sea (AS) gyring circulation through the West India Coastal Current (WICC) and the SC. During the summer monsoon (Figure 3), the SEC and EACC supply the then northward-flowing SC. The offshore-induced Ekman transport creates a coastal upwelling. South of the equator, the water recirculates in the Southern Gyre (SG) and north of the equator, in the Great Whirl (GW). A third anticyclonic cell can be observed off Socotra (SE). An anticyclonic circulation takes place in the AS, forming a coastal upwelling off Oman, deepening the mixed layer in central part of the AS, and reversing the WICC then flowing southward. The Southwest monsoon current (SMC) bifurcates into the Bay of Bengal (BoB) to supply the East Indian coastal current (EICC). The northward flow east of Sri Lanka forms a cyclonic dome shoaling the mixed layer (SD). Along Java, the reversed westward flow generates a coastal upwelling in the middle of the year.



### 3.1.2 Seasonal pattern of the sea surface temperature (SST)

In January, the winter cooling caused by southerly continental winds spreads from the northwestern Arabian Sea (23.5°C) to the Somali basin (26.5°C) whereas the SST in the rest of the equatorial region is above 28°C. Another low in SST in the WIO, west of 60°E, occurs in July when the summer monsoon is well established, due to coastal upwelling (Somali and Oman) with SST below 22°C. In contrast the central and eastern part of the equatorial Indian Ocean form a warm pool at all seasons.

The Mozambique Channel, from 12°S to 22°S, is also a warm area with low seasonal variability. The subtropical front (15 to 20°C) oscillates seasonally between 30°S-45°S in summer and 30-40°S in winter.

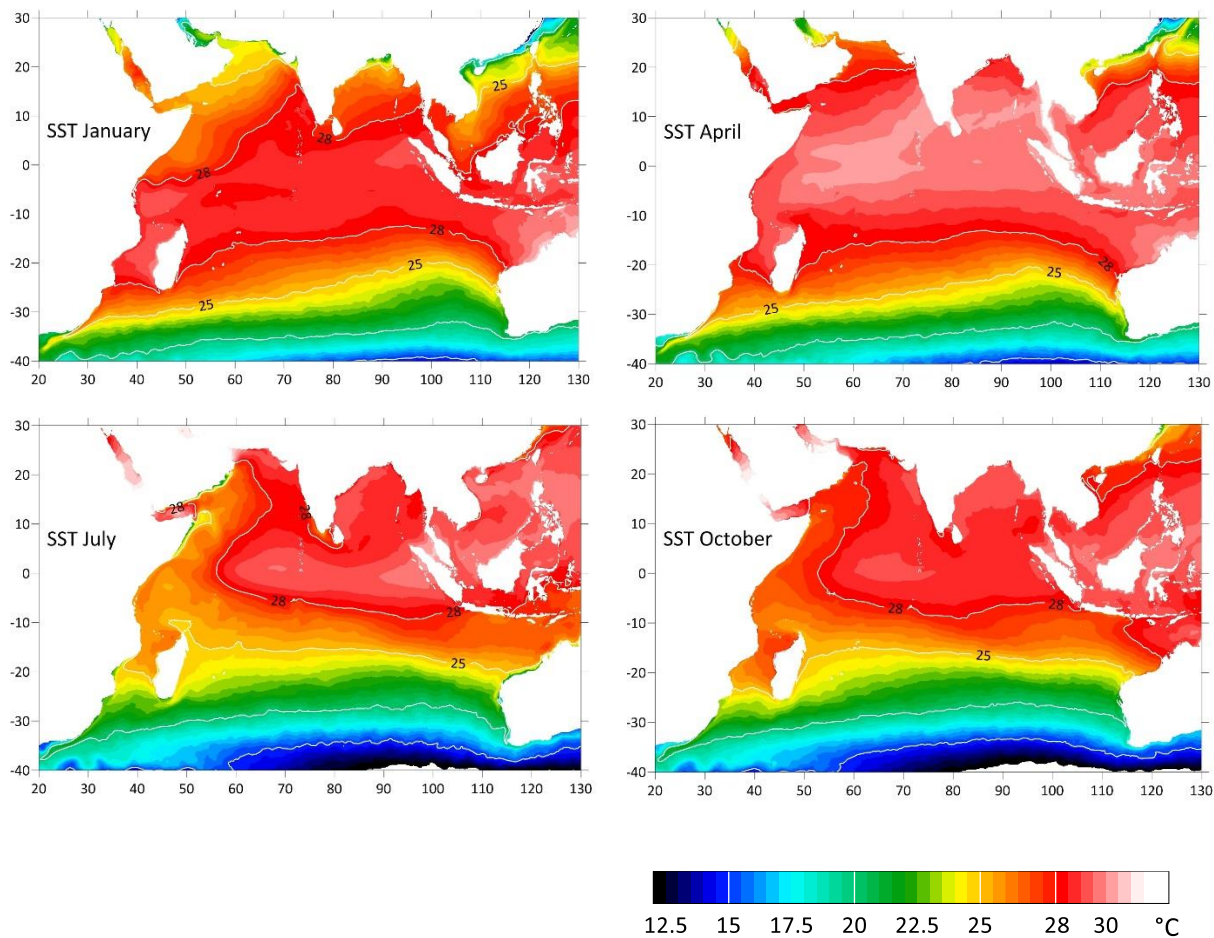


Figure 4. Sea surface temperature average for January, April, July and October. White contours for isotherms 15°, 20°, 25° and 28°C. Maps generated from outputs of the Global Ocean Physical reanalysis product at 1/12° resolution, EU Copernicus Marine Environment Monitoring Service.

### 3.1.3 Seasonal pattern of the surface salinity

There is a large contrast in surface salinity between the northwestern part where saline content is above 36.5 g.kg<sup>-1</sup> (Arabian Sea water) and the eastern Bay of Bengal (BoB) where salinity is below 33 g.kg<sup>-1</sup>. The AS is under the influence of the Persian Gulf water mass whereas the BoB low salinity is driven by large river runoff and excess precipitation. Such main contrast prevails all year round. The latitudinal range 0°-20°S is under the influence of the South Equatorial current flowing relatively low saline waters (34 to 34.5 g.kg<sup>-1</sup>) taking their origin in the Indonesian throughflow (Australian Mediterranean water). Its maximum westward extension occurs during the April-May inter-monsoon. The core of the subtropical gyre along 30°S, east of 70°E, contains the highest saline waters in the Southern hemisphere, as a consequence of evaporation exceeding precipitation (subtropical surface water). This feature is fully developed during the first semester and shrinks during the second semester (Figure 5).

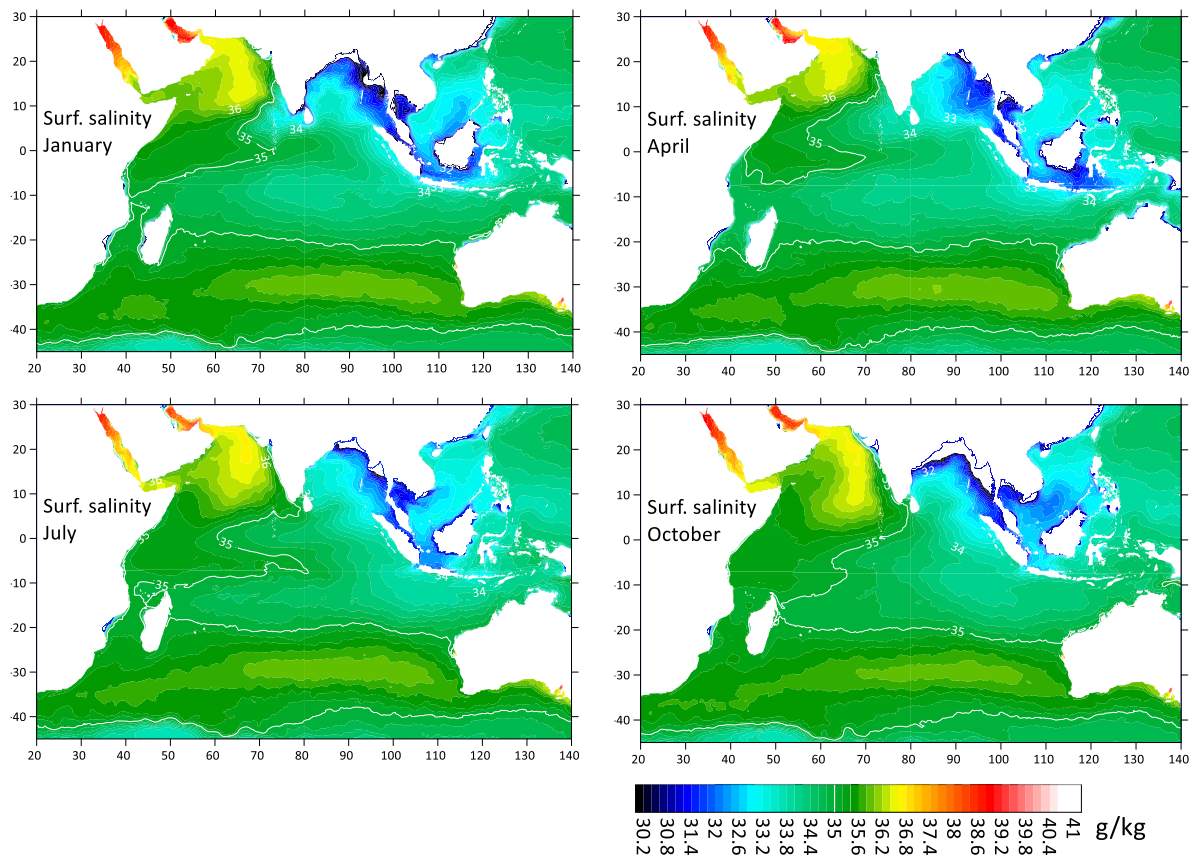


Figure 5. Surface salinity average for January, April, July and October. The white contour denotes the isohaline 35 g.kg<sup>-1</sup>. Maps generated from outputs of the Global Ocean Physical reanalysis product at 1/12° resolution, EU Copernicus Marine Environment Monitoring Service.

### 3.1.4 Dissolved oxygen pattern at depth

The Northern Indian Ocean, AS and BoB combined, form the second largest oxygen minimum zones (OMZ) in the world (after the East Pacific) but the thickest among the two other OMZs (Banse et al., 2013). The oxygen depletion (< 1 ml/l) is caused by remineralization of sinking biogenic particles (as a result of high biological productivity in the surface layer) and insufficient ventilation of the subsurface layers by the northward spread of the Southern hemisphere oxygenated waters (McCreary et al, 2013). OMZs create hypoxic stress for many marine organisms, and are not inhabited by active pelagic predators such as tunas and billfishes. In our maps (Figure 6) we represent only dissolved oxygen concentrations at 100 m and 300 m depth. At 100m, during January (winter monsoon) the less-oxygenated waters are distributed along the coast of AS and BoB, between 0°-5°N where BoB waters are advected by the NMC, and on the Seychelles-Chagos thermocline ridge. In July (summer monsoon), the only difference is the formation of a relative maximum in the south AS and an increase of the oxygen depletion east of Sri Lanka and south BoB. At 300 m, the oxygen concentration does not vary seasonally and reach anoxic levels north of 10°N. At all seasons and in the upper 1000m of the ocean (Boyer et al. 2013), the latitude 15°S marks a frontier between highly-oxygenated waters in the southern side (> 4 ml.l<sup>-1</sup>) and oxygen-depleted waters northwards (< 3 ml.l<sup>-1</sup>).



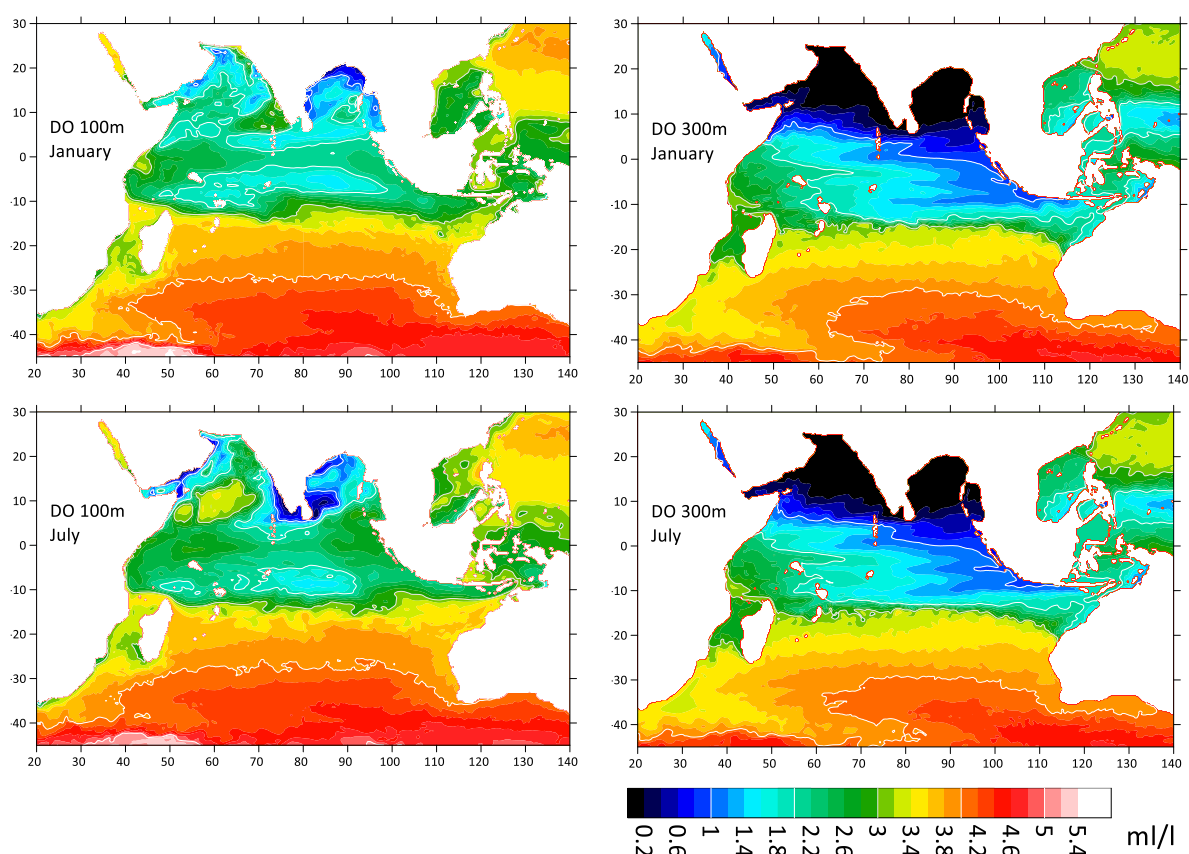


Figure 6. Average dissolved oxygen (DO) concentration at two depth levels, 100 m (left column) and 300 m (right column) for January and July. Maps generated from outputs of the Global Ocean Biogeo-chemistry Hindcast product at  $1/4^\circ$  resolution, EU Copernicus Marine Environment Monitoring Service.

## 3.2 Time series by sampling area, 2017-2019

Time series of parameters describing tuna and billfish habitats and supporting the otolith microchemistry analyses are presented for the sampling period of the project. As the MCEMS reanalysis products used for average maps were only available until 2018, we used different products to compute time series for 2017-2019. In detail, the two products are:

- for temperature and salinity, the operational Mercator Ocean global ocean analysis and forecast system at  $1/12^\circ$  (product identifier: GLOBAL\_ANALYSIS\_FORECAST\_PHY\_001\_024)
- for dissolved oxygen, the operational Mercator Ocean biogeochemical global ocean analysis and forecast system at  $1/4^\circ$  (product identifier: GLOBAL\_ANALYSIS\_FORECAST\_BIO\_001\_028)

Differences exist in the quantities calculated by the reanalyses and by the operational models. However, they depict the same seasonal pattern. The plots are the monthly values for surface temperature, surface salinity and dissolved oxygen concentrations at 100 m and 200 m.

### 3.2.1 Spatial stratification

Plots are provided for 12 boxes representing the different sampling sites of the project (Table 5). They are distributed from the southwestern Indian Ocean (start in South Africa) to the southeastern IO (end in Great Australian Bight). The South African sampling site is split in two sub -ones, north (N-SAF) and south (S-SAF).

Table 5. Summary of sampling areas with boundaries used for the time series

Area code	Area name	North and South limits in latitude	East and West limits in longitude
N-SAF	South Africa (north)	32°S – 33.5°S	16°E – 18°E
S-SAF	South Africa (south)	33.5°S – 35°S	16°E – 19°E
SWI	South West IO (Reunion)	18°S – 26°S	48°E – 60°E
WCI	West Central IO (Seychelles)	0° - 10°S	45°E – 60°E
NWI	North-West IO (Pakistan)	25°N – 20°N	60°E – 70°E
MAL	Maldives	10°N – 0°	70°E – 75°E
NCI	North Central IO (Sri Lanka)	10°N – 5°N	75°E – 85°E
NEI	North East IO (Sumatra)	10°N – 5°S	90°E – 100°E
ECI	East Central IO (Java)	5°S – 15°S	90°E – 120°E
ARA	Arafura Sea (Northern Australia)	10°S – 15°S	125°E – 140°E
SEI	South East IO (Western Australia)	20°S – 35°S	105°E – 120°E
GAB	Great Australian Bight (Southern Aust)	30°S – 35°S	130°E – 135°E

The inter-annual comparison of environmental conditions as defined by oxygen at depth and surface temperature and salinity was tested in each area by a non-parametric test (Kruskal-Wallis, KW). The tested groups were the annual distribution of values for each parameter. The test indicates whether or not the three years when sampling took place display comparable values and distributions. The test is not comparing one specific year to another. Results were visualized by box and whisker plots (not shown here) to check the anomalous years.

### 3.2.2 Analysis by sampling site

In the South African sites (Figure 7a), the seasonal cycle for all factors were in synchrony in both north and south sub-areas. The northern area showed reduced dissolved oxygen concentrations relative to the south (a difference of ~0.15 ml/l), in relation to the coastal Benguela upwelling (Pitcher et al., 2014). The two sub-areas are under the influence of the warm and saline water mass flowing from the Indian Ocean through the Agulhas Current (Lutjeharms, 2006) and their extension into the Atlantic Ocean. This influence is stronger in the south sub-area relative to the north, because of its closer connection with the Agulhas Current. The surface flow is increased during austral summer where both temperature and salinity reach their maximum values. In South Africa, sampling took place only in April 2018, when quantities were 35.31 g.kg<sup>-1</sup>, 18.36°C and 3.53 ml.l<sup>-1</sup> for salinity, temperature and dissolved oxygen respectively in the north-sub-area, and 35.34 g.kg<sup>-1</sup>, 18.69°C and 3.61 ml.l<sup>-1</sup> in the south-sub-area. Over the three years, **oxygen at 200 m was the only parameter showing significantly different values** both in the north (KW=14.55, p<0.001) and in the south (KW=8.41, p< 0.05) sub-areas.

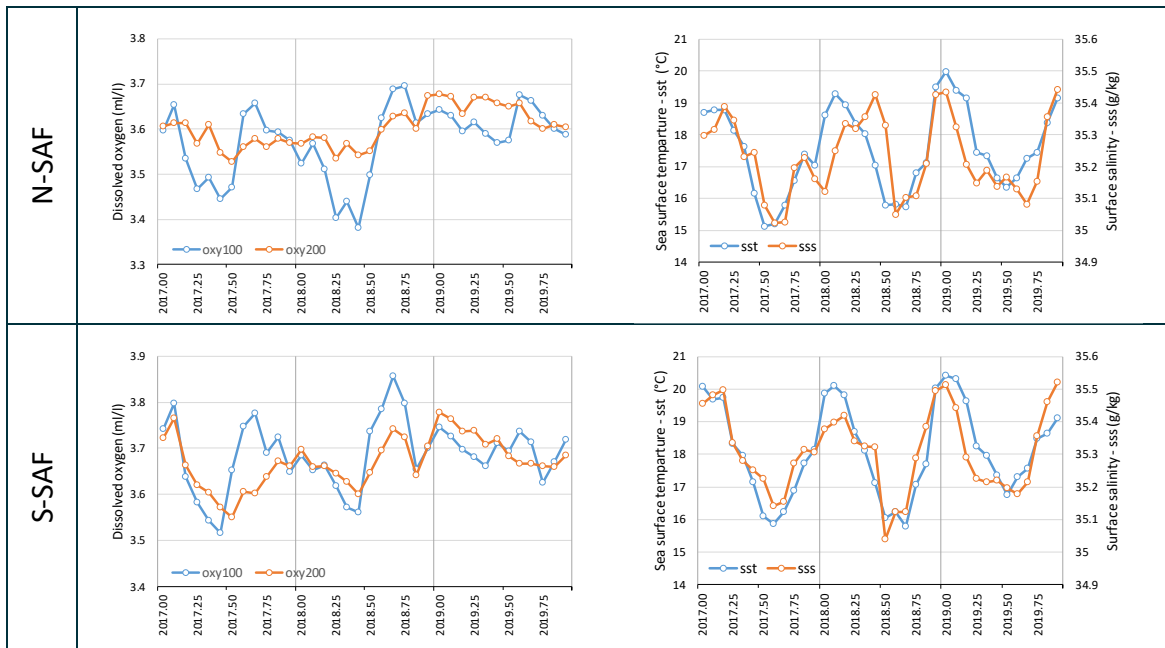


Figure 7a. Monthly time series for (left column) dissolved oxygen at 100 m (oxy100) and 200 m (oxy200) and (right column) surface temperature (sst) and salinity (sss) in the South African sampling sites of the project

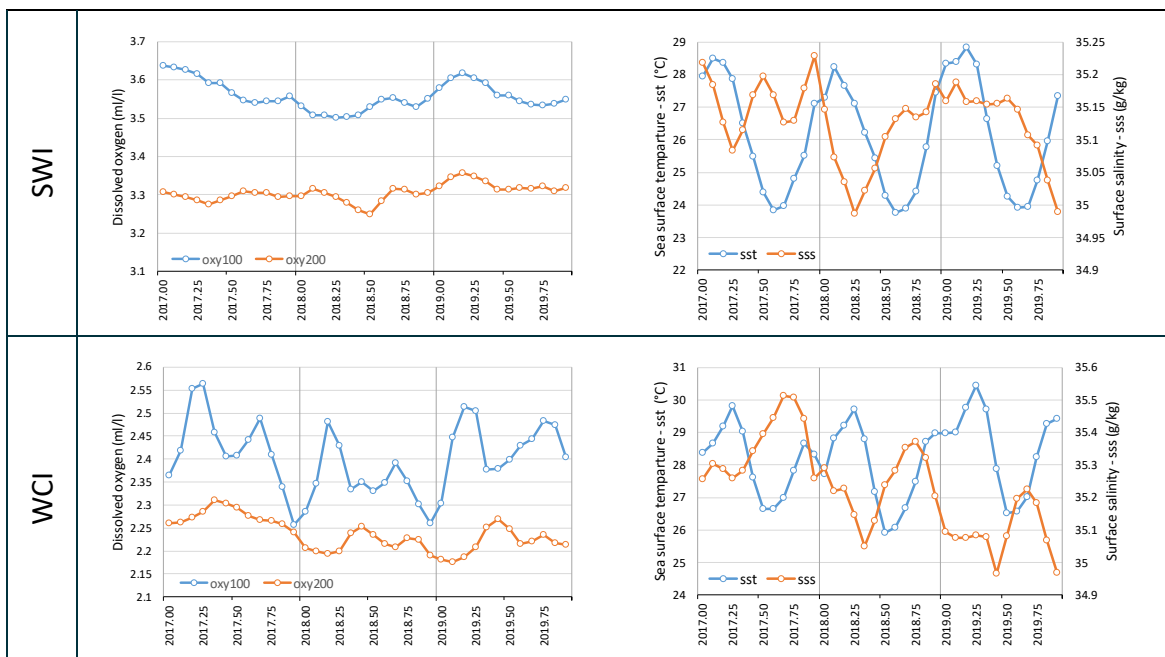


Figure 8b. Same as above, for the two western Indian Ocean sampling sites

Due to their south tropical origin, the dissolved oxygen concentration was higher (3.50 to 3.65 ml.l<sup>-1</sup> at 100 m) in the SWI (Reunion) compared to the WCI (Seychelles), the latter being under the influence of the North IO less-oxygenated waters (2.25 to 2.55 ml.l<sup>-1</sup> at 100 m, and 1ml.l<sup>-1</sup> less than SWI at 200 m). The oxygen seasonality was weak in the SWI. In the WCI, oxygen depicted a semi-annual cycle with maximum values in March-April and September. In both areas, 2018 showed lower average oxygen concentrations at 100 m than 2017 and 2019 (which were similar). Oxygen at 200 m was significantly higher in 2019 for SWI, and in 2017 for WCI. The amplitude of the temperature



signal was stronger in SWI than in WCI. Surface salinity was everywhere above 35 g.kg<sup>-1</sup>, so at intermediate levels, however a significant decline was noted from 2017 to 2019 in WCI. Overall, the environmental conditions showed **significant year-to-year differences**, specifically for oxygen ( $p<0.01$ ) and salinity ( $p<0.05$ ), whilst surface temperature remained comparable throughout the sampling period (Figure 8b).

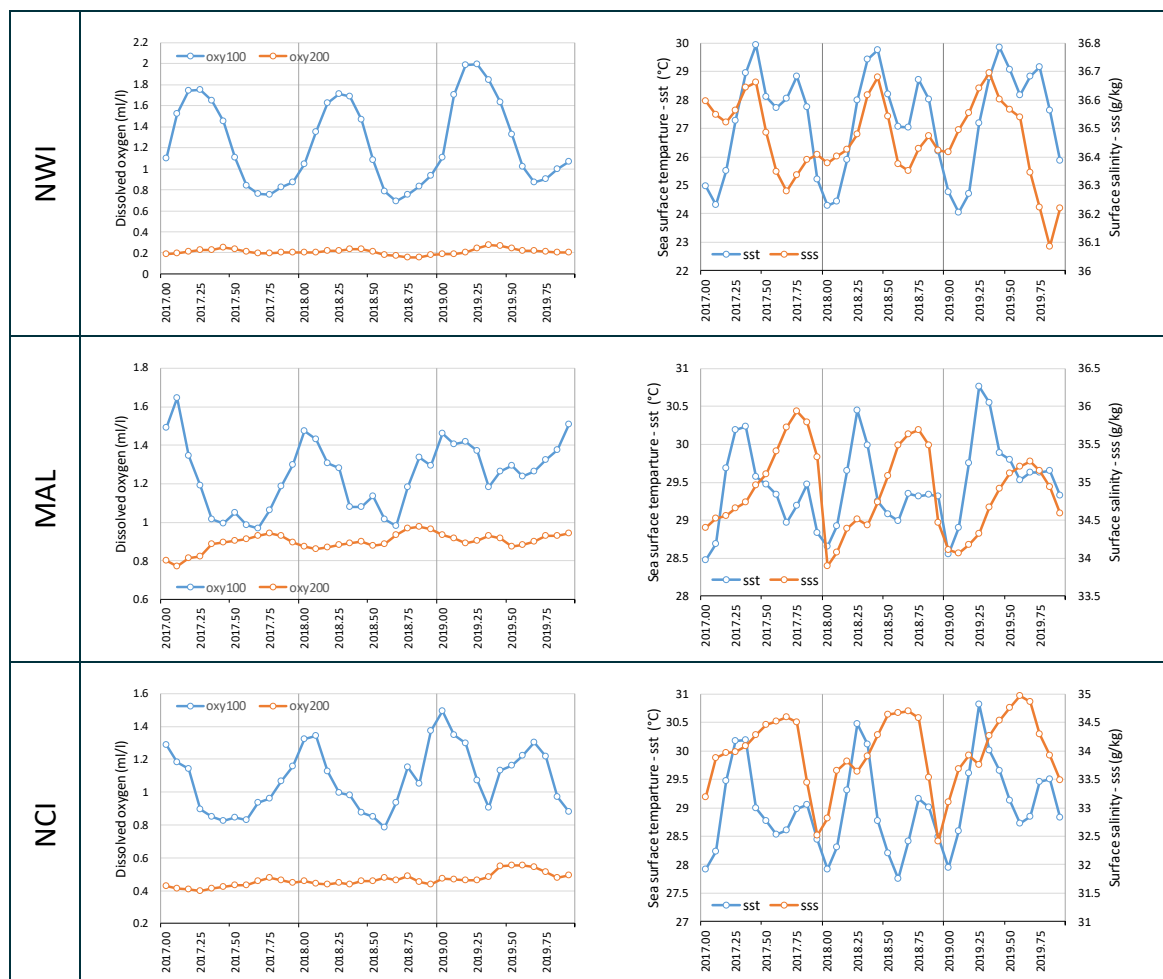


Figure 9c. Same as above, for the three Northern Indian Ocean sampling sites

The three northern Indian Ocean sampling sites are characterized by overall oxygen-depleted water masses (Figure 9c) and the depletion is the most developed in the Arabian Sea (NWI). There, the concentration did not fluctuate seasonally at 200 m, whereas the upper layer (100 m) underwent a clear seasonal signal, with maximal values in March-April and minimal in September-October. The temperature seasonal signal was also strong in the NWI, driven by summer warming and winter cooling with an amplitude of ~6°C. The saline content in the NWI was the highest among all sampling sites (up to 26.7 g.kg<sup>-1</sup>), as the area is under the influence of the outflow of the Persian Gulf water mass. In MAL and NCI, the seasonal cycles in oxygen concentration were in synchrony, with maximum values at 100 m at the turn of the year. In both areas, the temperature amplitude was lesser than in NWI (2° to 2.5°C). In these two areas, the surface salinity during the first semester is influenced by the less saline water masses advected by the winter monsoon current from the BoB and the East Indian Ocean. In contrast, the salinity during the second semester (summer monsoon) is influenced by the southward spread of Arabian Sea water alongshore India. Overall, the

environmental conditions in the northern sampling sites **did not exhibit drastic inter-annual changes** during the period of the project, **except for oxygen concentration at 200 m in NCI** which gradually increased during the study period (KW=20.2,  $p<0.001$ ).

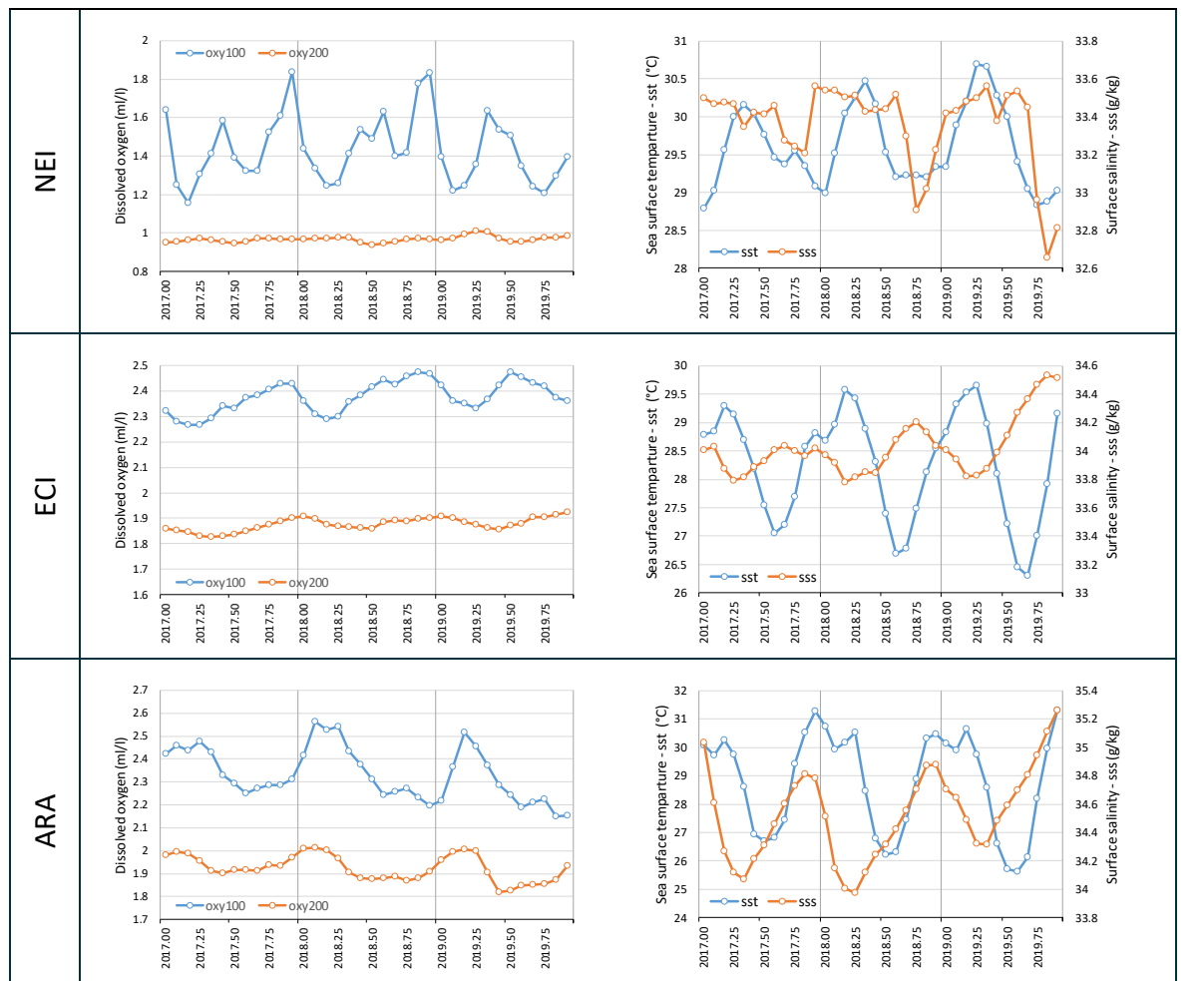


Figure 10d. Same as above, for the three eastern sampling sites

In the eastern sampling sites, the oxygen concentration gradually increased from NEI (off Sumatra and Thailand) to ARA (Arafura Sea). The NEI is under the influence of the BoB waters, with a semi-annual cycle in the oxygen concentration at 100 m (main maximum in December and secondary maximum around June). In contrast, an annual cycle characterized ECI and ARA. In the latter, the oxygen at 200 m showed some seasonal variability whilst this was not occurring neither in NEI nor ECI. The seasonal amplitude in temperature was weak in NEI ( $\sim 1.5^{\circ}\text{C}$ ), becoming larger in ECI and ARA as a consequence of local upwelling occurring in the middle of the year and of an intense summer warming at the turn of the year. The saline content in NEI was less ( $<33.6 \text{ g.kg}^{-1}$ ) than in the two other areas, because of the influence of BoB waters. Overall, in the eastern sampling sites, **oxygen at 200 m changed significantly in NEI** (higher in 2019, KW=6.81,  $p<0.05$ ) **and ECI** (lower in 2017, KW=12.15,  $p<0.01$ ) **but the rest of parameters showed identical range of values** over the project's years (Figure 10d).

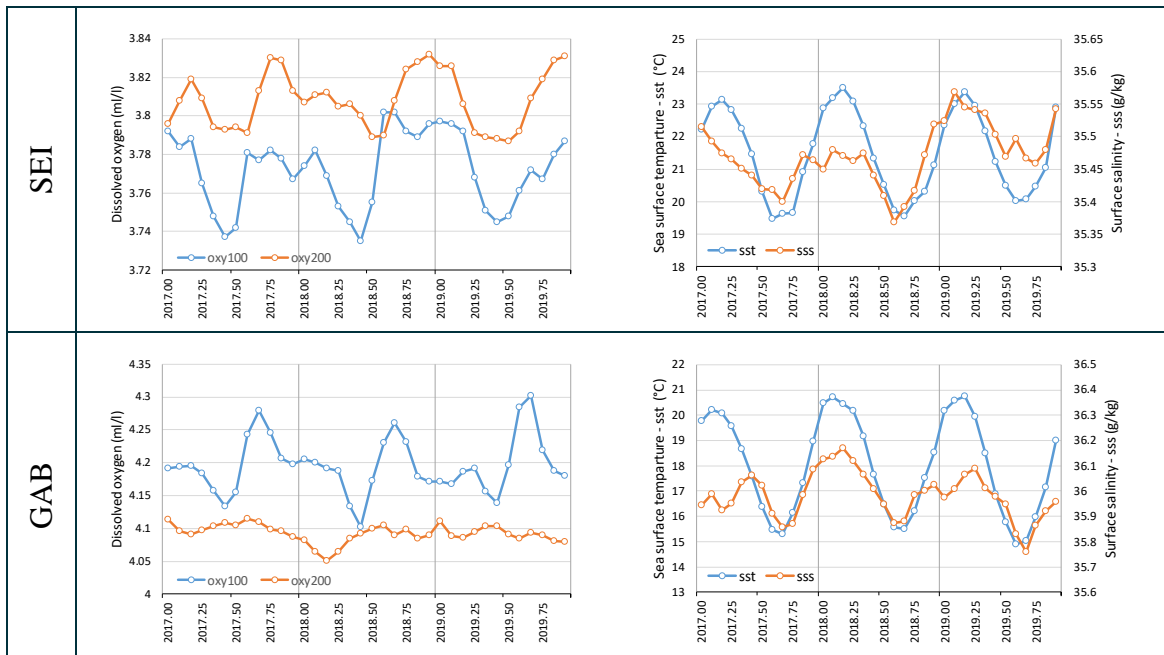


Figure 11e. Same as above, for the south-eastern sampling sites

In the two south-eastern sites, the oxygen concentration was generally high, above 3.74 ml.l<sup>-1</sup> in SEI and above 4 ml.l<sup>-1</sup> in GAB (Figure 11e). A peculiarity in SEI is the oxygen concentration at 200 m exceeding that of oxygen at 100 m (up to 0.6 ml.l<sup>-1</sup> higher). In both areas, a seasonal signal was observed, notably for oxygen at 100 m, with minimal values in June and maximal values during in September-October in GAB and over a longer period (September to March) in SEI. The temperature seasonal cycles were in synchrony in both areas, with a larger amplitude in GAB (~6°C) than in SEI (~4°C). The saline content was significantly higher in GAB (35.8 - 36.2 g.kg<sup>-1</sup>) than in SEI (35.35 - 35.55 g.kg<sup>-1</sup>). **Significant inter-annual changes concerned salinity in SEI (higher in 2019, KW=12.96, p<0.01) and oxygen at 200 m in GAB (higher in 2017, KW=9.86, p<0.01), whereas the conditions for the other parameters remained comparable in the southeastern sites.**

## 4 Genetic sequencing and otolith micro-chemistry methods

This section provides a description of the genetic and micro-chemistry methods that were common across species. More specific details of the methods for individual species are provided in the species-specific appendices.

### 4.1 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 an 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. For this project, 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.

#### 4.1.1 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 Eppendorf 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 *Sph*I. The *Pst*I 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.

#### **4.1.2 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 or 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.

## **4.2 Post-processing of DArT Sequencing data for neritics, albacore tuna, yellowfin tuna, billfish, and blue shark**

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 blue shark) and AZTI (bigeye tuna) for downstream processing.

### **4.2.1 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.

#### **4.2.2 Quality control filtering of loci and individuals**

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 poor quality individual samples (i.e. low DNA quality/quantity or DNA contamination. Quality control of DArTseq (and many other next generation sequencing) data involves multiple inter dependent steps and requires expert knowledge to make decision on threshold values for each filter. To our knowledge Radiator (Gosselin 2017) is currently the most adequate software to perform this task. Its development was based on the analysis of dozens of reduced representation genotyping datasets and lot of thinking went into the implementation of the different steps and their order. Removing SNP loci has an impact on individual samples apparent quality and vice versa, therefore some filtering criteria on individuals are assessed twice, a first pass early on to remove obvious low-quality samples and a second finer pass after most of the poor-quality SNP have been removed. Also, every time samples are removed, all monomorphic loci (those presenting only one allele) are removed automatically. Radiator also provides graphical output to help end-users deciding on threshold values. The threshold values used during our quality control process using Radiator are presented in Appendix 3.

#### **4.2.3 Quality control filtering of loci**

Filtering of SNP loci included examination of:

- 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), low call rate is either due to the presence of a mutation on one of the restriction enzyme sites or reflects poor amplification. In any case, it is best to remove markers with low call rate
- minor allele count (how many times the allele in lowest frequency was observed), this allows to discard rare allele that are probably sequencing error or true but uninformative variants at best
- coverage (how many times a locus was observed in average across all individuals), the lower the less confidence there is in the genotype call
- position of the SNP on the fragment being sequenced needs to be checked as the probability of sequencing errors is not random
- 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. Loci can deviate from HWE because true biological reason like population structure, but it can also be a sign of paralogy. This filter was used to remove loci deviating from HWE in multiple sampling location with high probability

- Missingness (proportion of times a loci was reported as NA). Too many NA values may indicate poor consistency among the individuals. Any loci with a proportion less than 0.15 were removed.

MALF (minor allele frequency) per sampling region. Genetic markers included in the analysis had to be found in at least 5% of individuals within a set of individuals representing a minimum of one sampling region. This removes those rare alleles that are spread throughout the sampling regions, and may also include private alleles that are below significant detection levels.

#### 4.2.4 Quality control filtering of individuals

Filtering of individuals included examination of:

- missingness (proportion of missing values), which reflects the overall DNA quality/quantity of the samples, the lower the better
- heterozygosity (proportion of loci exhibiting two different alleles). Whilst high heterozygosity can be due to normal biological processes, like hybridisation or introgression, our experience with dozens of species is that it is generally due to cross contamination between samples. Therefore, we discarded samples deviating from their sampling location average.
- total coverage (total number of sequences obtained per individual), similar to missingness, this reflects the DNA quality/quantity, the higher the better. However, individuals with exactly 100% coverage are unusual and may indicate contamination or measurement failure. These individuals are removed.
- genetic distance in order to detect and remove duplicate samples introduced on purpose (technical replicates) or by accident (same fish sampled twice). These were detected by removing one individual of any pair that had excessively small genetic distance between them. In this case, excessively small is defined to be smaller than the 0.0001 quantile.

### 4.3 Processing of RAD FASTQ (skipjack) and DArT FASTQ (bigeye) data

#### 4.3.1 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*.

#### 4.3.2 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, were 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.

## 4.4 Analysis of SNP data for population structure

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. We altered the choice of starting values very slightly to add confidence in the estimation process (details are given in Appendix 3).

Methods to choose the number of genetic groups is a difficult problem. It is well beyond the scope of this work to resolve this debate. The issues include:

- Philosophical. Is there a single correct K? Are the differences just categorisation of a continuum?
- Genetical. What actually constitutes a genetical group? Is it a stock? Population?
- Statistical. How to determine the number of K, even when there are natural groupings? This is an ongoing research topic (see Hui et al 2015 for a recent contribution).

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 inspect two other measures for choosing the number of groups. The first source is 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 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 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, only in presentation. 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.

## 4.5 Microchemistry methods

### 4.5.1 Sample selection and preparation

Representative individuals from each of the species and sampling locations were chosen from the total sample pool collected for the project (Table 4) and distributed to relevant lead partners for microchemistry (Table 6).

Sagittal otoliths were selected by fish length. One of the paired otoliths was analysed for trace element chemistry and, in some cases, the second otolith was analysed for stable isotopes,  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$ .

Transverse sections that contained the primordium were cut and then polished until the core (or primordium) 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 the fish inhabited in the recent period before being caught (weeks or months depending on the age of the fish).

### 4.5.2 LA-ICP-MS analysis

The 3 partners led the otolith analysis of different species as follows:

- CSIRO: longtail tuna, kawakawa, narrow-barred Spanish mackerel and bigeye tuna;
- AZTI: skipjack tuna, yellowfin tuna;
- IRD: albacore tuna, swordfish

Each partner used different facilities to analyse their assigned species. The use of different facilities (and platforms) means that the comparisons among locations for individual species will be consistent, but the potential for differences in the platforms used and 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. Further details of the methods and results from each of the partners are listed in Table 6.

**Table 6. Details of LA-ICP-MS analysis of otolith sections from 8 species conducted by the 3 partners.**

Species	Partner	Facility	Instrument	Laser method	Beam size	Position along otolith section	Elements measured	Other information
skipjack tuna	AZTI	Trace elements: Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux Université de Pau et des Pays de l'Adour/CNRS	Trace elements: High-resolution laser ablation inductively coupled plasma mass spectrometer (LA-HR-ICPMS)	Spot (trace elements)	30 µm	Trace elements: 1. Primordium 2. 65 µm from primordium 3. 200 µm from primordium 4. edge	Trace elements: Ba, Mg, Mn, Sr, Zn, Cu, Fe, Li	200 µm from primordium (±13-15 days of life) was selected as natal origin signature  Zn, Cu, Fe and Li were below the limit of detection and therefore discarded for analyses
yellowfin tuna	AZTI	Trace elements: Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux Université de Pau et des Pays de l'Adour/CNRS	Trace elements: High-resolution laser ablation inductively coupled plasma mass spectrometer (LA-HR-ICPMS)	Spot (trace elements)	Trace elements: 30 µm	Trace elements: 1. Primordium 2. 65 µm from primordium 3. 200 µm from primordium 4. edge	Trace elements: Ba, Mg, Mn, Sr, Zn, Cu, Fe, Li  Stable isotopes: δ13C, δ18O	65 µm from primordium (±14-16 days of life) was selected as natal origin signature  Zn, Cu, Fe and Li were below the limit of detection and therefore discarded for analyses
bigeye tuna	CSIRO	Centre for Ore Deposits and Earth Sciences (CODES), at the University of Tasmania (UTAS)	Resonetics RESOLution S-155 system with a Coherent 110 Compex Pro ArF excimer laser	Spot	30 µm	1. 65 µm from primordium 2. 110 µm from primordium 3. 800 µm from primordium 4. edge	Li, Na, Mg, P, K, Mn, Fe, Cu, Zn, Rb, Sr, Ba, Pb	
longtail tuna	CSIRO	Centre for Ore Deposits and Earth Sciences (CODES), at the University of Tasmania (UTAS)	Resonetics RESOLution S-155 system with a Coherent 110 Compex Pro ArF excimer laser	Spot	30 µm	1. 65 µm from primordium 2. 110 µm from primordium 3. 800 µm from primordium 4. edge	Li, Na, Mg, P, K, Mn, Fe, Cu, Zn, Rb, Sr, Ba, Pb	
Spanish mackerel	CSIRO	Centre for Ore Deposits and Earth Sciences (CODES), at the University of Tasmania (UTAS)	Resonetics RESOLution S-155 system with a Coherent 110 Compex Pro ArF excimer laser	Spot	30 µm	1. 65 µm from primordium 2. 110 µm from primordium 3. 1000 µm from primordium 4. edge	Li, Na, Mg, P, K, Mn, Fe, Cu, Zn, Rb, Sr, Ba, Pb	
kawakawa	CSIRO	Centre for Ore Deposits and Earth Sciences (CODES), at the University of Tasmania (UTAS)	Resonetics RESOLution S-155 system with a Coherent 110 Compex Pro ArF excimer laser	Spot	30 µm	1. 65 µm from primordium 2. 110 µm from primordium	Li, Na, Mg, P, K, Mn, Fe, Cu, Zn, Rb, Sr, Ba, Pb	

						3. 450 µm from primordium 4. edge		
swordfish	IRD	Montpellier University, Plateforme AETE-ISO	ThermoFisher Scientific <i>Element XR</i> <i>Laser</i> Télédynes G2 Excimer (193nm)	Transect	20µm	Mean of 3 spots after the core corresponding to the distance between 10 and 40µm from the core	Li, B, Mg, P, Mn, Cu, Zn, Rb, Sr, Ba	Above detection limits so retained for analysis: B, Mg, P, Zn, Sr, Ba
albacore	IRD	Montpellier University, Plateforme AETE-ISO	ThermoFisher Scientific <i>Element XR</i> <i>Laser</i> Télédynes G2 Excimer (193nm)	Transect	20µm	Mean of 3 spots after the core corresponding to the distance between 10 and 40µm from the core	Li, B, Mg, P, Mn, Cu, Zn, Rb, Sr, Ba	Above detection limits so retained for analysis: B, Mg, P, Zn, Sr, Ba B, Mg, P, Cu, Zn, Sr, Ba

### 4.5.3 General statistical approaches for otolith microchemistry

In this report, we concentrated on analysing elemental signatures from the otolith core, since these data should reflect the fish's spawning origins. However, in some circumstances, it was also be useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

Analyses that were performed on the core or edge data:

- Univariate tests for each element to test for differences among locations. The exact test used depended on whether the data were normally distributed (assessed using a Shapiro-Wilks test) and had equal variances among locations (assessed using a Fligner-Killeen test. Specifically: (1) when the data were normally distributed and location variances were equal, a t-test (for 2 locations) or ANOVA test (for 3+ locations) was used; (2) when the data were not normally distributed but location variances were equal, a Wilcox test (for 2 locations) or Kruskal-Wallis rank sum test (for 3+ locations) was used; and (3) when the data were not normally distributed and the location variances differed, a Yuen test (for 2 locations) or heteroscedastic one-way ANOVA for trimmed means (for 3+ locations) was used. When significant differences were found in the case of 3+ locations, different post-hoc comparisons were applied according to which category the data belonged: (1) Tukey's test, (2) Dunn's test or (3) Lincon test.
- Principal Component Analysis (PCA) was used to reduce dimensionality and help visualize the data as this can be difficult with so many elements, and to determine which elements account for most of the variability in the data.
- Permutational Multivariate Analysis of Variance (PERMANOVA) was used to test for differences in the multi-elemental signatures of fish among locations (based on Euclidean distance and 999 random repeats). When significant differences were found, post-hoc multilevel pairwise comparisons were used to identify which locations differed (using a Benjamini-Hochberg correction method to adjust the p-values).
- Clustering was used in some circumstances for identifying the most likely number of separate spawning origins, and to investigate whether spawning origins differed among fish from different sampling locations. Hierarchical clustering using Ward's agglomeration method was applied to dissimilarity data calculated using Euclidean distance. Once clusters were identified, PERMANOVAs were performed to test for differences in the multi-elemental signatures among clusters.

Note that prior to any multivariate analyses, the data were standardised (i.e., for each element, the data was centred by subtracting the mean and scaled by dividing by the standard deviation).

## 5 Results and discussion

### 5.1 Sample collection

Overall, the performance of the sample collection as compared to the original design is satisfactory, however, differences appear between species (Figure 12). The rating is excellent for tropical tunas, for both genetic tissues and otoliths (all over 100%). For neritic tunas, the rating is good for kawakawa (>75% in genetic tissues, > 100% for otoliths) and moderately satisfactory for longtail and narrow-barred Spanish mackerel. For albacore, the rating is moderately satisfactory in terms of numbers, however less satisfactory in terms of spatial coverage (unbalanced sampling between the West and the East Indian Ocean). Among the billfish, only swordfish had a good rating, the remainder being very low. Finally, the rating is good for sharks (>75%) and quasi null for scalloped hammer shark due to CITES constraints.

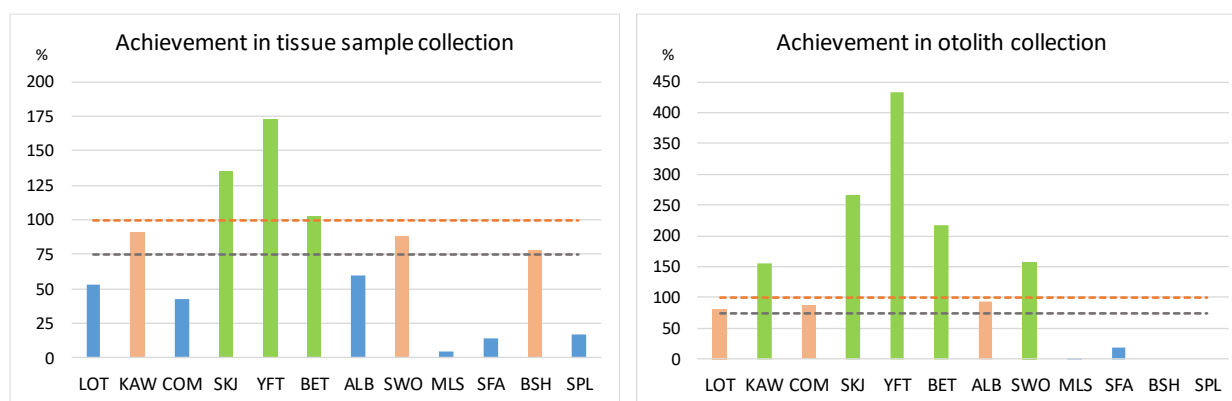


Figure 12. Summary of sampling performance for tissue (left) and otolith (right) collection relative to planned sample size.

### 5.2 Status of results and opportunity for further input

The timing of completion of the analysis means there has been limited time for review and interpretation by the IOTC Scientific Committee and the wider scientific community. Hence, the results presented are preliminary. As will be apparent from the results and discussion, in most cases they should be considered as formative hypotheses on which more specific studies can be designed and implemented to test. We make no definitive conclusions about what the results may mean in terms of potential implications for assessment and management, as this would be premature and needs to await the results of more detailed and comprehensive consideration of the analyses completed to date and the considered input of the Scientific Committee and Working Groups.

In light of the above, we invite and encourage feedback from the IOTC scientific community on these 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 results and preparation of working papers to

the appropriate Working Party and Scientific Committee meetings (Table 7) and planned peer review papers later in 2020.

Meeting	Dates	Species
WPNT10	6-10 Jul	LOT, KAW, COM
WPB18	2-5 Sep	SWO, STM, SFA
WPEB16	7-11 Sep	BSH
WPTT22	21-26 Oct	SKJ, YFT, BET
SC meeting	5-9 Dec	Final report
WPTmT08	TBA (2021)	ALB

Table 7. Papers to be submitted to IOTC meetings

### 5.3 Presentation of genetic population structure results

We present the results in two forms: A summary of the main results, the substantive conclusions for each method (genetics and microchemistry) and our conclusions and recommendations for future work for each species are provided in the next section of the report. Additional technical detail, related to the methods and results and previous relevant work, is provided for each species in the Appendix 1.

For the summary section, for each species we include: i) a map with the distribution of the samples that were processed for each method; a summary of the fit of the population structure model (where one was used), and; a representation of the probability of group membership ( of 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.

It is important to recognise, that the results of these analyses are used as a guide, rather than a definitive result for the “true” underlying population structure. Other considerations, including the geographic distribution of the samples and the pattern of genetic heterogeneity in the available samples need to be taken into consideration in the interpretation of the results. Hence, these results should be considered as working hypotheses.

The bottom plot illustrates the probability of group membership of individual fish/sharks among the sample locations (Figure 14). For most species, the result for the most likely number of groups is shown in the body of the report, while for some species multiple alternatives are shown. More detail is given for each species in the detailed results in Appendix 1. In cases where the most likely number of groups is one, then the plot is omitted, as it is uninformative: in that all locations and individuals are shown as the one colour.

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 investigation.

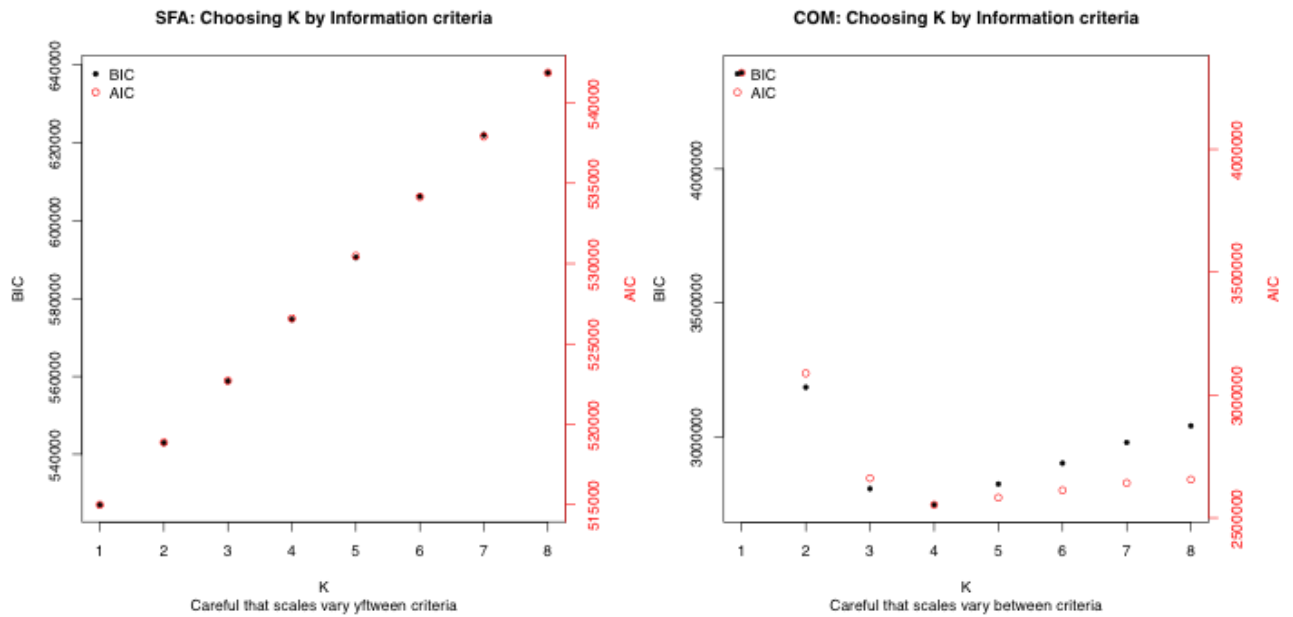


Figure 13. The Information Criterion plots summarise the results of model fits for the most likely number of genetic groups from the distribution of SNP data in the sample. The AIC and BIC are two forms of statistic used to summarise how well the model fits the data with the lower the value the more 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. These results should be interpreted subjectively with the geographical distribution of the groups. It seems that the information criteria may give a lower bound for the number of genetic groups.

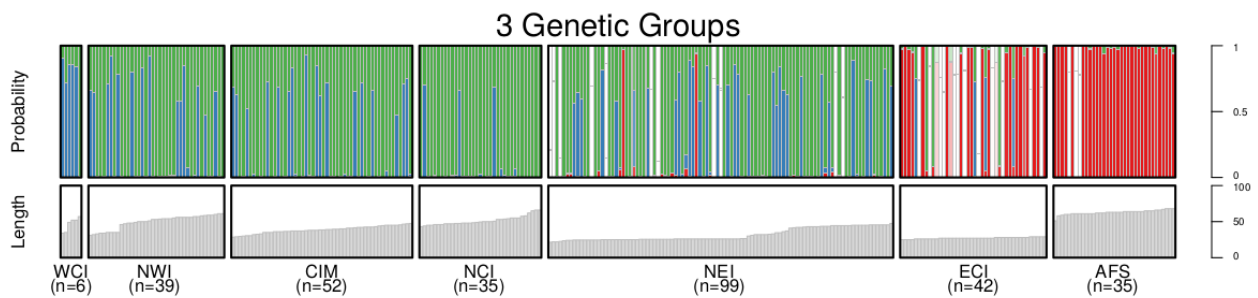


Figure 14. Assignment probabilities for Kawakawa into  $K=3$  genetic groups. Different colours represent different groups. Uncertainty is displayed as 'whiteness' -- those fish where the estimated group assignment probabilities is large are faded progressively to white. The bottom plots give the length distributions of the individuals sampled at that location. In this plot, we suggest that  $K=3$  is overfitted and that a smaller  $K$  is probable. This is due to the fact that the 'blue' genetic group is present in non-trivial proportions throughout many sampling locations.

## 5.4 Neritic tuna

### 5.4.1 Longtail tuna (*Thunnus tonggol*)

#### Genetics

- 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 221 past quality control and were included in the analysis of population structure (Figure 15);
- The population analysis based on StockR indicates three genetic groupings, one per sampled location within the Indian Ocean, which is supported by the AIC model fits (Figure 15).

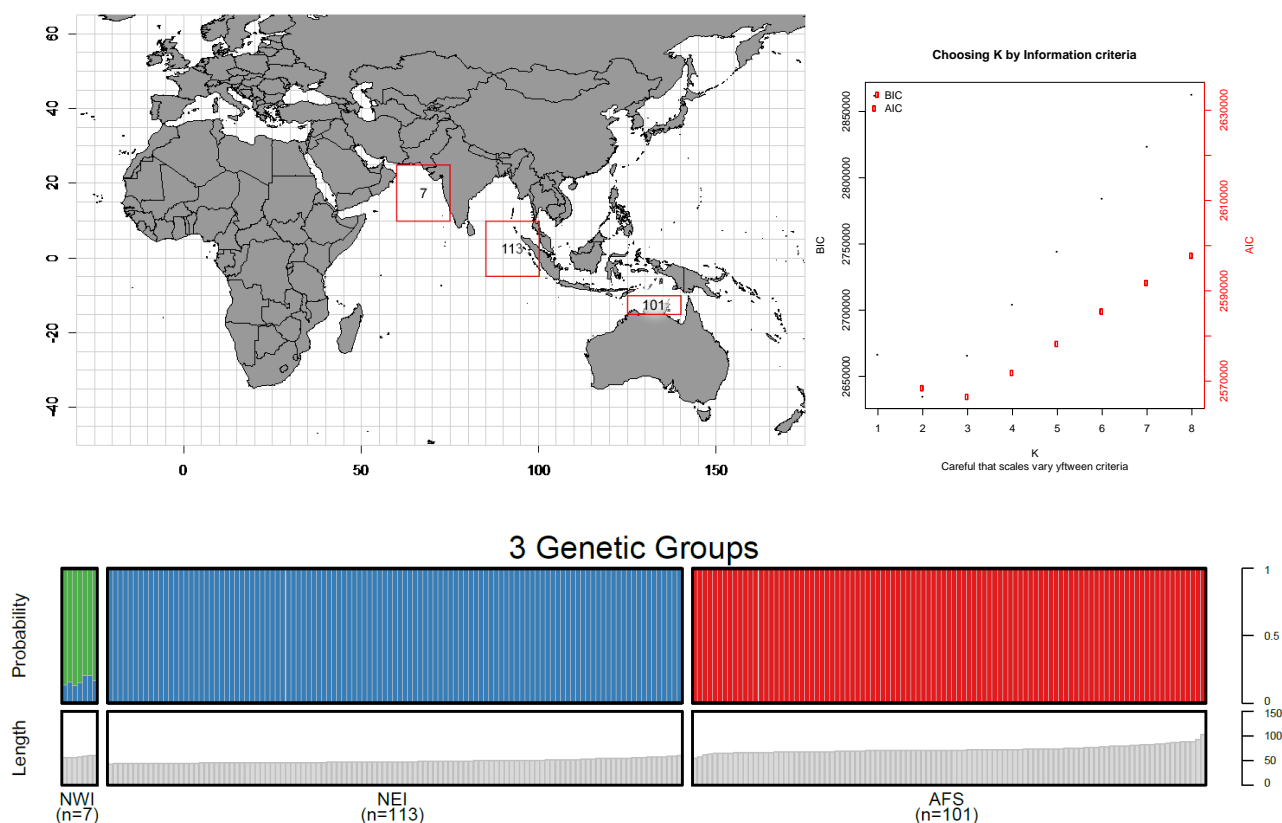


Figure 15. Top Left: Number of samples of longtail tuna (*Thunnus tonggol*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for longtail tuna for 3 genetic groups.



## Microchemistry

The longtail otolith samples analysed were from three sampling locations (Figure 16), referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI) and Arafura Sea (Arafura) (Figure 2). The sample sizes for NWI and Arafura are very small (Figure 16, Table 8), which makes it difficult to draw conclusions when comparing data among locations. Although it was not intended in the original design, samples from different locations were collected during different periods due to several reasons (see Section 2):

Fish from NWI and NEI ranged in length from 43-60 cm FL; those from Arafura were larger, ranging from 69-88 cm FL (Table 8). While age at length information is very limited for longtail, we estimate fish from NWI and NEI to have been 1-3 years old at capture, and those from Arafura to be 3+ years. This means that fish from all locations will have been spawned over several years, and those from Arafura will have been spawned in earlier years than those from NWI and NEI. Thus, differences observed in core signatures among locations may be partly due to cohort effects.

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However, it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

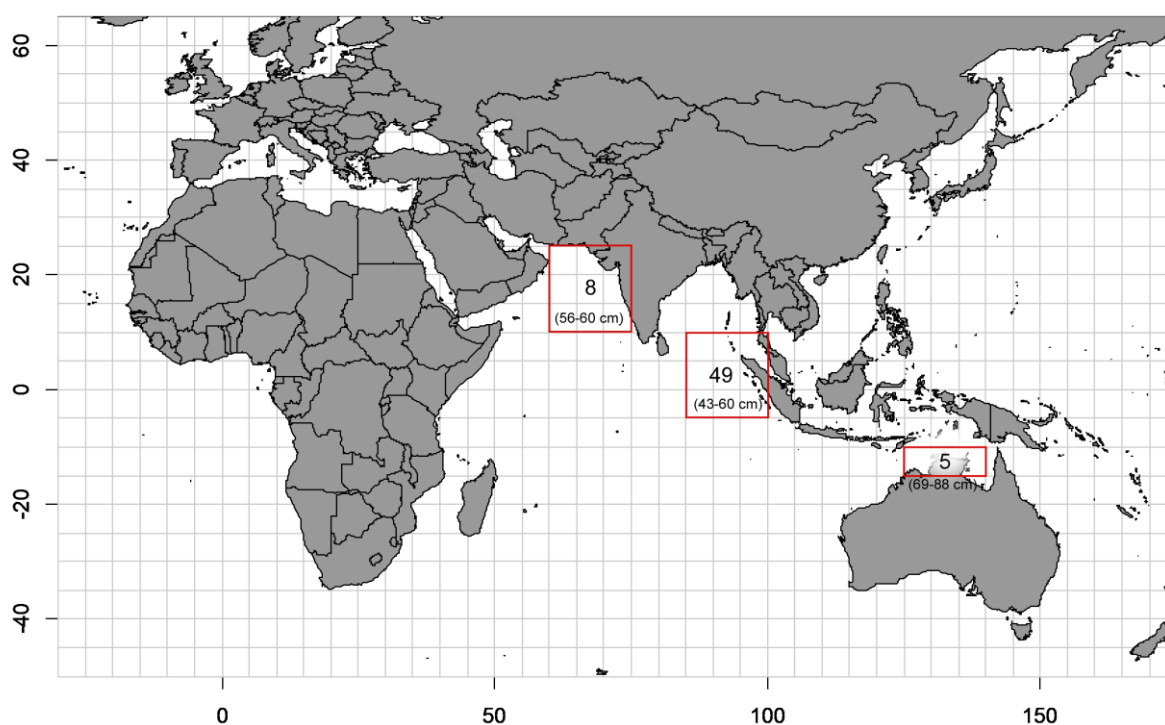


Figure 16. Map showing the number of longtail otoliths analysed for each of three sampling locations, referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI) and Arafura Sea (Arafura), and the size range of fish at each location.

Table 8. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
North-West Indian Ocean (NWI)	8	August - September 2018	56-60	2-3
North-East Indian Ocean (NEI)	49	April and October 2018	43-60	1-3
Arafura Sea (Arafura)	5	Dec 2017	69-88	3-7

\* based on results from Abdussamad et al. (2012) and Griffiths et al. (2010).

## Core Results

Core signatures were similar between locations for most elements. Univariate tests indicate that  $^{39}\text{K}$ ,  $^{23}\text{Na}$  and  $^{85}\text{Rb}$  differ significantly among locations ( $p < 0.05$ ), but the low sample sizes mean these results must be treated with caution.

Results from the PERMANOVA and pairwise tests indicate core signatures differ significantly between NWI and NEI.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA supports the finding that the core data from fish captured in NWI may differ from NEI (Figure 17).

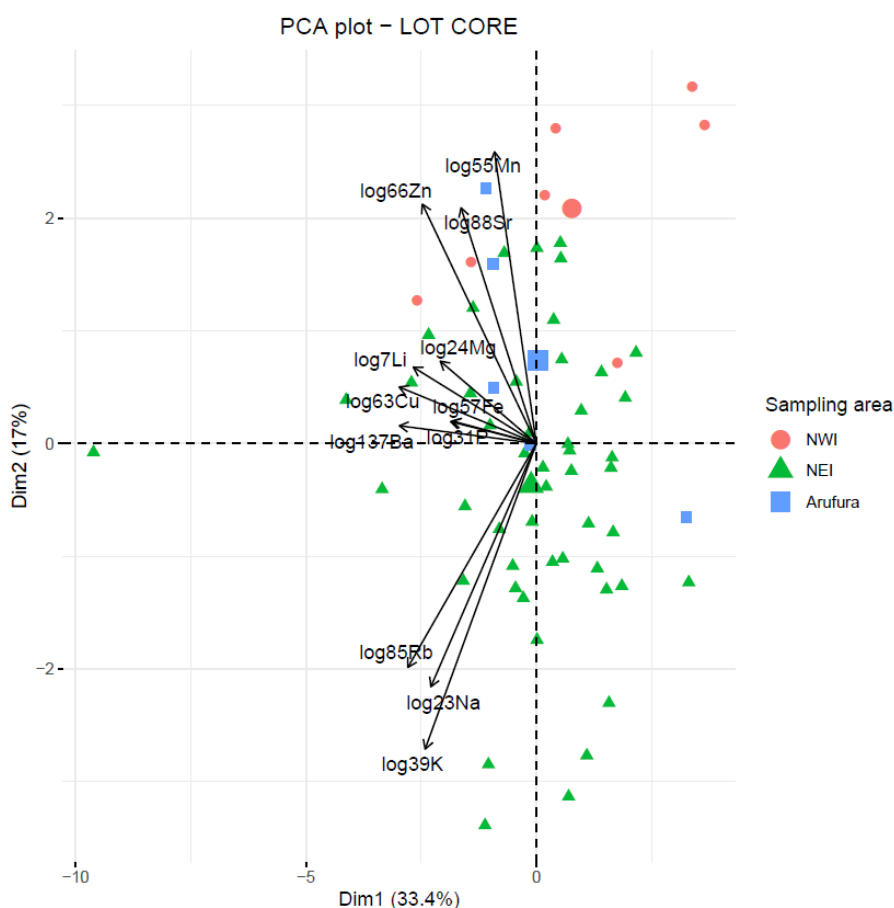


Figure 17. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the longtail otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge Results

Edge signatures differed between locations for many elements. Based on univariate tests,  $^{23}\text{Na}$ ,  $^{24}\text{Mg}$ ,  $^{39}\text{K}$ ,  $^{55}\text{Mn}$ ,  $^{66}\text{Zn}$  and  $^{137}\text{Ba}$ , all differ significantly among locations ( $p < 0.05$ ). Results from the PERMANOVA model also suggest edge signatures are not equal among all locations ( $p = 0.001$ ). Based on subsequent pairwise tests between locations, the edge signatures differ significantly between NWI and NEI, and between Arufura and NWI. The result for Arufura vs NEI is less clear but there is some evidence of a weak difference.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data confirms and helps to visualize these findings (Figure 18).

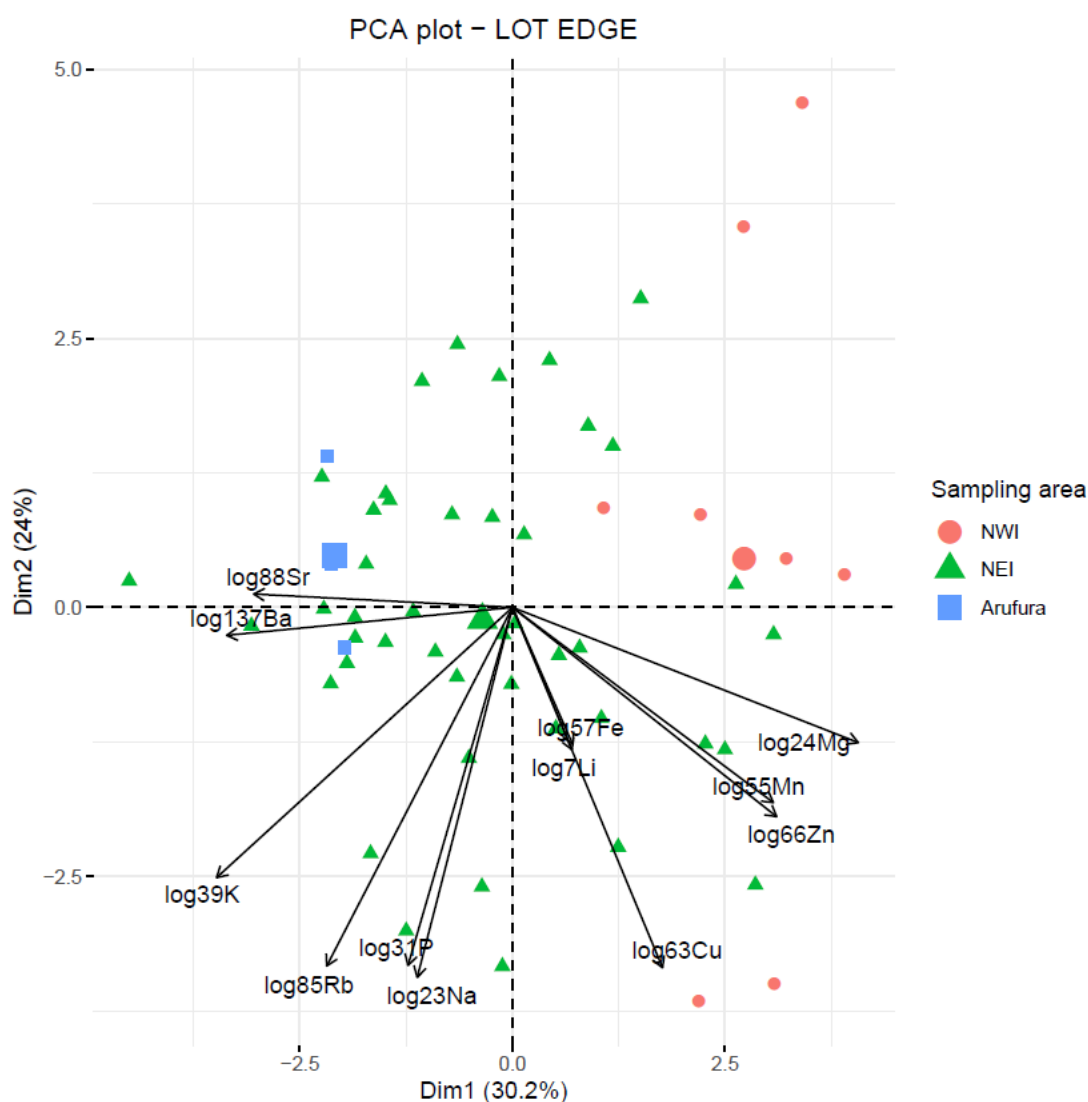


Figure 18. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the longtail otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Microchemistry summary

As noted, small sample sizes for two of the three locations means it is not possible to confidently draw any conclusions. Nevertheless, the fact that the core signatures between fish caught in NWI

and NEI show evidence of being different suggests these fish may have been spawned in different locations.

The core signatures of the five fish caught in Arafura overlap with those from NWI and NEI, so it is difficult to know if:

- i) they are from a distinct spawning area,
- ii) some fish were spawned in the same location as those caught in NEI, and others in the same location as those caught in NWI, or
- iii) they are from the same spawning ground as fish caught in one of the other locations but have slightly different core signatures due to differences in ocean chemistry between the years they were spawned (recall that the Arafura fish are older and were spawned in earlier years than fish from the other two locations).

The fact that the edge signatures appear to differ between the three capture locations suggests the ocean chemistry differs enough between these locations to be useful for classifying fish to them using otolith chemistry.

## 5.4.2 Kawakawa (*Euthynnus affinis*)

### Genetics

- 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, 308 past quality control and were included in the analysis of population structure (Figure 19);
- The population analysis based on StockR indicates a preference for 2 genetic groupings within the Indian Ocean based on the model fits (Figure 19).

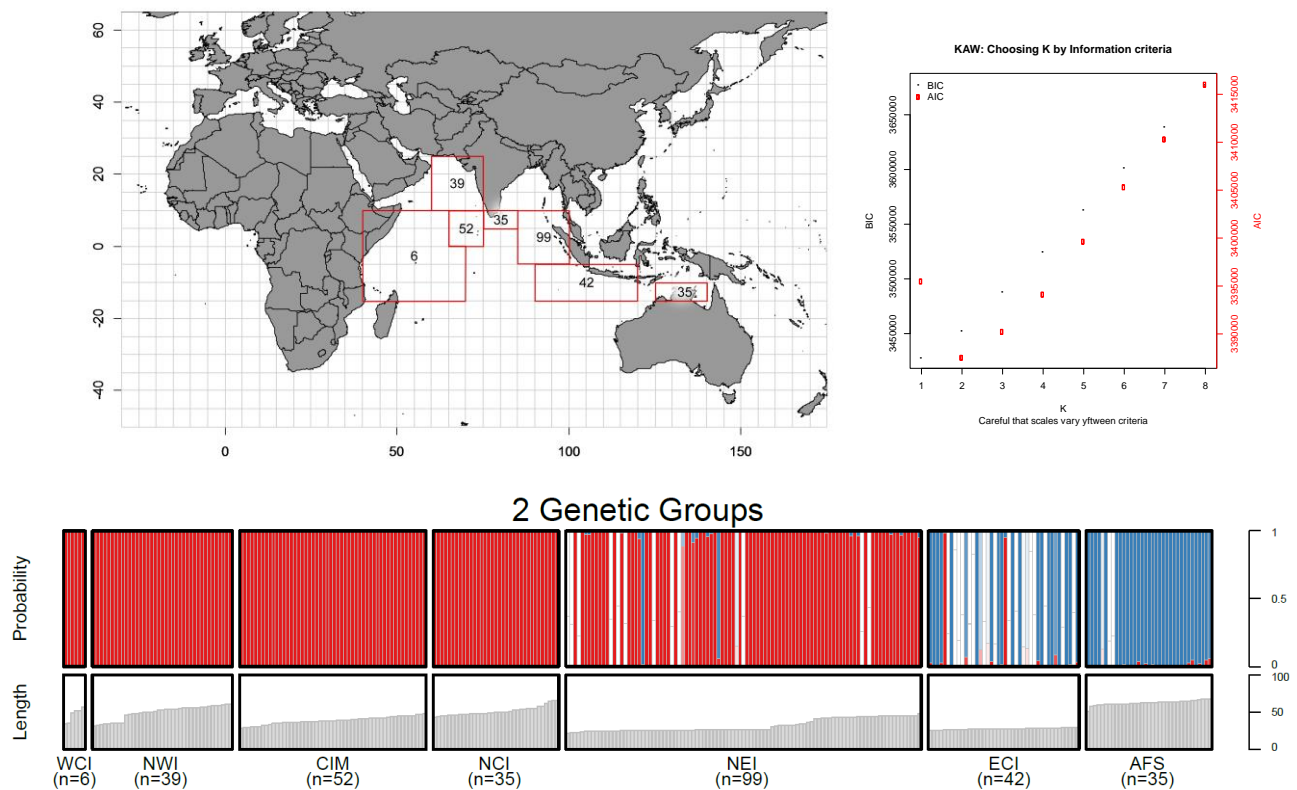


Figure 19. Top Left: Number of samples of kawakawa (*Euthynnus affinis*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for kawakawa suggesting for 2 genetic groups within the Indian Ocean.

## Microchemistry

The kawakawa otolith samples analysed are from four sampling locations (Figure 20), referred to as North-West Indian Ocean (NWI), Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI) (Figure 2). Although it was not intended in the original design, samples from different locations were collected during different periods, due to several reasons (see Section 4.5.2):

Fish ranged in fork length from 20 to 55 cm across all locations, with the fish from NEI being slightly smaller on average (Figure 20, Table 9). Juveniles grow rapidly and are estimated to be 50-65 cm by age 3, meaning fish sampled from all locations are likely to have been spawned over a few years. Thus, differences observed in core signatures among locations may be partly due to cohort effects.

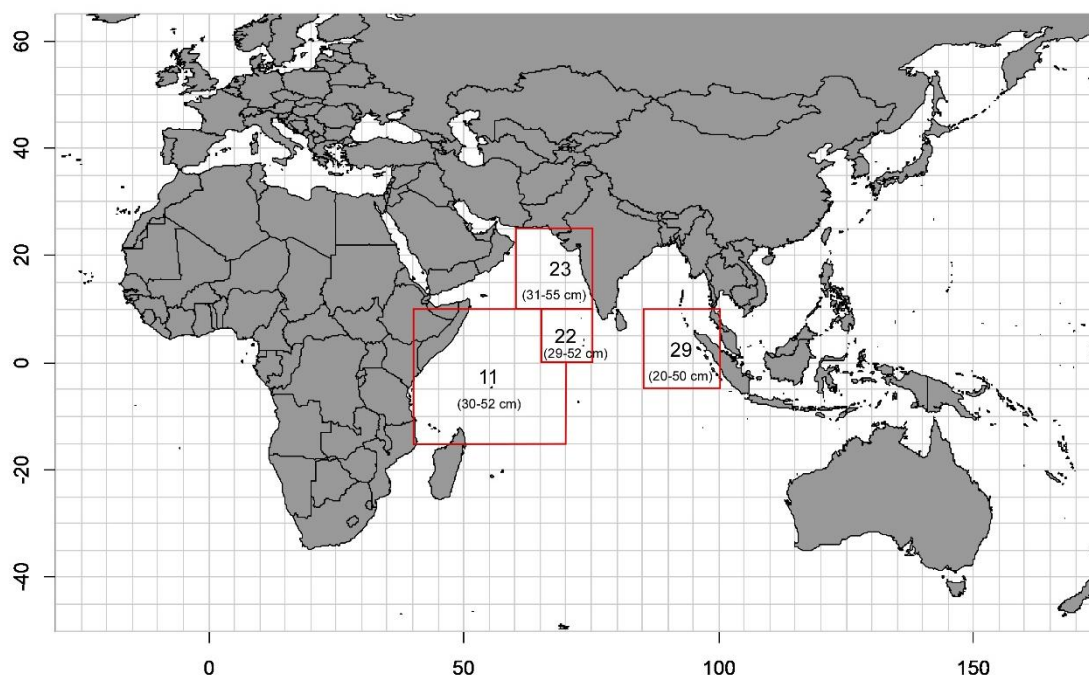


Figure 20. Map showing the number of kawakawa otoliths analysed for each of four sampling locations, referred to as North-West Indian Ocean (NWI), Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI), and the size range of fish at each location.

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However, it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

**Table 9. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.**

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
North-West Indian Ocean (NWI)	23	Apr-May 2018	31-55	0+ to 3
Western Central Indian Ocean (WCI)	11	Feb-Apr 2018	30-52	0+ to 3
Maldives (CIM)	22	Aug 2018 and Feb 2019	29-52	0+ to 3
North-East Indian Ocean (NEI)	29	Apr and Nov 2018	20-50	0+ to 2

\* based on results from Kahn (2004)

### Core Results

Core signatures were similar between locations for most elements but there was large variability in the data. Univariate tests indicate that only  $^{137}\text{Ba}$ , and to a lesser extent  $^7\text{Li}$ , differ significantly among locations ( $p < 0.05$ ).

Although results from fitting a PERMANOVA model to the core data provide weak evidence that the multi-elemental core signatures are not equal among all locations ( $p = 0.033$ ), subsequent pairwise tests between locations showed no significant differences.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA indicates that the core data does not differ significantly among locations (Figure 21).

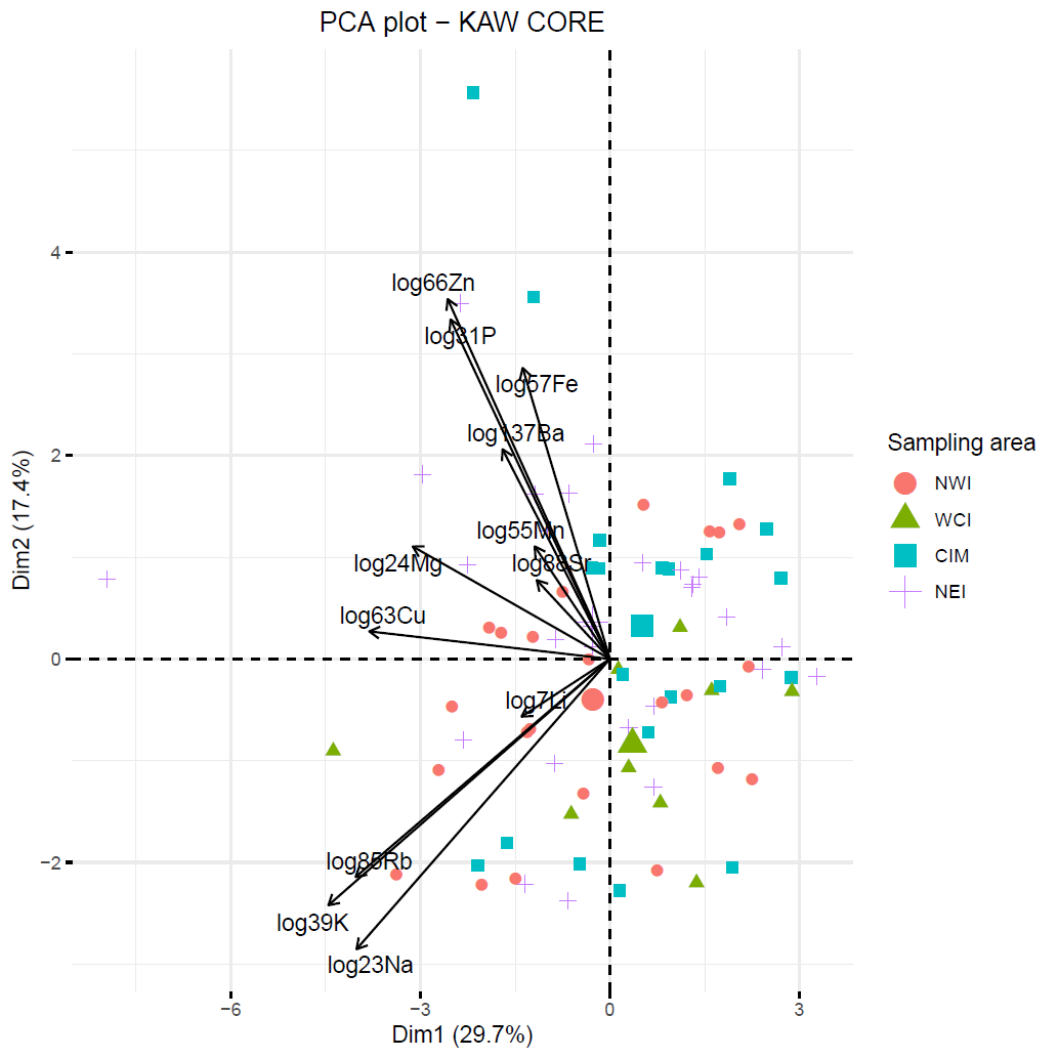


Figure 21. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the kawakawa otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge Results

Results from the PERMANOVA and pairwise tests suggest edge signatures differ significantly between CIM and NEI, CIM and NWI, and NEI and NWI.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data helps to visualize these findings, although the difference between NEI and NWI is not obvious in this figure (Figure 22).



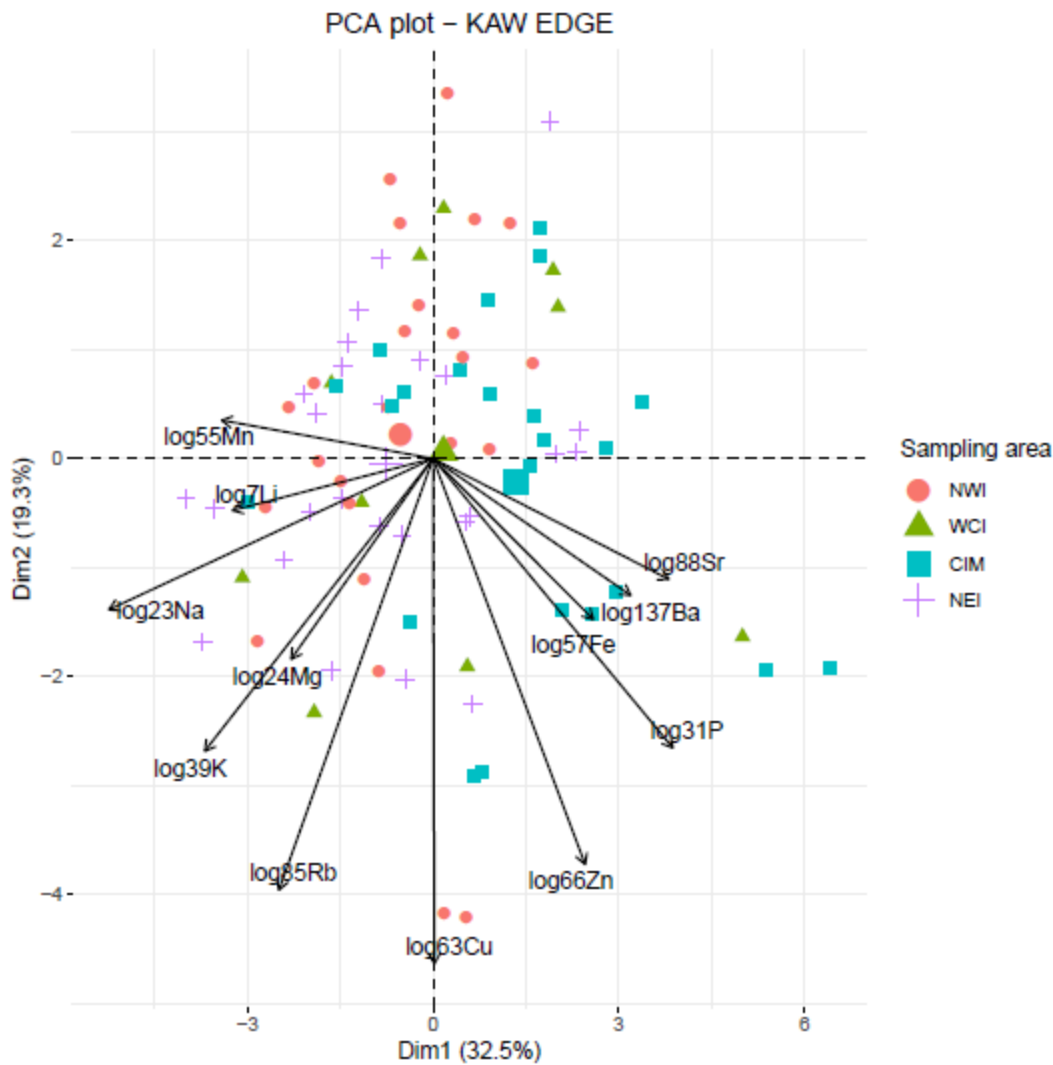


Figure 22. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the kawakawa otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

### Microchemistry summary

The fact that the core signatures do not show significant differences among locations suggests either that these fish all originated from the same location, or that they originated from different locations with similar water chemistry.

The fact that the edge signatures appear to differ between the some of the capture locations (particularly CIM and NEI, and CIM and NWI) suggests that using otolith chemistry to classify fish to these locations may be possible.

### 5.4.3 Narrow-barred Spanish mackerel (*Scomberomorus commerson*)

#### Genetics

- The sample coverage for narrow-barred Spanish mackerel is good for the eastern and reasonable for northern regions of the Indian Ocean range of the species and good for the Pacific outlier location. The latter were known spawning fish collected from the annual spawning run on the Great Barrier Reef. Samples have been collected in the SWI, but unfortunately were not available for inclusion in this analysis;
- A total of 256 samples from 5 Indian Ocean sampling regions and one Pacific outlier location were collected. A total of 207 were sequenced using DArTSeq and 189 and 153 past quality control and were included in the primary and finer-scale (see Appendix 8.1.5 for details of hierarchical approach to the analysis specific to this species) analysis of population structure (Figure 23);
- The primary analysis of population analysis indicated a strong preference for four genetic groupings within the Indian Ocean based on StockR model fits (Figure 23); the second round of analyses partitioned one of these groups (AFS and WCS) into two separate genetic groups (Figure 24);
- Based on these analyses, 4 genetic groups were identified within the Indian Ocean (NWI+NCI, NEI, ECI and AFS) and one consisting of WCS in the Pacific Ocean.

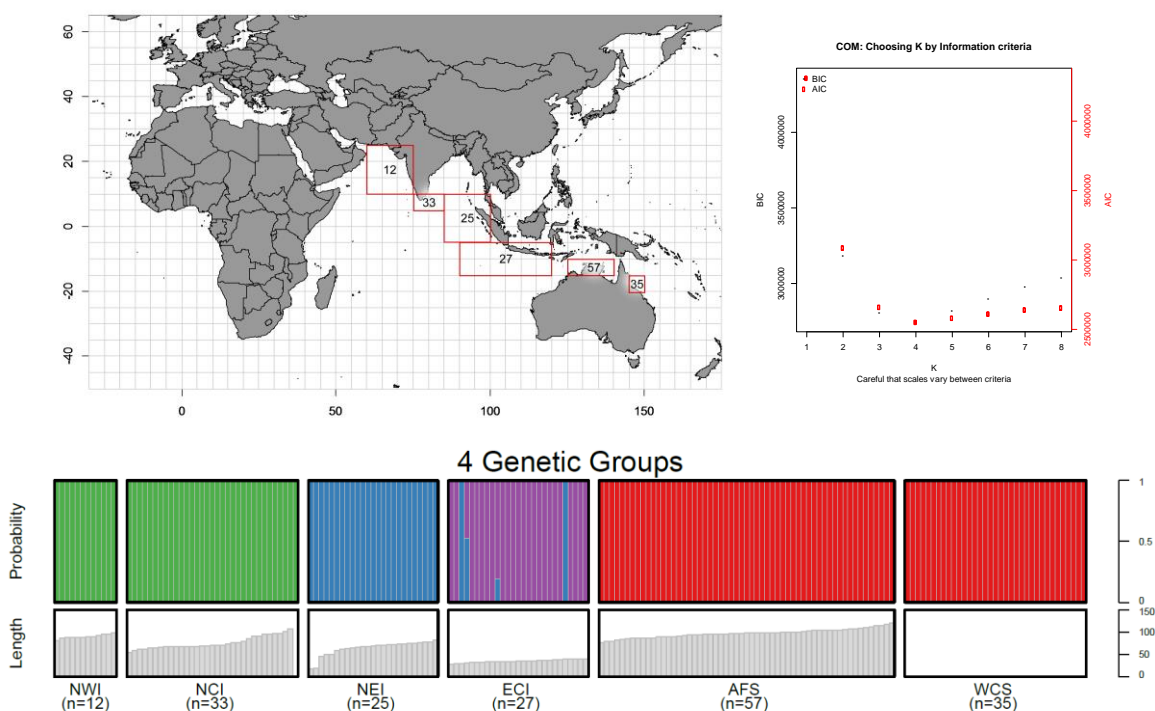


Figure 23. Top Left: Number of samples of narrow-barred Spanish mackerel sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Middle: results of first round of population structure analysis of DArTSeq using StockR for narrow-barred Spanish mackerel for 4 genetic groups. Bottom: results of second round of population structure analysis of DArTSeq using StockR for each potential cryptic species for 4 genetic groups.

## 2 Genetic Groups

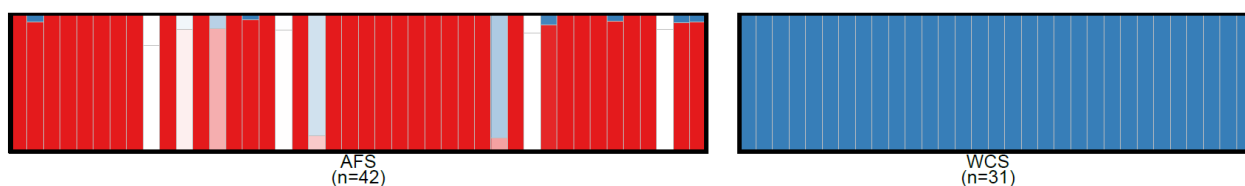


Figure 24. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for the AFS and WCS subset assuming 2 genetic groups.

## Microchemistry

The Spanish mackerel otolith samples analysed were from four sampling locations (Figure 25), referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI), Eastern-Central Indian Ocean (ECI) and Arafura Sea (Arafura) (Figure 2).

Although it was not intended in the original design, samples from different locations were collected during different periods, due to several reasons (see Section 2). The size of fish varied greatly between locations, with the fish from ECI being smallest (<50 cm FL) and those from NWI and Arafura being largest (mostly >80 cm FL) (Figure 25, Table 10). Because of the large variation in fish size, and thus age, they will have been spawned over many years; therefore, differences observed in core signatures among locations may be due, at least in part, to cohort effects.

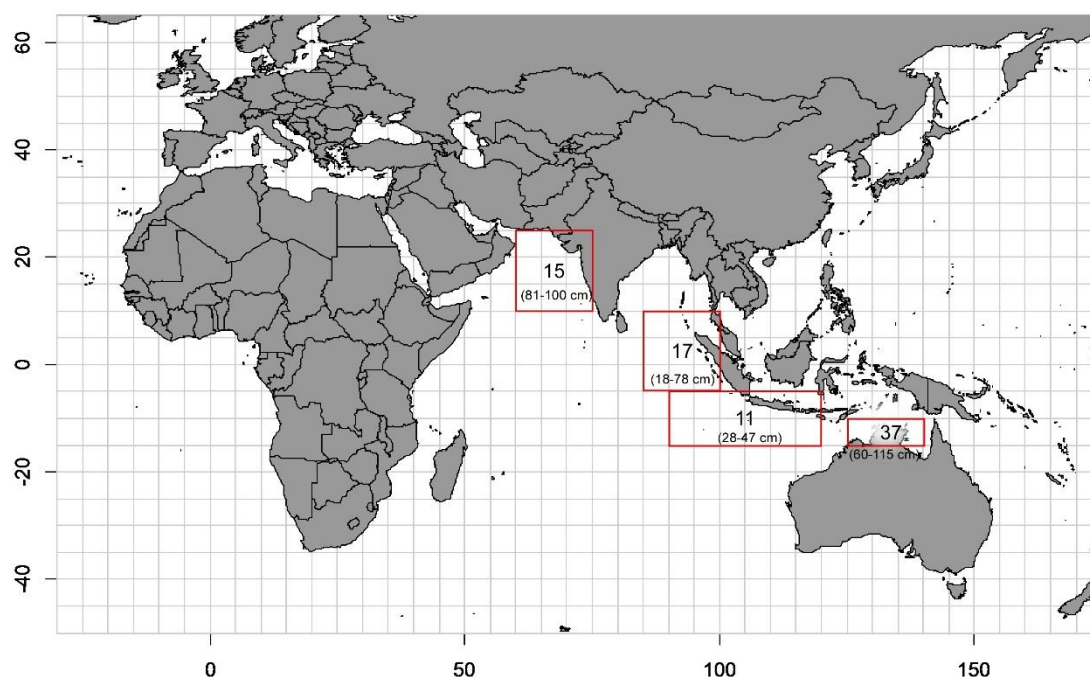


Figure 25. Map showing the number of Spanish mackerel otoliths analysed for each of the sampling locations, referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI), Eastern Central Indian Ocean (ECI) and Arafura Sea (Arafura), and the size range of fish at each location.

**Table 10. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.**

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
North-West Indian Ocean (NWI)	15	September 2018	81-100	2-3
North-East Indian Ocean (NEI)	17	April and November 2018	18-78	0+ -2
Eastern-Central Indian Ocean (ECI)	11	Nov 2018	28-47	0+
Arafura Sea (Arafura)	37	Dec 2017 and Mar-May 2019	80-115	2-6

\* The ranges in ages are for male and females combined (McPherson 1992, McIlwain et al. 2005).

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However, it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

### Core Results

Results from the PERMANOVA and pairwise tests between locations suggest that ECI differs significantly from the other three locations (NWI, NEI and Arafura), and that NWI differs from Arafura.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA supports and helps visualise these findings (Figure 26).

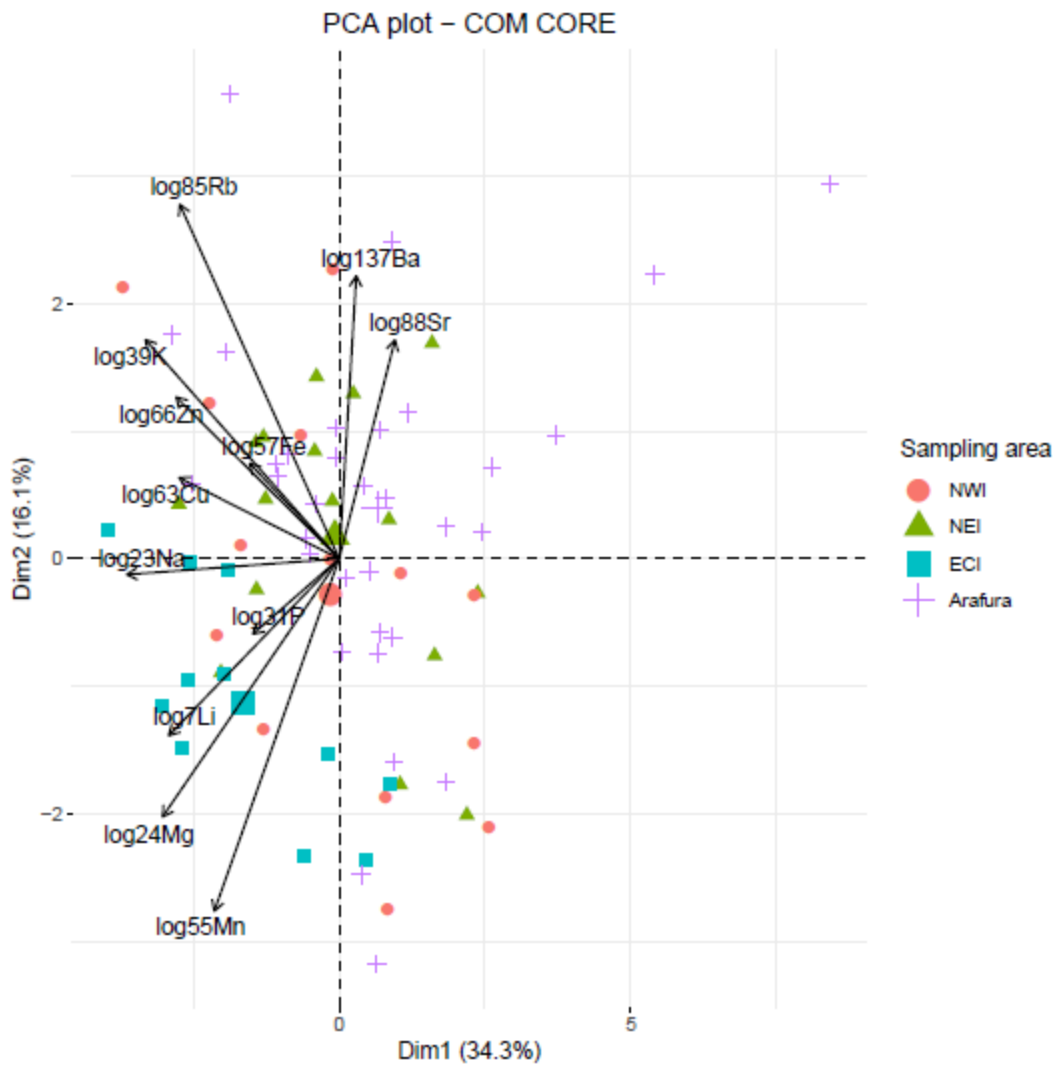


Figure 26. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the Spanish mackerel otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge Results

Results from the PERMANOVA and pairwise tests between locations suggest the edge signatures differ significantly between all locations.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data helps to visualize these findings, with the difference between ECI and the other locations being most obvious (Figure 27).

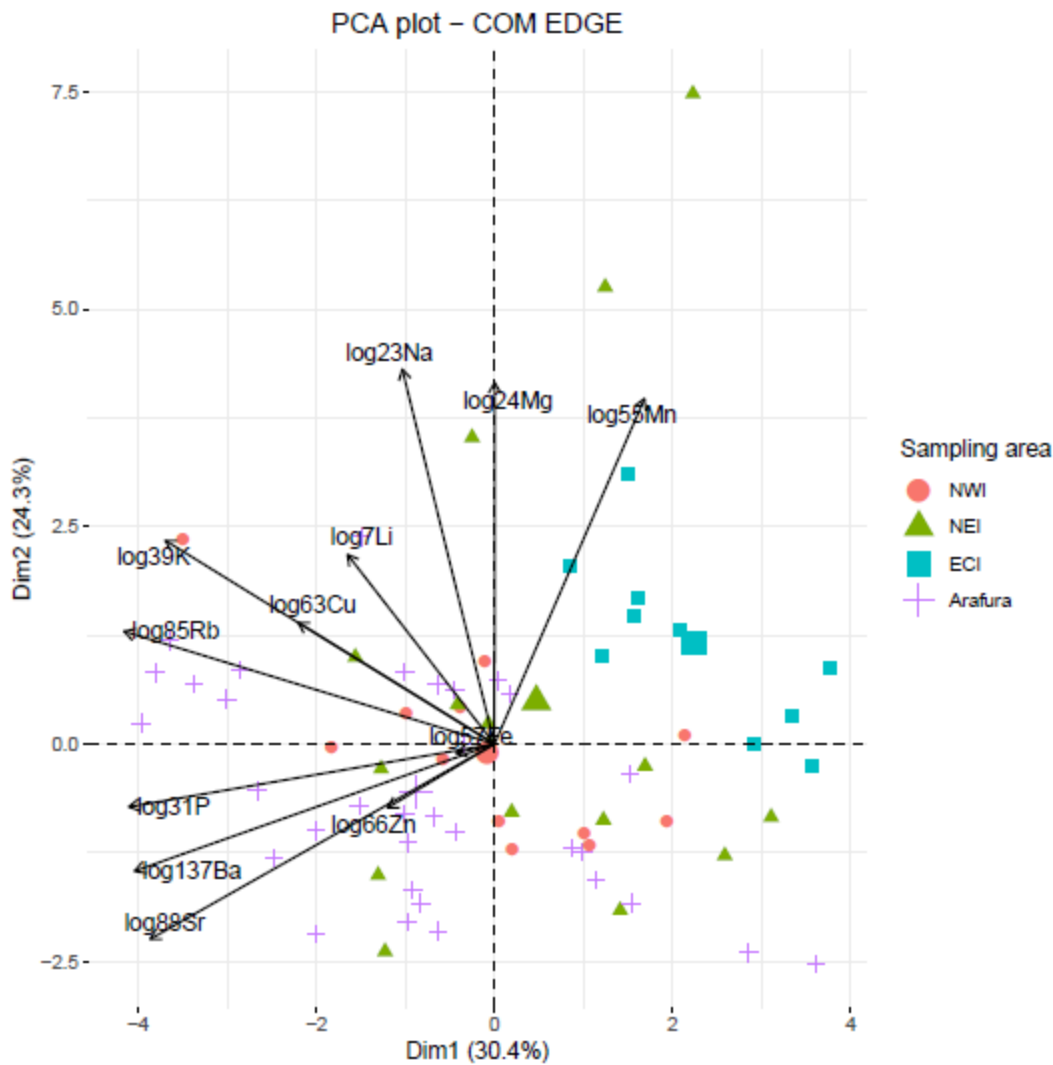


Figure 27. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the Spanish mackerel otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

### Microchemistry summary

The fact that the core signatures showed significant differences between ECI and the other locations may mean fish from ECI originated from a different spawning ground than fish from the other locations. However, because the fish from ECI are much smaller/younger than those from the other locations, and will have been spawned in more recent years, the differences could be due to temporal differences in ocean chemistry at the same spawning location.

The core signatures between fish from NWI and Arafura also differed significantly. These fish are more similar in size so will have been spawned over a similar range of years; thus there is more support for the hypotheses that fish from these two locations originated from different spawning grounds.

The fact that the edge signatures differ between capture locations suggests that otolith chemistry can be useful for classifying fish to these locations.

## 5.5 Tropical tunas

### 5.5.1 Skipjack tuna (*Katsuwonus pelamis*)

#### Genetics

- 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. Amongst these, 524 were selected for genotyping using RADSeq and 393 passed quality control and were included in population structure analyses. After SNP filtering and related individual removal, the final datasets resulted in 368 individuals (Figure 27);
- PCA, Admixture and stockR analyses support the existence of multiple genetic groups, although these do not coincide with the sampling locations. Within the Indian Ocean, this corresponded to NWI being different to the other Indian Ocean locations.

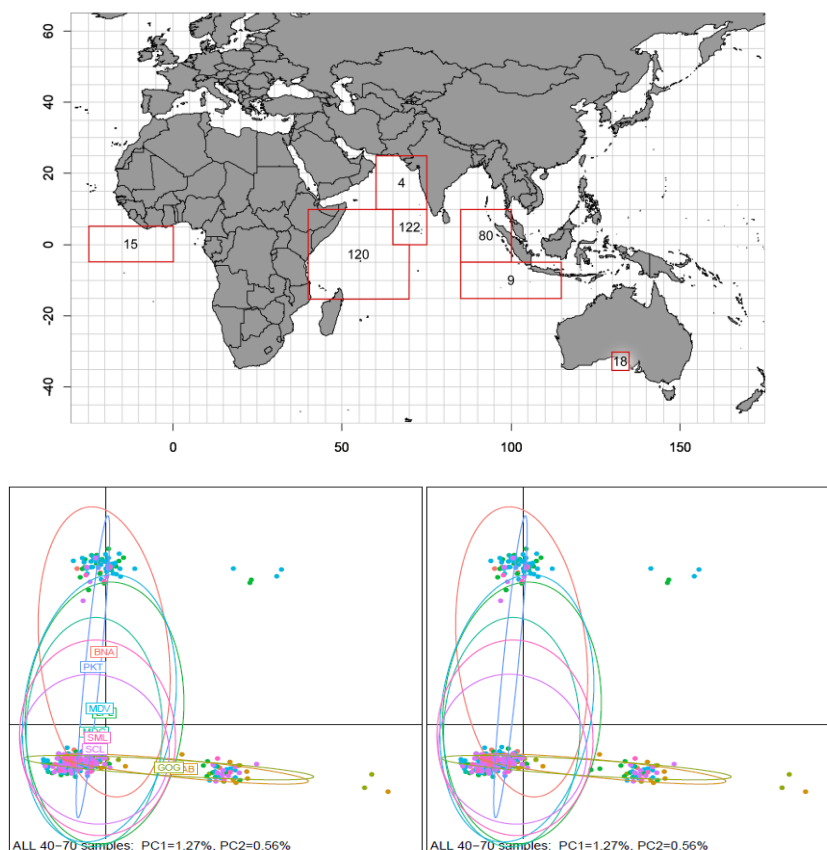


Figure 28. Top: Number of samples of skipjack (*Katsuwonus pelamis*) sequenced using RAD-Seq and included in the analysis by sampling region. Bottom: Principal component analysis results using the filtered datasets. Left plots are identical to right plots but removing labels for clarity.



## Microchemistry

Otoliths from 134 skipjack from 3 locations were analysed: 55 from the western central Indian Ocean (WCI), 38 from Maldives (CIM) and 41 from the north east Indian Ocean (NEI) (Figure 29). The skipjack were all young-of-the-year (YOY, <35 cm FL) and were obtained in two consequent years, 2018 and 2019 (Figure 29, Table 11).

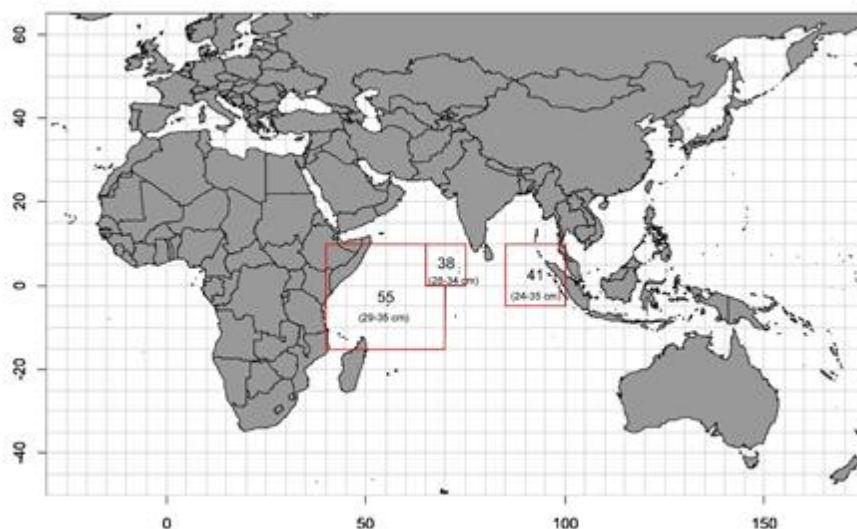


Figure 29. Map showing the number of skipjack otoliths analysed for each of three sampling locations, referred to as Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI), and the size range of fish at each location.

Table 11. Number, sampling period and estimated ages of fish for each of the three sampling locations WCI, CIM and ECI.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
Western Central Indian Ocean (WCI)	55	March 2018 and April 2019	29-35	0+
Maldives (CIM)	38	August 2018 and February 2019	28-34	0+
North East Indian ocean (NEI)	41	April 2018 and November 2018	24-35	0+

\* The ranges in ages are for male and females combined (Eveson et al., 2015).

Laser ablation ICP-MS was performed along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Eight trace elements were measured (see Section 4.5.2) but four were below the limit of detection and therefore discarded for analyses. The remaining four (Ba, Mg, Mn and Sr) were included in analyses of the core data, but only Ba and Sr were included in analyses of the edge data due to inconsistencies found in the distribution of Mg and Mn along the edge for other tropical tuna species.

The spot 200  $\mu\text{m}$  from the core ( $\pm 13\text{-}15$  days of life) was assumed to represent the natal origin signature. The spot at the edge of the otolith was also analysed, since these data reflect the fish's known capture location, and can be used for validation purposes.

### CORE vs Edge Results

Core versus edge signatures were significantly different for Ba and Sr even though fish are estimated to be only about 3-5 months old and assumed not to have changed locations between spawning and capture.

### CORE Results

Based on PERMANOVA and pairwise tests, there were significant differences in core signatures among locations for fish sampled in both 2018 ( $p=0.002$ ) and in 2019 ( $p=0.013$ ). In 2018, core signatures from NEI differed from core signatures of WCI and CIM. In 2019, core signatures from WCI differed from those of CIM.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data helps to visualize these findings, but also shows the large degree of overlap between locations (Figure 30).

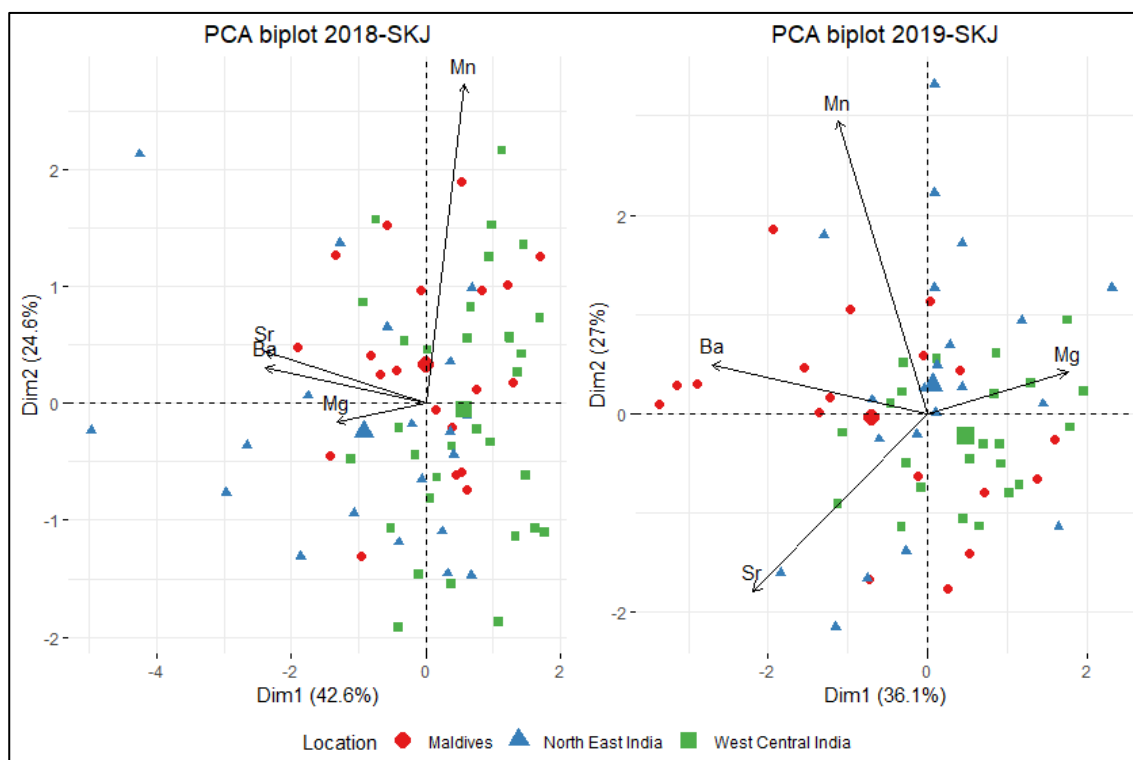


Figure 30. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the skipjack otolith core signatures for the 2018 and 2019 samples respectively. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## EDGE results

Based on PERMANOVA and pairwise tests, there were significant differences in edge signatures among locations ( $p=0.008$ ). Edge signatures differed between CIM and WCI in both years.

A biplot showing edge Ba and Sr signatures of individuals is consistent with and helps to visualize these findings (Figure 31).

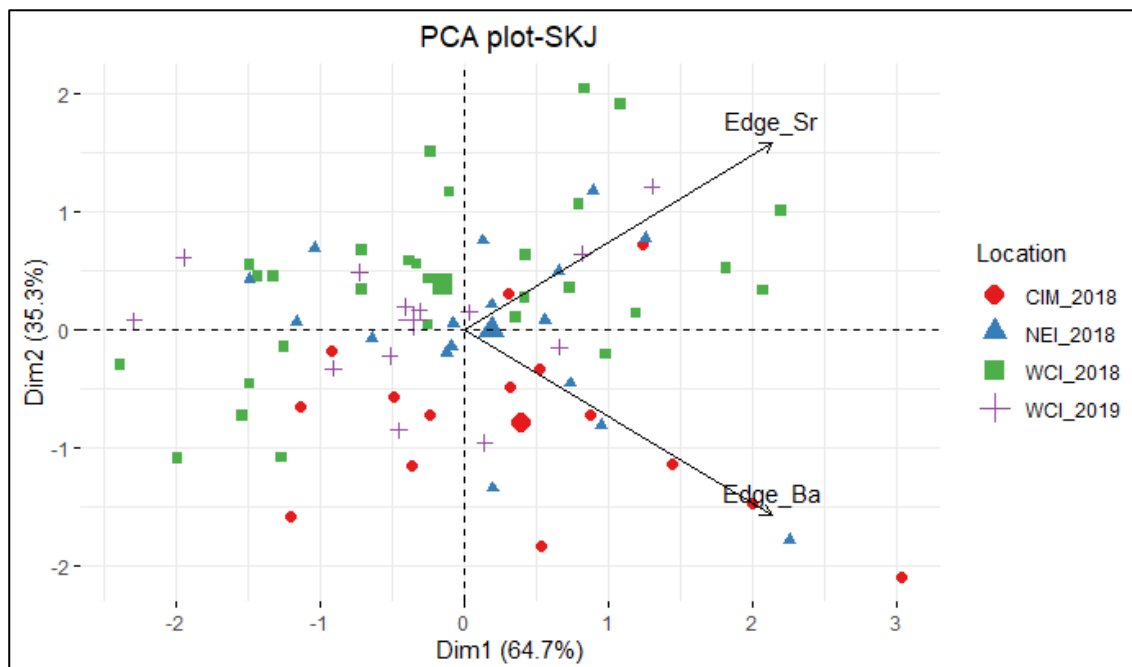


Figure 31. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the skipjack tuna otolith edge signatures. Individuals are coded by their sampling location and year. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Microchemistry summary

As the skipjack were YOY, we assumed they had not moved substantial distances from their spawning location. The significant differences in core and edge otolith signatures may be due to ontogenetic changes during the first months of life, which influenced the chemical composition of the otolith and/or seasonal changes in oceanography over the same period.

For samples collected in 2018, core signatures from NEI differed from those from WCI and CIM suggesting that NEI fish originated from a distinct spawning ground. However, the NEI fish were caught at a different time of year to the CIM fish so the difference between these locations could also be due to seasonal variability in ocean conditions. Similarly, the WCI and CIM samples were caught at different times of the same year, hence seasonal variation in oceanography could be obscuring regional differences.

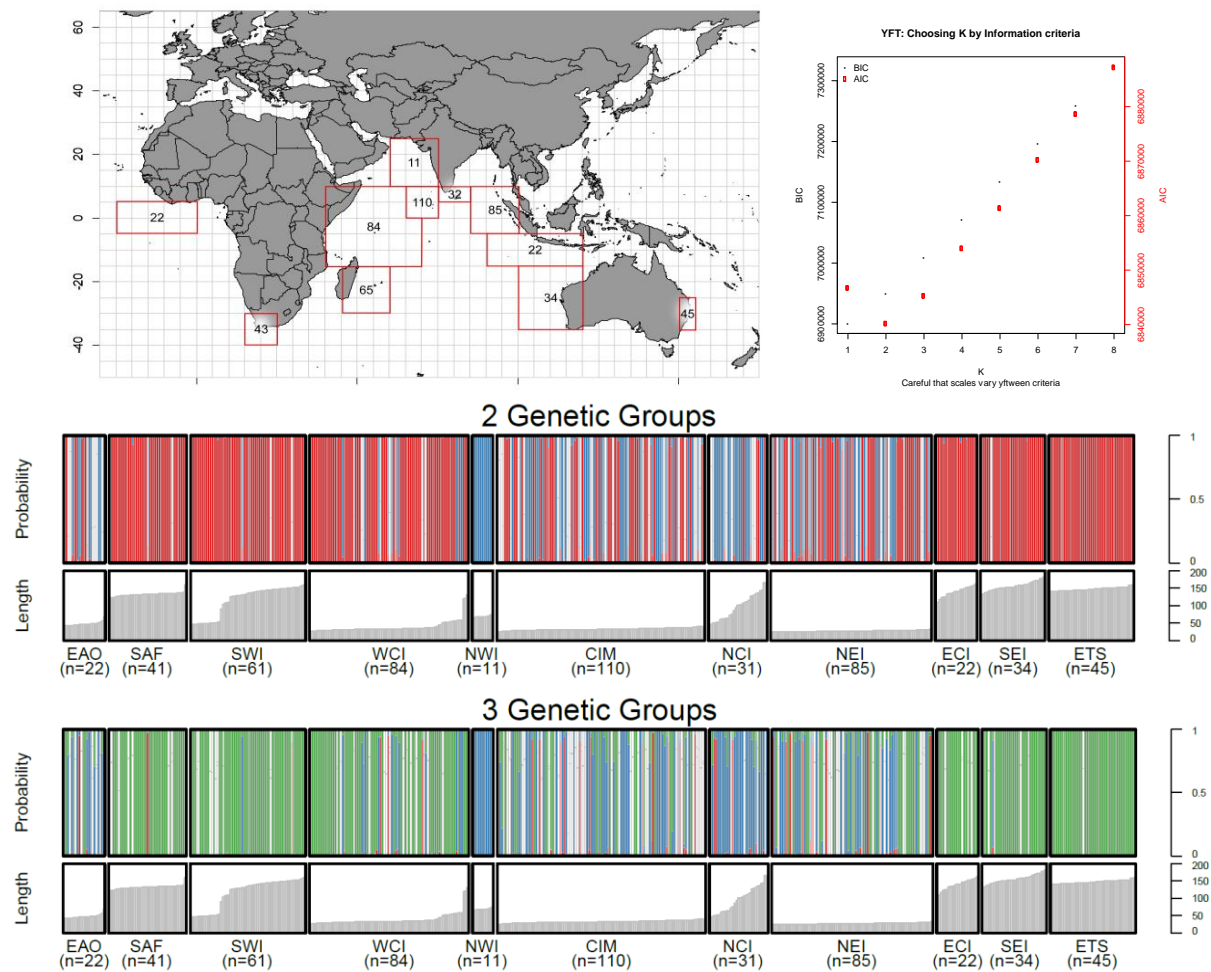
For samples collected in 2019, the only significant differences in core signatures were between WCI and CIM suggesting these fish originated from distinct spawning grounds. The lack of detectable differences in core signatures between NEI otoliths and the other two locations suggests that ocean chemistry does not differ significantly between these locations. However, as the NEI samples were collected at a different time of the same year, seasonal variation in oceanography may again be obscuring regional differences.

Edge signatures differed between WCI and CIM in both years but as they were collected at different times, hence the observed differences could be due to either regional or seasonal differences in ocean chemistry. The lack of detectable difference between WCI and NEI edge signatures, which were collected at the same time, suggests that ocean chemistry did not differ significantly between these locations at the time of capture.

### 5.5.2 Yellowfin tuna (*Thunnus albacares*)

#### Genetics

- The sample coverage for yellowfin was generally very good with a total of 1206 individuals collected from 9 Indian Ocean areas and two outlier locations (east Atlantic Ocean and south-west Pacific Ocean). The samples consist of a mix of YoY fish and mature adults with predominantly YoY in the equatorial regions and adult fish in the sub-tropical and temperate regions;
- A total of 664 samples, which matched sampling design parameters and passed initial DNA quality control checks, were sequenced using DArTSeq and included in the analysis of population structure (Figure 32).
- Model selection criteria using StockR indicate that 2 genetic groupings within the Indian Ocean are more likely than 1, with the likelihood for 1 and 3 groups being similar (Figure 32).
- Samples collected for this project, which are representative of major fishing areas within the Indian Ocean, are likely to be composed of a minimum of two (but likely more) genetically differentiated groups of yellowfin tuna. The most prominent difference is evident between groups sampled north and south of the equator. The fish sampled north of the equator appear to consist of at least two genetic groups. The samples from the southern regions cannot be statistically differentiated into more than a single group, although there is some indication of more than one group in these regions.
- Population analysis of the two outgroups of yellowfin tuna show evidence of restricted gene flow indicating the Indian Ocean is genetically isolated from the Atlantic and Pacific Oceans, which are likely the result of environmentally induced physiological barriers to migration.



**Figure 32.** Top Left: Number of samples of yellowfin tuna (*Thunnus albacares*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (K). Bottom: Individual bar plot probabilities 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 to a specific K group is below 80%. Fish lengths of individuals are plotted below each bar and sample size at location is in brackets.

## Microchemistry

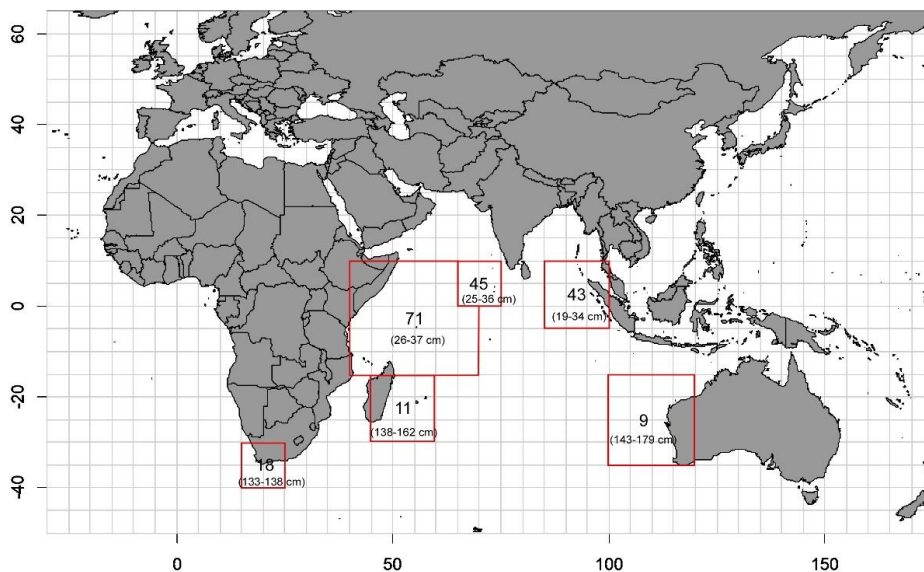
Otoliths from 197 yellowfin tuna from 6 locations (Figure 33) were analysed: 18 from South Africa (SAF), 11 from south west Indian Ocean (SWI), 71 from western central Indian Ocean (WCI), 45 from Maldives (CIM), 43 from north east Indian Ocean (NEI) and 9 from south eastern Indian Ocean (SEI). In this section, we will refer to WCI, CIM and NEI as the three northern locations, and SAF, SWI and SEI as the three southern locations.

All northern samples were young-of-the-year (YOY, 19.5-37.5 cm FL) and southern samples were adults (133-179 cm FL). The samples were obtained in two consequent years, 2018 and 2019 (Figure 33, Table 12). Samples from different locations and years were collected during different months (Table 12). YOY samples were analysed separately for each sampling year.

The otoliths were analysed using laser ablation ICP-MS along the otolith from the core (earliest-deposited material) to the edge (the most recently deposited material). Eight trace elements were measured (see Section 4.5.2), but four were below the limit of detection and therefore discarded for analyses. The remaining four (Ba, Mg, Mn and Sr) were included in analyses of the core data, but only Ba and Sr were included in analyses of the edge data due to inconsistencies found in the distribution of Mg and Mn along the edge.

The spot 65 µm from the core ( $\pm 13$ -15 days of life) was assumed to represent the natal origin signature. The spot at the edge of the otolith was also analysed for all YOY 2018 samples and WCI YOY 2019 samples, since these data reflect the fish's known capture location, and can be used for validation purposes.

Stable isotopes ( $d^{18}O$  and  $d^{13}C$ ) were measured in YOY 2018 samples. This analysis was performed at the Environmental Isotope Laboratory, Dept. of Geosciences, at the University of Arizona using an Automated carbonate preparation device (KIEL-III) coupled to a gas-ratio mass spectrometer (Finnigan MAT 252). The drill path was 400 x 600 µm along the dorsal and ventral arms.



**Figure 33.** Map showing the number of yellowfin tuna otoliths analysed for each of six sampling locations, referred to as South Africa (SAF), South-West Indian Ocean (SWI), Western Central Indian Ocean (WCI), Maldives (CIM), North-East Indian Ocean (NEI), and south east Indian Ocean (GAB); and the size range of fish at each location.



**Table 12. Number, sampling period and estimated ages of fish for each of the six sampling locations SAF, SWI, WCI, CIM, NEI and SEI.**

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
South Africa (SAF)	18	March 2018	133-138	5+
South West Indian Ocean (SWI)	11	March 2018 and Feb 2019	138-162	5+
West central Indian Ocean (WCI)	71	March-April 2018 and April 2019	26-37	0+
Maldives (CIM)	45	August 2018 and February 2019	25-36	0+
North east Indian (NEI)	43	April 2018 and November 2018	19-34	0+
South east Indian (SEI)	9	May 2019	143-179	5+

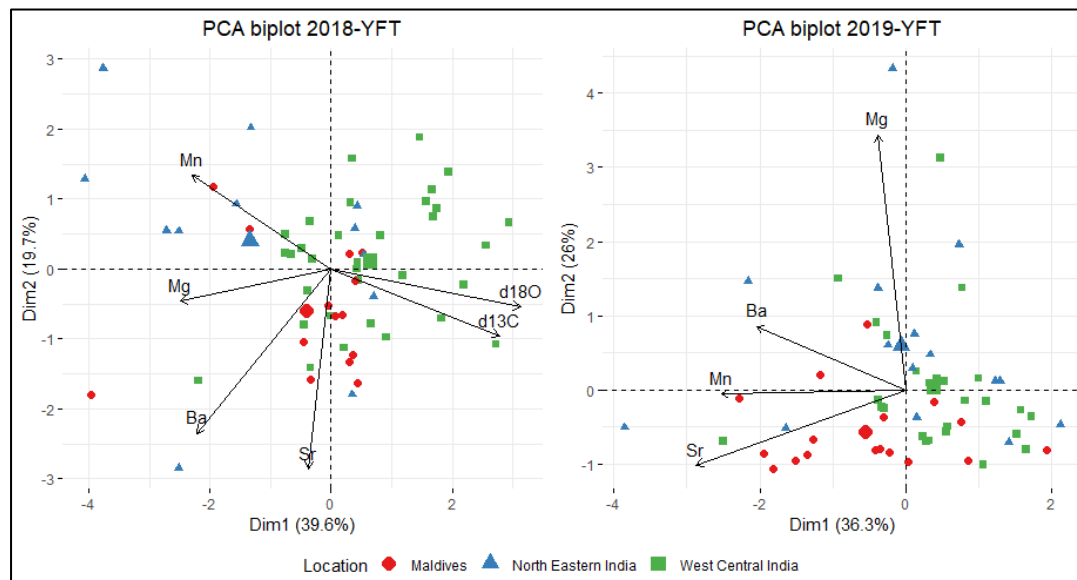
\* The ranges in ages are for male and females combined (Eveson et al., 2015).

### Core vs Edge Results - YOY

Core versus edge signatures for YOY fish were significantly different for Ba and Sr in both years even though these fish are assumed not to have changed locations between spawning and capture.

### CORE Results - YOY

Based on PERMANOVA and pairwise tests of core signatures of YOY, all pairs of locations differed significantly in both 2018 ( $p=0.001$ ) and 2019 ( $p=0.001$ ). A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data is consistent with and helps to visualize these findings (Figure 34).



**Figure 34. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the yellowfin tuna otolith core signatures for the 2018 and 2019 samples respectively. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.**

## CORE Results - Adults

Based on PERMANOVA and pairwise tests, core signatures of adult yellowfin tuna did not differ significantly among sampling locations ( $p > 0.05$ ). A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data is consistent with and helps to visualize these findings (Figure 35). A biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the adult yellowfin tuna otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. SAF= South Africa, SWI= South West Indian Ocean and SEI= South East Indian Ocean).

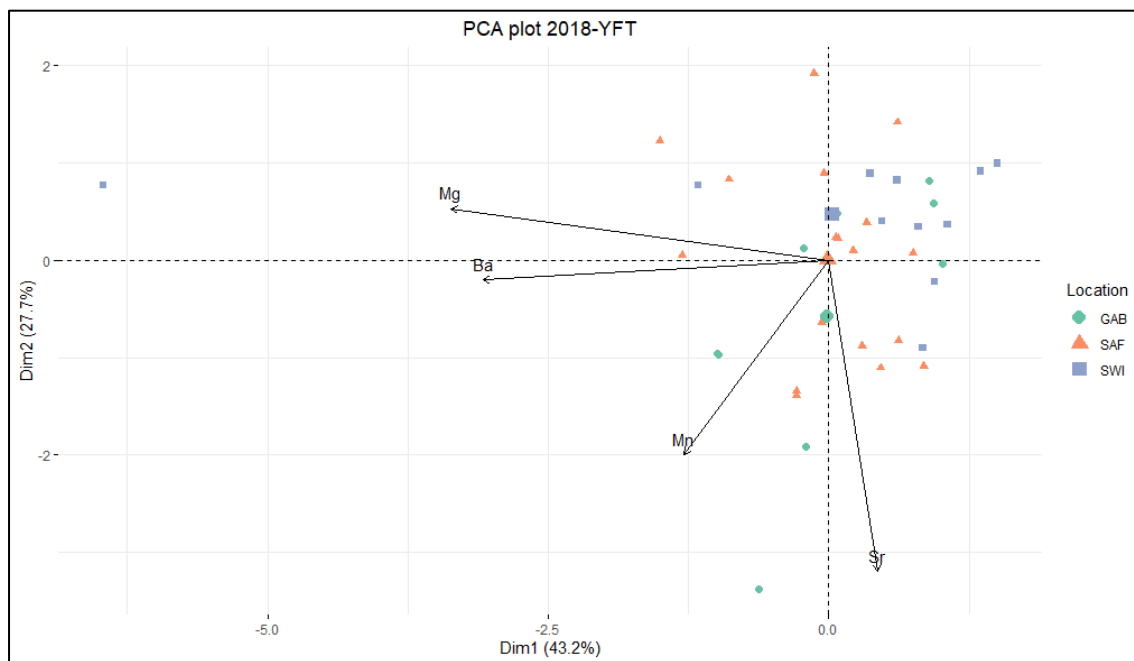


Figure 35. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the adult yellowfin tuna otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. SAF= South Africa, SWI= South West Indian Ocean and SEI= South East Indian Ocean). (Note SEI is shown as “GAB” in this figure).

## EDGE results – YOY

Results from the PERMANOVA and pairwise tests suggest that edge signatures for CIM differed from other locations ( $p = 0.001$ ). A biplot showing edge Ba and Sr signatures of individuals is consistent with and helps to visualize these findings (Figure 36).

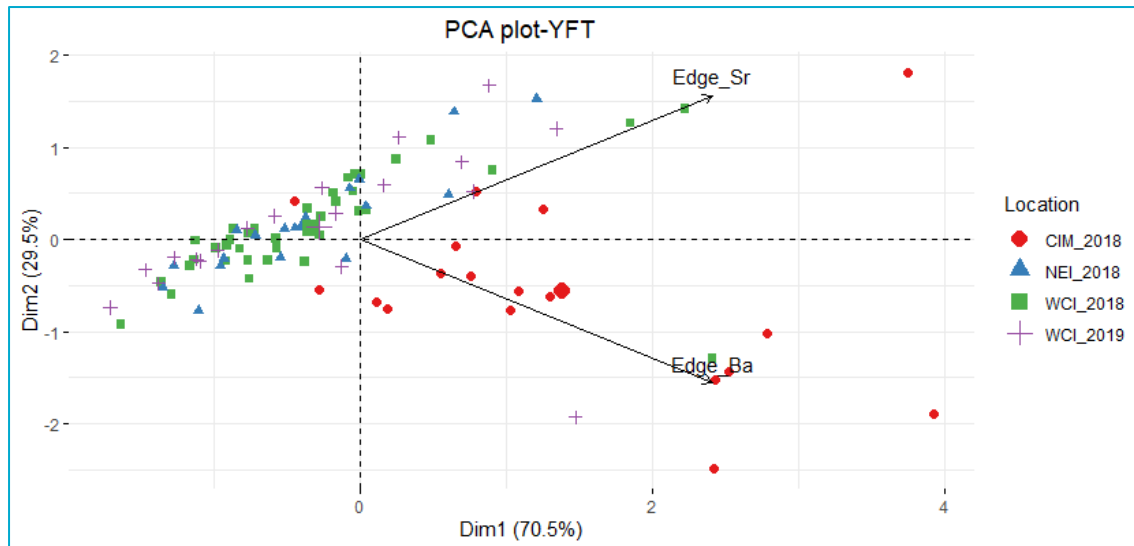


Figure 36. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the yellowfin tuna otolith edge signatures. Individuals are coded by their sampling location and year. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

### Microchemistry summary

The fact that YOY core signatures differed between the three capture locations in both 2018 and 2019 provides evidence that the fish originated from distinct spawning grounds. However, as the CIM samples in 2018 and the NEI samples in 2019 had been collected at different times of the year compared to the other two locations, the differences could be due, at least in part, to seasonal variability in ocean conditions.

Adult yellowfin tuna core signatures did not differ among capture locations, which suggests that all the fish had a common origin. However, since the adult samples were collected from a wide range of cohorts, differences among years could be larger than differences among locations.

The fact that edge signatures of YOY fish from the WCI and NEI were not different suggests that the ocean chemistry did not differ significantly between these locations at the time of capture (April). The edge signatures of YOY from CIM did differ significantly from the other two northern locations. As the CIM fish were collected at a different time of the year, the differences may be due to seasonal differences in ocean chemistry.

Core and edge signatures of YOY were different in all locations in both years, which could be due to ontogenetic changes during the first months of life strongly influencing the chemical composition of the otolith, or seasonal variation in oceanography over that time period.

### 5.5.3 Bigeye tuna (*Thunnus obesus*)

#### Genetics

- 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 8 Indian Ocean sampling regions and two outlier locations (east Atlantic and southwest) were collected. Amongst these, 496 were selected for genotyping using RADSeq and 472 passed quality control and were included in population structure analyses (Figure 37);
- PCA, Admixture and stockR analyses support the existence of three genetic groups across the sampled area (Atlantic, Indian and Pacific Oceans) but no evidence of intra-oceanic differentiation was detected inside the Indian Ocean (Figure 37).

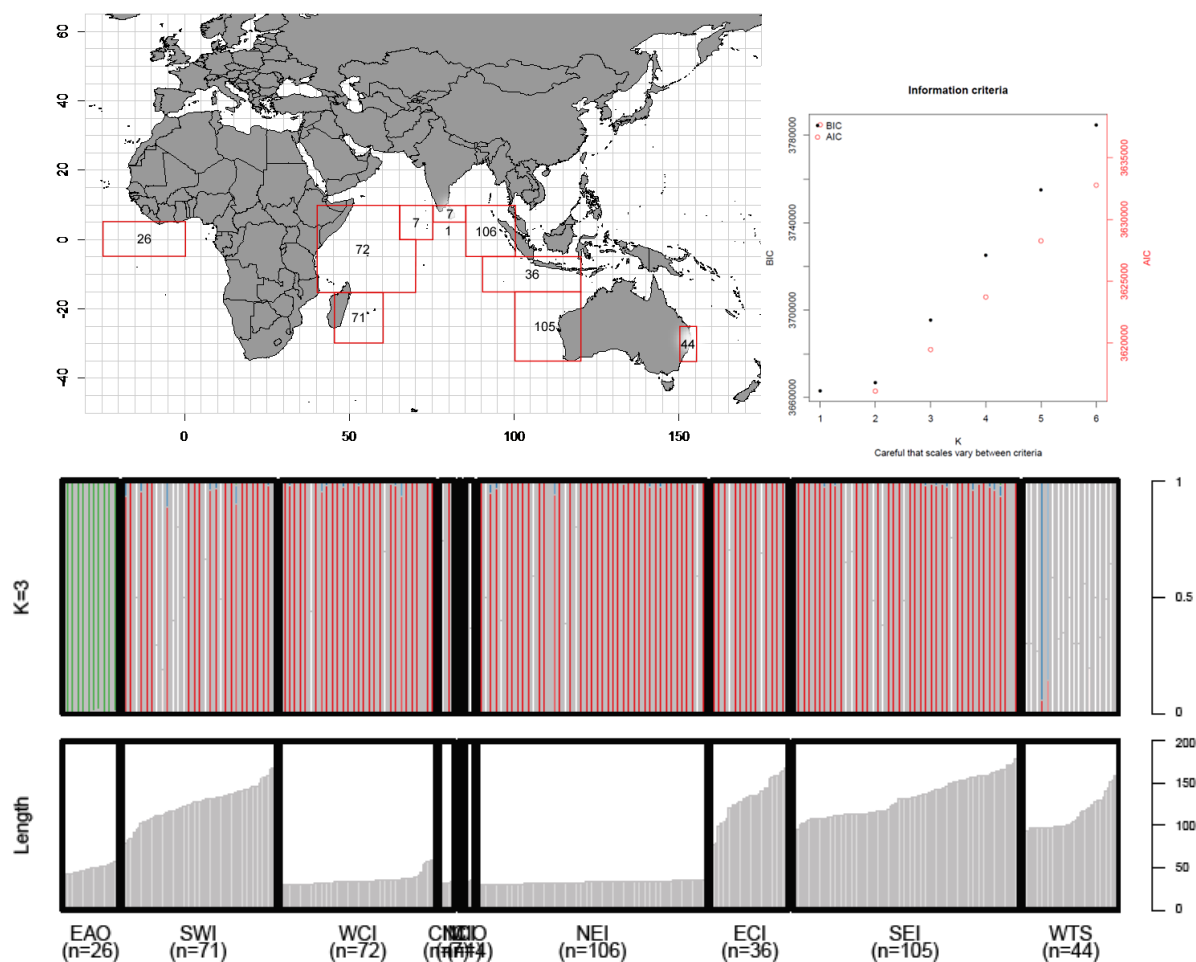


Figure 37. Top: Distribution of samples of bigeye (*Thunnus obesus*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: Results of population structure analysis of DArTSeq using StockR for bigeye tuna for 3 genetic groups

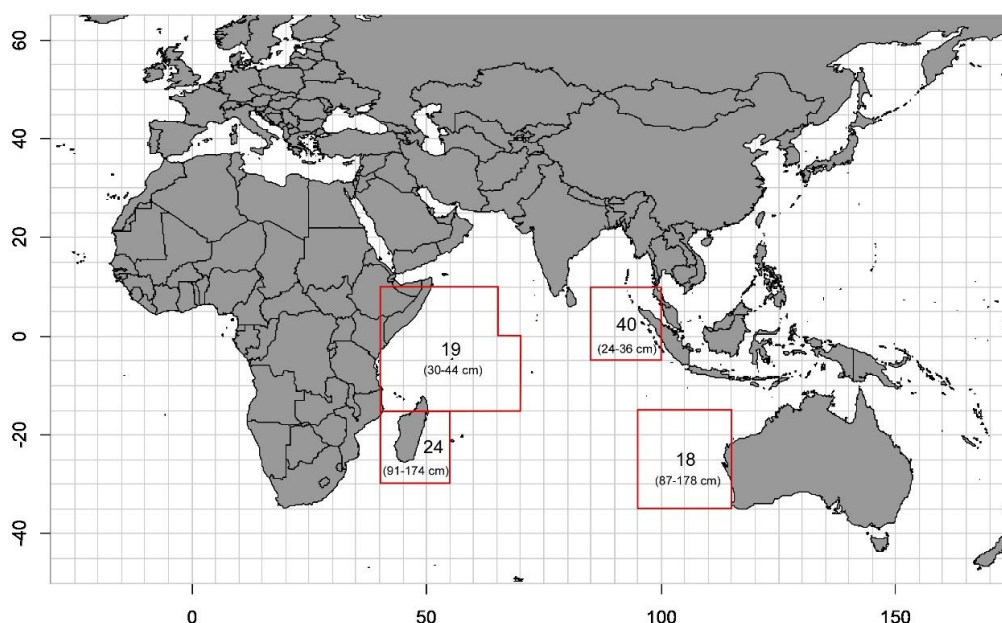
## Microchemistry

The bigeye otolith samples analysed are from four sampling locations (Figure 38), referred to as Western Central Indian Ocean (WCI), North-East Indian Ocean (NEI), South-West Indian Ocean (SWI) and South-East Indian Ocean (SEI) (see Figure 2). In this section, we will refer to WCI and NEI as the two northern locations, and SWI and SEI as the two southern locations.

Although it was not intended in the original design, samples from different locations were collected during different periods, due to several reasons (see Section 2). All fish from the two northern sites were juveniles (24-44 cm FL) (Zudaire et al. 2016) while those from the southern sites were larger (87-178 cm FL) (Figure 38, Figure 37, Table 13). The two northern locations are known spawning sites for bigeye (Nishikawa 1985, Stequert and Marsac 1989, Suman et al. 2015), so the aim of the sampling design was to obtain a spawning ground signature for these locations, then see whether the core signatures from the larger fish sampled at the southern locations corresponded to either of these spawning sites.

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However, it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.



**Figure 38.** Map showing the number of bigeye otoliths analysed for each of four sampling locations, referred to as Western Central Indian Ocean (WCI), North-East Indian Ocean (NEI), South-West Indian Ocean (SWI) and South-East Indian Ocean (SEI); and the size range of fish at each location.

**Table 13. The sampling period for bigeye otoliths analysed, and the estimated ages for bigeye from each of the sampling locations.**

Location	N	Sampling dates	FL (cm)	**Estimated age range (years)
Western Central Indian Ocean (WCI)	19	February-April 2018 (primarily)	30-44	0+
North-East Indian Ocean (NEI)	40	April and November 2018	24-36	0+
South-West Indian Ocean (SWI)	24	August-October 2017 (primarily)	91-174	2-15
South-East Indian Ocean (SEI)	18	May 2019	87-178	2-15

\*\* based on results from Eveson et al. (2015), Farley et al. (2006), Sardenne et al. (2015).

### Core vs Edge Results

Core and edge signatures were significantly different for most elements. This is clear in the otolith data for the northern locations, where even though fish are estimated to be only about 3 months old and assumed not to have changed locations between spawning and capture, their core and edge signatures are still significantly different.

### Core Results

Based on PERMANOVA and pairwise tests, there were significant differences in the core signatures among locations:

- Core signatures do not differ between the two northern (spawning) locations;
- Core signatures do not differ between the two southern locations;
- However, core signatures do differ between the northern locations and the southern locations.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data confirms and helps to visualize these findings (Figure 39).

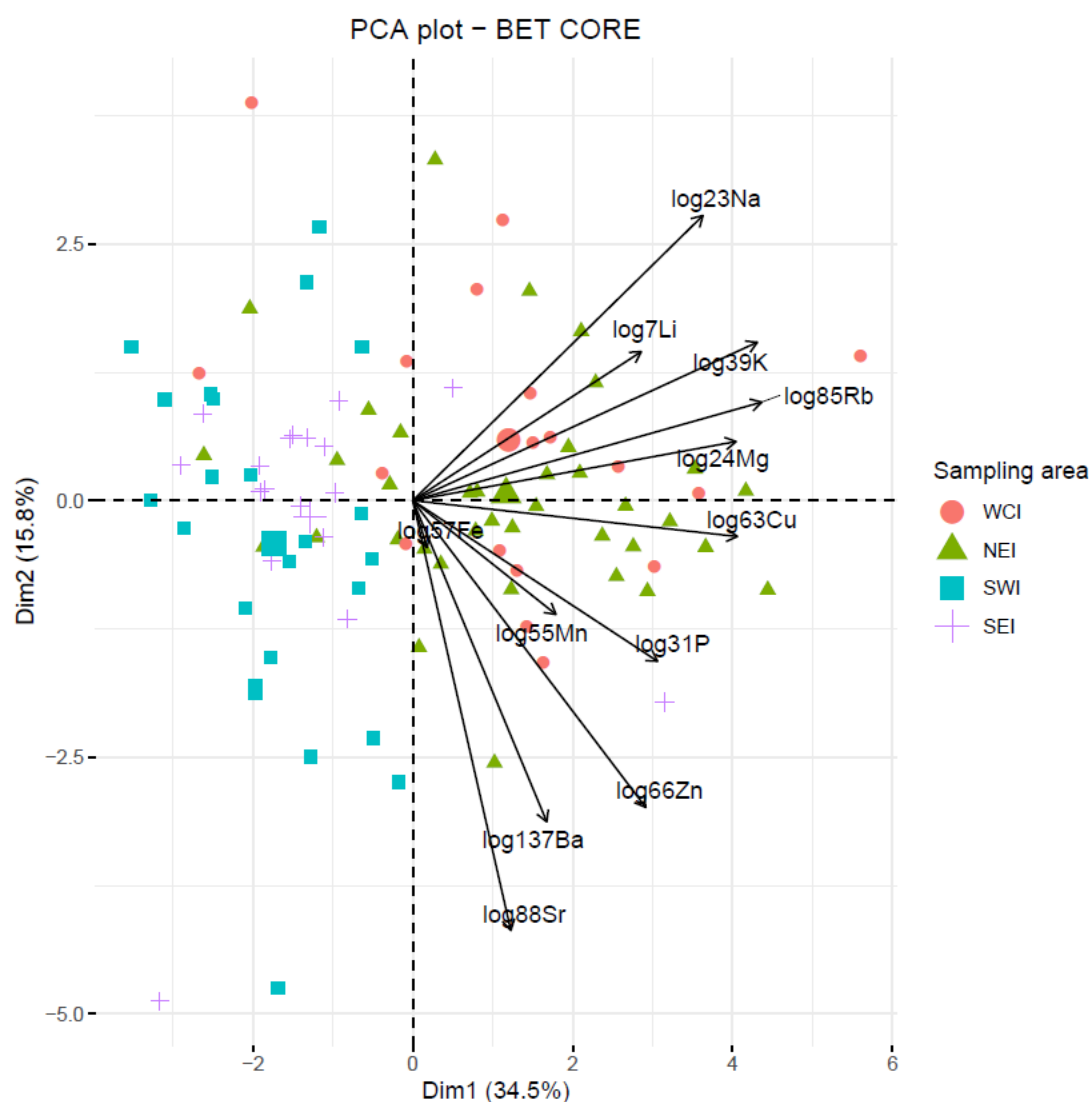


Figure 39. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the bigeye otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge results

Based on the PERMANOVA and pairwise tests, there were significant differences in edge signatures among locations:

- Edge signatures do not differ between the two northern (spawning) locations;
- Edge signatures do not differ between the two southern locations;
- However, edge signatures do differ between the northern spawning locations and the southern locations.

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data confirms and helps to visualize these findings (Figure 40).



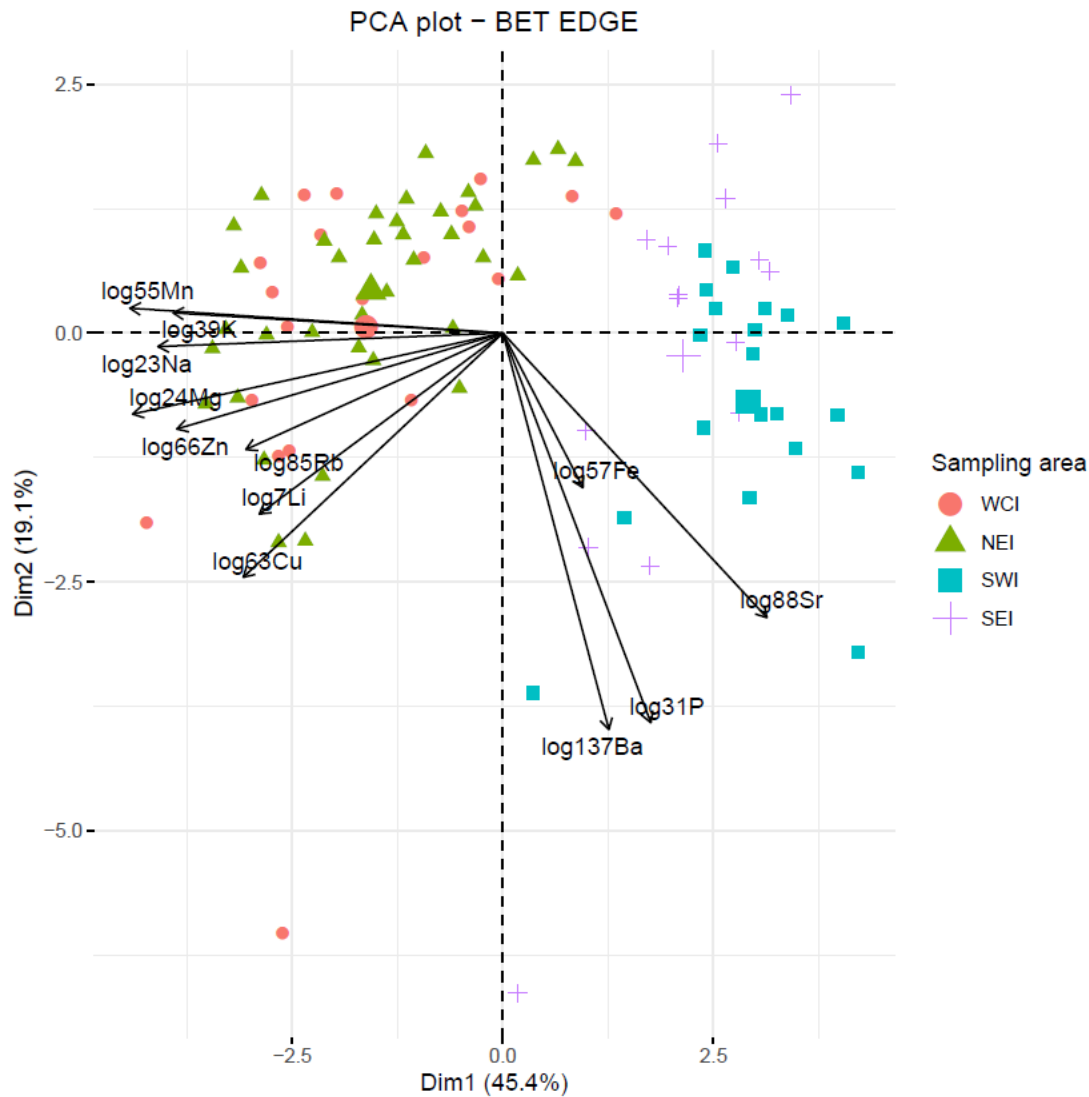


Figure 40. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the bigeye otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

### Microchemistry summary

The fact that the core signatures do not differ between the two northern spawning locations suggests that either:

- the ocean chemistry does not differ significantly between these locations, or
- fish were spawned in a single location and their larvae were transported within the first few days of life to separate locations.

We were able to investigate the first hypothesis by examining whether the edge signatures differed between fish from the two locations. We found that they did not, suggesting that the ocean chemistry is indeed similar. Unfortunately, this means that these data are not useful for distinguishing which of the two northern spawning locations fish sampled in the

south originated from. The fact that the core signatures do not differ between the two southern locations suggests that fish from these regions were spawned in waters with similar ocean chemistry.

Interestingly, the core signatures for the southern locations differed significantly from the core signatures of the northern spawning locations, indicating either: (i) the fish from the south were not spawned in either of the northern locations; or (ii) they were in fact spawned in one of the northern locations but the ocean chemistry was very different in the years they were spawned (estimated to cover a wide range of years from the mid-2000s to the mid-2010s), than in the years that the fish from the northern locations were spawned (2017-2018). If the latter hypothesis was true, we might have expected the core signatures of the southern fish to have much greater variability than the northern fish but still to overlap substantially, which is not the case. However, we cannot rule out that the ocean chemistry changed significantly between the two periods.

The fact that the edge signatures do not differ between the two southern locations suggests the ocean chemistry does not differ significantly between these locations. The edge signatures do differ significantly between the northern and southern locations. There is the potential, however, that this could in part be due to ontogenetic effects as the fish from the north are all young of the year whereas fish from the south are estimated to be ages 2-15.

## 5.6 Temperate tunas

### 5.6.1 Albacore (*Thunnus alalunga*)

#### Genetics

- A total of 288 individuals caught in the Northeast and Southeast Atlantic, Southwest Pacific and Indian Ocean were sampled, measured, sexed, and sequenced using DArTSeq on 103,676 SNPs (Figure 40);
- The population analysis based on AIC indicates poor genetic grouping (Figure 42) due to very low global genetic differentiation also confirmed through low pairwise  $F_{st}$  values (Table 14).
- By contrast, PCA, StockR, assignment, and hierarchical Bayesian results revealed a minimum of three genetic clusters corresponding to the three oceans: Atlantic, Indian and Pacific Oceans.
- Sampling areas in Indian Ocean (southwest and Indonesia) were genetically undifferentiated, as well as those in the Southwest Pacific (Tasmania and Australia).
- Alternatively, clustering with STRUCTURE and Puechmaille approach (adapted on uneven sampling), pairwise  $F_{st}$  including outliers (Table 14, bottom table) and evolutionary history (reconstructed with  $\delta a \delta i$ ) indicated 4 potential groups: North Atlantic, South Atlantic (South Africa), Indian Ocean (southwest and Indonesia), and Southwest Pacific (Australia and Tasmania) (Figure 42).
- South African individuals were more similar to the samples from the North Atlantic samples than to those from the Indian Ocean. Nevertheless, the low number of individuals genotyped (19 individuals) and then retained after bioinformatic filtration (12 individuals) in North Atlantic is insufficient to provide an accurate answer on the differentiation.

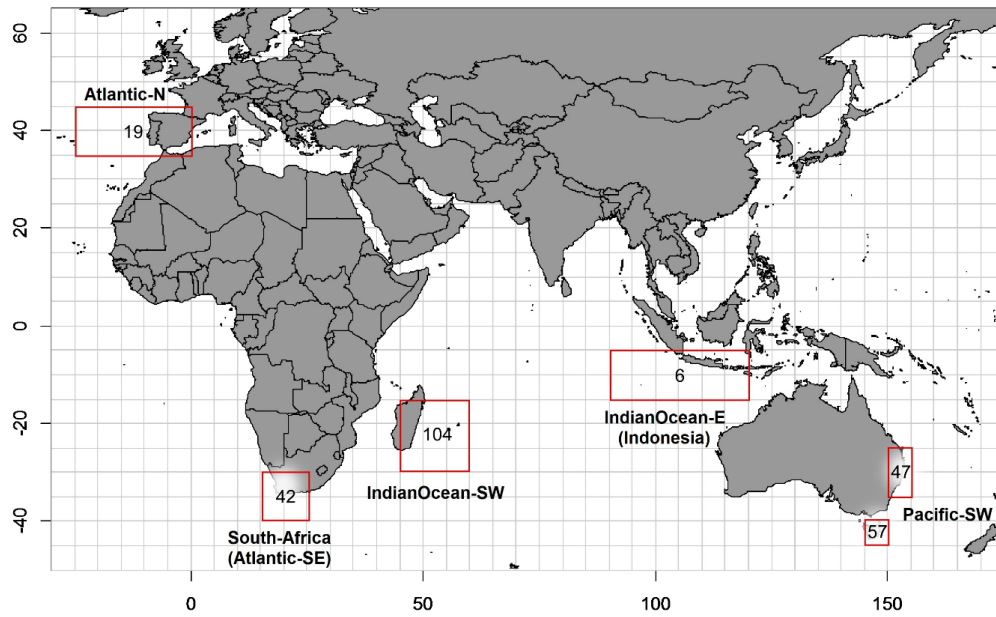
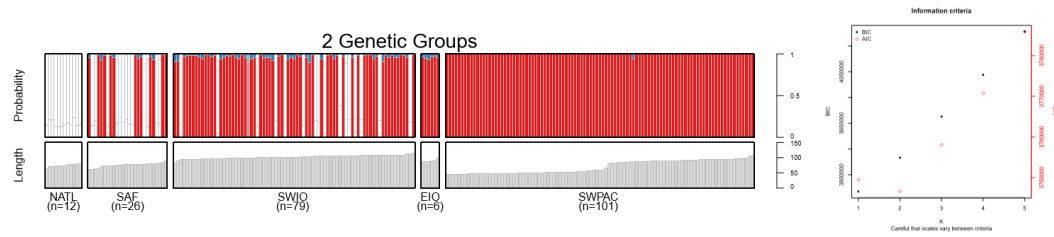
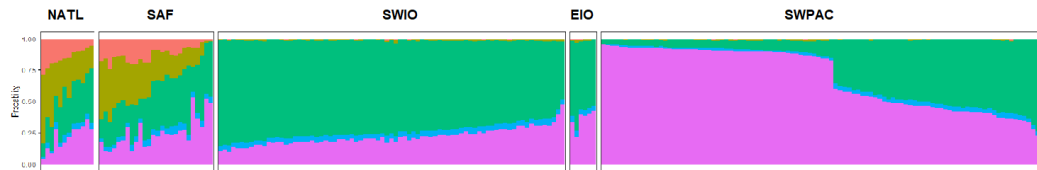


Figure 41. Map of albacore samples genotyped with number of individuals for each of the 5 main sampling regions: Northeast Atlantic (Atlantic-N), Southeast Atlantic (Atlantic-SE or South-Africa), Southwest Indian Ocean (IndianOcean-SW), East Indian Ocean (IndianOcean-E or Indonesia), and Southwest Pacific (Pacific-SW).

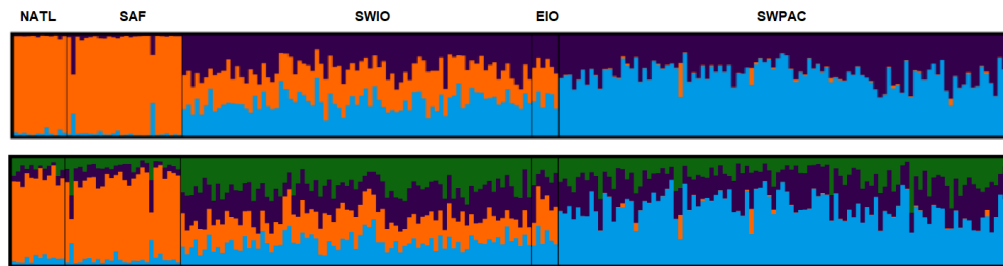
**A) Best K = 2 from AIC and stockR**



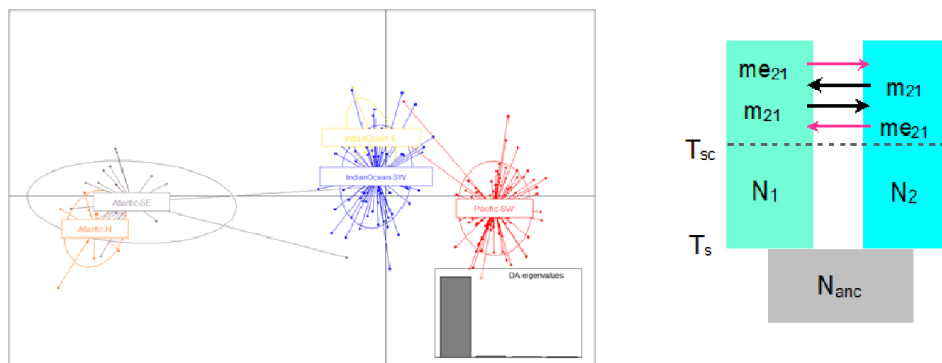
**B) Best K = 3 from AssignPop (genetic and morphometric data)**



**C) Best K = 3 and 4 from FST and hierarchical bayesian population analysis (STRUCTURE)**



**D) Best = 4 from evolutionary history analysis and DAPC**



**Figure 42.** Results of population structure analysis of DArTSeq using StockR (A), AssignPop (B), STRUCTURE and  $F_{st}$  (C), and DAPC and Evolutionary history from observed joint allele frequency spectrum (JAfS) presented the fittest model (D) for albacore samples over 20,220 SNPs (neutrals and with outliers) and 224 individuals. Each color represents an area. Bottom right: Information criterion used to assess one likelihood of different numbers of genetic groups ( $k$ ), lower indicating more likely.  $N_{anc}$ : ancient population size.  $N$ : population size.  $m_{12}$ ,  $m_{21}$ : migration rate from the population 1 to the population 2 and vice versa.  $me_{12}$ ,  $me_{21}$ : reduced effective migration rates for loci influenced by selection when semi-permeability was assumed.  $T_{sc}$ : duration of the secondary contact.  $T_s$ : time from the initial split to present (Append Figure 69).

Table 14. Top table: Pairwise  $F_{st}$  below diagonal with p-value (<0.05) in bold based on 1,000 bootstrap of albacore samples over 20,220 SNPs (including outliers) and 224 individuals with Arlequin software. Bottom table: Pairwise  $F_{st}$  below diagonal with p-value (<0.05) in bold based on 1,000 bootstrap of albacore samples over 20,038 SNPs (excluding 182 outliers) and 224 individuals with Arlequin software.

	Atlantic-N	Atlantic-SE	IndianOcean-SW	IndianOcean-E	Pacific-SW
Atlantic-N	<b>0</b>				
Atlantic-SE (South Africa)	<b>0.00609</b>	<b>0</b>			
IndianOcean-SW	<b>0.01586</b>	<b>0.01355</b>	<b>0</b>		
IndianOcean-E	<b>0.02059</b>	<b>0.01562</b>	0.00511	<b>0</b>	
Pacific-SW	<b>0.03033</b>	<b>0.0276</b>	<b>0.00426</b>	<b>0.01149</b>	<b>0</b>

	Atlantic-N	Atlantic-SE	IndianOcean-SW	IndianOcean-E	Pacific-SW
Atlantic-N	0.00000				
Atlantic-SE	0.00468	0.00000			
IndianOcean-SW	<b>0.00964</b>	<b>0.00758</b>	0.00000		
IndianOcean-E	<b>0.01554</b>	<b>0.01154</b>	0.00344	0.00000	
Pacific-SW	0.01761	<b>0.01519</b>	<b>0.00253</b>	0.00485	0.00000

## Microchemistry

Otoliths from eighty albacore were analysed (55-113 cm FL). Twenty were collected from the southwest Tasman Sea (SWT), 40 from the south west Indian Ocean (SWI) and 20 from South Africa (SAf) in the south west Atlantic Ocean at 2 sites: Hout Bay (SAS) and Saldana Bay (SAN) (Figure 43). The samples were collected between February 2018 and February 2019 (Table 15).

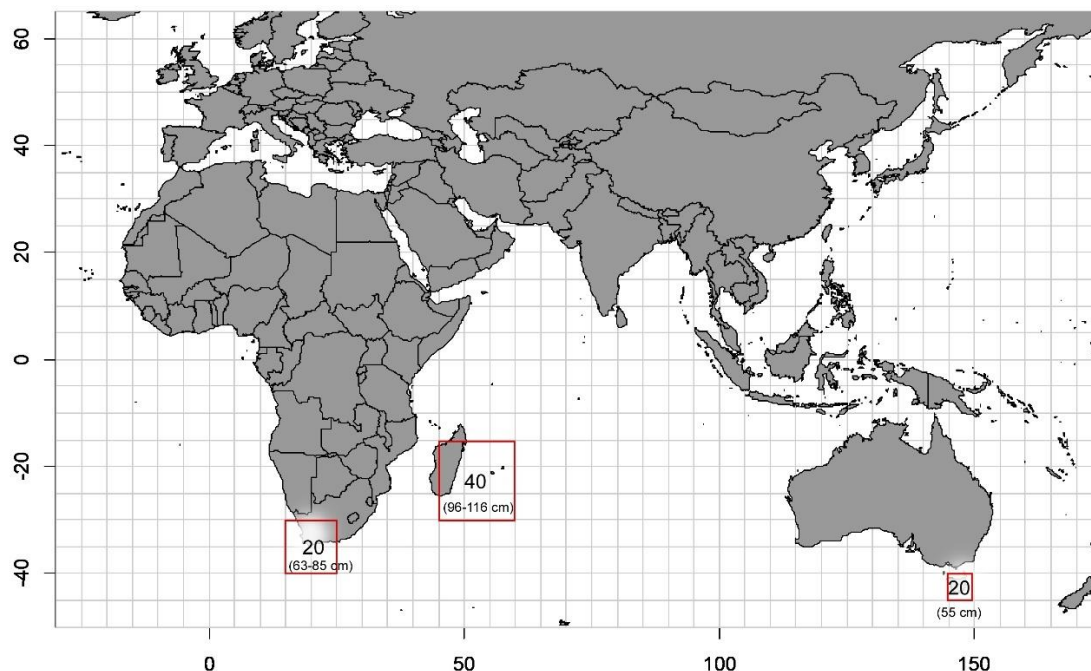


Figure 43. Map showing the number of albacore otoliths analysed for each of the three sampling locations, referred to as south west Indian Ocean (SWI), South Africa (SAf) and south west Tasman Sea (SWT); and the size range of fish at each location.

Table 15. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
South west Indian Ocean (SWI Feb 18)	12	Feb 2018	96-104	7-10
South west Indian Ocean (SWI May 18)	8	May 2018	98-113	7-15+
South west Indian Ocean (SWI 2019)	20	Dec 2018	96-116	7-15+
South Africa – Hout Bay (SAS)	6	Mar-Apr 2018	63-85	2-5
South Africa – Saldana Bay (SAN)	14	Mar-Apr 2018	63-85	2-5
South west Tasman Sea (SWTS)	20	Feb 2019	55	1

\* The ranges in ages are for male and females combined (Xu et al. 2014).

The otoliths were analysed at Montpellier University, Plateforme AETE-ISO (France) using LA-ICP-MS. The laser ablated along a transect between the otolith core and the edge, therefore acquiring a chemical signal from material deposited throughout the life of the fish.

The portion of the otolith transect near the core was examined to identify the chemical signatures deposited during the first weeks of life. These are most likely to reflect the fish spawning origins, i.e. the physio-chemical characteristics of the water masses in which



spawning occurred. The first laser point, on the core, was not included, in order to avoid any maternal influence on otolith composition. The mean of the next 3 points, between 10 and 40 microns after the core and corresponding to the first weeks of life, was used for analysis.

Fifteen chemical elements were measured. Those elements where more than 75% of results were above the limit of detection were retained for further analysis.

### **Core Results**

Following examination of the levels of detection for the 15 elements analysed by ICPMS, 7 elements were retained for further analysis (B, Mg, P, Cu, Zn, Sr and Ba).

The otolith core signatures from all areas largely overlapped, however, a PERMANOVA identified some variation in spawning origin according to fish capture location, although this was not significant ( $p = 0.057$ ).

PCA results show that although there is some overlap between the core signatures of SAN and SAS (South Africa north and south), they appear as mostly distinct groups (Figure 44). The core signature of SAN overlaps with that of SWI-Feb 18 and the core signature of SAS overlaps with those from SWI-May18 and SWI 2019 (Figure 44).

Clustering was performed on the P, Zn, Sr, and B core signatures (which were the elements contributing >20% to the first two dimensions of the PCA), and the most relevant number of clusters (i.e. distinct spawning origins) was found to be three. The three corresponding putative spawning origins (SpO) had significantly distinct multi-elemental signatures (PERMANOVA,  $p = 0.002$ ).

All three SpO contributed to the stocks of the three areas investigated, but in varied proportions (Figure 45; Table 16), except SpO-1 which did not contribute to the SWTS samples. SpO-2 was the main spawning source for the albacore analysed, particularly for those fish captured in SWTS and SWI (Table 16). SpO-1 was the second greatest spawning source, with most fish captured in SA originating from there (but, as noted above, no fish from SWTS).

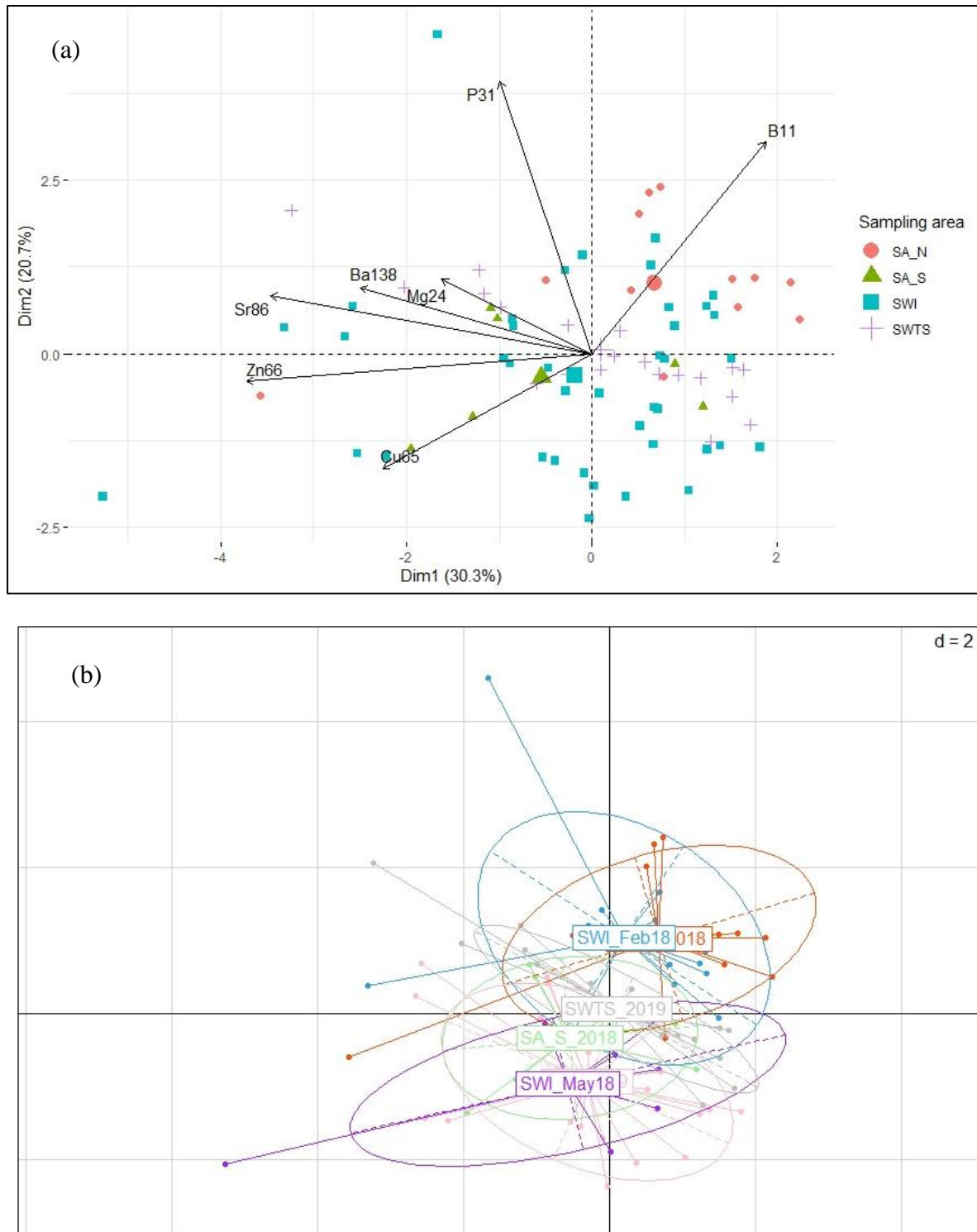


Figure 44. (a) Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the multi-elemental (B, P, Cu, Zn, Sr, Ba) signatures of the otolith cores of the 80 albacore analysed. Individuals are coded by their sampling location -- South Africa north (SA-N); South Africa south (SA-S), South West Indian Ocean (SWI) and Southwest Tasman Sea (SWTS). Their size on the graph is proportional to the quality of their representation in this plane. For the variables, the length of the arrow reflects the % of contribution to the total inertia. (b) Shows the same results as (a) but the colour coding for SWI is now broken down into the two sampling periods, Feb 2018 and May 2018.

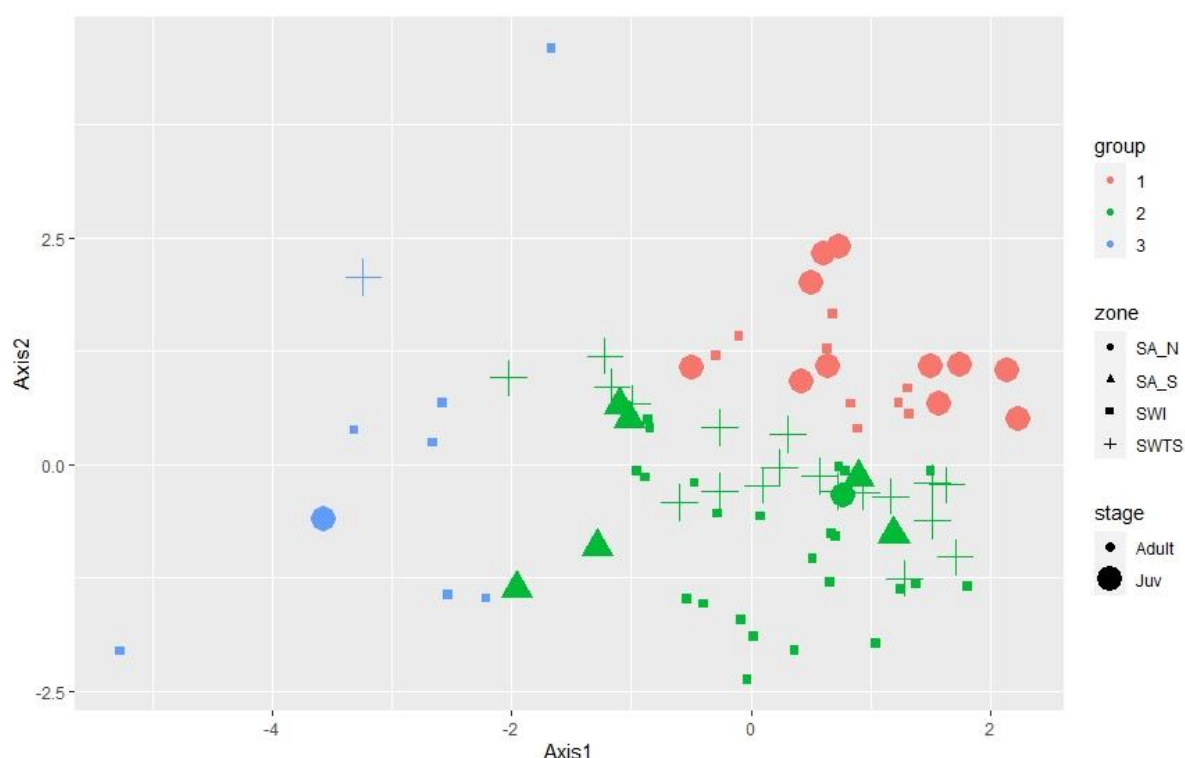


Figure 45. Projection of individuals on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr) signatures of the otolith core of the 80 albacore analysed. Colours on the graph represent the spawning origins (1 to 3) for each fish and symbols represent its final sampling area.

Table 16. Percent of individuals assigned to each spawning origin (SpO 1 to 3) in the albacore analysed (total N of fish tested = 80, FL = 55-116 cm) for each of the 3 sampling locations: South Africa (SA), South-Western (SWI) and Southwest Tasmanian Sea (SWTS).

	SA	SWI	SWTS
SpO-1	58%	22.5%	0%
SpO-2	37%	60%	95%
SpO-3	5%	17.5%	5%
Total	<b>100%</b>	<b>100%</b>	<b>100%</b>

Fish caught in the north and south South African locations had distinct signatures, the northern signature was found only in adults caught in SWI in February 2018 and the southern signature was found in adults caught in SWI in May 2018. This could indicate two spawning sources for south African fish or two spawning times, with different environmental conditions.

The samples collected do not allow us to confirm that these different spawning origins correspond to spatially discrete geographic areas in the Indian Ocean. Ideally, young-of-the-year fish would have been sampled from any areas of the IO where albacore may spawn, in order to get spawning ground signatures for known locations. However, this is logistically difficult to do in most regions of the IO, with the exception of South Africa.

Lastly, we note that there was a large range in the size (age) the fish analysed, so the results must be interpreted with caution as differences in otolith core signatures among locations or clusters might be due in part to cohort effects.

## 5.7 Billfish

### 5.7.1 Swordfish (*Xiphias gladius*)

#### Genetics

- The sample coverage for swordfish across the Indian Ocean was good with a total of 616 samples from 6 Indian Ocean sampling regions and a southwest Pacific outlier location in Coral Sea;
- A total of 417 samples were sequenced using DArTSeq and 309 passed quality control and were included in the analysis of population structure (Figure 46);
- Model selection criteria for different numbers of genetic groups (levels of K) using the program StockR suggests a single genetic grouping across all sample locations (Figure 46, top right). However, the geographical distribution of allele frequencies among sampling locations suggests a minimum partitioning of k=2 genetically differentiation groups roughly partitioned into northern (WCI, NCI, and NEI) and southern (SWI, SEI, and ECI) regions within the Indian Ocean (Figure 46).
- More focussed analysis specifically comparing the WTS (Coral Sea, south-west Pacific Ocean) samples with ECI and SEI (eastern Indian Ocean region) samples indicated no significant connectivity between the Indian and Pacific Oceans (Figure 47).
- Combined data from the current and previous studies indicates lack of significant gene flow from the Atlantic and SW Pacific into the Indian Ocean. Further investigation is required to determine likelihood of genetic connectivity from NW Pacific into the Eastern Indian via transport mediated by the Indonesian through flow.
- In summary, the demonstrated differences in gene-frequencies between the Indian and i) south-west Pacific Ocean, and ii) south-east Atlantic Ocean supports the assessment and management of swordfish within the Indian Ocean as an independent unit, although 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 further investigation via additional structured sampling and analysis that includes and expands the areas covered in this project and, ideally, includes sampling of spawning adults.

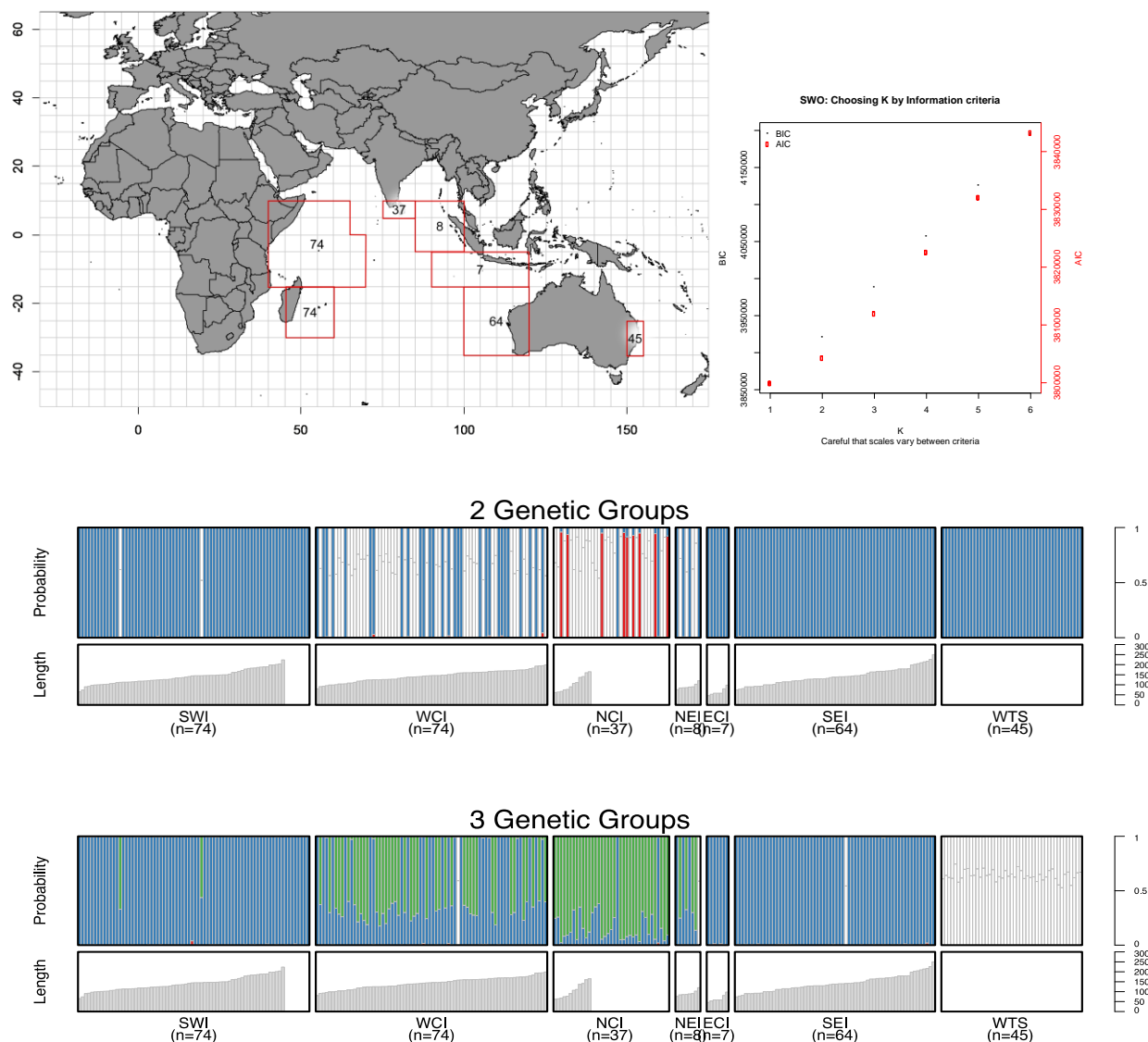


Figure 46. Top Left: Number of samples of swordfish (*Xiphias gladius*) sequenced using DARTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k). Bottom: Individual bar plot probability of assignment to a particular K genetic group when modelled at K=2 and K=3 genetic groups. Bars lacking colour indicate uncertainty in assignment 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.

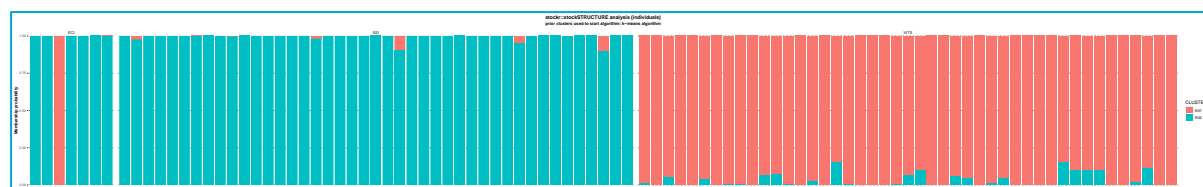


Figure 47. Comparison of eastern Indian Ocean (ECI and SEI) and western Pacific (WTS) Individual bar plot probability of assignment to a particular K genetic group when modelled at K=2 (cyan and red). From left to right are ECI (east central Indian), SEI (south east Indian), WTS (west Tasman Sea).

## Microchemistry

Otoliths from seventy swordfish were analysed (80-226 cm LJFL): 30 collected from the south west Indian Ocean (SWI), 20 from the western central Indian Ocean (WCI) and 20 from the south east Indian Ocean (SEI) (Figure 48). The fish were sampled during three periods: Nov-Dec 2017, March-May 2018 and December 2018 (Table 17).

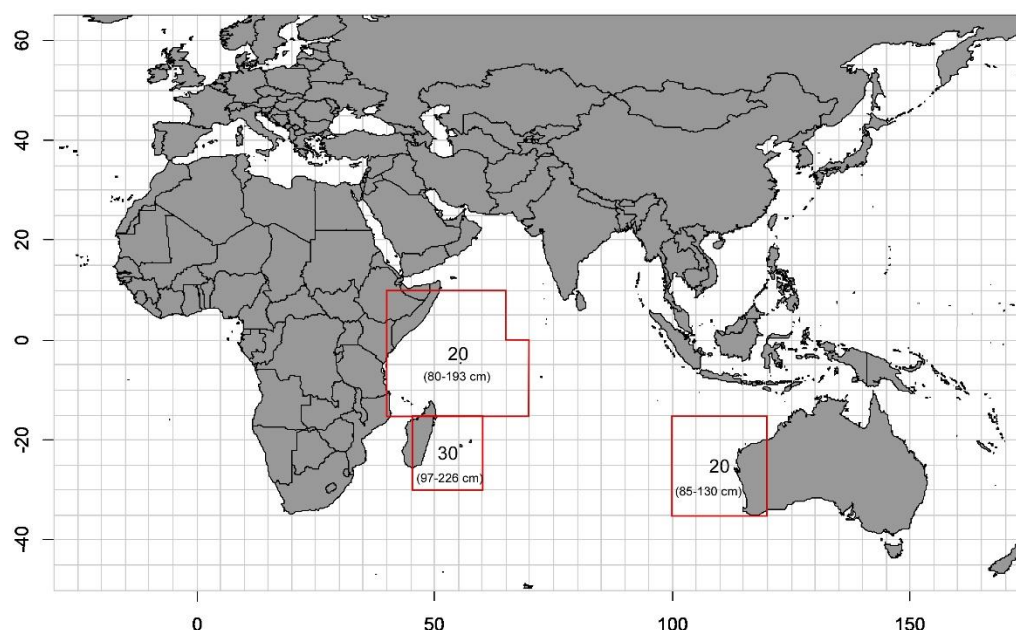


Figure 48. Map showing the number of swordfish otoliths analysed for each of the three sampling locations, referred to as south west Indian Ocean (SWI), west central Indian Ocean (WCI) and south east Indian Ocean (SEI); and the size range of fish at each location.

Table 17. Number, sampling period, size range and estimated ages of fish for each of the three sampling locations SWI, WCI and SEI.

Location	N	Sampling dates	LJFL (cm)	* Estimated age range (years)
south west Indian Ocean (SWI)	20	Nov-Dec 2017	97-226	1-10
south west Indian Ocean (SWI)	10	Dec 2018	111-204	2-8
west central Indian Ocean (WCI)	20	Mar-May 2018	80-193	1-7
south east Indian Ocean (SEI)	20	Mar-May 2018	85-130	1-3

\* The ranges in ages are for male and females combined (Varghese et al. 2013, Wang et al. 2010).

The otoliths were analysed at Montpellier University, Plateforme AETE-ISO (France) using LA-ICP-MS. The laser ablated 20 micron spots along a transect between the otolith core and the edge, therefore acquiring a chemical signal from material deposited throughout the life of the fish.

The portion of the otolith transect near the core was examined to identify the chemical signatures deposited during the first weeks of life. These are most likely to reflect the fish spawning origins, i.e. the physio-chemical characteristics of the water masses in which spawning occurred. The first laser point, on the core, was not included, in order to avoid any maternal influence on otolith composition. The mean of the next 3 points, between 10 and 40 microns after the core, considered to correspond to the first weeks of life, was used for analysis.

Fifteen chemical elements were measured. Those elements where more than 75% of results were above the limit of detection (LOD) were retained for further analysis.

### Core Results

After examination of the levels of detection for the 15 elements analysed by ICPMS, 6 elements were retained for further analysis (B, Mg, P, Zn, Sr and Ba).

The otolith core signatures from all areas overlapped, however, a PERMANOVA suggested some variation in elemental composition at the core according to fish capture location, although this was only just significant ( $p = 0.047$ ). PCA results show that although there is some overlap between the core signatures of SWI and SEI fish, they form fairly distinct groups (Figure 49).

Clustering was performed on the Mg, P, Sr, Ba and B core signatures (which were the elements contributing  $> 20\%$  to the first two dimensions of the PCA), and the most relevant number of clusters (i.e. distinct spawning origins) was found to be three. The three corresponding putative spawning origins (SpO) had significantly distinct multi-elemental signatures (PERMANOVA,  $p = 0.002$ ).

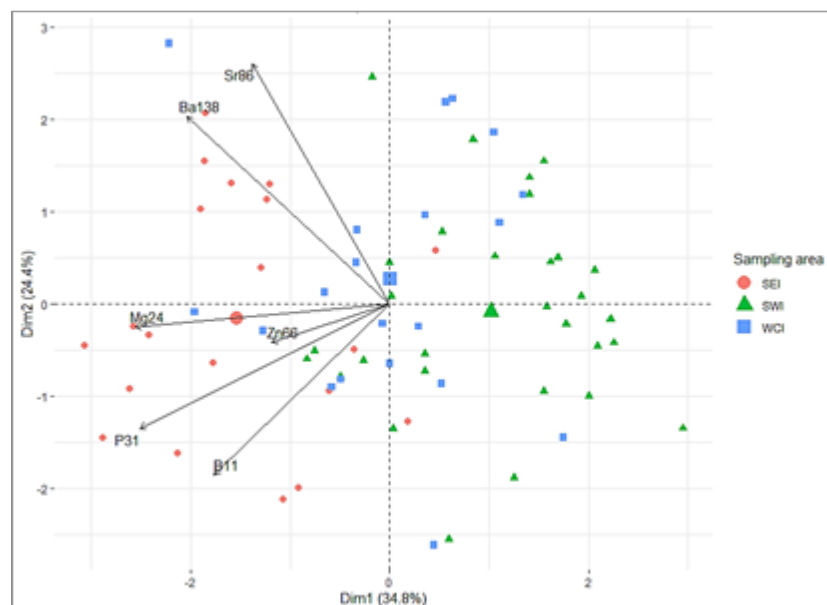


Figure 49. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the swordfish otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.



All three SpO contributed to the stocks of the three areas investigated, but in varied proportions (Figure 50; Table 18). SpO-1 was the main spawning source for the swordfish analysed; it was the origin of 40-60% of the fish analysed in all the areas sampled (Table 18). SpO-3 was the second greatest spawning source, particularly for fish from SWI and WCI (Table 18). SpO-2 contributed the least, however it was still an important source of fish captured in SEI (Table 18).

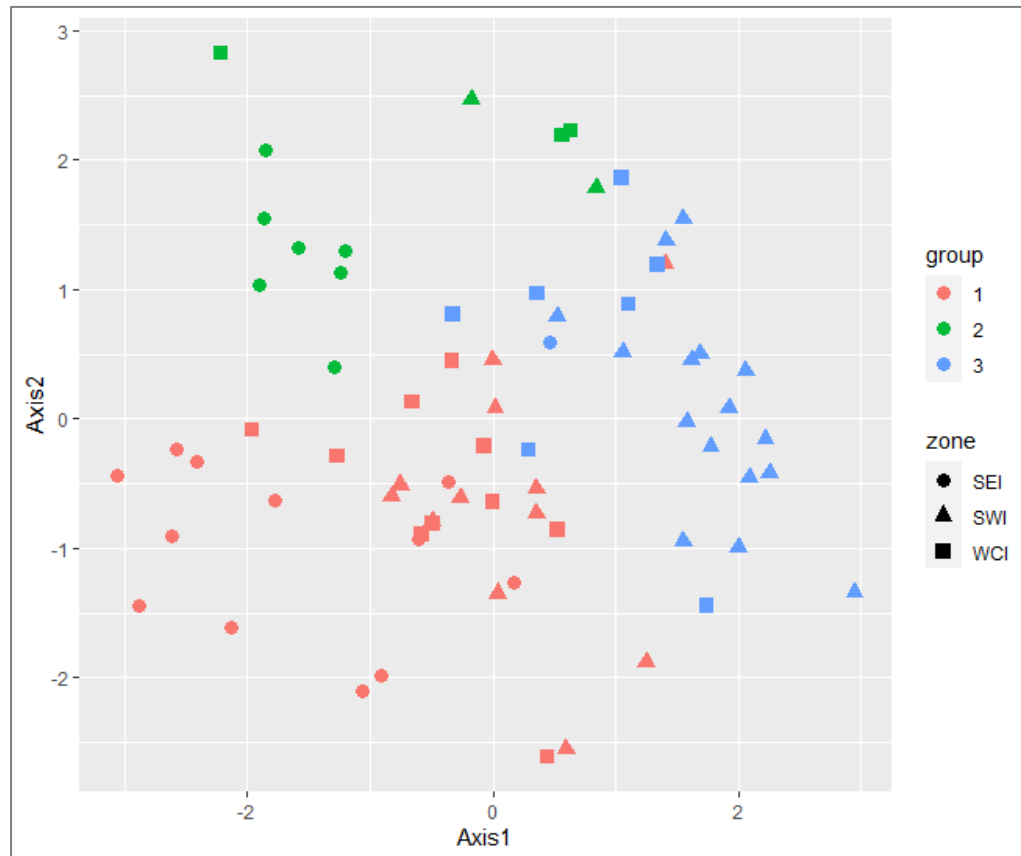


Figure 50. Projection of individuals on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr, Ba) signatures of the otolith core of the 70 swordfish analysed. Colours on the graph represent the spawning origins (1 to 3) for each fish and symbols represent its final sampling area.

Table 18. Percent of individuals assigned to each spawning origin (SpO 1 to 3) in the sub-adult and adult swordfish analysed (total N of fish tested = 70, FL = 80-226 cm) for each of the 3 areas investigated in the Indian Ocean: South-Est (SEI), South-Western (SWI) and Central-Western (CWI).

	SEI	SWI	WCI
SpO-1	60%	40%	50%
SpO-2	35%	7%	15%
SpO-3	5%	53%	35%
Total	100%	100%	100%

Before drawing any conclusions about swordfish spawning origin and mixing of stocks in the Indian Ocean, these preliminary results must be tested by analysing more samples. Ideally, samples would be collected from young-of-the-year in all areas of the IO where spawning may occur. This will provide evidence to test whether the hypothetical spawning ground signatures identified in this study correspond to geographically distinct spawning areas.

Lastly, we note that there was a large range in the size (age) the fish analysed, so differences in otolith core signatures among locations or clusters might be due in part to cohort effects.



### 5.7.3 Indo-Pacific sailfish (*Istiophorus platypterus*)

#### Genetics

- 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 65 past quality control and were included in the analysis of population structure (Figure 52);
- Population analysis, based on StockR, suggests a single genetic grouping across all sample locations in the Indian Ocean (Figure 52).

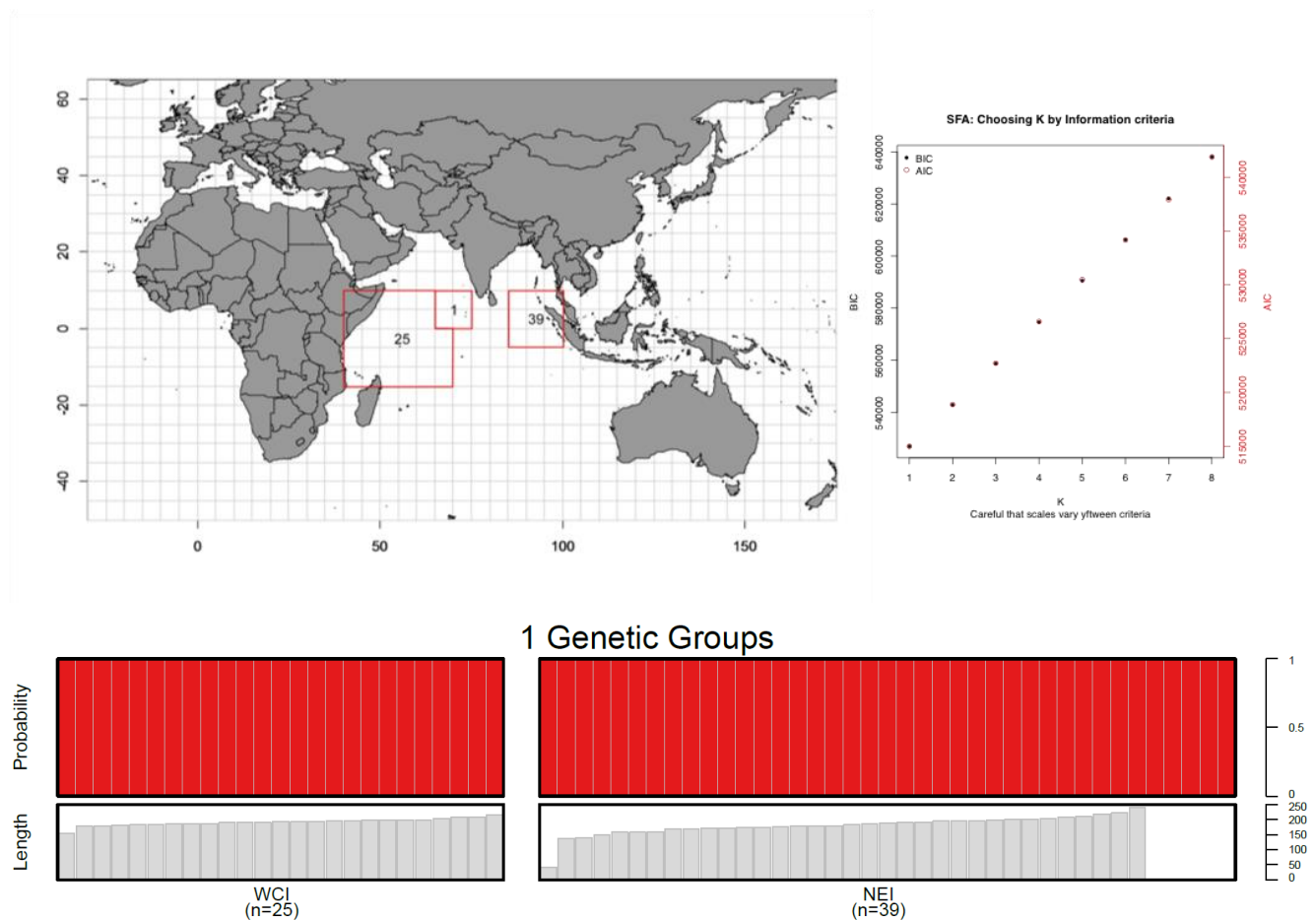


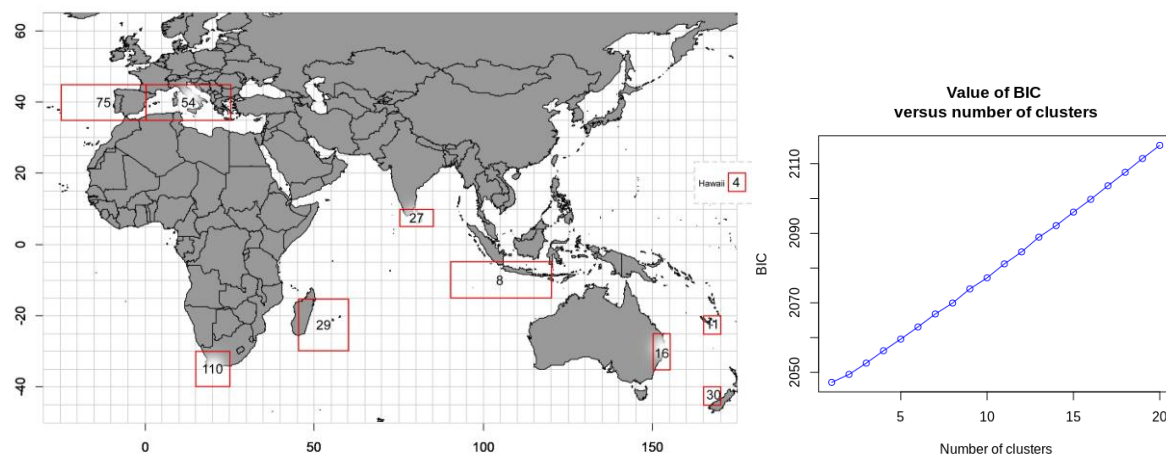
Figure 52. Top Left: Number of samples of Indo-Pacific sailfish (*Istiophorus platypterus*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k).

## 5.8 Sharks

### 5.8.1 Blue shark (*Prionace glauca*)

#### Genetics

- The samples of the blue shark cover nearly all oceans, in order to encompass its worldwide distribution range and replace the IO stocks in their broad context. In fact, this species has been shown, thus far, to travel large distances and to exhibit no apparent genetic differentiation across oceanic basins. The sampling coverage is fair for the Eastern Northern and Western regions of the Indian Ocean and for the Pacific outlier location.
- Due to the synergy with a previous project from the partner team (IRD-Ifremer) and new collaborations through CSIRO, a total of 585 samples were available from three locations of the Indian Ocean, the SW Pacific, North Pacific, South Atlantic, North Atlantic and Mediterranean Sea.
- A total of 376 samples were submitted to DArT, of which 364 samples passed DArTSeq QC and were included in the population structure analyses (Figure 53).
- The population analysis based on StockR indicates two genetic groups between North-Atlantic/Mediterranean and the Indian-Pacific Oceans (Figure 53, Append Figure 90). Figure 53. Top Left: Number of samples of blue shark (*Prionace glauca*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups ( $k$ ), indicating in fact no more likely group than a single 1. Bottom: results of population structure analysis of DArTSeq using StockR for blue shark for 2 genetic groups), due to very low global genetic differentiation, which was confirmed through low pairwise  $F_{st}$  values (Table 19).
- Additionally, both the careful analysis of pairwise  $F_{st}$  and of clustering plots obtained through PCA show a differentiation of groups of individuals from each Ocean and some Oceanographic basins (Mediterranean, North Atlantic, South Atlantic, Indian Ocean and South Pacific; Append Figure 89).



stockR: Blue shark data without outliers for K=2

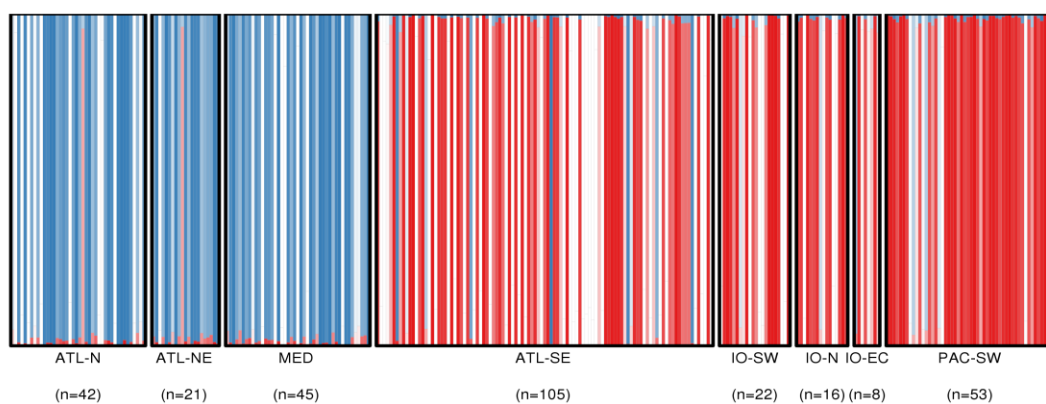


Figure 53. Top Left: Number of samples of blue shark (*Prionace glauca*) sequenced using DArTSeq and included in the analysis by sampling region. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups ( $k$ ), indicating in fact no more likely group than a single 1. Bottom: results of population structure analysis of DArTSeq using StockR for blue shark for 2 genetic groups.

Table 19. Pairwise  $F_{ST}$  values and their level of significance after correction with q-value (\*  $p < 0.01$ , \*\*  $p < 0.001$ , \*\*\*  $p < 0.0001$ ).

$F_{ST}$	MED (45)	ATL-N (42)	ATL-NE (21)	ATL-SE (105)	IO-EC (8)	IO-N (16)	IO-SW (22)	PAC-SW (53)
MED		0.0007***	0.0010***	0.0015***	0.0023***	0.0017***	0.0017***	0.0022***
ATL-N			0.0006**	0.0013***	0.0010*	0.0011***	0.0015***	0.0017***
ATL-NE				0.0018***	0.0014**	0.0015***	0.0015***	0.0020***
ATL-SE					0.0001	0.0000	0.0000	0.0000
IO-EC						0.0004	0.0000	0.0000
IO-N							0.0000	0.0000
IO-SW								0.0001
PAC-SW								

## 6 Summary and conclusions

### 6.1 Sampling

Sampling was completed between late 2017 and early 2019 with a total of 5,767 tissue samples and 3,635 otoliths collected or made available to the project from partner archives, representing ~73% and 188% of the planned total sample sizes for tissue and otoliths respectively. Of these, 3,635 tissue samples have been genotyped and 689 processed and analysed for otolith microchemistry following sample selection and quality control protocols.

The spatial coverage of tissue sampling achieved has been good for the majority of species, including the three neritic species, the three tropical species and swordfish. The coverage for these species would be improved through extending future sampling further into the Arabian Gulf, east Africa and Bay of Bengal, through increased engagement of coastal states in those regions. In the case of swordfish, contemporaneous samples from outlier locations in the south-west Pacific and south Atlantic would be valuable to refine the level of inter-ocean connectivity. There are gaps in tissue sample coverage for albacore and blue shark in the south-east and central north-east and for swordfish in the central north-east. These gaps are likely to be most effectively addressed through continuing the development of systematic tissue and otolith sampling programs in Indonesia through, for example, the Research Institute of Tuna Fisheries in Benoa for samples of adult spawning fish, and coordination of biological sampling through observer programs of relevant distant water fleets in the central eastern and south-eastern Indian Ocean. In general, it was difficult to obtain the target sample sizes for the other two billfish using the targeted “temporal window” sampling design adopted for this project due to their relatively low occurrence in landings and fish markets in many areas. This may best be addressed by establishing regular tissue sampling regimes in key ports where these species are landed and/or observer programs for fleets that take significant catches, in order to accumulate the required sample sizes for these species.

The CITES listing of scalloped hammerhead sharks shortly after the approval of the Expression of Interest for the project created administrative issues for the collection and international transport of samples during the first round of sampling and further collection of samples for this species ceased at that point. These issues are not insurmountable but require in country project partners with appropriate certification for dispatch and receipt of tissue samples of listed species. Given the recent listing of another IOTC shark species (shortfin mako), ensuring the necessary arrangements are established in the preparatory phases of future projects should be a priority.

The sample coverage for otolith microchemistry was often less complete for each species than for tissue due to the additional logistic difficulty in obtaining otoliths, particularly in the case of larger, more valuable adults, or for operations where fish were frozen at sea. Good otolith sample coverage was achieved for kawakawa and Spanish mackerel, the three tropical

tunas and swordfish, whereas the lack of otolith samples from the south-east Indian Ocean for albacore, in particular, limited the scope of inferences that could be made based on microchemistry for that species. Analysis of microchemistry of vertebrae for blue shark was abandoned due to the logistic difficulty associated with collection of sufficient samples across the range. The combination of the additional logistical difficulty and cost associated with obtaining otolith (or vertebrae) samples and the relative strength of the inferences that can be made based on them should be carefully considered as part of planning and design phases of any future stock structure work proposing to use microchemistry of hard parts.

The project has provided a sound foundation for exploring hypotheses for population structure for many of the study species. For some species, such as albacore, improvements in sample coverage, or additional samples from locations with low sample sizes, are required before substantive interpretations and conclusions about population structure across the Indian Ocean can be made. It will be important, therefore, that the samples and data collected through this project are appropriately archived and curated to ensure they are available for use in future studies. The project partners will work with the IOTC secretariat and Scientific Committee to finalise arrangements for archiving, access and management of the samples and the data arising from the project, so that this foundation data set is available to build the understanding of population structure of these species into the future.

## 6.2 Population structure

There are a number of important considerations in considering the results reported here. Firstly, while the project team has taken great care in conducting the analysis and interpreting the output, the detailed results are yet to be peer reviewed outside the project team. Nor have the various working parties of the IOTC had the opportunity to review the analysis, interpretation and conclusions. As noted, the project team will prepare and submit working papers for the appropriate IOTC Working Party meetings through 2020 and, in parallel, prepare and submit scientific papers for the broader peer review literature. Hence, the results presented in this report should be considered preliminary until these standard scientific reviews have been completed.

Second, to reiterate the point made earlier, in the case of the population genetic analysis, failure to detect population structure for a particular species with a particular set of samples and method, does not prove absence population structure that may be important from an assessment or monitoring perspective. It is possible that more comprehensive sampling across the range of the species than was possible in this project and the application of new methods may provide additional information on the structure and connectivity of these populations, which is not evident in the current results.

In this context, the results of this project and their interpretation should be seen as working hypotheses on population structure. They provide a solid foundation for initial deliberations by the IOTC Working Parties and Scientific Committee on their implications for assessment and management of the stocks under IOTC purview, and design of future targeted studies, for



individual or multiple species, to test their validity and further refine our understanding and inform stock assessment and management of these important fisheries.

### **6.2.1 Connectivity with other oceans**

With the exception of blue shark, striped marlin and Indo-Pacific sailfish, the results are consistent with the Indian Ocean being considered a closed unit for stock assessment and management purposes. In the case of blue shark there is evidence of restricted gene flow with the south-east Atlantic, but there is potential for some level of connection with the south-west Pacific. In the case of the two billfish, the sampling coverage was not sufficient to resolve this question unequivocally. For Indo-Pacific sailfish, there were no out-group samples available to the project. In the case of striped marlin, the number of samples was too low to address this question with any confidence, but the results were consistent with the small number of Indian Ocean samples being from a different genetic group than the south-west Pacific samples and the recently published work on global population structure of striped marlin (Mamoozadeh et al., 2020).

### **6.2.2 Genetic differentiation north and south of the equator**

There was a general tendency for genetic differentiation between samples from locations north and south of the equator for those species where the required sampling coverage was achieved. This included the three neritic species, skipjack and yellowfin tuna, but not bigeye and swordfish. This genetic differentiation was not evident for blue shark, where there was one location with a reasonable sample size north of the equator and a number of locations with good sample sizes to the south. This result is perhaps not surprising, given the very weak levels of genetic variation observed for blue shark across its range (see Appendix 3). This north-south genetic structure was also not apparent for albacore, striped marlin or Indo-Pacific sailfish. However, the sampling coverage for these species north and south of the equator was not sufficient to address this question directly.

### **6.2.3 Pronounced genetic differentiation between north-west Indian Ocean and other locations**

Each of the four species for which samples were available from the north-west Indian Ocean (skipjack, longtail and yellowfin tuna and Spanish mackerel) demonstrated strong genetic differentiation from other locations sampled for the project. In the case of yellowfin tuna, this NWI “signature” was also evident in samples from other locations north of the equator. As noted above, it was not possible to obtain samples from other locations in the Arabian Sea or in the region of north-eastern Africa to determine whether this strong “north-western” signal was more widespread. This should be a high priority for immediate future research to test the generality of the results presented here and determine the extent of the genetic differentiation evident in the samples available from this region.

#### **6.2.4 Stronger genetic differentiation in neritic species**

The strongest genetic differentiation and greatest extent of partitioning was evident for two of the neritic species, with three genetic groups within the Indian Ocean identified for longtail tuna and four for Spanish mackerel. This genetic partitioning also corresponded closely with the sample locations. Given the results of previous population structure studies for these species, this result was not unexpected (see Appendix 3 for these species). The strength of the differentiation does, however, provide a very clear example of the power of the SNP markers to identify population structure. The results from the otolith microchemistry for these two species were not as clear, partly due to the reduced sample coverage relative to the genetics, but they could be interpreted as being generally consistent with the pattern seen in the genetics. Previous studies of population structure on these species have also demonstrated the utility of microchemistry for detecting finer-scale population structure (e.g. Buckworth et al. 2007). However, as noted in the detailed results for each of these species, care needs to be taken in the spatial interpretation of these results given the differences in the time of the sampling and the different age classes of fish in the different locations.

Interestingly, the level of genetic differentiation evident for longtail tuna and Spanish mackerel was not apparent for kawakawa. There was evidence of genetic structure between locations north and south of the equator, as noted above, but not the finer scale structure evident for the other two species. This is despite the fact that the sample coverage for genetics for kawakawa was the best of all of the neritics and extended into pelagic environments in the central and western-central Indian Ocean outside the range of the other two species, where most of the samples came from purse-seine catches associated with targeted fishing for skipjack. The otolith core microchemistry for kawakawa did not provide additional evidence for population structure from a spawning ground perspective. The results of the otolith edge microchemistry did indicate some structure, similar to that for longtail tuna, but again, the extent to which this reflected spatial structure versus temporal differences in sampling cannot be determined with the available samples.

The clear genetic differentiation among locations for long-tail tuna and Spanish mackerel is strong evidence for separate stocks within the sampling range covered by this project. Despite the best efforts of the project team, it does not cover the full range of these species within the Indian Ocean. The results for both species provide a compelling case design and implement future studies to: i) extend the sampling coverage to the full range of these species within the Indian Ocean; ii) increase sample sizes and temporal coverage (i.e. establish sampling capacity for multi-year program); and, iii) extend to the other identified priority neritic species (bullet and frigate tuna and Indo-Pacific king mackerel) to provide comprehensive description of population structure of important neritic species for stock assessment and management purposes.

### 6.3 Extension and communication of results

The project has delivered a substantial contribution to the information base on population structure of neritic and pelagic tunas, billfish and blue shark in the Indian Ocean. It has also engaged a wide range of researchers, fisheries scientists (and fishers), beyond the project team, in the design and implementation of the project. The next stage is to engage more directly with the wider IOTC scientific community, through the working parties in the first instance, in the interpretation of the results presented in this report for population structure and their implications for how research, monitoring, stock assessment and management should be conducted in the future. This will be done through preparation of working papers on the results for each species, which will include further examination of the results and analyses completed to date, and additional papers examining the potential implications for monitoring and assessment. The papers will be presented at the appropriate working party meetings through 2020, with synthesis papers presented to the Scientific Committee at the end of 2020 (see Table 7). In parallel to this engagement with the working parties, peer review papers for the wider scientific literature will be prepared and published, consistent with the publication protocols for the project.

## 7 Literature Cited

- Abaunza, P., Murta, A. G., Campbell, N., Cimmaruta, R., Comesana, A. S., Dahle, G., . . . Zimmermann, C. (2008). Stock identity of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean Sea: Integrating the results from different stock identification approaches. *Fisheries Research*, 89(2), 196-209. doi:DOI 10.1016/j.fishres.2007.09.022
- Abaunza, P., Murta, A. G. and Stransky, C. (2013). Sampling for Interdisciplinary Analysis. In, S.X. Cadrin, L.A. Kerr and S. Mariani (Eds.), *Stock Identification Methods: Applications in Fishery Science*, Academic Press, pp. 477-500.
- Abedi, E., Zolgharnein, H., Salari, M. A., & Qasemi, A. (2012). Genetic differentiation of narrow-barred Spanish mackerel (*Scomberomorus commerson*) stocks using microsatellite markers in Persian Gulf. *American-Eurasian Journal of Agricultural & Environmental Sciences*, 12(10), 1305-1310.
- Abdussamad EM, Said Koya KP, Ghosh S, Rohit P, Joshi KK, Manojkumar B, Prakasan D, Kemparaju S, Elayath MNK, Dhokia HK, Sebastine M, Bineesh KK (2012) Fishery, biology and population characteristics of longtail tuna, *Thunnus tonggol* (Bleeker, 1851) caught along the Indian coast. *Indian J Fish* 59(2):7–16.
- Alexander, DH, J Novembre, K Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19:1655-1664.
- Allendorf FW, Luikart GH, Aitken SN (2012). *Conservation and the Genetics of Populations*. John Wiley & Sons.
- Arai, T., Kotake, A., and Kayama, S. (2005). Movements and life history patterns of the skipjack tuna *Katsuwonus pelamis* in the western Pacific, as revealed by otolith Sr: Ca ratios. *Journal of the Marine Biological Association of the United Kingdom* 85, 1211–1216. doi:10.1017/S0025315405012336
- Artetxe-Arrate et al. (2019). Otolith microchemistry: a useful tool for investigating stock structure of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean. *Marine and Freshwater Research*. <https://doi.org/10.1071/MF19067>
- Artetxe-Arrate I., Fraile I., Pecheyran C. and Murua H. (2019). Natal signatures in otoliths of young-of-the-year yellowfin (*Thunnus albacares*) and skipjack (*Katsuwonus pelamis*) tuna from the Indian Ocean. *Front. Mar. Sci. Conference Abstract: XVI European Congress of Ichthyology*. doi: 10.3389/conf.fmars.2019.07.00086
- Appleyard, SA, RD Ward, PM Grewe. (2002). Genetic stock structure of bigeye tuna in the Indian Ocean using mitochondrial DNA and microsatellites. *Journal of Fish Biology* 60:767-770.
- Bailleul, D., Mackenzie, A., Sacchi, O., Poisson, F., Bierne, N. , and Arnaud-Haond, S. (2018) Large-scale genetic panmixia in the blue shark (*Prionace glauca*): A single worldwide population, or a genetic lag-time effect of the "grey zone" of differentiation? *Evolutionary Applications*, 11, 614-630.

- Banse, K., Naqvi, S.W.A., Narvekar, P.V., Postel, J.R., Jayakumar, D.A. (2013). Oxygen minimum zone of the open Arabian Sea: variability of oxygen and nitrite from daily to decadal time scales. *Biogeosciences Discuss*, 10, 15455–15517.
- Bierne, N., D. Roze, and J. J. Welch. (2013). Pervasive selection or is it . . .? why are FST outliers sometimes so frequent? *Molecular Ecology* 22:2061–2064.
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20(10), 2044–2072. doi:DOI 10.1111/j.1365-294X.2011.05080.x
- Bitencourt, A., Silva, D.A., Carvalho, E.F., Loiola, S., and Amaral, C.R.L. (2019) Study of genetic variability of the Blue Shark *Prionace glauca* (Linnaeus, 1758). *Forensic Science International Genetics Supplement Series*, 7, 594–596.
- Bonhommeau S, Evano H, Huet J, Chapat M, Varenne F, Le Foulgoc L, Richard E, Tessier E, Chanut J, Nieblas AE, Brisset B. (2019). Biometric data for large pelagic fish in the Indian Ocean. SEANOE. <https://doi.org/10.17882/62757>
- Bougeard S, Dray S. (2018). “Supervised Multiblock Analysis in R with the ade4 Package.” *Journal of Statistical Software*, 86(1), 1–17.
- Boyer, T.P., Antonov, J.I., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Johnson, D.R. Locarnini, R.A., Mishonov, A.V., O'Brien, T.D., Paver, C.R., Reagan, J.R., Seidov, D., Smolyar, I.V. and Zweng, M.M. 2013 World Ocean Database 2013. Silver Spring, MD, NOAA Printing Office, 208pp. (NOAA Atlas NESDIS, 72). <http://hdl.handle.net/11329/357>
- Buckworth, R. C., Newman, S. J., Ovenden, J. R., Lester, R. J. G., and McPherson, G. R. (2007). The Stock Structure of Northern and Western Australian Spanish Mackerel. Final Report, Fisheries Research & Development Corporation Project 1998/159. Department of Primary Industry, Fisheries and Mines, Northern Territory Government, Australia. *Fishery Report* **88**, i-vi, 225 p.
- Cadrin, Steven & Kerr, Lisa & Mariani, Stefano. (2014). Interdisciplinary Evaluation of Spatial Population Structure for Definition of Fishery Management Units. 10.1016/B978-0-12-397003-9.00022-9.
- Carvalho, G. R., & Hauser, L. (1994). Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries*, 4(3), 326–350. doi:10.1007/BF00042908
- Catchen, J, PA Hohenlohe, S Bassham, A Amores, WA Cresko. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22:3124–3140.
- Charrad M., Ghazzali N., Boiteau V. and Niknafs A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61, pp. 1–36.
- Chen, K.-Y. et al. (2018). assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution* 9, 439–446.
- Chessel D, Dufour A, Thioulouse. (2004). “The ade4 Package – I: One-Table Methods.” *R News*, 4(1), 5–10. <https://cran.r-project.org/doc/Rnews/>.

- Chiang, H-C, C-C Hsu, GC-C Wu, S-K Chang, H-Y Yang. (2008). Population structure of bigeye tuna (*Thunnus obesus*) in the Indian Ocean inferred from mitochondrial DNA. *Fisheries Research* 90:305-312.
- Compagno, L. J. V. (1984). FAO species catalogue. Sharks of the world. An annotated and illustrated catalogue of shark species known to date. Part 1. Hexanchiformes to Lamniformes. FAO Fish Synop., 4(125), 250.
- Dammannagoda, S, D Hurwood, P Mather. (2011). Genetic analysis reveals two stocks of skipjack tuna (*Katsuwonus pelamis*) in the northwestern Indian Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* 68:210-223.
- da Silva, C., Kerwath, S.E., Wilke, C.G., Meyer, M., and Lamberth, S.J. (2010). First documented southern transatlantic migration of a blue shark *Prionace glauca* tagged off South Africa. *African Journal of Marine Science*, 32, 639-642.
- Davies, C.A., D. Brophy, T. Jeffries, E. Gosling. (2011). Trace elements in the otoliths and dorsal spines of albacore tuna (*Thunnus alalunga*, Bonnaterre, 1788): An assessment of the effectiveness of cleaning procedures at removing postmortem contamination. *Journal of Experimental Marine Biology and Ecology* 396 162–170.
- De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* 22:1383–1399.
- De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, 56(6), 879-886. doi:10.1080/10635150701701083
- Dhurmeea Z, Zudaire I, Chassot E, Cedras M, Nikolic N, Bourjea J, et al. (2016) Reproductive Biology of Albacore Tuna (*Thunnus alalunga*) in the Western Indian Ocean. *PLoS ONE* 11(12): e0168605. doi:10.1371/journal.pone.0168605
- Dray S, Dufour A. (2007). “The ade4 Package: Implementing the Duality Diagram for Ecologists.” *Journal of Statistical Software*, 22(4), 1–20.
- Dray S, Dufour A, Chessel D. (2007). “The ade4 Package – II: Two-Table and K-Table Methods.” *R News*, 7(2), 47–52. <https://cran.r-project.org/doc/Rnews/>.
- Ely, B, J Viñas, JR Alvarado Bremer, D Black, L Lucas, K Covello, AV Labrie, E Thelen. (2005). Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*). *BMC Evolutionary Biology* 5:19.
- Etter, PD, S Bassham, PA Hohenlohe, EA Johnson, WA Cresko. (2011). SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. In: V Orgogozo, MV Rockman, editors. *Molecular Methods for Evolutionary Genetics*. Totowa, NJ: Humana Press. p. 157-178.
- Evanno, G. & Regnaut Sand Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14:2611-2620.
- Eveson, Paige, Julien Million, Fany Sardenne, c and Gaël Le Croizier. (2015). Estimating growth of tropical tunas in the Indian Ocean using tag-recapture data and otolith-based age estimates. *Fisheries Research* 163:58–68.

- Fauvelot, C. & Borsa, P. (2011). Patterns of genetic isolation in a widely distributed pelagic fish, the narrow-barred Spanish mackerel (*Scomberomorus commerson*). *Biological Journal of the Linnean Society* 104, 886–902.
- Farley JH, Clear NP, Leroy B, Davis TL and McPherson G. (2006). Age, growth and preliminary estimates of maturity of bigeye tuna, *Thunnus obesus*, in the Australian region. *Mar Freshwater Res.* 57:713–724.
- Fourcade, Y., A. Chaput-Bardy, J. Secondi, C. Fleurant, and C. Lemaire. (2013). Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology* 22:2065–2073.
- 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.
- Fraille, I., Arrizabalaga, H., Santiago, J., Goni, N., Arregi, I., Madinabeitia, S., Wells, R. J. D., and Rooker, J. R. (2016). Otolith chemistry as an indicator of movements of albacore (*Thunnus alalunga*) in the North Atlantic Ocean. *Marine and Freshwater Research* 67(7) 1002–1013. doi:[10.1071/MF15097](https://doi.org/10.1071/MF15097)
- Frankham, R. (1995). Conservation genetics. *Annual Review of Genetics*, 29, 305-327.
- 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.
- Gagnaire, P. A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., . . . Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8(8), 769-786. doi:[10.1111/eva.12288](https://doi.org/10.1111/eva.12288)
- 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.
- Georges A, Gruber B, Pauly GB, White D, Adams M, Young MJ, Kilian A, Zhang X, Shaffer HB, Unmack PJ. (2018). Genome-wide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: Emydura) of eastern Australia. *Mol. Ecol.*, 27, pp. 5195-5213.
- Gosselin, T., Lamothe, M., Devloo-Delva, F., and Grewe, P. (2019). RADseq Data Exploration, Manipulation and Visualization using R. DOI:[10.5281/zenodo.2595083](https://doi.org/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](https://doi.org/10.5281/zenodo.3687060)



- Graves, JE, SD Ferris, AE Dizon. (1984). Close genetic similarity of Atlantic and Pacific skipjack tuna (*Katsuwonus pelamis*) demonstrated with restriction endonuclease analysis of mitochondrial DNA. *Marine Biology* 79:315-319.
- Grewe P, Grueger C, Aquadro CF, Bermingham E. (1993). Mitochondrial DNA Variation among Lake Trout (*Salvelinus namaycush*) Strains Stocked into Lake Ontario. *Can. J. Fish. Aquat. Sci.* 50, 2397–2403.
- Griffiths, S.P., Fry, G.C., Manson, F.J. and Lou, D.C. (2010). Age and growth of longtail tuna, *Thunnus tonggol*, in tropical and temperate waters of the central Indo-Pacific. *ICES Journal of Marine Science* 67(1): 125-134.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *Plos Genetics* 5, e1000695.
- Grünwald NJ, Goss EM. (2011). Evolution and population genetics of exotic and re-emerging pathogens: Novel tools and approaches. *Annual Review of Phytopathology* 49:249–267.
- Habib, A., Hjh Zohrah & Zohrah Sulaiman. (2012). Population genetic structure of longtail tuna (*Thunnus tonggol*) along Wallace’s Line.
- Habib, A., & Sulaiman, Z. (2017). Mitochondrial DNA analyses of narrow-barred Spanish mackerel (*Scomberomorus commerson*) sampled from the Arabian Sea, the Bay of Bengal, and the Indo-Malay Archipelago. *Zoology and Ecology*, 27(3-4), 245-250.
- Hedgecock, D., Barber, P. H., & Edmands, S. (2007). Genetic Approaches to Measuring Connectivity. *Oceanography*, 20(3), 70-79.
- Hellberg, M. E. (2009). Gene Flow and Isolation among Populations of Marine Animals. *Annual Review of Ecology Evolution and Systematics*, 40, 291-310. doi:DOI 10.1146/annurev.ecolsys.110308.120223
- Hellberg, M. E., Burton, R. S., Neigel, J. E., & Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, 70(1), 273-290.
- Hoolihan, J. P., Anandh, P. & van Herwerden, L. (2006). Mitochondrial DNA analyses of narrow-barred Spanish mackerel (*Scomberomorus commerson*) suggest a single genetic stock in the ROPME sea area (Arabian Gulf, Gulf of Oman, and Arabian Sea). *ICES Journal of Marine Science* 63, 1066–1074.
- 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.
- Hoyle, Simon D, Rishi Sharma, Miguel Herrera. (2014). IOTC–2014–WPTmT05–24\_Rev1. Stock assessment of albacore tuna in the Indian Ocean for 2014 using Stock Synthesis.
- 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.
- Ianelli, J.N. (1993). Studies on the population structure of skipjack tuna *Katsuwonus pelamis* in the central and eastern Pacific Ocean. Ph.D. Thesis. University of Washington, Seattle. 215 p.
- Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405.
- Jombart, T., Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 2011 Nov 1;27(21):3070-1. doi: 10.1093/bioinformatics/btr521. Epub 2011 Sep 16.
- Jombart T, Collins C. (2015). Analysing genome-wide SNP data using adegenet 2.0.0. Imperial College London MRC Centre for Outbreak Analysis and Modelling June 23, 2015. <http://adegenet.r-forge.r-project.org/files/tutorial-genomics.pdf>
- 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.
- Kahn, M Z. (2004). Age and growth, mortality and stock assessment of *Euthynnus affinis* (Cantor) from Maharashtra waters. *Indian J. Fish.*, 51(2): 209-213, Apr. Jun. 2004.
- 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.
- King, J.R., Wetklo, M., Supernault, J., Taguchi, M., Yokawa, K., Sosa-Nishizaki, O., and Withler, R.E. (2015) Genetic analysis of stock structure of blue shark (*Prionace glauca*) in the north Pacific Ocean. *Fisheries Research*, 172, 181-189.
- Kitchens, L., Rooker, J., Reynal, L., Falterman, B., Saillant, E., and Murua, H. (2018). Discriminating among yellowfin tuna *Thunnus albacares* nursery areas in the Atlantic Ocean using otolith chemistry. *Marine Ecology Progress Series* 603, 201–213. doi:[10.3354/MEPS12676](https://doi.org/10.3354/MEPS12676)
- Kohler, N.E., and Turner, P.A. (2008). Stock structure of the blue shark (*Prionace glauca*) in the North Atlantic Ocean based on tagging data. *Sharks of the Open Ocean: Biology, Fisheries and Conservation*, 339-350.
- Kohler, N.E., Turner, P.A., Hoey, J.J., Natanson, L.J. and Briggs, R. (2002). Tag and recapture data for three pelagic shark species: Blue Shark (*Prionace glauca*), Shortfin Mako (*Isurus xyrinchus*), and Porbeagle (*Lamna nasus*) in the North Atlantic Ocean. *Collective Volume of Scientific Papers*, 54, 1231-1260.
- Kunal, S. P., Kumar, G., Menezes, M. R. & Meena, R. M. (2014). Genetic homogeneity in longtail tuna *Thunnus tonggol* (Bleeker, 1851) from the northwest coast of India inferred from direct sequencing analysis of the mitochondrial DNA D-loop region. *Marine Biology Research* 10, 738–743.
- Kunal, S. P. (2018). Population genetic structure of longtail tuna *Thunnus tonggol* Bleeker 1811 and yellowfin tuna *thunnus albacares* bonnaterre 1788 from the Indian region.

- Lawson DL, Dorp Lv, Falush D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 3258.
- Leone A, Urso I, Damalas D, Martinsohn J, Zanzi A, Mariani S, Sperone E, Micarelli P, Garibaldi F, Megalofonou P, Bargelloni L, Franch R, Macias D, Prodöhl P, Fitzpatrick S, Stagioni M, Tinti F, Cariani A. 2017. Genetic differentiation and phylogeography of Mediterranean-North Eastern Atlantic blue shark (*Prionace glauca*, L. 1758) using mitochondrial DNA: panmixia or complex stock structure? *PeerJ* 5:e4112 <https://doi.org/10.7717/peerj.4112>
- Li, W.W., Dai, X.J., Zhu, J.F., Tian, S.Q., He, S., and Wu, F. (2017). Genetic differentiation in blue shark, *Prionace glauca*, from the central Pacific Ocean, as inferred by mitochondrial cytochrome b region. *Mitochondrial DNA Part A*, 28, 575-578.
- Lotterhos, K. E., and M. C. Whitlock. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* 23:2178–2192.
- 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
- Lutjeharms, J.R.E (2006). *The Agulhas Current*. Springer Books.
- Macbeth, G. M., Broderick, D., Buckworth, R. C. & Ovenden, J. R. (2013). Linkage Disequilibrium Estimation of Effective Population Size with Immigrants from Divergent Populations: A Case Study on Spanish Mackerel (*Scomberomorus commerson*). *G3: Genes, Genomes, Genetics* 3, 709-717.
- Macdonald, J. I., Farley, J. H., Clear, N. P., Williams, A. J., Carter, T. I., Davies, C. R., and Nicol, S. J. (2013). Insights into mixing and movement of South Pacific albacore *Thunnus alalunga* derived from trace elements in otoliths. *Fisheries Research* 148, 56-63. doi:[10.1016/J.FISHRES.2013.08.004](https://doi.org/10.1016/J.FISHRES.2013.08.004)
- McIlwain, J.L., Claereboudt, M., Al-Oufi, H., Zaki, S., and Goddard, J. (2005). Spatial variation in age and growth of the kingfish (*Scomberomorus commerson*) in the coastal waters of the Sultanate of Oman. *Fisheries Research* 73:283-298.
- McPherson, G.R. (1992). Age and growth of the narrow-barred Spanish mackerel, *Scomberomorus commerson* (Lacepède, 1800) in north-eastern Queensland waters. *Australian Journal of Marine & Freshwater Research* 43:1269-1282.
- 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.
- Maxwell, S.M., Scales, K.L., Bograd, S.J., Briscoe, D.K., Dewar, H., Hazen, E.L., Lewison, R.L., Welch, H., and Crowder, L.B. (2019). Seasonal spatial segregation in blue sharks (*Prionace glauca*) by sex and size class in the Northeast Pacific Ocean. *Diversity and Distributions*, 25, 1304-1317.
- McCreary, J.P., Yu, Z., Hood, R., Vinaychandran, P.N., Furue, R., Ishida, A., Richards, K.J. (2013). Dynamics of the Indian-Ocean oxygen minimum zones. *Prog. Oceanogr.* 112–113, 15–37.

- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology* 21:2839–2846.
- Menezes, MR, M Ikeda, N Taniguchi. (2006). Genetic variation in skipjack tuna *Katsuwonus pelamis*(L.) using PCR-RFLP analysis of the mitochondrial DNA D-loop region. *Journal of Fish Biology* 68:156-161.
- Menezes, MR, G Kumar, SP Kunal. (2012). Population genetic structure of skipjack tuna *Katsuwonus pelamis* from the Indian coast using sequence analysis of the mitochondrial DNA D-loop region. *Journal of Fish Biology* 80:2198-2212.
- Menezes, MR, D Noguchi, M Nakajima, N Taniguchi. (2008). Microsatellite development and survey of genetic variation in skipjack tuna *Katsuwonus pelamis*. *Journal of Fish Biology* 73:463-473.
- Meyers G., Bailey, R., and Worby, A. (1995). Geostrophic transport of Indonesian throughflow. *Deep-Sea Research*. 42: 1163-1174.
- Muthiah, C. (1986). Maturation and spawning of *Euthynnus affinis*, *Auxis thazard* and *A. rochei* in the Mangalore inshore area during 1979-1982. In: Tuna fisheries of the Exclusive Economic Zone of India. Biology and stock assessment. Bull. Cent. Mar. Fish. Res. Inst., 36: 71-85, E.G. Silas (Ed.), Central Marine Fisheries Research Institute Cochin.
- Nakano, H., and Seki, M. (2003) Synopsis of biological data on the blue shark, *Prionace glauca* Linnaeus. *Bulletin-Fisheries Research Japan*, 18-55.
- Nakano, H., and Stevens, J.D. (2008). The Biology and Ecology of the Blue Shark, *Prionace Glauca*. *Sharks of the Open Ocean*, pp. 140-151. Blackwell Publishing Ltd.
- Nikolic N, Montes I, Lalire M, Puech A, Bodin N, Arnaud-Haong S, Corse E, Kerwath S, Gaspar P, Hollanda SJ, Bourjea J, West W, Bonhommeau S. Connectivity and population structure of albacore tuna across southeast Atlantic and western Indian Oceans inferred from multidisciplinary methodology. *Scientific Reports Nature*. In revision.
- Nishikawa, Y., Honma, M., Ueyanagi, S. and Kikawa, S. (1985). Average distribution of larvae of oceanic species of Scombroid fishes, 1956-1981. *Far Seas Fish. Res. Lab. S Ser.* 12:1-99.
- Pitcher, G.C, Probyn, T.A, du Randt, A., Lucas, A.J., Bernard, S., Evers-King, H. et al. (2014). Dynamics of oxygen depletion in the nearshore of a coastal embayment of the southern Benguela upwelling system. *J. Geophys. Res. Oceans* 119(4): 2183-2200.
- 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.
- Pritchard JK, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

- 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.
- Puechmaille SJ. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16:608–627.
- Purcell, S., B. Neale, K Todd-Brown, et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81:559-575.
- Queiroz N, Humphries NE, Noble LR, Santos AM, Sims DW (2012). Spatial Dynamics and Expanded Vertical Niche of Blue Sharks in Oceanographic Fronts Reveal Habitat Targets for Conservation. *PLoS ONE* 7(2): e32374. <https://doi.org/10.1371/journal.pone.0032374>
- Ramage, C.S. (1969). Indian Ocean surface meteorology. *Oceanogr. Mar. Biol.*, 7:11-30.
- Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, 10(4), 361-395. doi:DOI 10.1111/j.1467-2979.2008.00324.x
- Rooker, Jay R., Jaime R. Alvarado Bremer, Barbara A. Block, Heidi Dewar, Gregorio De Metrio, Aldo Corriero, Richard T. Kraus, Eric D. Prince, Enrique Rodriguez-Marin, and David H. Secor. (2007). Life History and Stock Structure of Atlantic Bluefin Tuna (*Thunnus thynnus*). *Reviews in Fisheries Science*, 15:265–310, ISSN: 1064-1262 print DOI: 10.1080/10641260701484135
- Rooker, J. R., David Wells, R. J., Itano, D. G., Thorrold, S. R., and Lee, J. M. (2016). Natal origin and population connectivity of bigeye and yellowfin tuna in the Pacific Ocean. *Fisheries Oceanography* 25, 277–291. doi:[10.1111/FOG.12154](https://doi.org/10.1111/FOG.12154)
- Rougeux, C., Bernatchez, L. and Gagnaire, PA. (2017). Modeling the Multiple Facets of Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8), 2057–2074. <http://doi.org/10.1093/gbe/evx150>
- Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings* 5:54.
- Sardenne, F., Dortel, E., Le Croizier, G., and Million, J. (2015). Determining the age of tropical tunas in the Indian Ocean from otolith microstructures. *Fisheries* 163, 44–57. doi:[10.1016/J.FISHRES.2014.03.008](https://doi.org/10.1016/J.FISHRES.2014.03.008)
- Schott, F.A. & McCreary, J.P. (2001). The monsoon circulation of the Indian Ocean. *Prog. Oceanogr*, 51, 1–123.

- Schouten, M.W., de Ruijter, W.P.M., van Leeuwen, P.J. and Ridderinkhof, H. (2003). Eddies and variability in the Mozambique Channel. *Deep-Sea Research II*, 50, 1987–2003.
- Shaklee, J. B., Phelps, S. R., & Salini, J. P. (1990). Analysis of fish stock structure and mixed-stock fisheries by the electrophoretic characterization of allelic isozymes. In: Whitmore DH, ed. *Application of electrophoretic and isoelectric focusing techniques in fisheries management*. Boca Raton, FL: CRC Press, 173-196.
- Sharp, G. D. (1978). Behavioural and physiological properties of tunas and their effects on vulnerability to fishing gear. In: Sharp, G. D., Dizon, A. E. (eds.) *The physiological ecology of tunas*. Academic Press, New York, p. 397-449.
- Shuford, R. L., Dean, J. M., Ste'quert, B., and LaBonne, M. (2007). Elemental fingerprints in otoliths of juvenile yellowfin tuna from spawning grounds in the Atlantic Ocean. *Collective Volume of Scientific Papers ICCAT* **60**, 314–329.
- 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.
- Stequert, B.; Marsac, F. (1989). Tropical tuna: surface fisheries in the Indian Ocean. FAO Fisheries Technical Paper, No. 282, 238p.
- Stift, M., Kolář, F. & Meirmans, P.G. (2019). STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* 123, 429–441. <https://doi.org/10.1038/s41437-019-0247-6>
- Sulaiman, Z. H., & Ovenden, J. R. (2010). Population genetic evidence for the east–west division of the narrow-barred Spanish mackerel (*Scomberomorus commerson*, Perciformes: Teleostei) along Wallace’s Line. *Biodiversity and Conservation*, 19(2), 563-574.
- Suman, Ali, Hari Eko Irianto, Khairul Amri, Budi Nugraha and Gatut Bintoro. (2015). Population structure and bioreproduction of bigeye tuna (*Thunnus obesus*) in western part of Sumatera and southern part of Java and Nusa Tenggara, Indian Ocean. *Ind. Fish. Res. J.*, Vol 21. No 2, Dec 2015:109-116.
- Sun, C. L., S. Z. Yeh, Y. J. Chang, H. Y. Chang and S. L. Chu. (2013). Reproductive biology of female bigeye tuna *Thunnus obesus* in the western Pacific Ocean. *Journal of Fish Biology* 83, 250–271.
- Taguchi, M., King, J.R., Wetklo, M., Withler, R.E., and Yokawa, K. (2015). Population genetic structure and demographic history of Pacific blue sharks (*Prionace glauca*) inferred from mitochondrial DNA analysis. *Marine and Freshwater Research*, 66, 267-275.

- Thorrold, S. R., Zacherl, D. C., & Levin, L. A. (2007). Population Connectivity and Larval Dispersal Using Geochemical Signatures in Calcified Structures. *Oceanography*, 20(3), 80-89.
- Vandeperre, F., Aires-da-Silva, A., Fontes, J., Santos, M., Santos, R. S., and Afonso, P. (2014). Movements of Blue Sharks (*Prionace glauca*) across Their Life History. *Plos One*, 9, 14.
- van Herwerden, L., McIlwain, J., Al-Oufi, H., Al-Amry, W. & Reyes, A. (2006). Development and application of microsatellite markers for *Scomberomorus commerson* (Perciformes; Teleostei) to a population genetic study of Arabian Peninsula stocks. *Fisheries Research* 79, 258–266.
- 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.
- Verissimo, A., Sampaio, I., McDowell, J.R., Alexandrino, P., Mucientes, G., Queiroz, N., da Silva, C., Jones, C.S., and Noble, L.R. (2017). World without borders-genetic population structure of a highly migratory marine predator, the blue shark (*Prionace glauca*). *Ecology and Evolution*, 7, 4768-4781.
- Vineesh, N., Kathirvelpandian, A., Divya, P. R., Mohitha, C., Basheer, V. S., Gopalakrishnan, A. and Jena, J. K. (2016). Hints for panmixia in *Scomberomorus commerson* in Indian waters revealed by mitochondrial ATPase 6 and 8 genes. *Mitochondrial DNA Part A* 27, 2822–2824.
- Vitalis R. (2020). DetSel: A Computer Program to Detect Markers Responding to Selection. <https://cran.r-project.org/web/packages/DetSel/index.html>
- Walther, Benjamin D. (2019). The art of otolith chemistry: interpreting patterns by integrating perspectives. *Marine and Freshwater Research* 70, 1643–1658. <https://doi.org/10.1071/MF18270>
- 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.
- Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89(5), 438-450.
- Waples, R. S. (2015). Testing for Hardy-Weinberg proportions: have we lost the plot? *J Hered*, 106(1), 1-19. doi:10.1093/jhered/esu062
- Waples, R. S., and Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6), 1419-1439. doi:10.1111/j.1365-294X.2006.02890.x
- Ward, R. D. (2000). Genetics in fisheries management. *Hydrobiologia*, 420, 191-201. doi:10.1023/a:1003928327503
- Wells, R. D., Rooker, J. R., and Itano, D. G. (2012). Nursery origin of yellowfin tuna in the Hawaiian Islands. *Marine Ecology Progress Series* 461, 187–196. doi:[10.3354/MEPS09833](https://doi.org/10.3354/MEPS09833)



- Werner, F. E., Cowen, R. K., & Paris, C. B. (2007). Coupled Biological and Physical Models Present Capabilities and Necessary Developments for Future Studies of Connectivity. *Population Oceanography*, 20(3), 54-69.
- Whitlock, M. C., Lotterhos, K. E., & Editor: Judith, L. B. (2015). Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F(ST). *The American Naturalist*, 186(S1), S24-S36. doi:10.1086/682949
- Willette, D. A., Santos, M. D., & Leadbitter, D. (2016). Longtail tuna *Thunnus tonggol* (Bleeker, 1851) shows genetic partitioning across, but not within basins of the Indo-Pacific based on mitochondrial DNA. *Journal of Applied Ichthyology*, 32(2), 318-323.
- Williams, R. and Lester, R. (2006). Stock structure of Spanish mackerel *Scomberomorus commerson* along the Australian east coast deduced from parasite data. *Journal of Fish Biology* 68, 1707–1712.
- Yang, J, SH Lee, ME Goddard, PM Visscher. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* 88:76-82.
- Zamroni, A., Suwarso, S. and Wibowo, A. (2018). Genetic characterization of longtail tuna *Thunnus tonggol* (Bleeker, 1851) based on partial sequence of 16S rRNA mitochondrial gene. *Indonesian Fisheries Research Journal*, 24(2), 83-89.
- Zudaire I., Chassot E., Murua H., Dhurmeea Z., Cedras M. and Bodin N. (2016). Sex-ratio, size at maturity, spawning period and fecundity of bigeye tuna (*Thunnus obesus*) in the western Indian Ocean. OTC-2016-WPTT18-37.

## 8 Appendix 1: Detailed species results

### 8.1 Neritic Tuna

#### 8.1.1 Longtail tuna (*Thunnus tonggol*) - population genetics

To the best of our knowledge, five attempts were made at detecting genetic population structure across the Indo-Pacific distribution of *Thunnus tonggol* prior to this study. All five attempts were based on mitochondrial markers, which provide much less resolution than the approach deployed in this study. Some population structure was detected by Willette et al. (2016) between the Indian and Pacific oceans, but they didn't detect any structure at a finer scale within the Pacific Ocean. Moreover, Habib et al. (2012) reported that the Wallace's line acts as a barrier to gene flow for longtail tuna. No population structure was found between two locations on the coast of India using mitochondrial markers (Kunal 2018). Similarly, no structure was detected for this species within the Indonesian archipelago (Zamroni 2018).

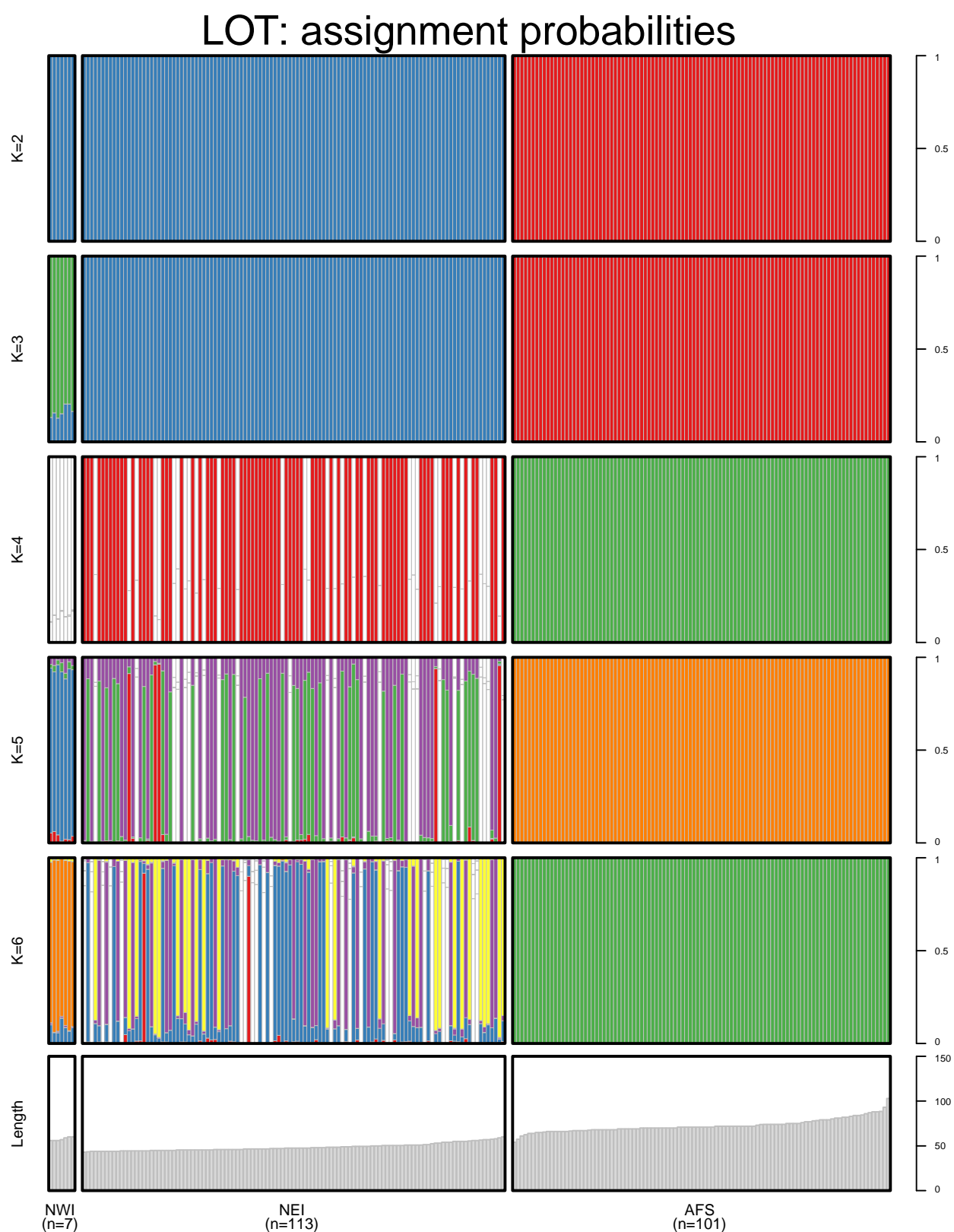
In this project we genotyped 353 Longtail tuna samples (including technical replicates) for a total of 123,600 SNPs. Our Radiator quality control left 221 unique individuals and 18,670 SNPs available for downstream analysis. This approach has been previously described in section 4.2.2 - 4.2.4 and details for each filter are presented in Append Table 1.

Our stockR analysis (Append Figure 1) revealed that each sampling location in the Indian ocean hosts a distinct genetic group, therefore providing unprecedented resolution into the population structure of Longtail tunas. These results both demonstrate the high detection power of our genotyping approach and highlight the need for further investigation of fine-scale population structure for this species, in particular the need for a sampling program with higher spatial resolution.



**Append Table 1. Radiator filtering steps for Longtail tuna *Thunnus tongol*, including threshold values and the number of individuals, locus and markers at the start of each step.**

<b>Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.95	353 / 76842 / 123600
Filter monomorphic markers		353 / 75515 / 120656
Filter markers in common		353 / 75515 / 120656
Filter individuals based on missingness	0.3	353 / 70846 / 115157
Filter monomorphic markers		299 / 70846 / 115157
Filter MAC	10	299 / 70390 / 113600
Filter coverage min / max	10 / 100	299 / 37319 / 47195
Filter genotyping	0.3	299 / 24046 / 31853
Filter SNPs position on the read	all	299 / 18702 / 24694
Filter markers snp number	4	299 / 18702 / 24694
Filter short Id	mac	299 / 18670 / 24531
detect mixed genomes	0 / 0.15	299 / 18670 / 18670
Filter monomorphic markers		267 / 18670 / 18670
detect duplicate genomes	0.25	267 / 18670 / 18670
Filter monomorphic markers		221 / 18670 / 18670



**Append Figure 1.** Individual length frequencies and results of population structure analysis of DArTSeq using StockR for all Longtail tuna assuming 2-6 genetic groups.

### 8.1.2 Longtail tuna (*Thunnus tonggol*) – otolith microchemistry

To date there have been no studies of longtail tuna otolith microchemistry. However the otoliths of longtail tuna are very similar in appearance and structure to those of the other *Thunnus* species so the methodology for this study was adopted from previous studies that have been successful in identifying natal origins and movements of other tunas (e.g. Artetxe-Arrate et al. 2019 for *Thunnus albacares*, Fraile et al. 2016 for *Thunnus alalunga*, and Rooker et al. 2016 for *Thunnus albacares* and *Thunnus obesus*).

#### Methods

The longtail otolith samples analysed were from three sampling locations (Append Figure 2), referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI) and Arafura Sea (Arafura) (Figure 2). The sample sizes for NWI and Arafura are very small (Append Figure 2, Append Table 2), which makes it difficult to draw conclusions when comparing data among locations. Although it was not intended in the original design, samples from different locations were collected during different periods due to several reasons (see Section 2).

Fish from NWI and NEI ranged in length from 43-60 cm FL; those from Arafura were larger, ranging from 69-88 cm FL (Append Figure 3, Append Figure 2, Append Table 2). While age at length information is very limited for longtail, we estimate fish from NWI and NEI to have been 1-3 years old at capture, and those from Arafura to be 3+ years. This means that fish from all locations will have been spawned over several years, and those from Arafura will have been spawned in earlier years than those from NWI and NEI. Thus, differences observed in core signatures among locations may be partly due to cohort effects.

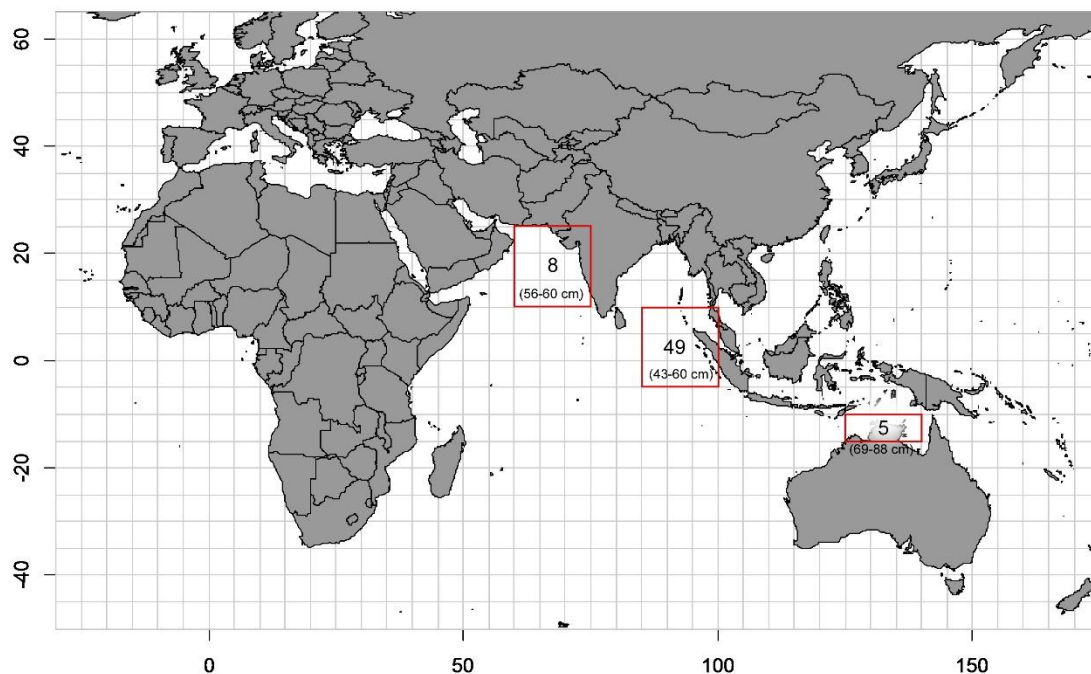
The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

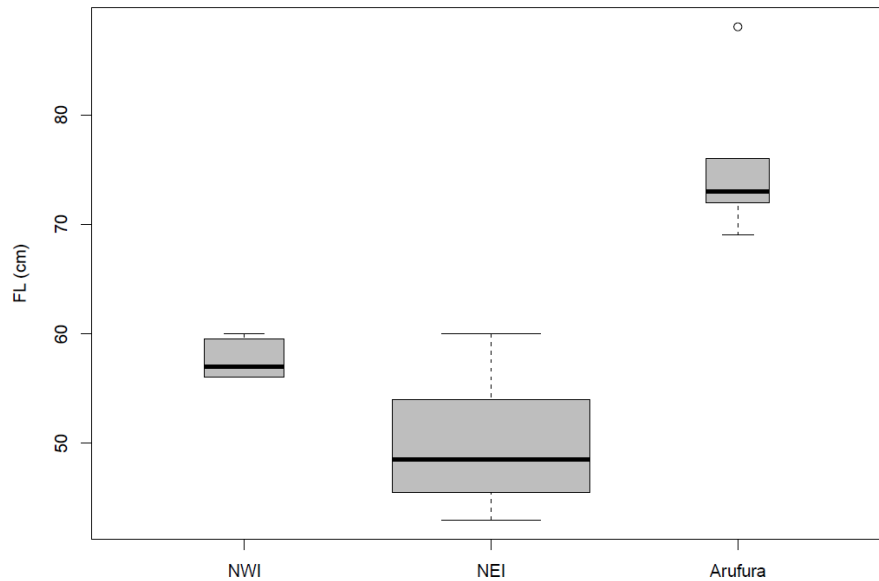
Analyses that were performed on the core or edge data:

- Univariate tests for each element to test for differences among locations. The exact test used depended on whether the data were normally distributed (assessed using `shapiro.test` in R) and had equal variances between locations (assesses using `fligner.test` in R), and whether there were two or more locations.
- PERMANOVA (non-parametric version of MANOVA): used to test for differences in the multi-elemental signatures of fish among locations and/or cluster groupings (results obtained using the `adonis` function from the `vegan` package in R).

- PCA: used to help visualize the data as this can be difficult with so many elements, and to determine which elements account for most of the variability in the data (results obtained using the `dudi.pca` function from the `ade4` package in R).
- Clustering: in some circumstances, used for identifying the most likely number of separate spawning origins, and to investigate whether spawning origins differ among fish from different sampling locations (results obtained using the `hclust` function in R with `method="ward.D2"`, and the dissimilarity matrix calculated using Euclidean distance).
- Note that prior to any multivariate analyses, the data were standardised (i.e., for each element, the data was centred by subtracting the mean and scaled by dividing by the standard deviation).



**Append Figure 2.** Map showing the number of longtail otoliths analysed for each of three sampling locations, referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI) and Arafura Sea (Arafura), and the size range of fish at each location.



**Append Figure 3.** Boxplots of longtail fork length (FL, cm) by sampling location, including only fish whose otoliths were selected for analysis.

**Append Table 2.** Number, sampling period, size range and estimated ages of fish for each of the sampling locations.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
North-West Indian Ocean (NWI)	8	August - September 2018	56-60	2-3
North-East Indian Ocean (NEI)	49	April and October 2018	43-60	1-3
Arafura Sea (Arafura)	5	Dec 2017	69-88	3-7

\* based on results from Abdussamad et al. 2012 and Griffiths et al. 2010.

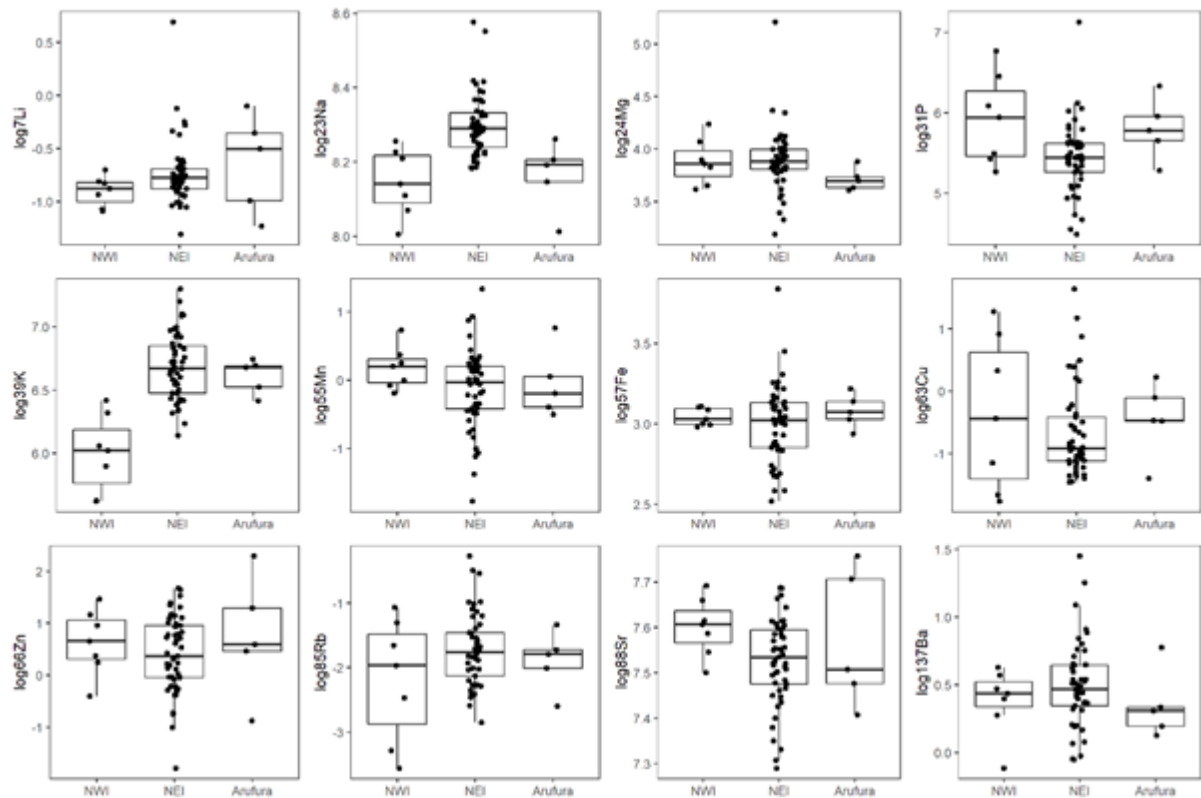
## Core Results

Core signatures were similar between locations for most elements (Append Figure 4). Univariate tests indicate that 39K, 23Na and 85Rb differ significantly among locations ( $p < 0.05$ ), but the low sample sizes mean these results must be treated with caution.

Results from fitting a PERMANOVA model to the core data suggests the multi-elemental core signatures of longtail are not equal among all locations ( $p = 0.003$ ). Based on subsequent pairwise tests between locations, we found this difference to be driven by a difference in core signatures between NWI and NEI (Append Table 3).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA supports the finding that the core data from fish captured in NWI may differ from NEI (Append Figure 5).

Clustering the data did not reveal further information about potential groups of fish. The core signatures did not separate into groups that may have represented different spawning locations; the optimal number of clusters was found to be one. When enforcing two clusters, one of the clusters was made up of one outlying fish.

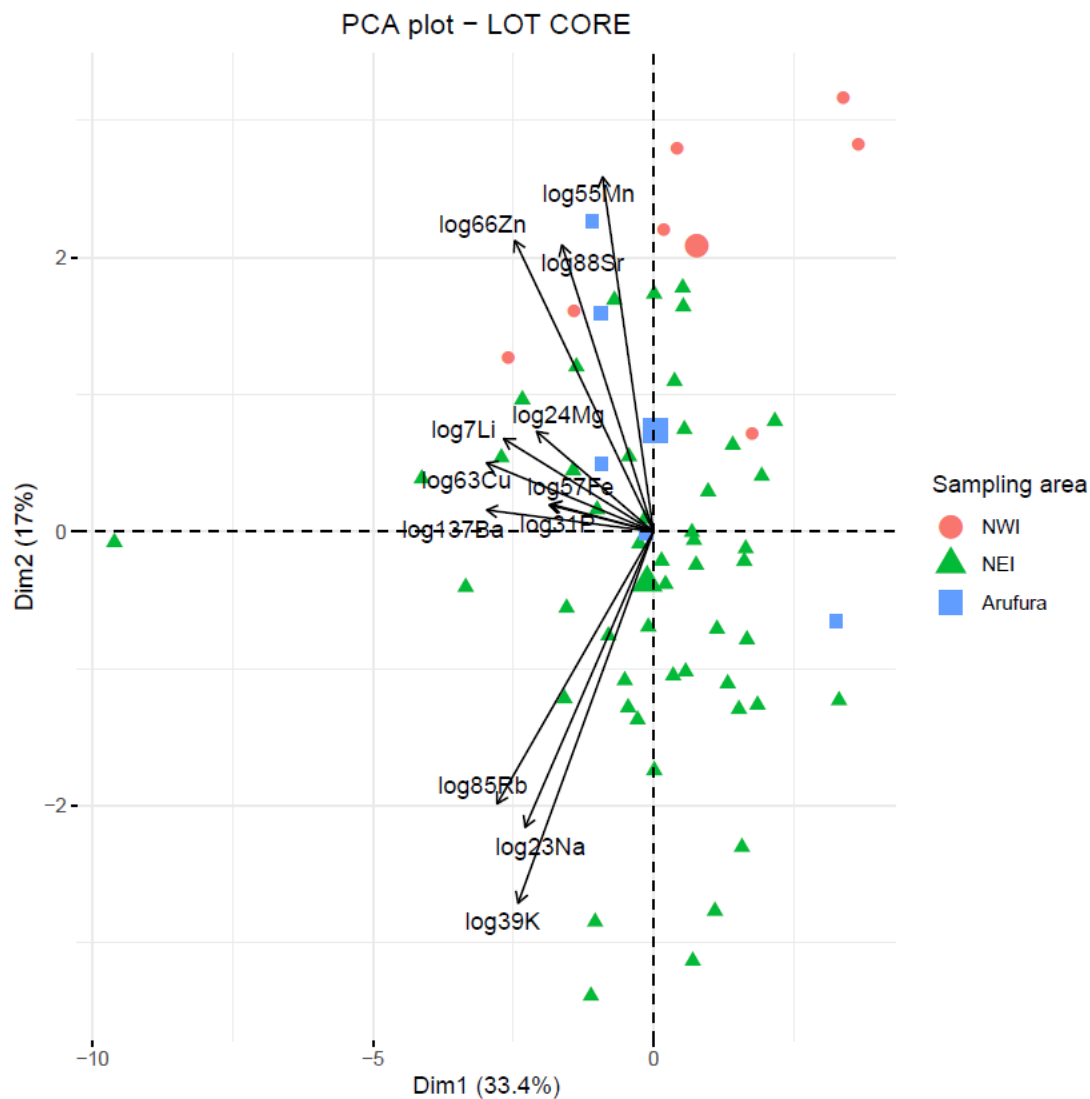


Append Figure 4. Boxplots comparing longtail otolith core data between locations for each element. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 3. Results of pairwise comparisons of longtail otolith core signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
NEI	vs	Arafura	1	17.455	1.563	0.029	0.163	0.227	
NEI	vs	NWI	1	59.087	5.319	0.090	0.001	0.003	*
Arafura	vs	NWI	1	14.993	1.315	0.116	0.227	0.227	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).

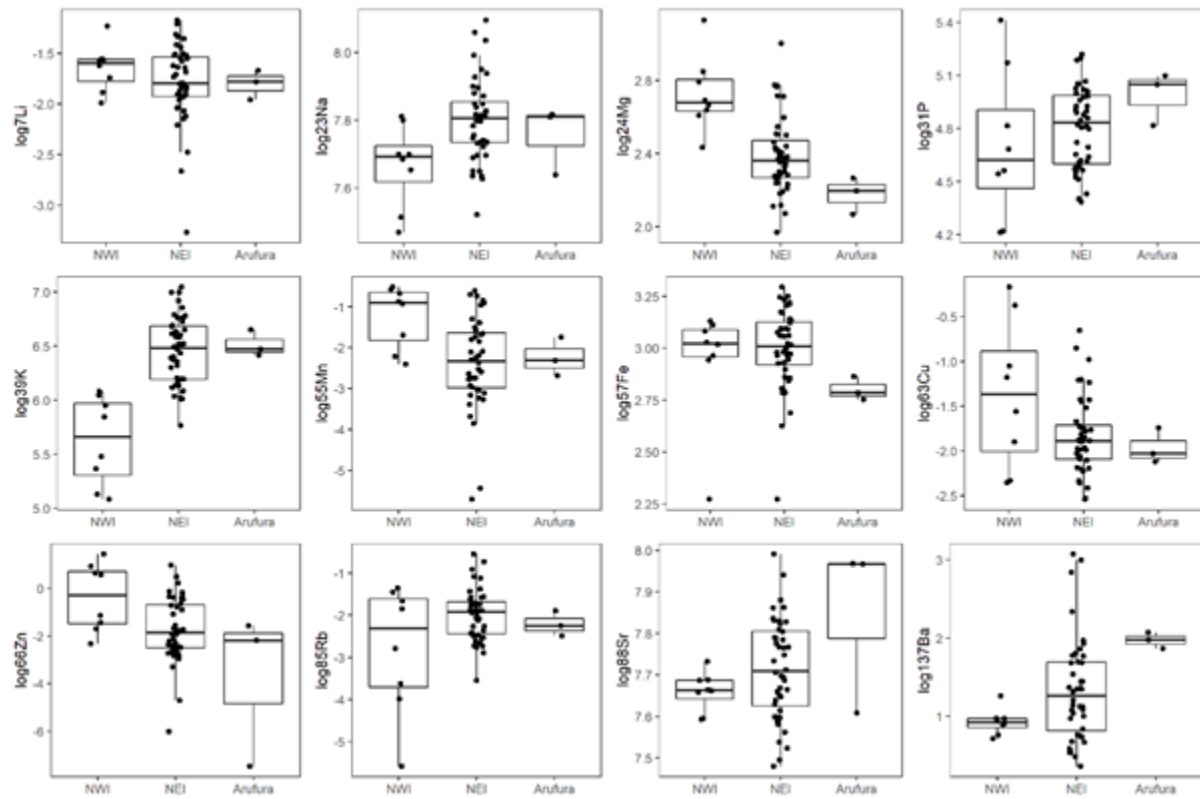


Append Figure 5. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the longtail otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge results

Edge signatures differed between locations for many elements (Append Figure 6). Based on univariate tests,  $^{23}\text{Na}$ ,  $^{24}\text{Mg}$ ,  $^{39}\text{K}$ ,  $^{55}\text{Mn}$ ,  $^{66}\text{Zn}$  and  $^{137}\text{Ba}$ , all differ significantly among locations ( $p < 0.05$ ). Results from the PERMANOVEdge ResultsA model also suggest edge signatures are not equal among all locations ( $p = 0.001$ ). Based on subsequent pairwise tests between locations, the edge signatures differ significantly between NWI and NEI, and between Arafura and NWI (Append Table 4). The result for Arafura vs NEI is less clear but there is some evidence of a weak difference (Append Table 4).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data confirms and helps to visualize these findings (Append Figure 7).



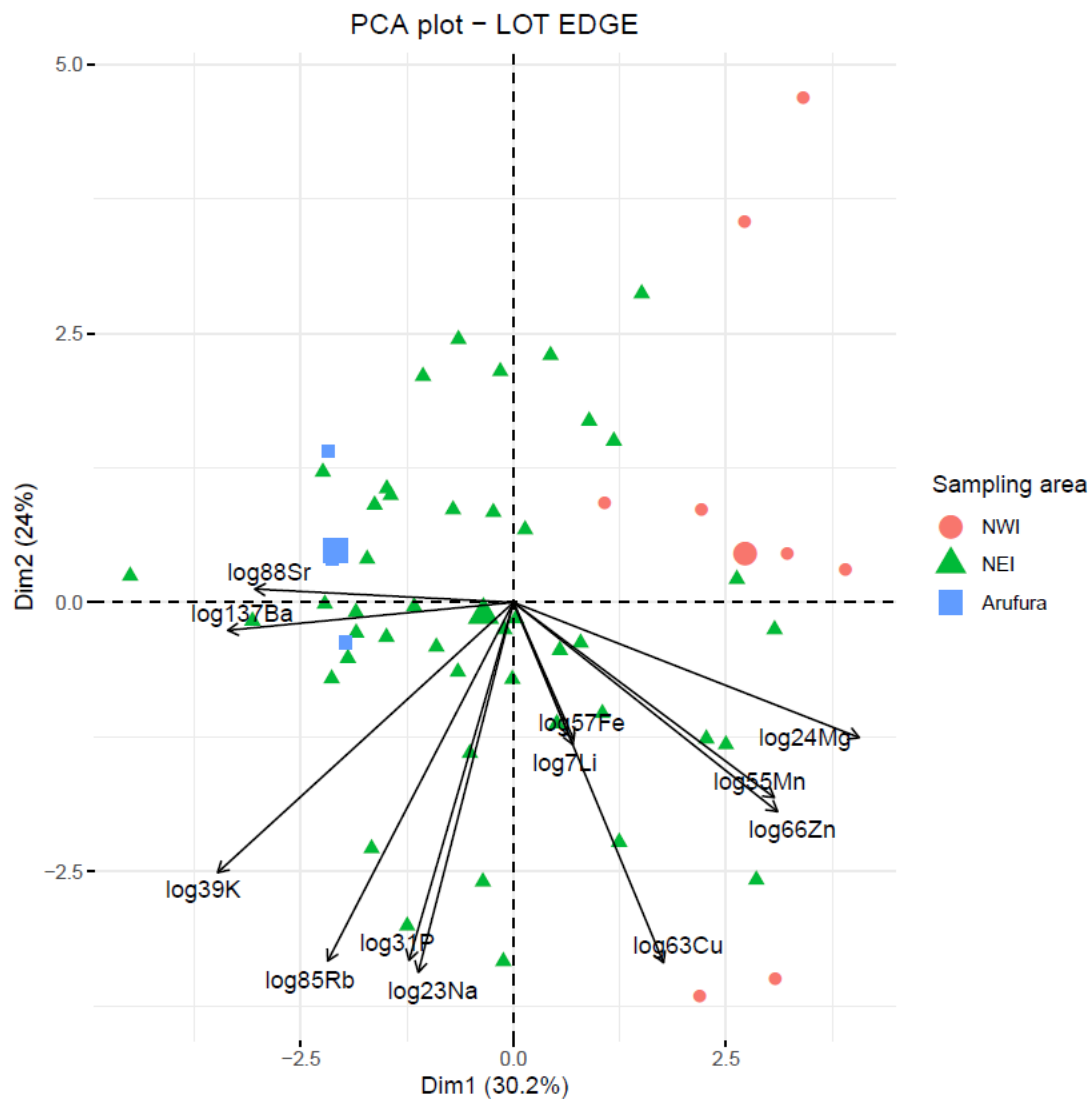
Append Figure 6. Boxplots comparing longtail otolith edge data for each element between locations. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 4. Results of pairwise comparisons of multi-elemental edge signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
NEI	vs	Arafura	1	17.773	1.796	0.038	0.089	0.089	
NEI	vs	NWI	1	76.963	7.240	0.126	0.001	0.003	*
Arafura	vs	NWI	1	52.677	3.877	0.301	0.007	0.011	.

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).





**Append Figure 7. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the longtail otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.**

## Summary

Core data should be representative of spawning locations. Thus we can test whether the core signatures differ among locations, which would indicate that fish caught in these discrete locations were spawned in different locations (recalling the caveat that differences could occur even if all fish were spawned in the same location if the water chemistry in that location differed significantly between the years they were spawned).

Because we do not have any longtail otoliths from very small/young fish, we cannot assume that the capture location is equal to the spawning location. However, we can test whether the core signatures differ among locations, which would indicate that fish caught in these fairly distant locations were spawned in different locations (recalling the caveat that differences could occur even if all fish were spawned in the same location if the water chemistry in that location differed significantly between the years they were spawned).

As noted previously, small sample sizes for two of the three locations means it is not possible to confidently draw any conclusions. Nevertheless, the fact that the core signatures between fish caught in NWI and NEI show evidence of being different suggests these fish may have been spawned in different locations.

The core signatures of the five fish caught in Arafura overlap with those from NWI and NEI, so it is difficult to know if:

- i) they are from a distinct spawning area,
- ii) some fish were spawned in the same location as those caught in NEI, and others in the same location as those caught in NWI, or
- iii) they are from the same spawning ground as fish caught in one of the other locations but have slightly different core signatures due to differences in ocean chemistry between the years they were spawned (recall that the Arafura fish are older and were spawned in earlier years than fish from the other two locations).

Edge data should be representative of capture locations. Thus, if we find differences in edge signatures between locations, this verifies that otolith chemistry can be useful for classifying fish to these locations. The fact that the edge signatures of longtail appear to differ between the three capture locations suggests the ocean chemistry differs enough between these locations to be useful for classifying fish to them using otolith chemistry.

### 8.1.3 Kawakawa (*Euthynnus affinis*) - population genetics

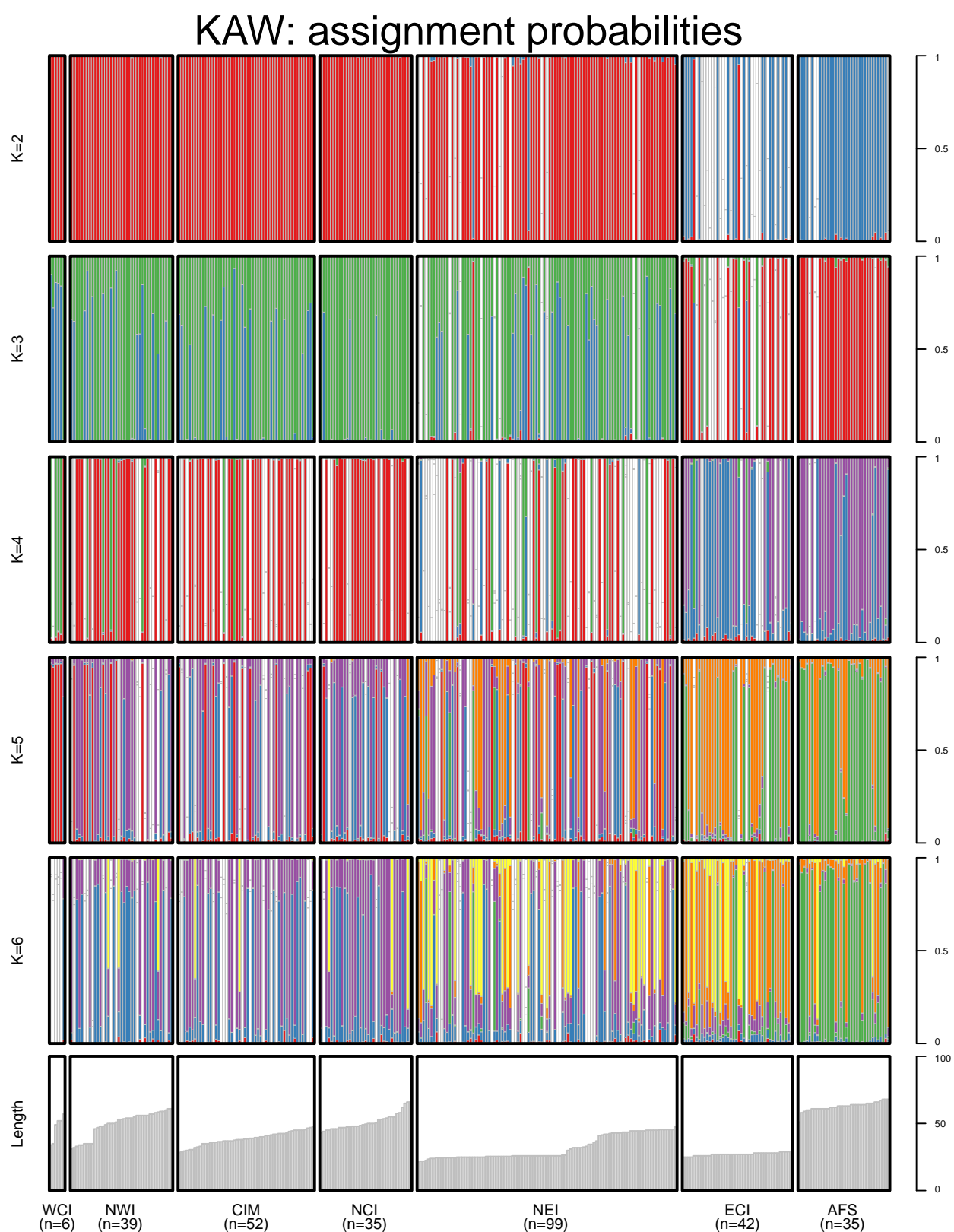
Investigation of the genetic population structure of Kawakawa *Euthynnus affinis* thus far has been limited to the analysis of mitochondrial DNA. Kumar et al. (2012) investigated the genetic population structure of Kawakawa along the coast of India, whereas Jonhson et al. (2016) compared two different localities on the northern coast of Tanzania and Masazurah et al. (2012) searched for population structure within the strait of Malacca. None of these studies detected any sign of genetic structure. In Southeast Asia, the analysis on the analysis of a portion of the control region by Santos et al. (2010) did not reveal any genetic structure. However, Jackson et al. (2014) also analysed a portion of the Control Region and detected multiple barriers to gene flow within the Indonesian Archipelago showing that genetic differentiation can happen at relatively small geographic scale in this species.

In this project we genotyped 432 Kawakawa samples (including technical replicates) for a total of 66,910 SNPs. Our Radiator quality control left 308 unique individuals and 10,334 SNPs available for downstream analysis. This filtering approach has been previously described in section 4.2.2 - 4.2.4 and details for each filter are presented in Append Table 5.

Our stockR analysis (Append Figure 8) detected the presence of at least two distinct genetic group amongst our seven sampling locations. The barrier to gene flow is located between the NEI and the ECI sampling locations. Possible explanations for the higher proportion of individuals with uncertain assignment in these two sampling locations includes: i) ongoing mixing between the two genetic groups still occurs or ii) the divergence is fairly recent and there is incomplete lineage sorting. This also shows that the gene pool is not completely homogeneous on each side of the genetic break. It is possible that more demanding genetic approach, such as Close-Kin Mark-Recapture (Feutry et al 2017) or the study of shared fragment of genomic fragments (Ralph & Coop 2013), would help resolve finer patterns of genetic structure in this species.

**Append Table 5. Radiator filtering steps for Kawakawa *Euthynnus affinis*, including threshold values and the number of individuals, locus and markers at the start of each step.**

<b>Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.95	432 / 48572 / 66910
Filter monomorphic markers		432 / 48068 / 66188
Filter markers in common		432 / 48068 / 66188
Filter individuals based on missingness	0.3	432 / 47358 / 65383
Filter monomorphic markers		416 / 47358 / 65383
Filter MAC	20	416 / 42040 / 58699
Filter coverage min / max	10 / 100	416 / 24562 / 31527
Filter genotyping	0.3	416 / 16935 / 22223
Filter SNPs position on the read	all	416 / 10383 / 14335
Filter markers snp number	4	416 / 10383 / 14335
Filter short Id	mac	416 / 10334 / 14074
detect mixed genomes	0 / 0.195	416 / 10334 / 10334
Filter monomorphic markers		390 / 10334 / 10334
detect duplicate genomes	0.25	390 / 10334 / 10334
Filter monomorphic markers		308 / 10334 / 10334



**Append Figure 8.** Individual length frequencies and results of population structure analysis of DArTSeq using StockR for all kawakawa tuna assuming 2-6 genetic groups.

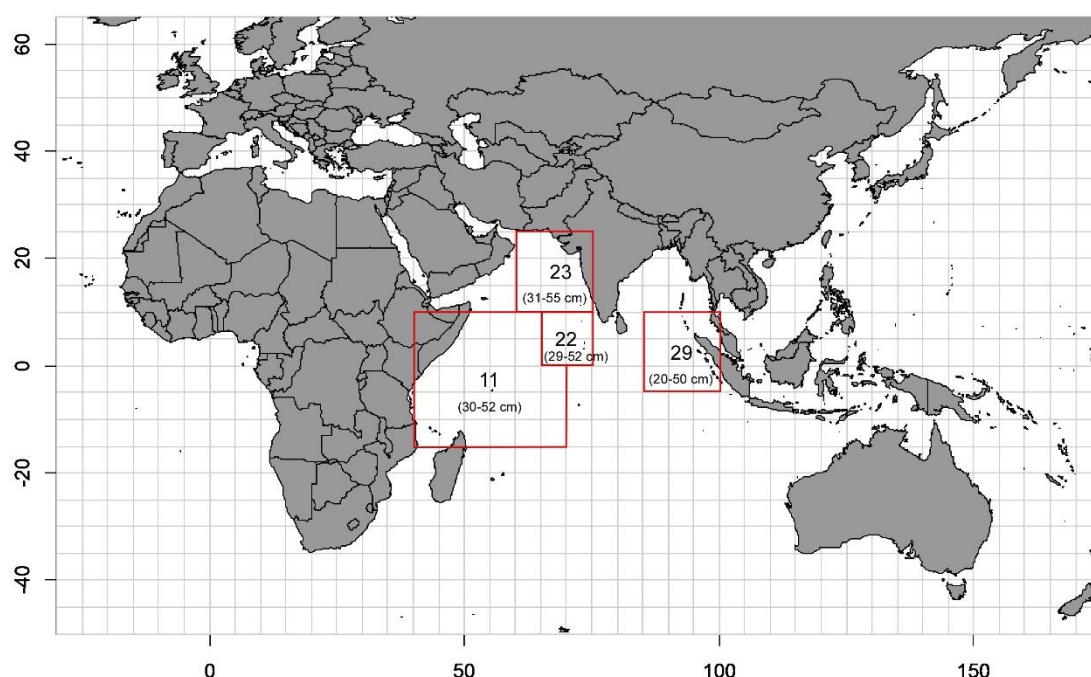
#### 8.1.4 Kawakawa (*Euthynnus affinis*) - otolith microchemistry

To date there have been no studies of longtail tuna otolith microchemistry. However the otoliths of kawakawa are similar in appearance and structure to those of the related tuna species so the methodology for this study was adopted from previous studies that have been successful in identifying natal origins and movements of the *Thunnus* species (e.g. Artetxe-Arrate et al. 2019 for *Thunnus albacares*, Fraile et al. 2016 for *Thunnus alalunga*, and Rooker et al. 2016 for *Thunnus albacares* and *Thunnus obesus*).

##### Methods

The kawakawa otolith samples analysed are from four sampling locations (Append Figure 9), referred to as North-West Indian Ocean (NWI), Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI) (Figure 2). Although it was not intended in the original design, samples from different locations were collected during different periods, due to several reasons (see Section 2).

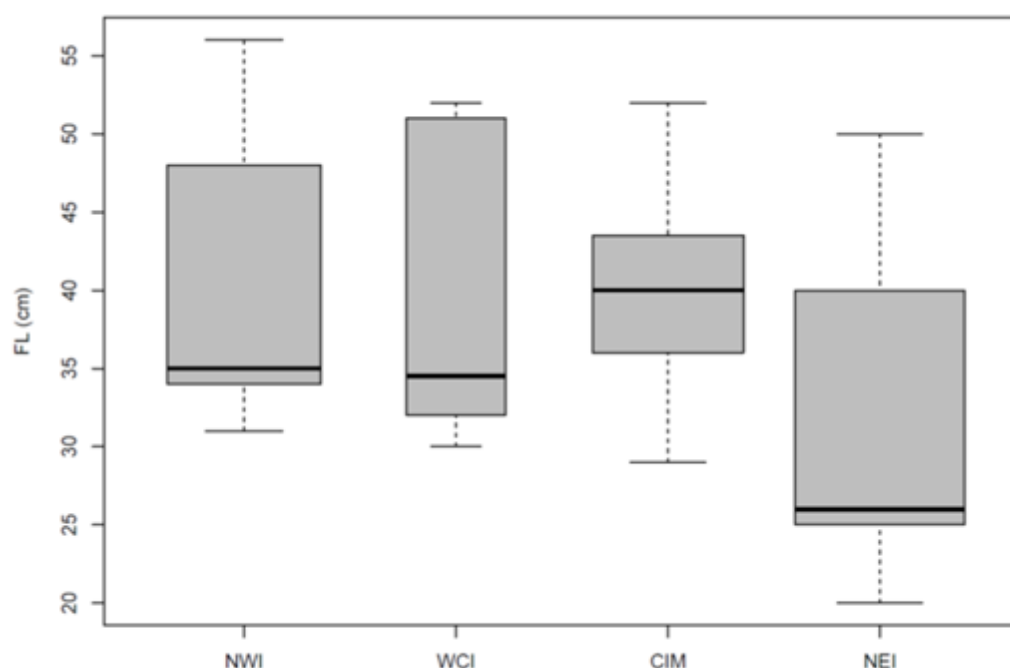
Fish ranged in fork length from 20 to 55 cm across all locations, with the fish from NEI being slightly smaller on average (Append Figure 9, Append Figure 10, Append Table 6). Juveniles grow rapidly and are estimated to be 50-65 cm by age 3, meaning fish sampled from all locations are likely to have been spawned over a few years. Thus, differences observed in core signatures among locations may be partly due to cohort effects.



Append Figure 9. Map showing the number of kawakawa otoliths analysed for each of four sampling locations, referred to as North-West Indian Ocean (NWI), Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI), and the size range of fish at each location.

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However, it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.



Append Figure 10. Boxplots of kawakawa fork length (FL, cm) by sampling location, including only fish whose otoliths were selected for analysis.

Append Table 6. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
North-West Indian Ocean (NWI)	23	Apr-May 2018	31-55	0+ to 3
Western Central Indian Ocean (WCI)	11	Feb-Apr 2018	30-52	0+ to 3
Maldives (CIM)	22	Aug 2018 and Feb 2019	29-52	0+ to 3
North-East Indian Ocean (NEI)	29	Apr and Nov 2018	20-50	0+ to 2

\* based on results from Kahn (2004)

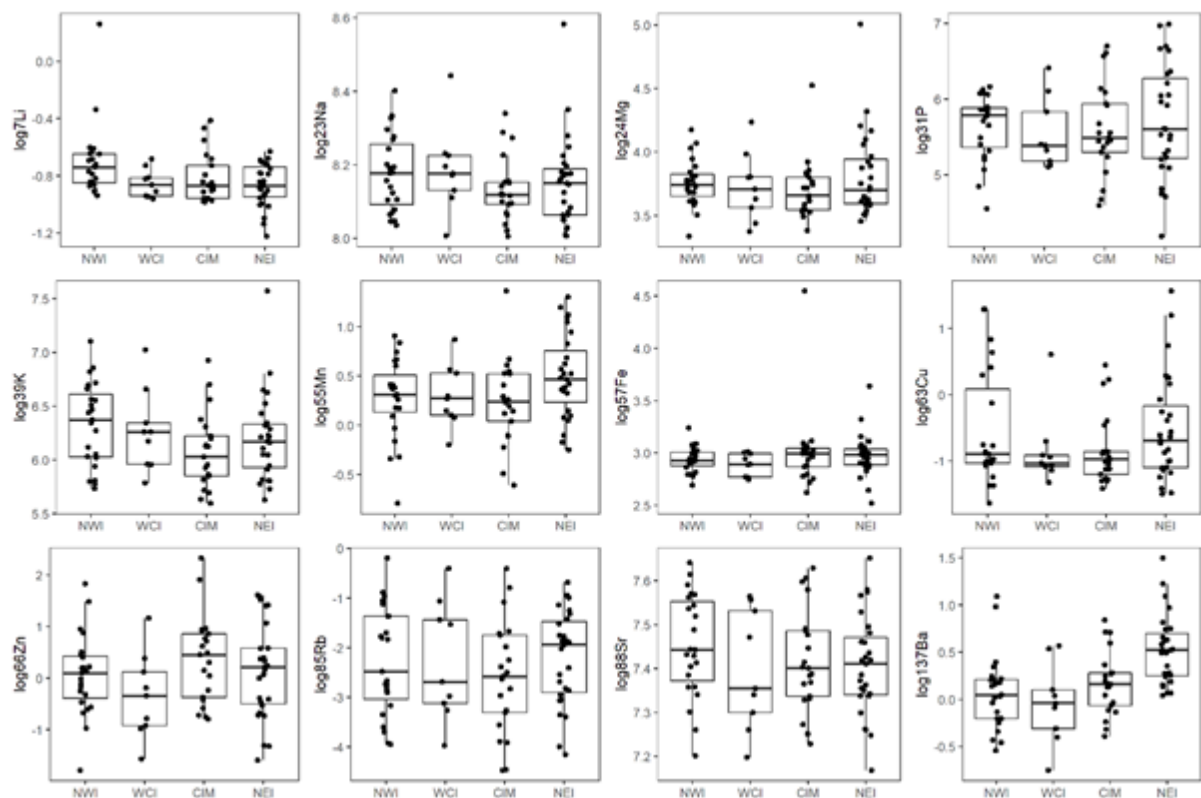
## Core results

Core signatures were similar between locations for most elements but there was large variability in the data (Append Figure 11). Univariate tests indicate that only  $^{137}\text{Ba}$ , and to a lesser extent  $^7\text{Li}$ , differ significantly among locations ( $p < 0.05$ ).

Although results from fitting a PERMANOVA model to the core data provide weak evidence that the multi-elemental core signatures of kawakawa are not equal among all locations ( $p = 0.033$ ), subsequent pairwise tests between locations showed no significant differences (Append Table 7).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA indicates that the core data does not differ significantly among locations (Append Figure 12).

Clustering the data did not reveal further information about potential groups of fish. The core signatures did not separate into groups that may have represented different spawning locations; the optimal number of clusters was found to be one.



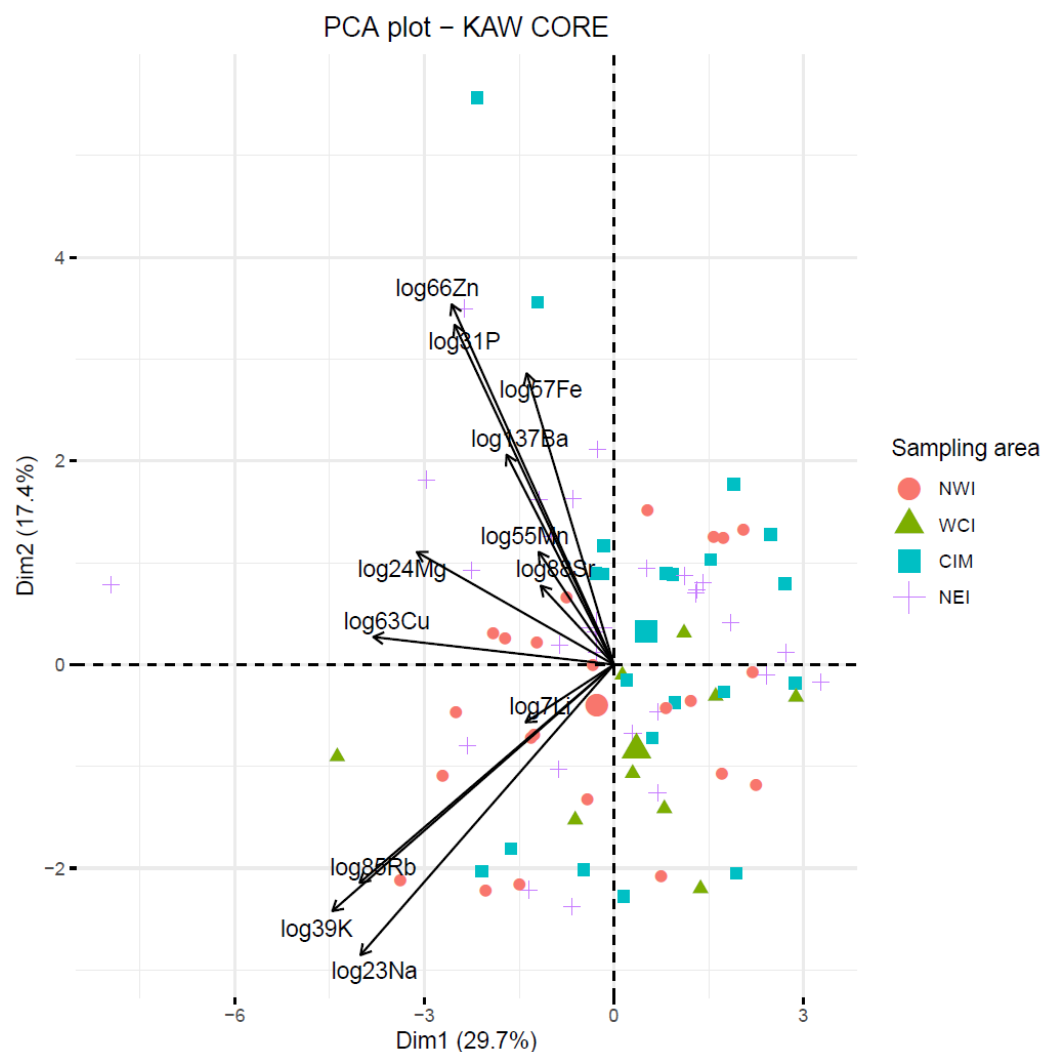
Append Figure 11. Boxplots comparing kawakawa otolith core data between locations for each element. Note that the data have been log-transformed to reduce skewness and facilitate comparison.



**Append Table 7. Results of pairwise comparisons of kawakawa otolith core signatures between locations.** In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
NEI	vs	CIM	1	22.678	1.843	0.039	0.081	0.219	
NEI	vs	WCI	1	20.043	1.664	0.047	0.110	0.219	
NEI	vs	NWI	1	31.213	2.620	0.052	0.019	0.114	
CIM	vs	WCI	1	11.873	1.051	0.036	0.393	0.472	
CIM	vs	NWI	1	17.642	1.548	0.036	0.146	0.219	
WCI	vs	NWI	1	9.002	0.838	0.027	0.525	0.525	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



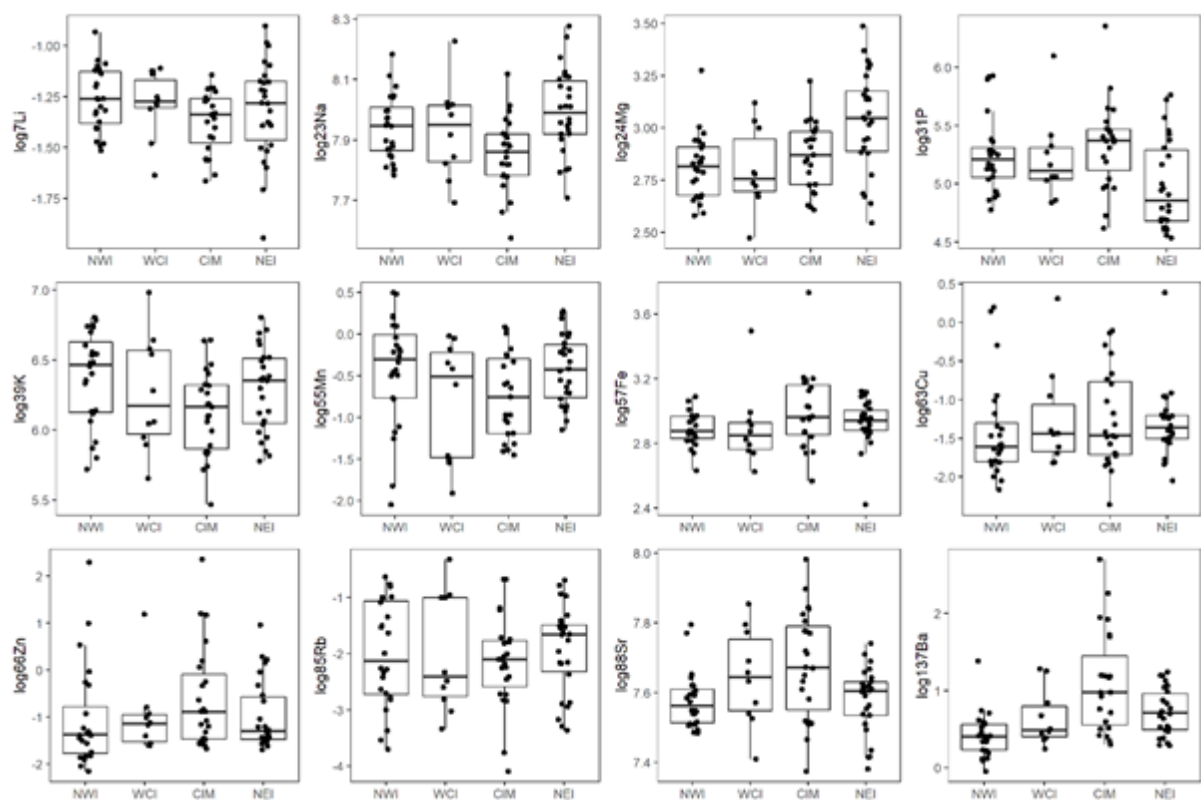
**Append Figure 12. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the kawakawa otolith core signatures.** Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge results

Boxplots comparing edge signatures between locations show there is large variability in the data, but that a few elements appear to differ among locations (Append Figure 13). Based on univariate tests,  $^{23}\text{Na}$ ,  $^{24}\text{Mg}$ ,  $^{31}\text{P}$  and  $^{137}\text{Ba}$  differ significantly among locations ( $p < 0.05$ ).

Results from the PERMANOVA model also suggest edge signatures are not equal among all locations ( $p = 0.001$ ). Based on subsequent pairwise tests between locations, the edge signatures differ significantly between CIM and NEI, CIM and NWI, and NEI and NWI (Append Table 8).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data helps to visualize these findings, although the difference between NEI and NWI is not obvious in this figure (Append Figure 14).

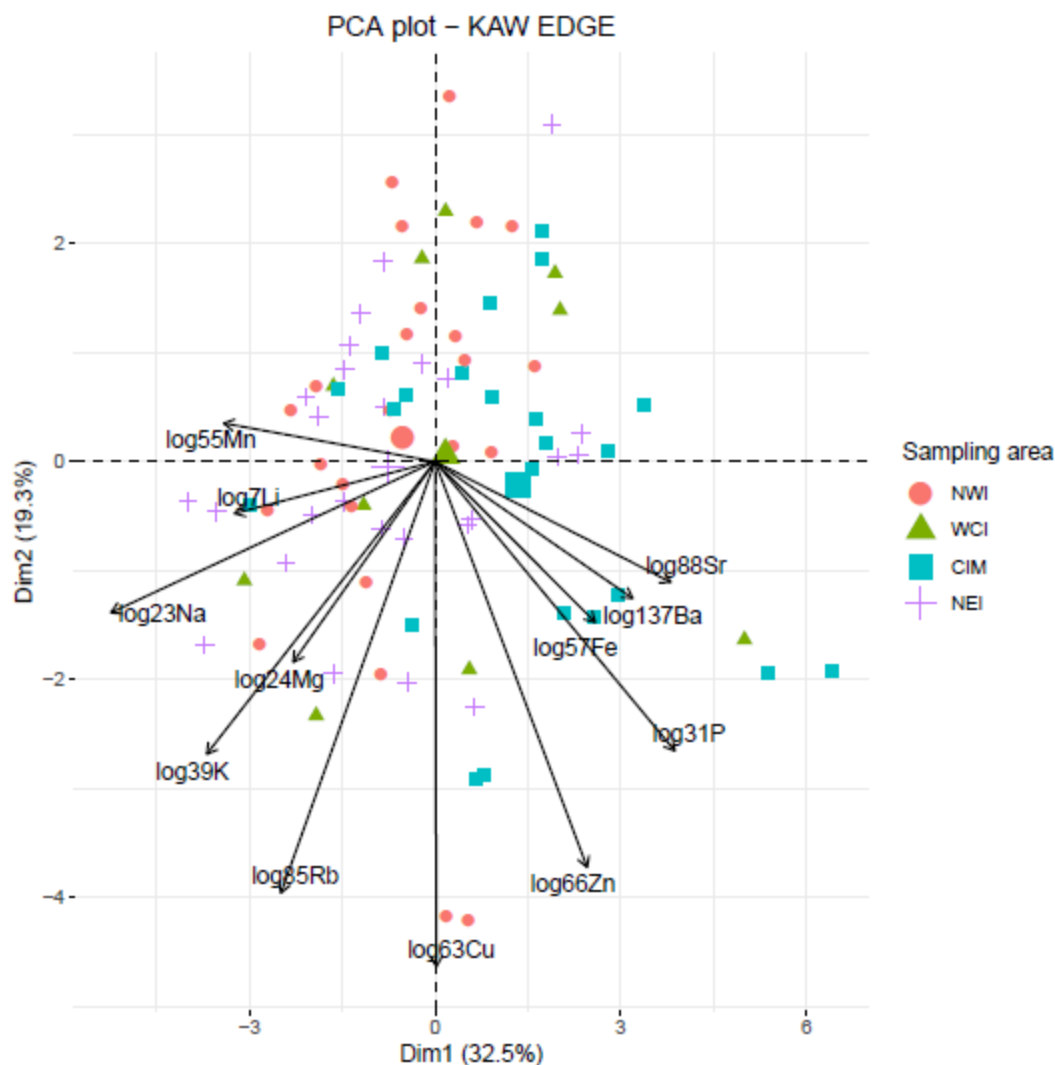


Append Figure 13. Boxplots comparing kawakawa otolith edge data for each element between locations. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 8. Results of pairwise comparisons of multi-elemental edge signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
NEI	vs	CIM	1	58.813	5.188	0.099	0.001	0.003	*
NEI	vs	WCI	1	18.839	1.724	0.048	0.108	0.162	
NEI	vs	NWI	1	28.902	2.966	0.058	0.007	0.014	.
CIM	vs	WCI	1	16.484	1.237	0.038	0.270	0.324	
CIM	vs	NWI	1	58.445	5.165	0.103	0.001	0.003	*
WCI	vs	NWI	1	7.936	0.730	0.022	0.634	0.634	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 14. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the kawakawa otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Summary

The fact that the core signatures do not show significant differences among locations suggests either that these fish all originated from the same location, or that they originated from different locations with similar water chemistry.

The fact that the edge signatures of kawakawa appear to differ between the some of the capture locations (particularly CIM and NEI, and CIM and NWI) suggests that using otolith chemistry to classify fish to these locations may be possible.

### 8.1.5 Narrow-barred Spanish mackerel (*Scomberomorus commerson*) - population genetics

Several studies investigated the genetic population structure of the Narrow-barred Spanish mackerel *Scomberomorus commerson* using either mitochondrial DNA or microsatellite nuclear markers. These studies generally report evidence strong genetic population structure in this species at the regional scale. At least three genetic stocks have been identified in the Pacific (Shaklee et al. 1990, JB Shaklee in Buckworth et al. 2007, Sulaiman & Ovenden 2010). In fact, the genetic differentiation is so high that possibility of having in fact a complex of cryptic species has been pushed forward (Fauvelot & Borsa 2011). In the Indian ocean samples from the Persian Gulf and the Oman Sea differ from those found in the Timor Sea (Fauvelot & Borsa 2011). It seems that the Bay of Bengal also hosts a distinct population but the sampling size available to Habib & Sulaiman (2017) was too low (N=5) to allow definite conclusions. Moreover, no genetic structure was reported between the Arabian Gulf, the Gulf of Oman and the Arabian Sea (Hoolihan et al. 2006) or along the coast of India (Vineesh et al. 2016) using mitochondrial DNA. Using microsatellite markers Abedi et al. 2012 did not find any genetic differentiation inside the Persian Gulf, but two different stocks were identified around the Arabian Peninsula (van Herwerden et al. 2006).

In this project we genotyped 258 Spanish Mackerel samples (including technical replicates) for a total of 194,591 SNPs. For this species we conducted two rounds of radiator quality control and stockR analyses, one on the entire dataset and one on three subsets of the data. The three subsets, each corresponding of two sampling locations (NWI/NCI, NEI/ECI and AFS/WCS), exhibit very large differences in heterozygosity and possibly represent cryptic species like it has been hypothesised by Fauvelot & Borsa (2011), which warrants dedicated quality control and analysis. After the first round of Radiator quality control on the entire dataset, there were 189 unique individuals and 17,500 SNPs left for analysis. After the second round of Radiator quality control, the NWI/NCI subset comprised 32 unique individuals and 7570 SNPs, the NEI/ECI subset comprised 48 unique individuals and 13,981 SNPs and the AFS/WCS comprised 73 unique individuals and 12010 SNPs. The quality control approach has been previously described in section 4.2.2 - 4.2.4 and details for each filter are presented in Append Table 9.

Out of all the species analysed as part of this project, Spanish Mackerel exhibited the strongest genetic differentiation within the Indian Ocean. In fact, our first round of analysis including six locations revealed large differences in heterozygosity between three regions, each including two sampling locations, consistent with the presence of multiple cryptic species. Whilst our data can't determine if these regions host distinct species or just highly differentiated population, this result warranted the analysis of each of these three regions separately. The first round of stockR analysis detected four clearly distinct genetic groups (Append Figure 15). The first one was made of the NWI and NCI sampling locations, the second and third one corresponded to the NEI and ECI sampling locations respectively and the fourth group comprised the AFS and WCS sampling locations. The second round of stockR analyses allowed to detect further spatial clustering within this fourth group (Append Figure

18), which there was a hint of at  $K = 5$  of the first round of analysis (Append Figure 15), but not within the other groups (Append Figure 16 and Append Figure 17). Having only 32 individuals that past quality control for NWI and NCI combined was possibly a limiting factor to detect further clustering in that region. Nevertheless, this study detected a total of five genetic groups amongst six sampling locations demonstrating further the limited genetic connectivity in Spanish Mackerel reported in other parts of its range (Shaklee et al. 1990, van Herwerden et al. 2006, Buckworth et al 2007, Sulaiman & Ovenden 2010, Fauvelot & Borsa 2011). The fact that almost each location sampled as part of this project host a distinct genetic group warrants the investigation of the spatial resolution of genetic connectivity at even finer scale for this species.

**Append Table 9. Radiator filtering steps for Spanish Mackerel *Scomberomorus commerson*, including threshold values and the number of individuals, locus and markers at the start of each step. A - First round of filtering on all locations. B - Second round of filtering: 1 – NWI and NCI only; 2 – NEI and ECI only; 3 – AFS and WCS only.**

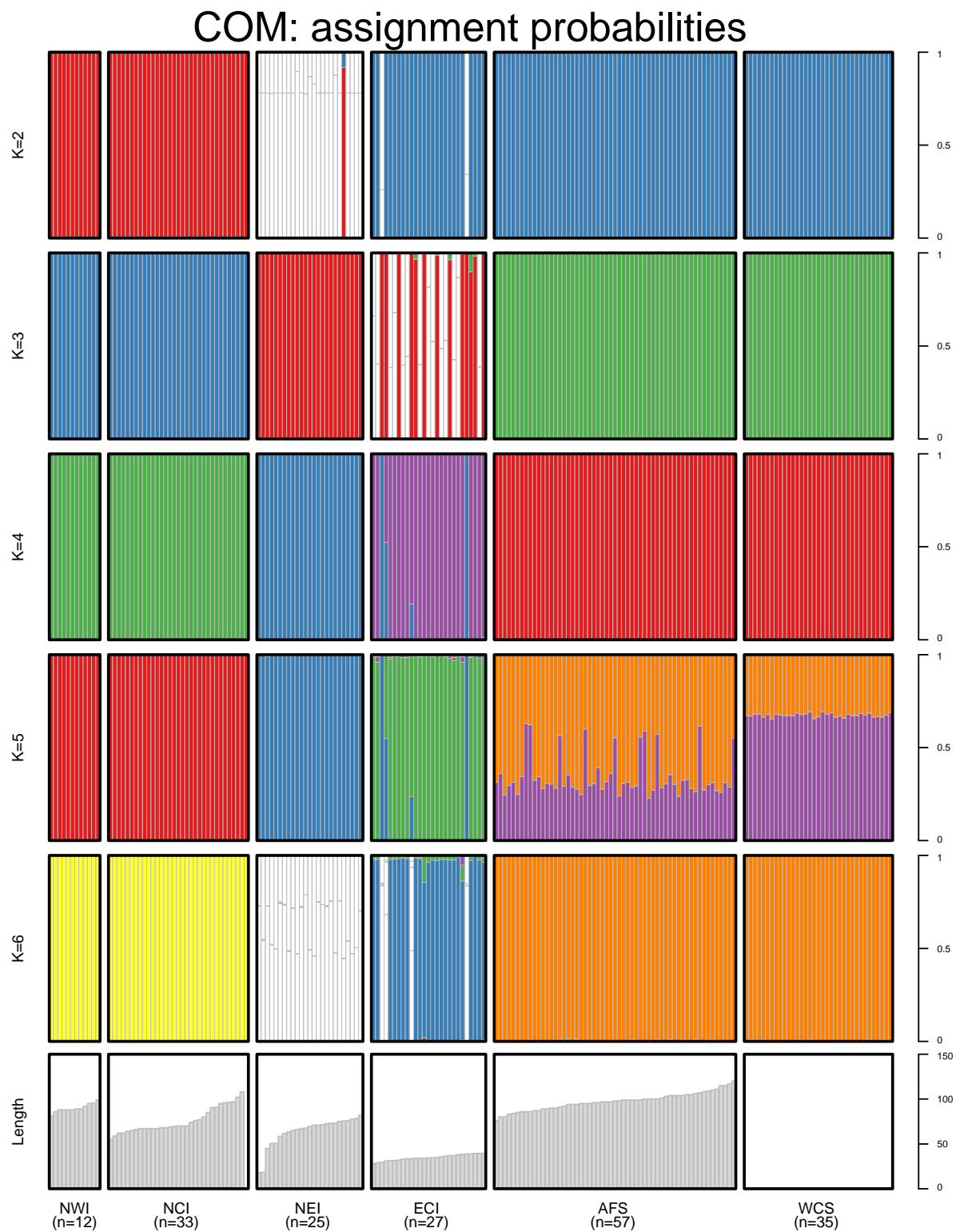
<b>A-Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.95	258 / 106487 / 194591
Filter monomorphic markers		258 / 103039 / 187335
Filter markers in common		258 / 103039 / 187335
Filter individuals based on missingness	0.3	258 / 87517 / 163895
Filter monomorphic markers		246 / 87517 / 163895
Filter MAC	20	246 / 80140 / 144145
Filter coverage min / max	10 / 100	246 / 38189 / 47617
Filter genotyping	0.3	246 / 21721 / 28002
Filter SNPs position on the read	all	246 / 17509 / 22465
Filter markers snp number	4	246 / 17509 / 22465
Filter short Id	mac	246 / 17500 / 22419
detect mixed genomes	0 / 0.2	246 / 17500 / 17500
Filter monomorphic markers		237 / 17500 / 17500
detect duplicate genomes	0.125	237 / 17500 / 17500
Filter monomorphic markers		189 / 17500 / 17500

<b>B1-Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.95	56 / 106487 / 194591
Filter monomorphic markers		56 / 103039 / 187335
Filter markers in common		56 / 67409 / 104197
Filter individuals based on missingness	0.4	56 / 58585 / 93435
Filter monomorphic markers		48 / 58585 / 93435
Filter MAC	10	48 / 50014 / 76560
Filter coverage min / max	10 / 100	48 / 15494 / 18779
Filter genotyping	0.3	48 / 8905 / 10908
Filter SNPs position on the read	all	48 / 7573 / 8736

Filter markers snp number	4	48 / 7573 / 8736
Filter short Id	mac	48 / 7571 / 8725
detect mixed genomes	0 / 0.3	48 / 7571 / 7571
Filter monomorphic markers		35 / 7571 / 7571
detect duplicate genomes	0.9	35 / 7570 / 7570
Filter monomorphic markers		32 / 7570 / 7570

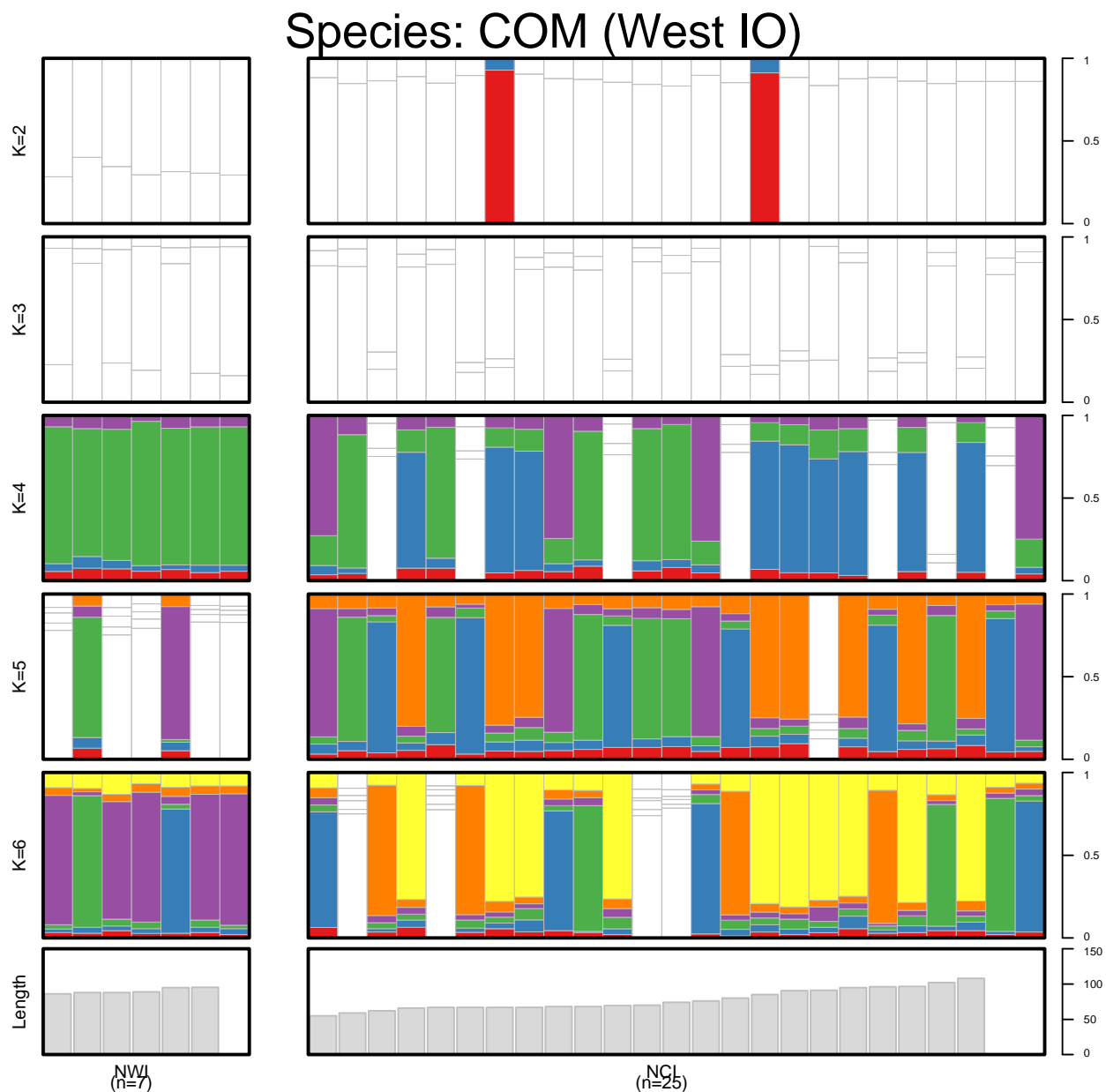
<b>B2-Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.95	77 / 106487 / 194591
Filter monomorphic markers		77 / 103039 / 187335
Filter markers in common		77 / 68227 / 101404
Filter MAC	10	77 / 67999 / 101126
Filter coverage min / max	10 / 100	77 / 30519 / 36140
Filter genotyping	0.3	77 / 17393 / 20845
Filter SNPs position on the read	all	77 / 13981 / 16636
Filter markers snp number	1E+12	77 / 13981 / 16636
Filter short Id	mac	77 / 13981 / 16636
detect mixed genomes	0.19 / 0.274	77 / 13981 / 13981
Filter monomorphic markers		65 / 13981 / 13981
detect duplicate genomes	0.25	65 / 13981 / 13981
Filter monomorphic markers		48 / 13981 / 13981
Filter HWE	20.01	48 / 13981 / 13981
Filter DArT reproducibility	0.95	77 / 106487 / 194591

<b>B3-Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.95	125 / 106487 / 194591
Filter monomorphic markers		125 / 103039 / 187335
Filter markers in common		125 / 63228 / 92466
Filter individuals based on missingness	0.25	125 / 61355 / 90357
Filter monomorphic markers		121 / 61355 / 90357
Filter MAC	10	121 / 55876 / 80593
Filter coverage min / max	10 / 100	121 / 22759 / 26736
Filter genotyping	0.3	121 / 14225 / 17018
Filter SNPs position on the read	all	121 / 12012 / 14239
Filter markers snp number	4	121 / 12012 / 14239
Filter short Id	mac	121 / 12011 / 14234
detect mixed genomes	0 / 0.245	121 / 12011 / 12011
Filter monomorphic markers		100 / 12011 / 12011
detect duplicate genomes	0.75	100 / 12010 / 12010
Filter monomorphic markers		73 / 12010 / 12010

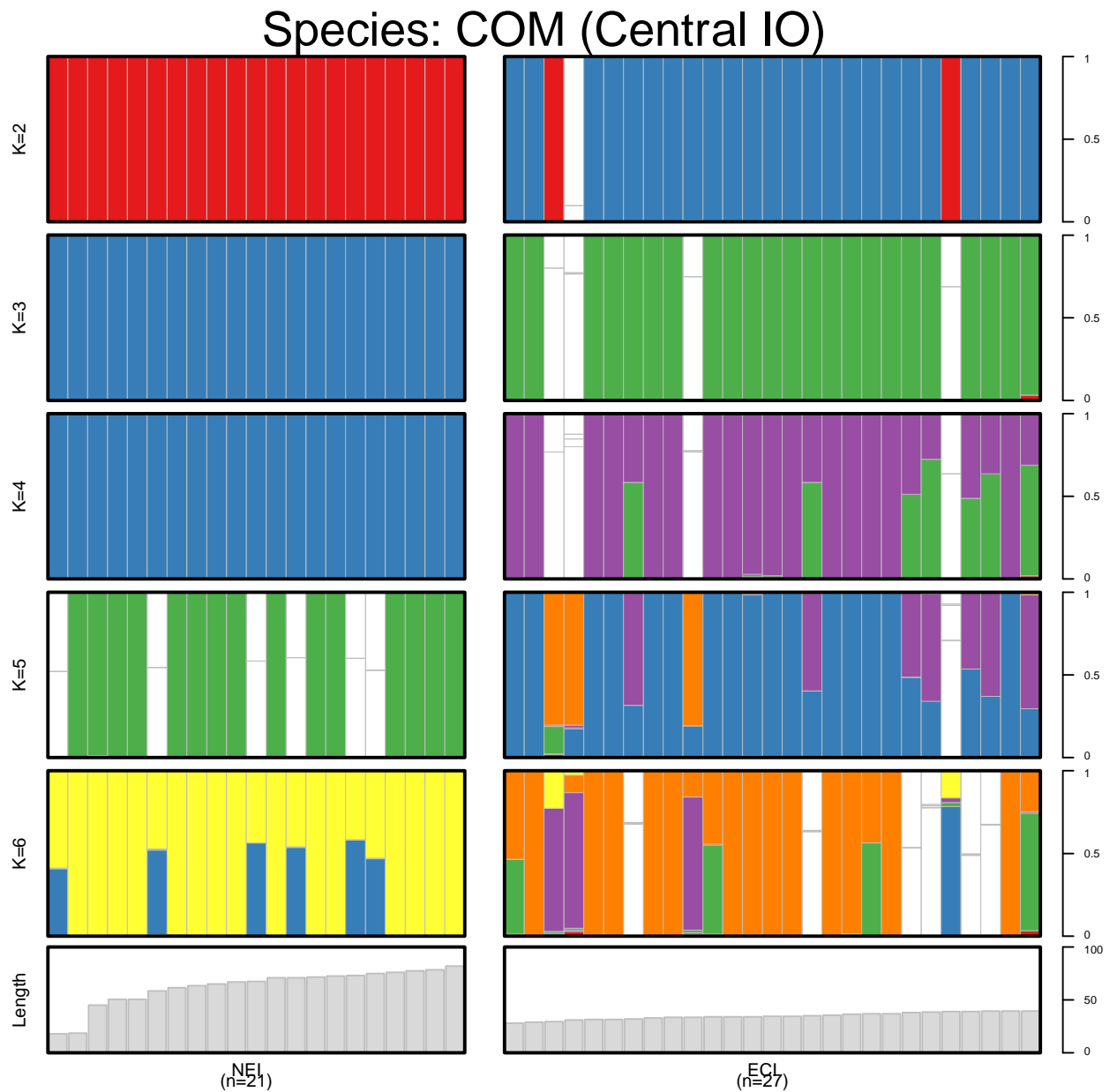


**Append Figure 15. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for all Spanish Mackerel assuming 2-6 genetic groups.**

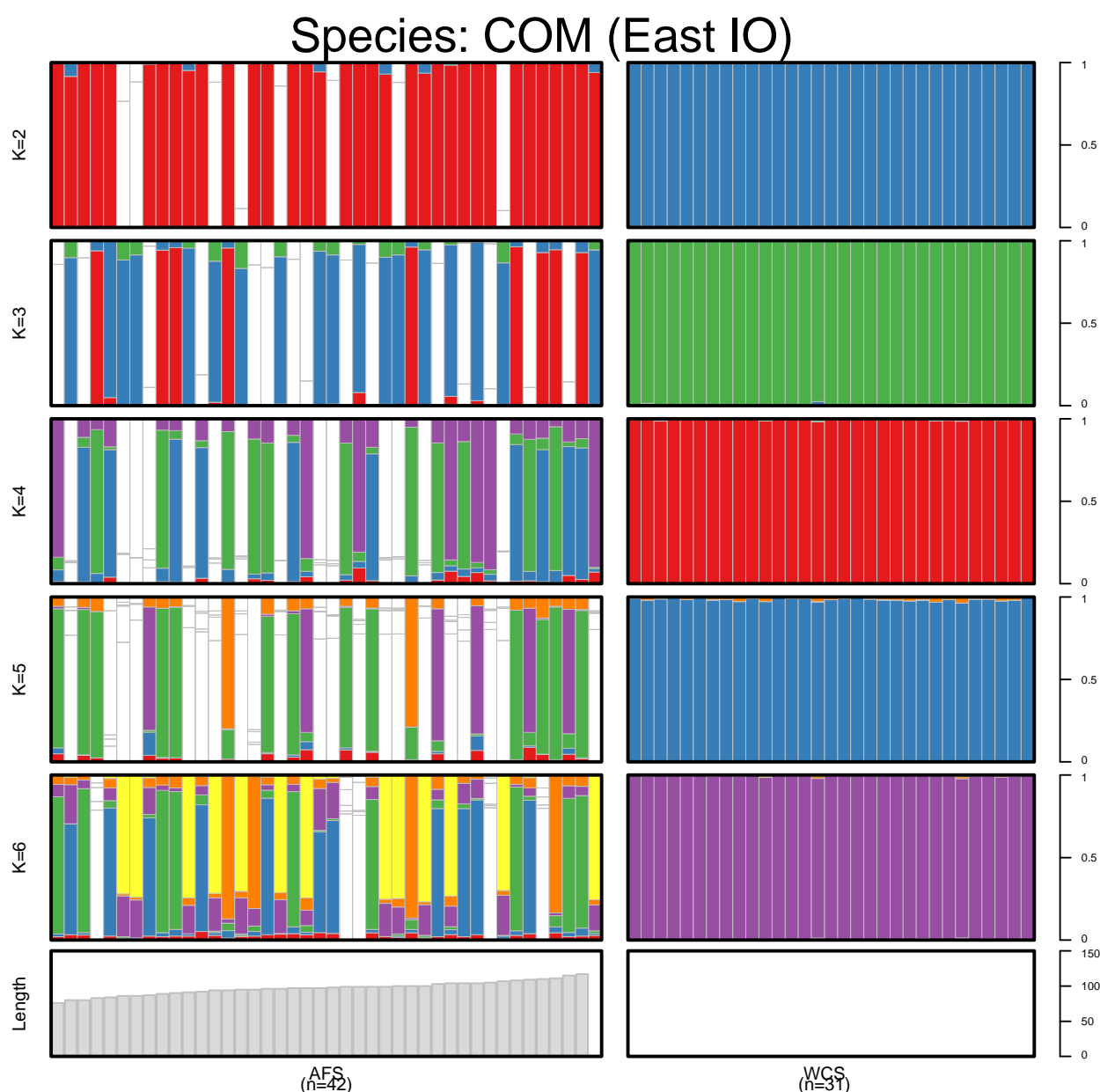




**Append Figure 16. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for the NWI and NCI subset assuming 2-6 genetic groups.**



**Append Figure 17.** Individual length frequencies and results of population structure analysis of DArTSeq using StockR for the NEI and ECI subset assuming 2-6 genetic groups.



Append Figure 18. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for the AFS and WCS subset assuming 2-6 genetic groups.

### 8.1.6 Narrow-barred Spanish mackerel (*Scomberomorus commerson*) - otolith microchemistry

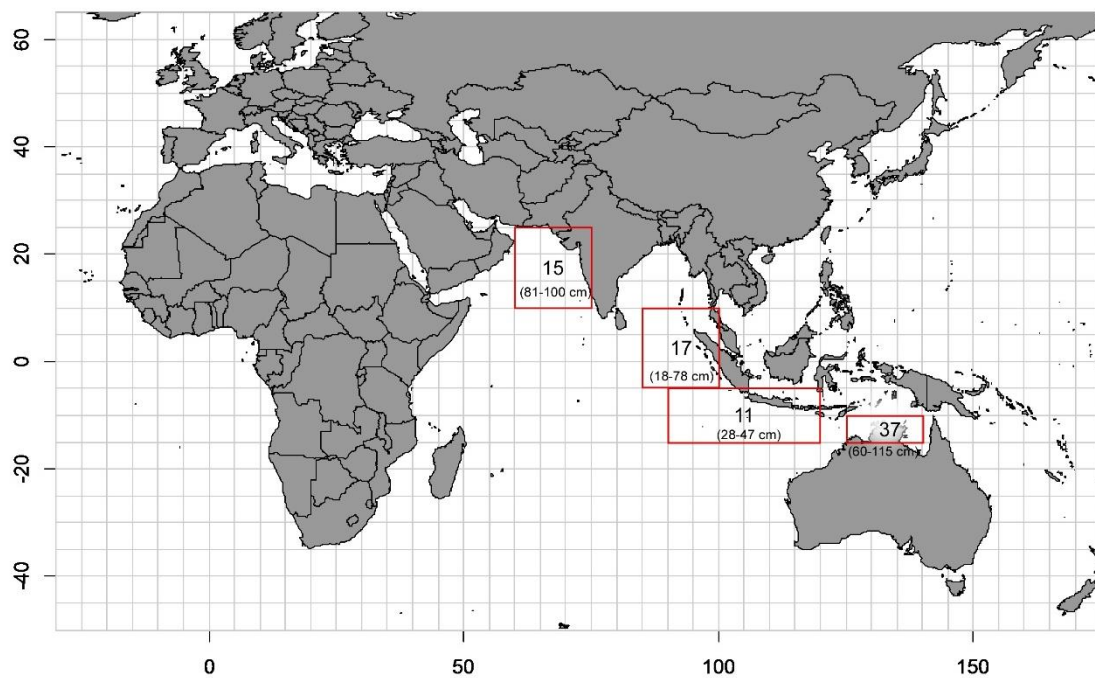
The investigation of otolith microchemistry of Spanish mackerel has thus far been limited to a study of otolith stable isotopes  $\delta^{18}\text{O}$  and  $\delta^{13}\text{C}$  in northern and western Australia (Buckworth et al. 2007). They found temporarily stable spatial structure in populations of Spanish mackerel and concluded that these could be considered separate stocks for the

purposes of fisheries management. Buckworth et al. (2007) did not investigate otolith elemental microchemistry.

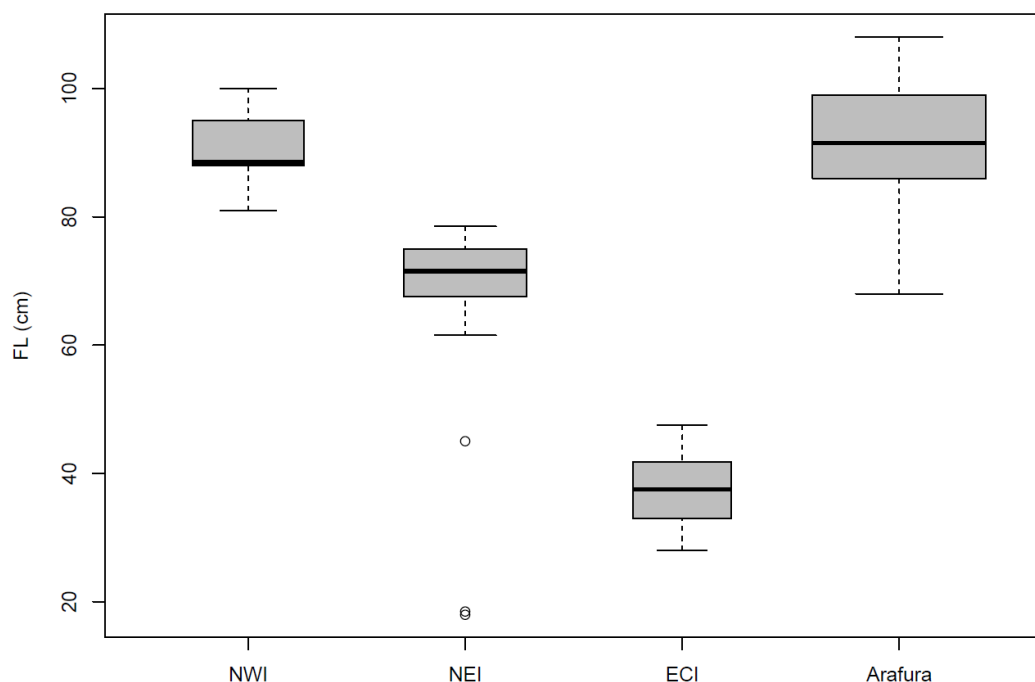
## Methods

The Spanish mackerel otolith samples analysed were from four sampling locations (Append Figure 19), referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI), Eastern-Central Indian Ocean (ECI) and Arafura Sea (Arafura) (Figure 2).

Although it was not intended in the original design, samples from different locations were collected during different periods, due to several reasons (see Section 2). The size of fish varied greatly between locations, with the fish from ECI being smallest (<50 cm FL) and those from NWI and Arafura being largest (mostly >80 cm FL) (Append Figure 19, Append Figure 20, Append Table 10). Because of the large variation in fish size, and thus age, they will have been spawned over many years; therefore, differences observed in core signatures among locations may be due, at least in part, to cohort effects.



Append Figure 19. Map showing the number of Spanish mackerel otoliths analysed for each of the sampling locations, referred to as North-West Indian Ocean (NWI), North-East Indian Ocean (NEI), Eastern Central Indian Ocean (ECI) and Arafura Sea (Arafura), and the size range of fish at each location.



**Append Figure 20.** Boxplots of Spanish mackerel fork length (FL, cm) by sampling location, including only fish whose otoliths were selected for analysis.

**Append Table 10.** Number, sampling period, size range and estimated ages of fish for each of the sampling locations.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
North-West Indian Ocean (NWI)	15	September 2018	81-100	2-3
North-East Indian Ocean (NEI)	17	April and November 2018	18-78	0+ -2
Eastern-Central Indian Ocean (ECI)	11	Nov 2018	28-47	0+
Arafura Sea (Arafura)	37	Dec 2017 and Mar-May 2019	80-115	2-6

\* The ranges in ages are for male and females combined (McPherson 1992, McIlwain et al. 2005).

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However, it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

Analyses that were performed on the core or edge data:

- Univariate tests for each element to test for differences among locations. The exact test used depended on whether the data were normally distributed (assessed using `shapiro.test` in R) and had equal variances between locations (assesses using `fligner.test` in R), and whether there were two or more locations.
- PERMANOVA (non-parametric version of MANOVA): used to test for differences in the multi-elemental signatures of fish among locations and/or cluster groupings (results obtained using the `adonis` function from the `vegan` package in R).
- PCA: used to help visualize the data as this can be difficult with so many elements, and to determine which elements account for most of the variability in the data (results obtained using the `dudi.pca` function from the `ade4` package in R).
- Clustering: in some circumstances, used for identifying the most likely number of separate spawning origins, and to investigate whether spawning origins differ among fish from different sampling locations (results obtained using the `hclust` function in R with `method="ward.D2"`, and the dissimilarity matrix calculated using Euclidean distance).

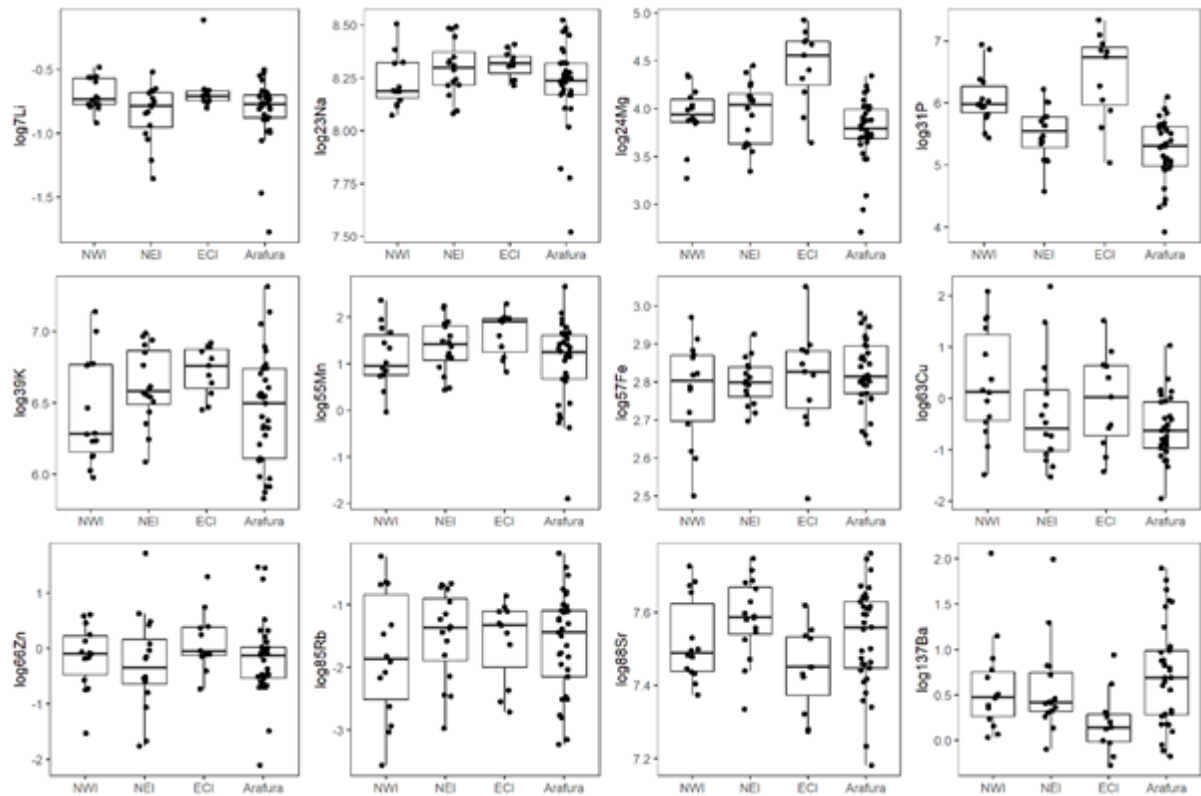
Note that prior to any multivariate analyses, the data were standardised (i.e., for each element, the data was centred by subtracting the mean and scaled by dividing by the standard deviation).

### Core results

Core signatures differ between locations for many elements (Append Figure 21). Based on univariate tests, 24Mg, 31P, 63Cu, 88Sr and 137Ba differ significantly between at least 4 locations ( $p < 0.05$ ).

Results from fitting a PERMANOVA model to the core data also suggest that the multi-elemental core signatures of Spanish mackerel are not equal among all locations ( $p = 0.001$ ). Subsequent pairwise tests between locations suggest that ECI differs significantly from the other three locations (NWI, NEI and Arafura), and that NWI differs from Arafura (Append Table 11).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA supports and helps visualise these findings (Append Figure 22).

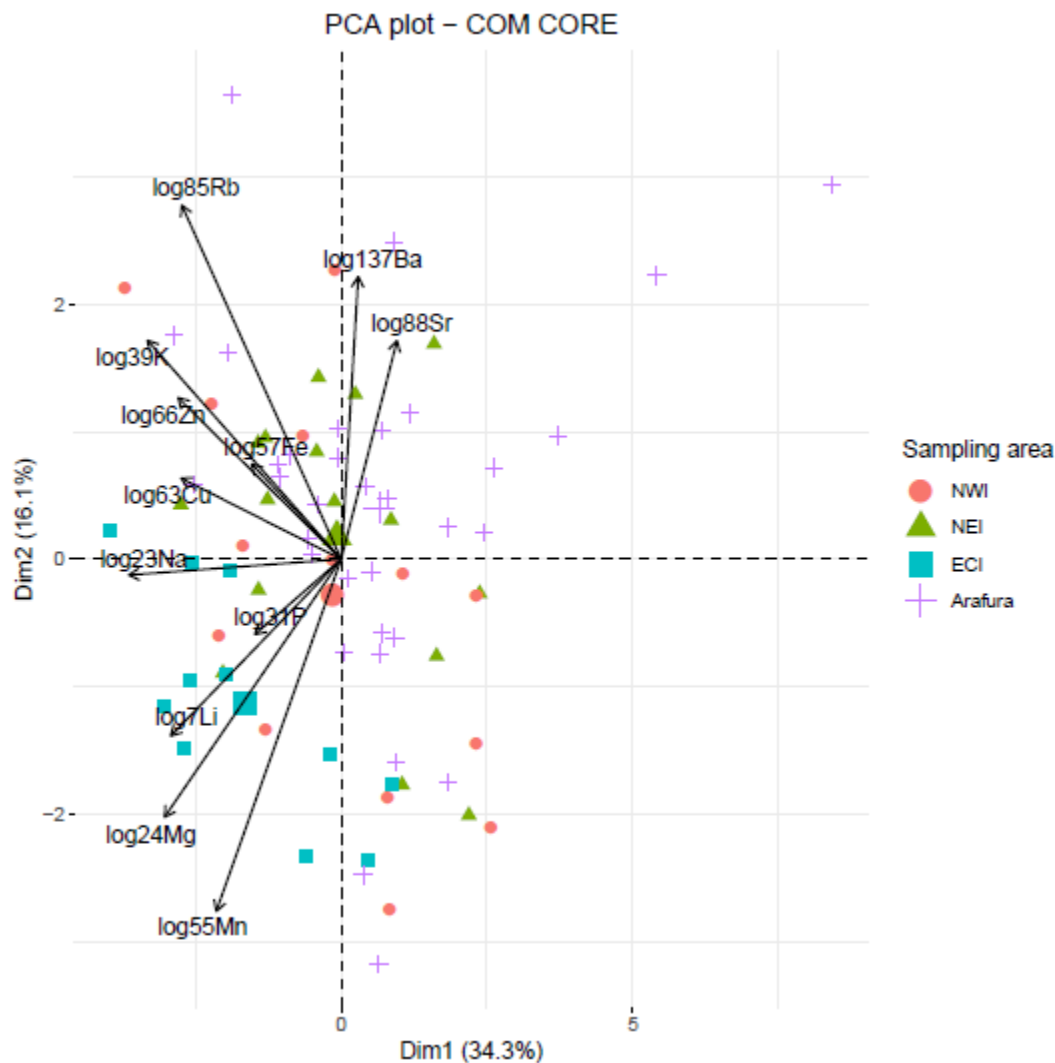


Append Figure 21. Boxplots comparing Spanish mackerel otolith core data between locations for each element. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 11. Results of pairwise comparisons of Spanish mackerel otolith core signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
NWI	vs	NEI	1	18.857	1.869	0.063	0.090	0.108	
NWI	vs	ECI	1	28.054	2.765	0.107	0.025	0.038	.
NWI	vs	Arafura	1	32.765	2.774	0.058	0.020	0.038	.
NEI	vs	ECI	1	41.201	4.451	0.151	0.001	0.003	*
NEI	vs	Arafura	1	16.333	1.449	0.030	0.189	0.189	
ECI	vs	Arafura	1	82.853	7.242	0.147	0.001	0.003	*

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 22. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the Spanish mackerel otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

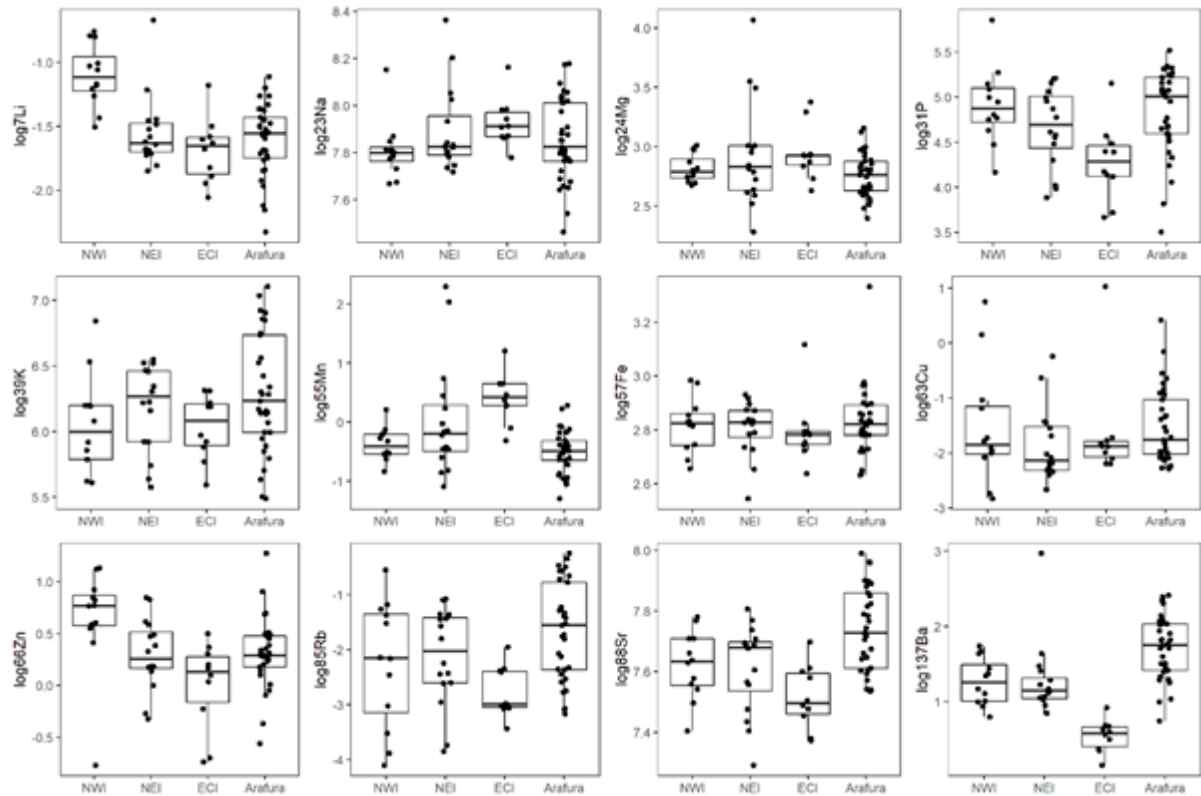
### Edge results

Edge signatures differed between locations for many elements (Append Figure 23). Based on univariate tests, 7Li, 24Mg, 31P, 55Mn, 66Zn, 85Rb, 88Sr and 137Ba all differ significantly among locations ( $p < 0.05$ ).

Results from the PERMANOVA model also suggest edge signatures are not equal among all locations ( $p = 0.001$ ). Based on subsequent pairwise tests between locations, the edge signatures differ significantly between all locations (Append Table 12).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the edge data helps to visualize these findings, with the difference between ECI and the other locations being most obvious (Append Figure 24).



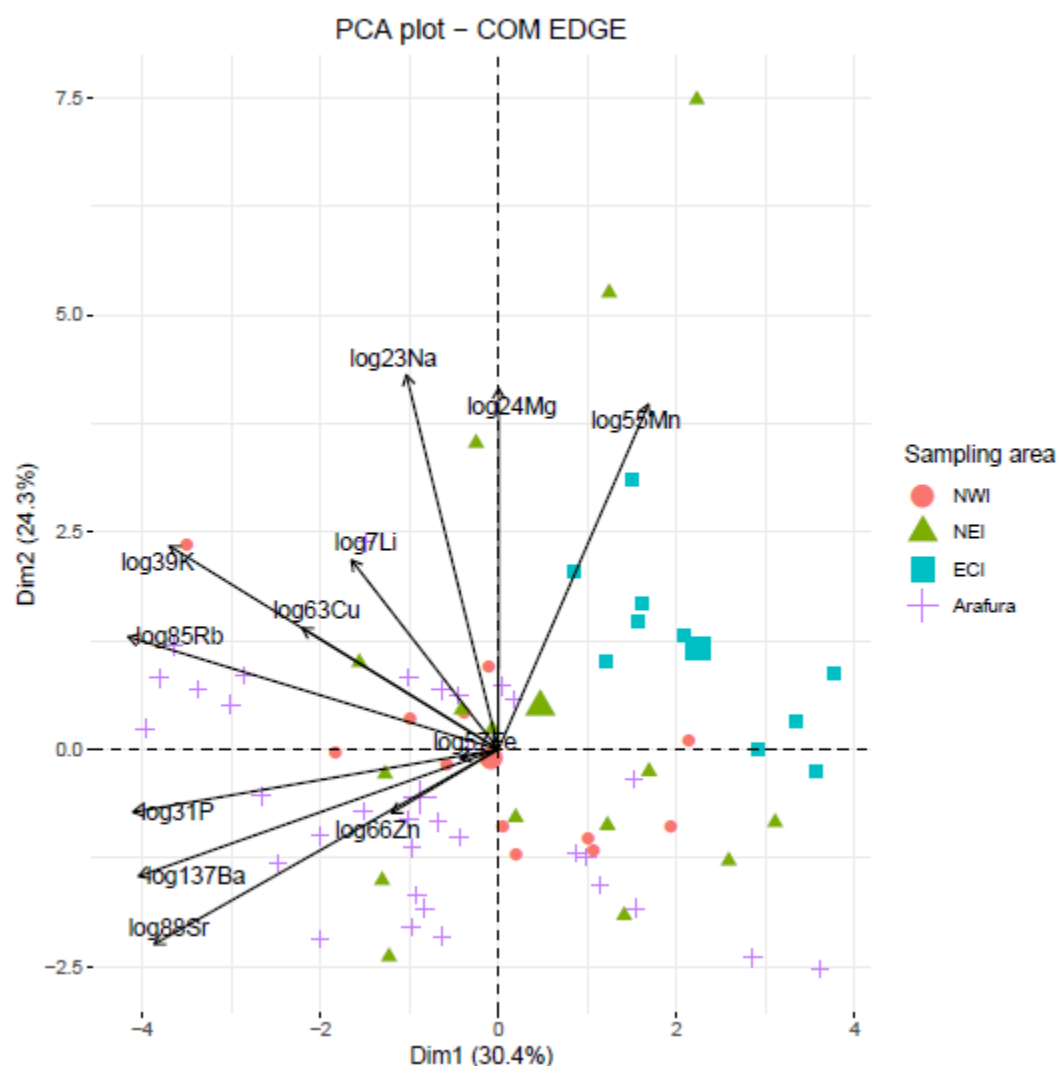


Append Figure 23. Boxplots comparing Spanish mackerel otolith edge data for each element between locations. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 12. Results of pairwise comparisons of multi-elemental edge signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
NWI	vs	NEI	1	27.480	2.362	0.083	0.027	0.032	.
NWI	vs	ECI	1	65.951	7.558	0.274	0.001	0.002	*
NWI	vs	Arafura	1	46.143	4.774	0.100	0.001	0.002	*
NEI	vs	ECI	1	27.805	2.482	0.094	0.039	0.039	.
NEI	vs	Arafura	1	34.926	3.219	0.064	0.006	0.009	*
ECI	vs	Arafura	1	100.314	10.767	0.208	0.001	0.002	*

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 24. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the Spanish mackerel otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Summary

The size of fish varied greatly between locations, with the fish from ECI being smallest (<50 cm FL) and those from NWI and Arafura being largest (mostly >80 cm FL) (Append Figure 19, Append Figure 20). Because the size/age range of adult fish analysed was wide, meaning fish were spawned in a wide range of years, differences observed in core signatures among locations may be due, at least in part, to cohort effects.

However, core data should be representative of spawning locations. Thus we can test whether the core signatures differ among locations, which would indicate that fish caught in these fairly distant locations were spawned in different locations (recalling the caveat that differences could occur even if all fish were spawned in the same location if the water chemistry in that location differed significantly between the years they were spawned).

The fact that the core signatures showed significant differences between ECI and the other locations may mean fish from ECI originated from a different spawning ground than fish from the other locations. However, because the fish from ECI are much smaller/younger than those from the other locations, and will have been spawned in more recent years, the differences could be due to temporal differences in ocean chemistry at the same spawning location.

The core signatures between fish from NWI and Arafura also differed significantly. These fish are more similar in size so will have been spawned over a similar range of years; thus there is more support for the hypotheses that fish from these two locations originated from different spawning grounds.

Edge data should be representative of capture locations. Thus, if we find differences in edge signatures between locations, this verifies that otolith chemistry can be useful for classifying fish to these locations. The fact that the edge signatures of Spanish mackerel differ between capture locations suggests that otolith chemistry can be useful for classifying fish to these locations.

## 8.2 Tropical tuna

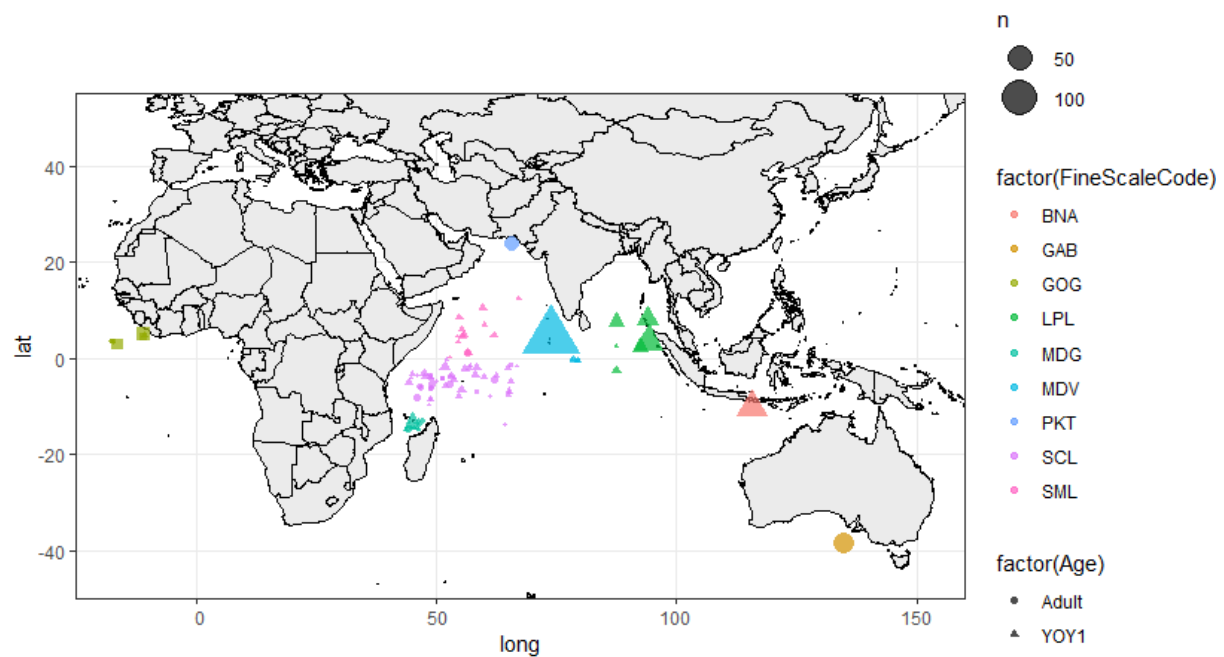
### 8.2.1 Skipjack tuna (*Katsuwonus pelamis*) - population genetics

Previous studies on the population structure of skipjack tuna have resulted in conflicting results regarding differentiation between and within oceans. For example, analyses based on the mitochondrial D-loop region (Menezes, Ikeda, Taniguchi 2006) or on nuclear microsatellite markers (Menezes et al. 2008) found respectively high levels or no genetic differentiation between skipjack tuna from the Indian and Pacific oceans. Additional analyses comparing Atlantic and Pacific Oceans skipjack tuna also revealed a high degree of genetic similarity among these two oceans (Graves, Ferris, Dizon 1984; Ely et al. 2005). Within the Indian Ocean, several studies have reported the existence of heterogeneous groups that were not related to any spatial pattern (Dammannagoda, Hurwood, Mather 2011; Menezes, Kumar, Kunal 2012).

#### Methods

##### *Sampling, DNA extraction and RAD-seq library preparation*

Skipjack tuna samples from the Indian and Atlantic Ocean were obtained from scientific surveys and commercial fisheries (Append Figure 25). The 525 samples were organized in 9 areas, named according to the closest land to where they were caught (Gulf of Guinea, Great Australian Bight, Benoa, Lampulo, Maldives, Pakistan, Madagascar, Somalia and Seychelles). Samples were classified as less than 6-month-old young of the year (<35 cm), more than 6-month-old young of the year (35-45 cm) and adults (>45 cm) according to the age-length keys available for this species. From each specimen, a ~1 cm<sup>3</sup> muscle tissue sample was excised and stored in 96% molecular grade ethanol at -20 °C prior to DNA extraction. Genomic DNA was extracted from about 20 mg of 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. Restriction-site-associated DNA libraries were prepared following the methods of Etter et al. (Etter et al. 2011). A range of 250-500 ng of genomic DNA from each sample was digested using the *Sbf*I restriction enzyme and then ligated to modified Illumina P1 adapters containing 5bp unique barcodes. Pools of DNA samples were sheared using the Covaris® M220 Focused-ultrasonicator™ Instrument (Life Technologies) and size-selected to 300-500 pb by cutting DNA fragments migrated through an agarose gel. Following the ligation of the Illumina P2 adapters, each genomic library was amplified using 12 PCR cycles. Finally, pools of 96 samples were paired-end sequenced (100 pb) on an Illumina HiSeq2000. \_ENREF\_5



**Append Figure 25.** Samples collected for this study. Each location is represented by one color (BNA – Benoa, GAB – Great Australian Bight, GOG – Gulf of Guinea, LPL – Lampulo, MDG – Madagascar, MDV – Maldivas, PKT – Pakistan, SCL – Seychelles, SML – Somalia) and shapes indicate if samples are less than 6 month old young of the year (YOY1), more than 6 month old young of the year (YOY2) or adults. Size of shapes are proportional to the number of samples collected per point.

### *RAD-tag assembly and SNP calling*

Generated RAD-tags were analyzed using *Stacks* version 2.4 (Catchen et al. 2013). Using *process\_radtags* reads with any uncalled base, with total low-quality scores or with quality score below 20 within 10bp size sliding windows were removed. The module *clone\_filter* was applied to reads whose forward and reverse pairs passed quality filtering to remove PCR duplicates. The module *ustacks* was then used to assemble orthologous tags (stacks) per individual, with a minimum coverage depth required to create a stack (parameter -m) of 3, and a maximum nucleotide mismatches allowed between stacks (parameter -M) of 2. RAD loci were then assembled using the module *cstacks* with a maximum of 3 mismatches allowed between sample tags when generating the catalog (parameter -n). Matches to the catalog for each sample were searched using *sstacks* and transposed using *tsv2bam* and the module *gstacks* was used to assembly paired-end reads into contigs, merging them to the single-end locus and identifying and genotyping SNPs. Only samples with more than 40,000 and less than 70,000 reads were selected for further analyses. The module *populations* was used to export from the catalog, the SNPs presented in RAD loci found in at least 90% of the individuals and in the first 100 bases of each contig. Using *PLINK* version 1.07 (Purcell et al. 2007), SNPs with more than 5% missing data and a minimum allele frequency (MAF) smaller than 0.05 as well as samples with more than 25% missing data were excluded from downstream analyses. The resulting genotype tables were exported into *Structure* and *Genepop* formats. Related individuals were identified using GCTA (Yang et al. 2011).

### *Genetic diversity and population structure analyses*

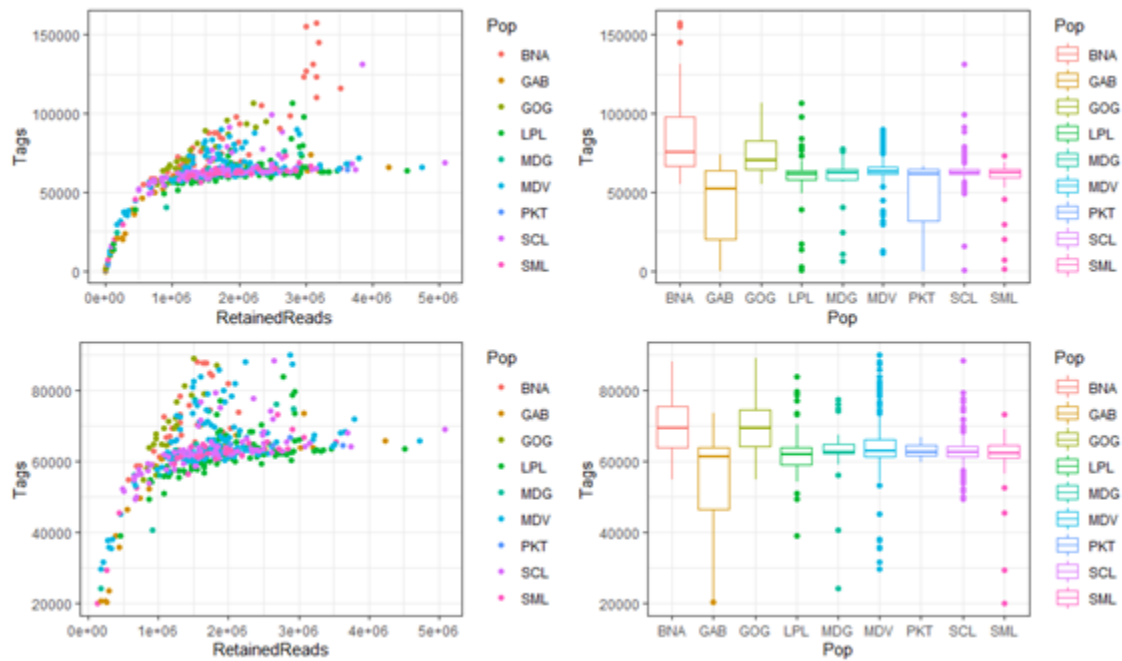
Principal Component Analysis (PCA) were performed using the *adeigenet* R package (Jombart, Ahmed 2011), and ADMIXTURE (Alexander, Novembre, Lange 2009) was run assuming from 2 to 4 ancestral populations (K) setting default parameters. The value of K with lowest associated error value was identified using ADMIXTURE's cross-validation procedure testing K values from 1 to 6. Groups were sought in the genetic data as implemented in the R package stockR (Foster et al. 2018). 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 (25 times in this initial analysis) the genetic data and see how well the groupings match those from the analysis of the full data. The groupings are displayed using probabilities of individual fish to each genetic group are obtained using bootstrap methods (Foster et al. 2018), using 100 resamples in this initial analysis.

## **Results**

### *Data quality*

Considering all 524 individuals, an average of 73% of the reads were retained for de novo assembly and average coverage was 26X. Number of tags per individual was variable, and some samples resulted in a too low (due to a low number of reads) or too high (potentially due to cross-contamination) number of tags (Append Figure 26). After removing individuals with too many or too low number of tags, the datasets consisted 393 individuals.

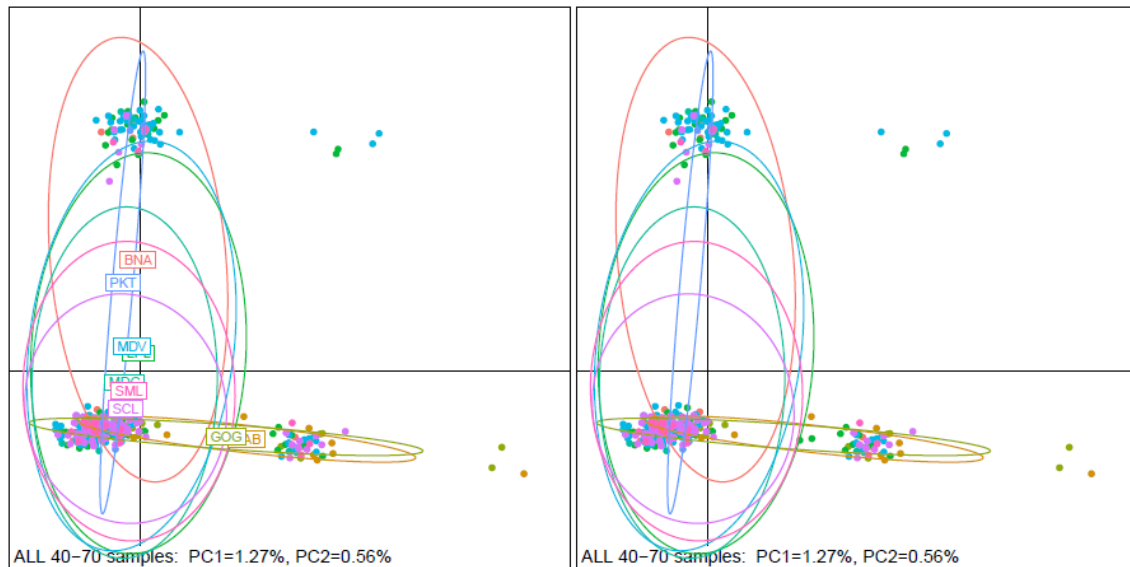
After SNP filtering and related individual removal, the final datasets resulted in 368 individuals and 14346 SNPs.



Append Figure 26. Relation between number of tags and number of reads per individual (left) and boxplot of the number of tags per location for the complete dataset (above) and de dataset after removing individuals with too high or too low number of tags.

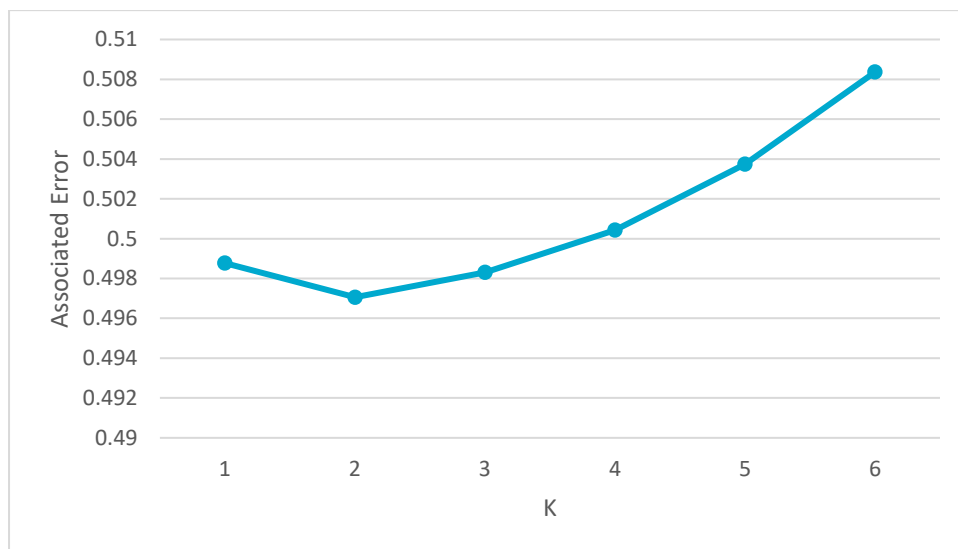
### *Genetic diversity and population structure*

Principal component analyses (Append Figure 27) based on all samples from the strict filtering dataset excluding those related pairs show five groups of samples: a “central” group that contains individuals form all locations, a group that contains individuals form all locations except from Pakistan, group that contains individuals from all locations except from Gulf of Guinea and Great Australian Bight and two small groups that contain respectively individuals from Maldives and Lampulo and from Gulf of Guinea and Great Australian Bight.



**Append Figure 27.** Principal component analysis using the filtered datasets. Left plots are identical to right plots but removing labels for clarity.

ADMIXTURE analyses show that the number of assumed ancestral populations with the lowest associated error ( $K$ ) is 2 (Append Figure 28). Individual ancestral proportions (Append Figure 29) showed two clusters, one which includes samples from all locations and a second cluster which contains individuals from all locations except from Gulf of Guinea and Great Australian Bight.



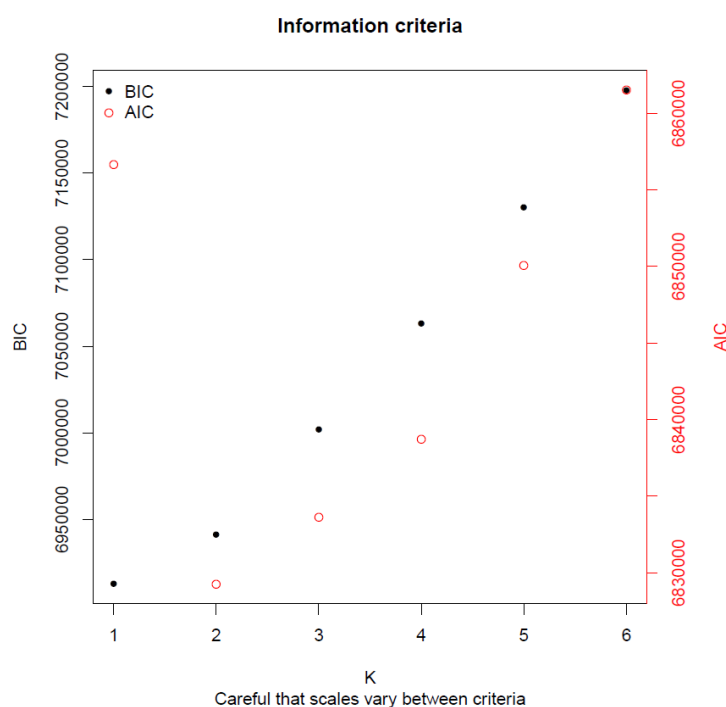
**Append Figure 28.** Associated error for each value of assumed ancestral population ( $K$ ) estimated by ADMIXTURE.



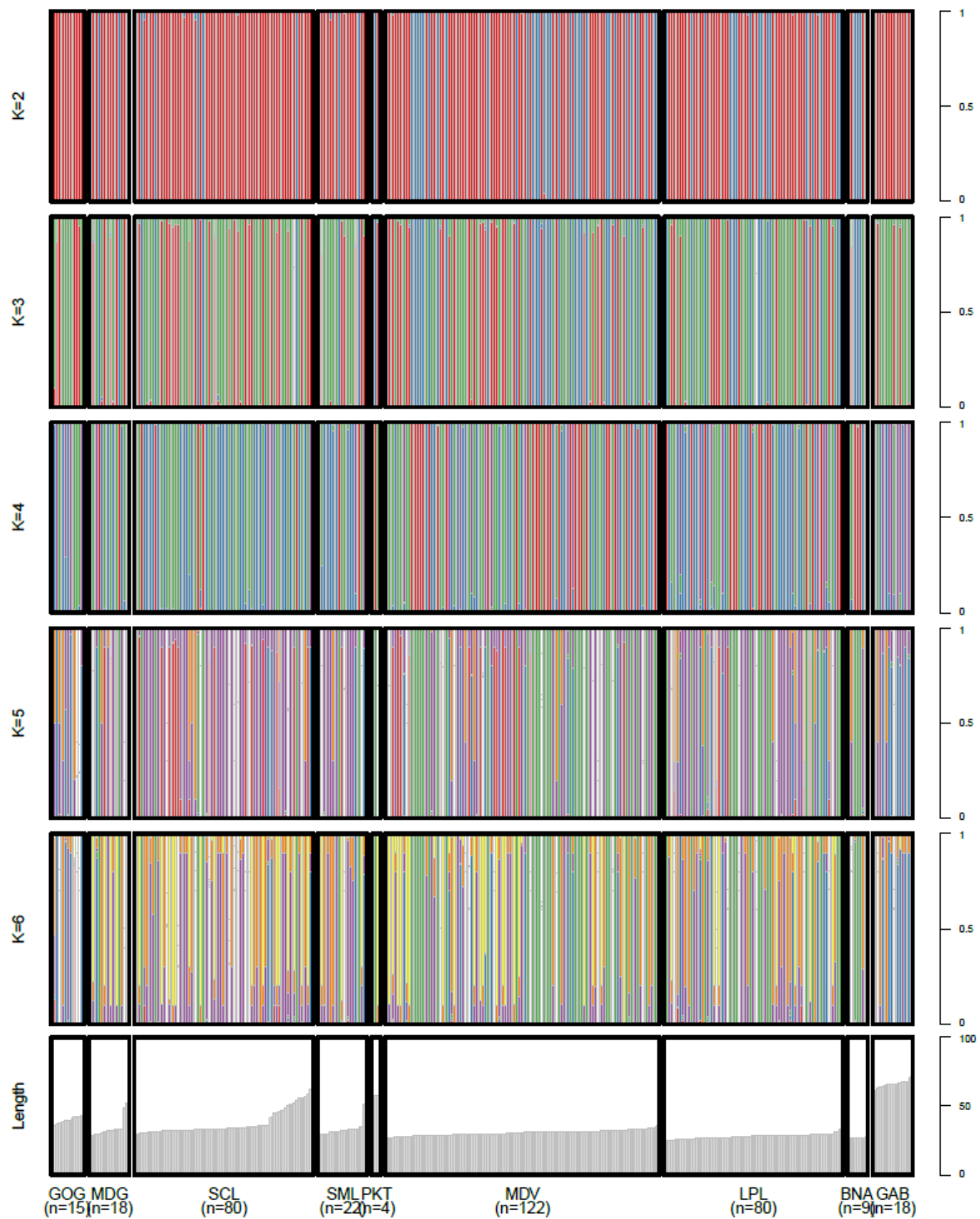


**Append Figure 29.** Individual ancestry proportions estimated by ADMIXTURE when assuming from two to four ancestral populations (K) of samples from different locations.

The StockR analyses support the presence of one or two groups (Append Figure 30). Similarly, to the previous analyses, a group that includes individuals from all locations except from the two most southern ones (Gulf of Guinea and Great Australian Bight), and a group that includes individuals from all locations except for Pakistan are obtained (Append Figure 31).



**Append Figure 30.** Information Criterion plot summarizing the results of model fits (using AIC and BIC) for the most likely number of genetic groups from the distribution of SNP data in the sample.



Append Figure 31. Assignment of fish to groups (assuming 1 to 6 possible groups) according to StockR.

## Summary

The three analyses performed support existence of genetic groups, although these do not coincide with the locations where samples were captured. All analyses support some separation between the north (Pakistan) and the south (Gulf of Guinea and Great Australian

Bight) by suggesting two groups, one that contains individuals from all locations except from the former and another that contain individuals from all locations including the latter two. Our findings are congruent with previous findings and confirm that the population structure of skipjack is complex and that it's understanding will require testing hypothesis such as the presence of stocks spawning at different locations, the effect of currents in larvae drifting and potential interbreeding of isolated populations.

### **8.2.2 Skipjack tuna (*Katsuwonus pelamis*) - otolith microchemistry**

Two studies of skipjack tuna (SKJ) otolith microchemistry have been undertaken in the Pacific Ocean. Ianelli (1993) detected population structure in the central and eastern Pacific Ocean using otolith microconstituent analysis. Arai et al (2005) studied patterns in Sr:Ca ratios to investigate residency, movements and mixing in skipjack from tropical western Pacific and off the coast of Japan. In the Indian Ocean, in a study related to this current project, Artetxe-Arrate et al (unpublished) have examined the core of young-of-the-year skipjack from four known spawning areas to investigate ontogenetic, interannual and spatial variability in otolith elemental chemistry.

For this study, samples were collected in three known skipjack spawning grounds in two different years, so the aim of the sampling design was to obtain a spawning ground signature from young-of-the-year fish and to determine if the signature was temporally stable. For that, our assumptions were that fish had not moved significantly from their natal area so that the analysed core and edge otolith material had been deposited while the fish were in their nursery ground.

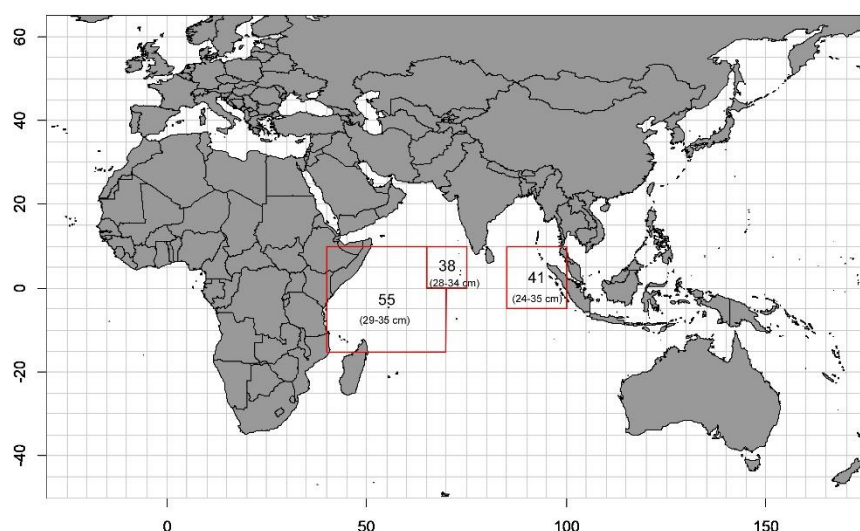
#### **Methods**

Otoliths from 134 skipjack from 3 locations were analysed: 55 from the western central Indian Ocean (WCI), 38 from Maldives (CIM) and 41 from the north east Indian Ocean (NEI) (Append Figure 32). The skipjack were all young-of-the-year (YOY, <35 cm FL) and were obtained in two consequent years, 2018 and 2019 (Append Figure 32, Append Table 13).

Although it was not intended in the original design, samples from different locations and years were collected during different periods, due to several reasons (see Section 2):

- WCI: March 2018 and April 2019
- CIM: August 2018 and February 2019
- NEI: April 2018 and November 2018

In this section, we will refer to the second round of NEI samples (those collected in November 2018) as the 2019 samples from this location.



**Append Figure 32.** Map showing the number of skipjack otoliths analysed for each of three sampling locations, referred to as Western Central Indian Ocean (WCI), Maldives (CIM) and North-East Indian Ocean (NEI), and the size range of fish at each location.

**Append Table 13.** Number, sampling period and estimated ages of fish for each of the three sampling locations WCI, CIM and ECI.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
Western Central Indian Ocean (WCI)	55	March 2018 and April 2019	29-35	0+
Maldives (CIM)	38	August 2018 and February 2019	28-34	0+
North East Indian ocean (NEI)	41	April 2018 and November 2018	24-35	0+

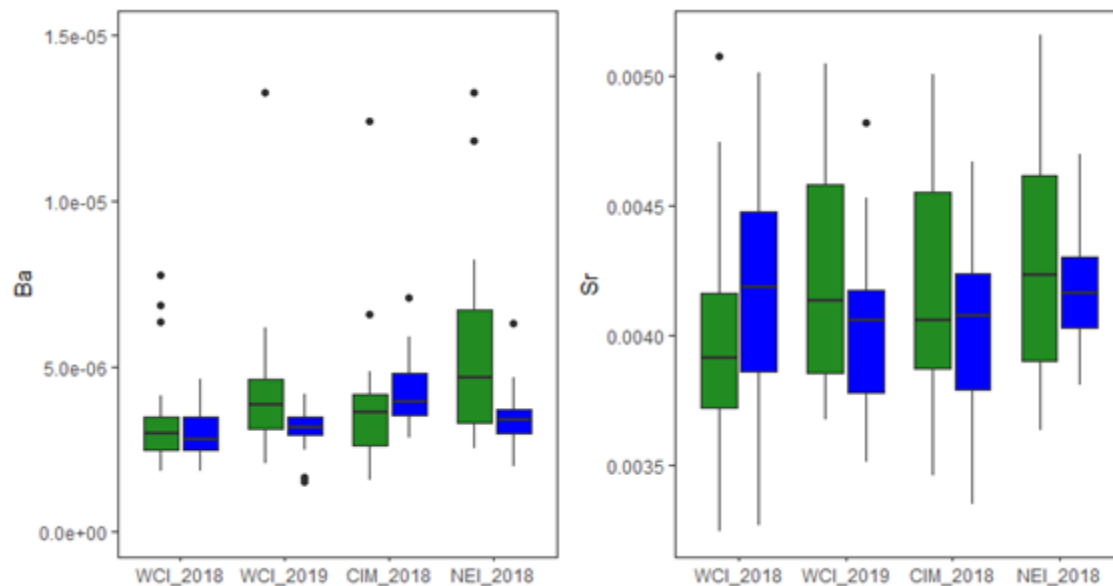
\* The ranges in ages are for male and females combined (Eveson et al., 2015).

The otoliths were analysed at the Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux Université de Pau et des Pays de l'Adour/CNRS using high resolution LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Eight trace elements were measured (see Section 4.5.2), 4 were below the limit of detection and therefore discarded for analyses.

The spot 200  $\mu\text{m}$  from the core ( $\pm 13$ -15 days of life) was selected as natal origin signature. The spot at the edge of the otolith was also analysed, since these data reflect the fish's known capture location, and can be used for validation purposes.

## Results

Core data included Ba, Mg, Mn and Sr, but edge data was only analysed for Ba and Sr. This is due to inconsistencies found in the distribution of Mg and Mn along the edge for other tropical tuna species (for more information, see Fig. Y2 in the yellowfin tuna otolith microchemistry section) (Append Figure 33).

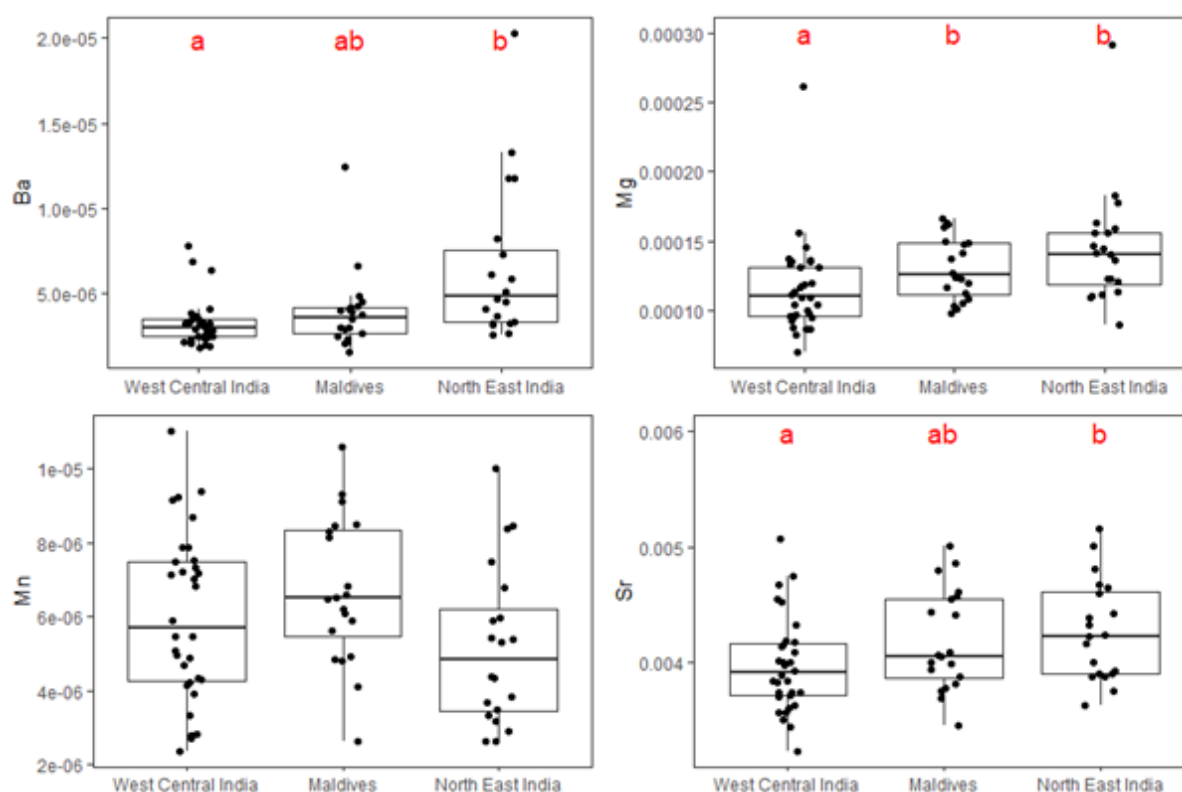


Append Figure 33. Boxplots comparing skipjack tuna core (green) and edge (blue) signatures at the three sampling locations. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian. For WCI edge was analysed both in 2018 and 2019.

## CORE results

Core and edge signatures were significantly different for both elements (Append Figure 34). This is clear in the otolith data even though fish are estimated to be only about 3 months old and assumed not to have changed locations between spawning and capture, their core and edge signatures are still significantly different.

Boxplots comparing core signatures between locations for skipjack tuna sampled in 2018 show that Ba, Mg and Sr varied between WCI and NEI, with CIM showing intermediate values, except for Mg, for which more closely resembles NEI (Append Figure 34).



Append Figure 34. Boxplots comparing skipjack tuna otolith core data between locations for each element for samples collected in 2018. Significant differences ( $p < 0.05$ ) among locations are shown with red letters.

Based on PERMANOVA, multielemental core signatures were not equal among all locations for fish sampled in 2018 ( $p = 0.002$ ). Based on subsequent pairwise tests between locations: Core signatures from NEI differ with core signatures of WCI and CIM (Append Table 14).

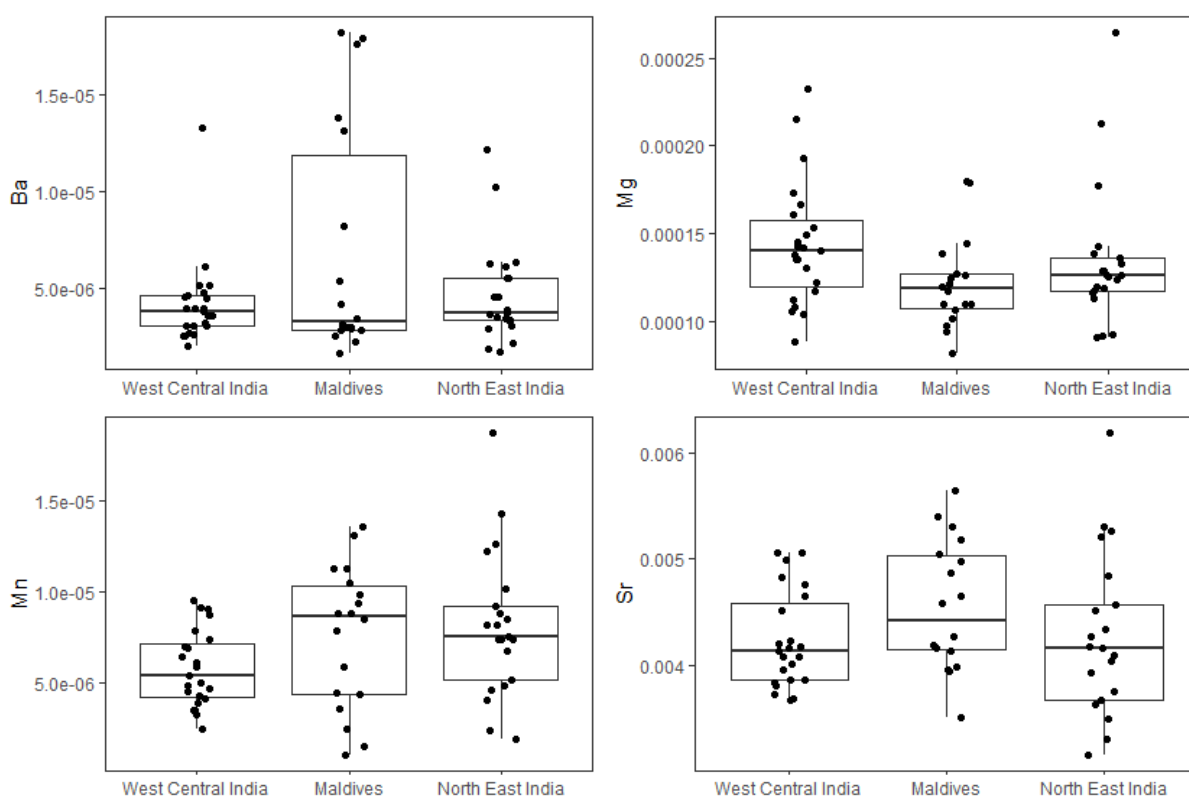
Based on PERMANOVA, multielemental core signatures were not equal among all locations for fish sampled 2019 ( $p = 0.013$ ). Based on subsequent pairwise tests between locations: WCI core signatures differ from those of CIM (Append Table 15).

Boxplots comparing core signatures between locations for skipjack tuna sampled in 2019 did not show any differences between locations (Append Figure 35).

Append Table 14. Results of pairwise comparisons of skipjack tuna otolith core signatures between locations of samples from 2018. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI	vs	CIM	1	6.792	2.324	0.044	0.052	0.052	
WCI	vs	NEI	1	29.372	7.437	0.129	0.001	0.003	**
CIM	vs	NEI	1	13.901	3.372	0.081	0.010	0.015	*

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



**Append Figure 35.** Boxplots comparing skipjack tuna otolith core data between locations for each element for samples collected in 2019. Significant differences ( $p < 0.05$ ) among locations are shown with red letters.

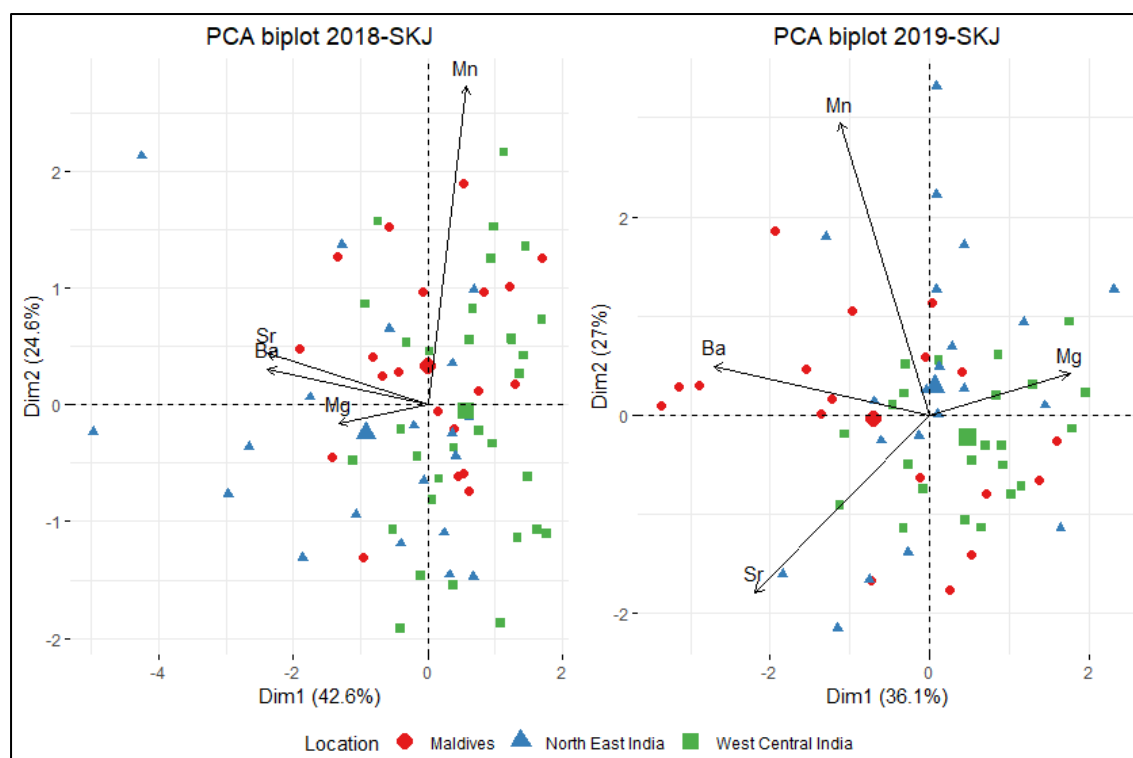
**Append Table 15.** Results of pairwise comparisons of skipjack tuna otolith core signatures between locations of samples from 2019. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI	vs	CIM	1	14.601	4.210	0.097	0.004	0.012	*
WCI	vs	NEI	1	5.846	1.770	0.040	0.151	0.188	
CIM	vs	NEI	1	7.044	1.468	0.382	0.188	0.188	

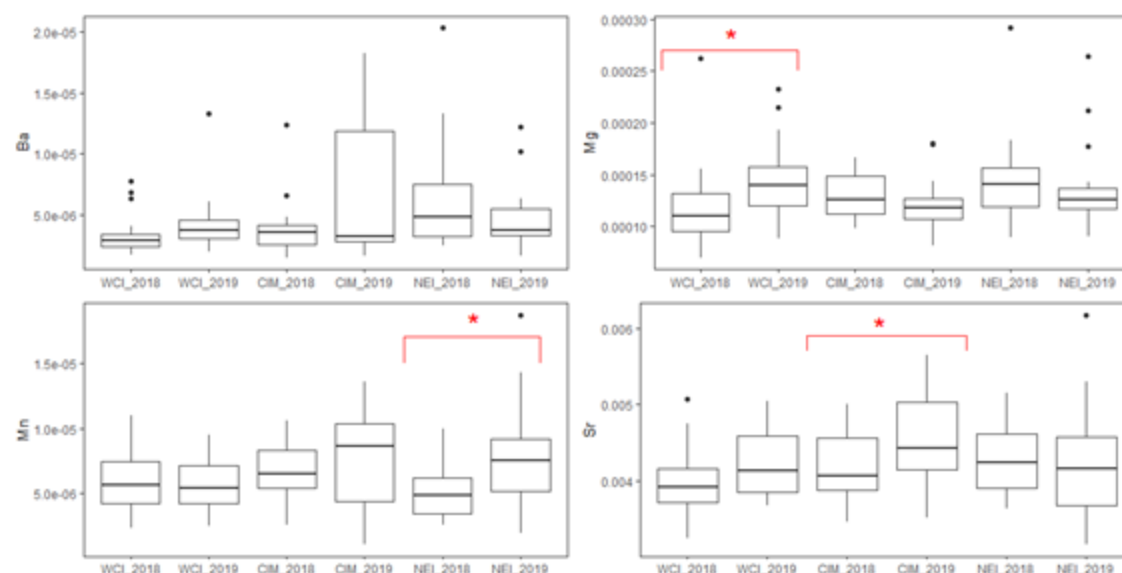
Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data confirms and helps to visualize these findings (Append Figure 36).

Boxplots comparing core signatures within locations between sampling years show that some elemental signatures were not consistent (Append Figure 37); Mg varied among years in the WCI, Mn in the NEI and Sr in the CIM. Ba did not vary among years within any location.



Append Figure 36. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the skipjack otolith core signatures for the 2018 and 2019 samples respectively. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.



Append Figure 37. Boxplots comparing skipjack tuna otolith core data between sampling years for each pair of locations. *Significant* differences ( $p<0.05$ ) among years withing locations are shown with red \*. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian.

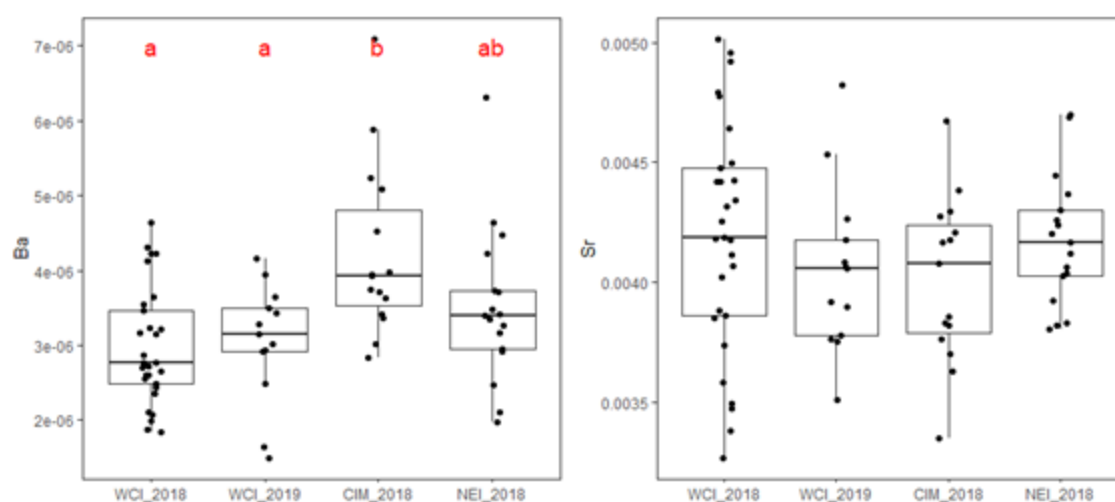


## EDGE results

Otolith edge signatures varied between locations for skipjack tuna, Ba varied among WCI and CIM (for both years of WCI signatures), with NEI showing intermediate values. Sr edge signatures did not differ among locations (Append Figure 38).

Based on the PERMANOVA, multielemental edge signatures are not equal among all locations ( $p=0.008$ ).

Based on subsequent pairwise tests between locations, edge signatures differed between CIM with WCI, in both (Append Table 16). A biplot showing edge Ba and Sr signatures of individuals confirms and helps to visualize these findings (Append Figure 39).

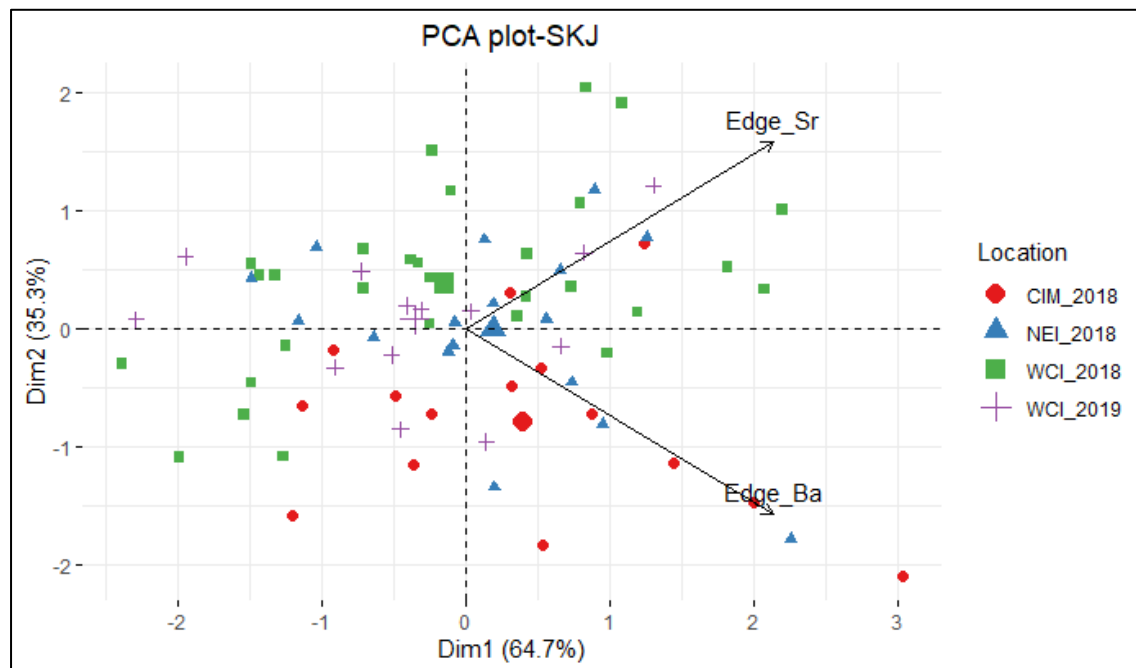


Append Figure 38. Boxplots comparing skipjack tuna otolith edge data between locations for each element for samples collected. Significant differences ( $p < 0.05$ ) among locations are shown with red letters. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian. For WCI edge was analysed both in 2018 and 2019

Append Table 16. Results of pairwise comparisons of skipjack tuna otolith edge signatures between locations. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian. For WCI edge was analysed both in 2018 and 2019.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI_2018	vs	CIM_2018	1	16.535	8.005	0.160	0.001	0.006	**
WCI_2018	vs	NEI_2018	1	2.811	1.509	0.033	0.212	0.272	
CIM_2018	vs	NEI_2018	1	5.278	3.042	0.092	0.080	0.160	
WCI_2018	vs	WCI_2019	1	1.226	0.654	0.016	0.464	0.464	
CIM_2018	vs	WCI_2019	1	9.201	5.310	0.170	0.013	0.04	*
NEI_2018	vs	WCI_2019	1	2.239	1.555	0.052	0.227	0.272	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



**Append Figure 39.** Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the skipjack tuna otolith edge signatures. Individuals are coded by their sampling location and year. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Summary

The skipjack otoliths were collected from three locations that are known spawning sites -- WCI, CIM and NEI -- and, by analysing spots near the otolith core, we aimed to obtain an otolith signature that was representative of the nursery locations. By analysing spots on the otolith edge, we aimed to obtain an otolith signature that was representative of capture locations, at time of capture.

As the skipjack were young-of-the-year at capture, estimated to be aged 3-5 months old, we assumed they had not moved significantly from their location of origin and therefore their capture location was the same as the location in which they were spawned. However, there were significant differences between core and edge otolith signatures. If our assumption that fish have not moved far from their spawning grounds is correct, then the differences could be due to ontogenetic changes during the first months of life, which influenced the chemical composition of the otolith. However, if oceanography varied during the first months of life, that is, the oceanography was different at the time of spawning and at time of capture, the seasonal effect could also have influenced the chemical composition of the otolith.

For samples collected in 2018, core signatures from NEI differed from those from WCI and CIM. The NEI and WCI otoliths were from fish of similar lengths that had been caught at the same time, providing evidence that fish from these two locations did, in fact, have distinct spawning ground otolith signatures. The NEI and CIM samples were from the same size fish but were collected at different times of the same year, indicating the CIM fish were spawned in a different season to the NEI and WCI fish. Therefore, the CIM difference in otolith core signature could be due, at least in part, to seasonal variability in ocean chemistry. Similarly, the WCI and CIM samples were from the same size fish but were collected at different times of the same year, however, their otolith cores were not different. The lack of detectable differences in otolith core signatures could be due to similar oceanography in both locations at the 2 spawning times and hence seasonal variation in oceanography could be obscuring regional differences.

For samples collected in 2019, the only significant differences in core signatures were between WCI and CIM. The fish from these 2 locations had similar lengths and collection dates, providing evidence that the otoliths have distinct spawning ground signatures. The NEI samples were from the same size fish but were collected at different times of the same year, indicating the NEI fish were spawned in a different season to the WCI and CIM fish. The lack of detectable differences in otolith core signatures could indicate that the otoliths from those locations are indistinguishable, or that seasonal variation in oceanography could be obscuring regional differences.

Edge signatures differed between WCI and CIM, in both years, but not with NEI. As WCI and CIM samples were collected at different periods, observed differences could be due to either regional or seasonal differences in ocean chemistry. The lack of detectable difference between WCI and NEI edge signatures, which were collected at the same time, suggests that ocean chemistry did not differ significantly between these locations at the time of capture.

We were able to investigate the seasonal variability hypothesis by looking at the edge and core signatures of the fish from WCI and NEI, which were captured at the same time in 2018. No differences were found between edge signatures of WCI and NEI samples, but the core signatures were significantly different. These fish were captured in March, when the regional oceanography is more similar among these locations than at other times of the year (see Figure 4, Figure 5 ). We estimated these fish were hatched around October of 2017, when the regional oceanographic differences are more marked between these locations (see Figure 4, Figure 5).

We also found interannual differences in the core signatures within locations. In the case of CIM and NEI, for which Sr and Mn differ, respectively, these differences could be due to seasonal effects, as fish were sampled at different periods in each year. In the case of WCI, where samples were collected in the same period in both years, Mg differed between years. Hence, for Mg, we suggest caution be used if different cohorts are pooled for spawning ground signature identification.

The fact that we observe differences in core signatures when comparing fish from different locations sampled at the same time, but not when comparing fish from locations sampled at

different times, suggests that seasonal variability could be obscuring regional differences in the core signatures. This issue highlights the importance of sampling individuals from different locations at the same time. In addition, collecting samples in different seasons would provide an opportunity to determine the level of influence that regional oceanography and seasonality have on otolith microchemistry.

Overall, the results for skipjack tuna microchemistry suggest that seasonality can strongly influence otolith signatures and should be considered in any further studies, ideally sampling fish from different locations at the same times. In this study, the results indicate that different nursery areas can be discriminated within the Indian Ocean using otolith microchemistry. It could be an effective tool to study the connectivity and mixing rates of older individuals within the Indian Ocean. However, due to seasonal and interannual variability, it is crucial that juveniles from different locations are collected at the same time of the year to assess the extent of seasonal and interannual variability, in order to establish a baseline to investigate the natal origin of adults.

### 8.2.3 Yellowfin tuna (*Thunnus albacares*) – population genetics

Yellowfin tuna are currently managed as a single stock within the Indian Ocean yet there is growing evidence to support the presence of population subdivision within the Indian Ocean that could potentially influence change in current assessment strategies. The presence of three yellowfin tuna stocks (western and two eastern) within the Indian Ocean was first muted by Kurogane and Hiyama (1958), while Morita and Koto (1970) suggested a more conservative structure of single eastern and single western stocks based on longline fishery data. Subsequently, a number of population studies, which have examined a variety of genetic markers, have reported varying degrees of population differentiation within the Indian Ocean from single to multiple stocks. A single Indian Ocean, Atlantic Ocean, and two Pacific Ocean (eastern and western) stocks were reported in a global analysis of allozymes from two different studies (Sharp et al 1978; Ward et al., 1997). Results from studies using mtDNA and DNA microsatellites failed to demonstrate population differentiation and supported conclusions of these earlier studies that were consistent with presence of a single stock of yellowfin tuna within the Indian Ocean (Nishida et al., 2001; Ely et al., 2005; Wu et al., 2010). In contrast, Demmannagoda (et al., 2008) and Kunal (et al., 2013) reported mtDNA variation among samples collected near India, Sri Lanka and the Maldives that supported the presence of up to three genetically discrete Indian Ocean stocks between the north western (Arabian Sea), north central Indian Ocean, and north east Indian Ocean. Similarly, analysis of SNP variation and parasite data has indicated a lack of connectivity or the potential for a localised bottleneck to geneflow among areas sampled from central Indian Ocean, eastern Indian Ocean, and Indonesian Archipelagic waters (Moore et al., 2019). Results from studies of both genetic and parasite data have provided evidence of population differentiation consistent with a minimum of two Indian Ocean stocks of yellowfin tuna however, none have demonstrated temporal stability nor have they managed to deliver broad spatial coverage of the Indian Ocean within a single study (Barth et al., 2017; Moore et al., 2019; Proctor et al., 2019). The current study uses the increased genetic resolution of high throughput SNP genotyping technology along with a more targeted sampling design to further investigate and resolve temporal and spatial population differentiation within the Indian Ocean.

## Results

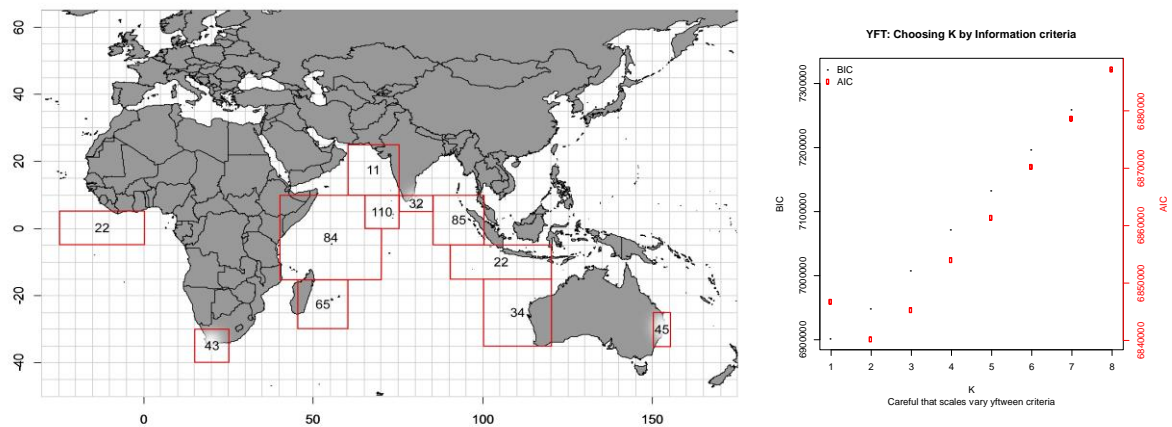
A total of 1206 yellowfin tuna were collected from 9 Indian Ocean sites and two outgroup sites (east Atlantic Ocean and south-west Pacific Ocean) which represents about one third of the species pantropical distribution (Append Figure 40). The samples consisted of a mix of young of year (YoY) fish and mature adults. DNA from 664 individuals (out of the 1206 total YFT collected), which matched sampling design (section 2.1) requirements and passed our internal quality control checks (DNA quality and species identity), were sequenced at Diversity Arrays Technology (DART) using their proprietary DARTSeq™ technology (see Section 4).

Analysis of the raw yellowfin tuna DARTSeq™ FASTQ files using the DARTSoft14™ analysis pipeline delivered a genotype data set of 54,733 SNP markers from 39,663 loci. Further

filtering of the genotyped loci using the program “radiator” (Gosselin 2019, <https://github.com/thierrygosselin/radiator>) focussed on two main aspects: i) removal of poor-quality loci; and ii) removal of compromised individuals (see section 4.4.2 for details), and produced a final filtered data set of 546 individuals genotyped at 15,562 SNP makers (one SNP per locus fragment) from 11 sampling locations (see Append Table 17).

Heterogeneity in the radiator filtered data, which was assessed among the sample locations using the program stockR (Foster et al., 2019), indicated a K of 2 genetic groups had the highest parsimony based on its AIC value (Append Figure 40; see also Section 4.4 for further details). Furthermore, the distribution of individuals belonging to each of K genetic groups (K2, K3, K4) appeared to have a non-random geographic partitioning (Append Figure 41). For a small proportion of individuals in each of these three K-models, the individual bar plots show up as white, rather than a solid colour, which indicates that for some fish the 80% CI of membership probability included both zero and one (both certain-membership and certain-not-membership). While this reflects the state of uncertainty in group membership for some individuals, more than 70% of fish were assigned with certainty for the K2, K3, and K4 models (Append Figure 41). Notably, the group membership plots for the K = 2, 3 and 4 (see Append Figure 40) consistently show that NWI fish are relatively genetically homogeneous with all 12 fish having higher than 95% assignment probability to single divergent group (Append Figure 41). Furthermore, this same genetic type is present at the four closest sampling locations (WCI, CIM, NCI, NEI), indicating that it may be primarily confined to the region of the Indian Ocean north of the equator. In addition, the K = 2, 3 and 4 models all showed a decline in the proportion of this genetic type with increased distance from the Arabian Sea sampling region and that it was virtually absent among adults sampled from the four locations south of the equator (SAF, SWI, ECI, SEI) and present in at <10% in the juveniles collected from WCI. Estimation of relatedness among the Indian Ocean sampling locations, based on analysis of  $F_{st}$  (Weir and Cockerham, 1984) and implemented in the program “assigner” (Gosselin 2020, [https://rdr.io/github/thierrygosselin/assigner/man/fst\\_WC84.html](https://rdr.io/github/thierrygosselin/assigner/man/fst_WC84.html)), also supports the uniqueness of the fish from NWI (Append Figure 42, a). The resulting  $F_{st}$  dendrogram demonstrates substantial differentiation between NWI and the rest of Indian Ocean sampling locations (Append Figure 42, b). The dendrogram also indicates greater differentiation exists between the cluster of sampling locations to the north of the equator and those to the south, as well as highlighting closer genetic relationships between western (SAF and SWI separated by 1800nm) and eastern Indian (ECI and SEI separated by 1200nm). Despite a minimum 5000 km separating south western pair (SAF and SWI) from eastern paired (ECI and SWI) locations, the  $F_{st}$  analysis indicated that these adult pairs were more similar to each other than either was to juvenile samples taken from locations directly to the north (Append Figure 43 a,b). Notably, the genetic makeup of the ECI adults sampled from the eastern Indian Ocean were quite different from the NEI juvenile sample taken from Lampulo just 5000km to the north. In contrast the juveniles from WCI Madagascar were not greatly dissimilar from the SWI and SAF adult samples taken to the south (1500km and 3700km respectively) although there was a small proportion (6 fish out of 84 or 7%) of the genotype, which predominated in the Arabian Sea sample location, that was at even lower frequency among the southern adults (1 out of 67 or 1.6%). Examination of the Atlantic Ocean and Pacific Ocean outgroup populations

also suggests limited connectivity from either into the Indian Ocean. The SAF and SWI sample locations were found to be substantially different.

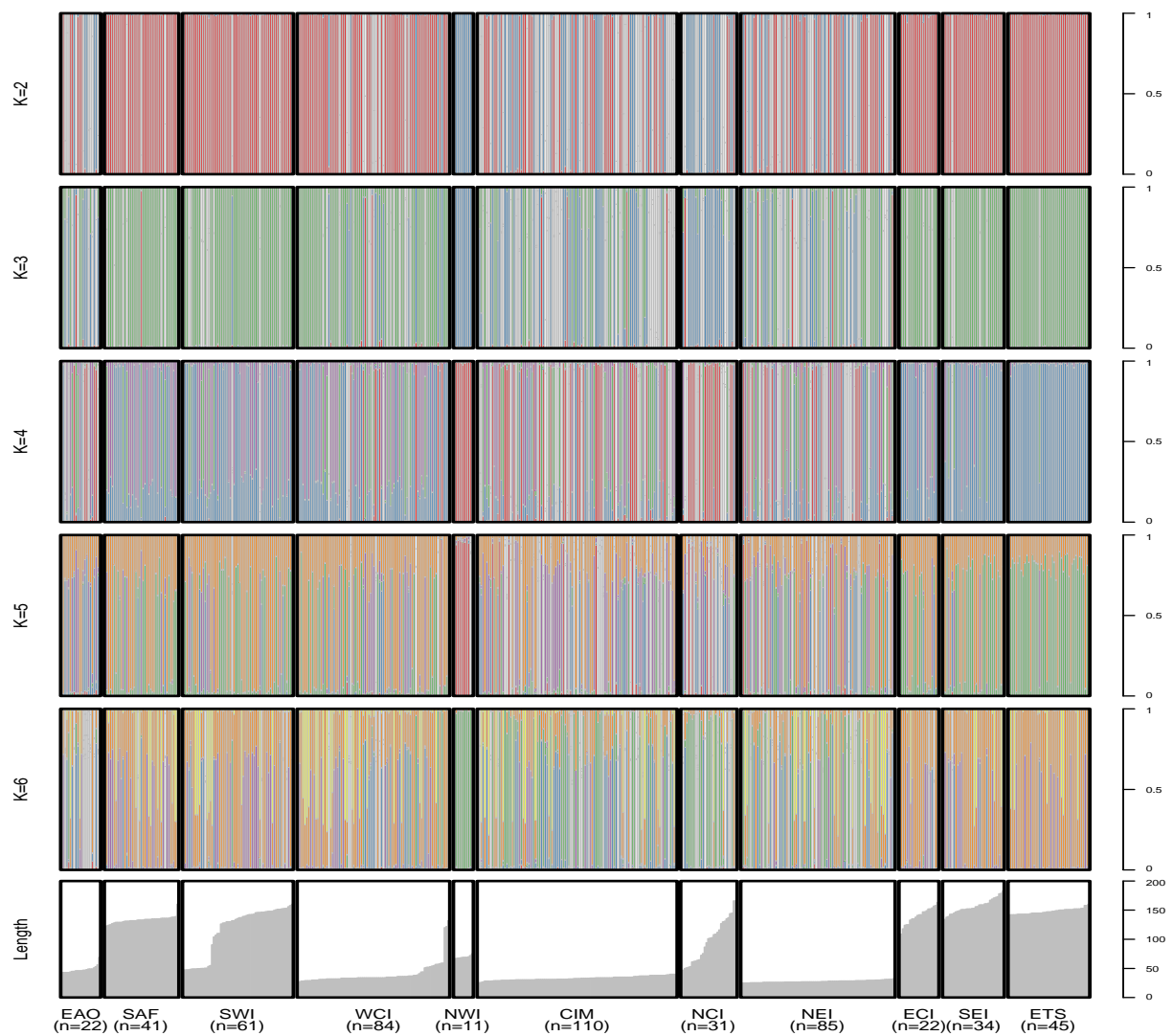


Append Figure 40. Left: Distribution of samples (N= 664) of yellowfin tuna (*Thunnus albacares*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Right: Information criterion used to assess the likelihood of different numbers of genetic groups (K).

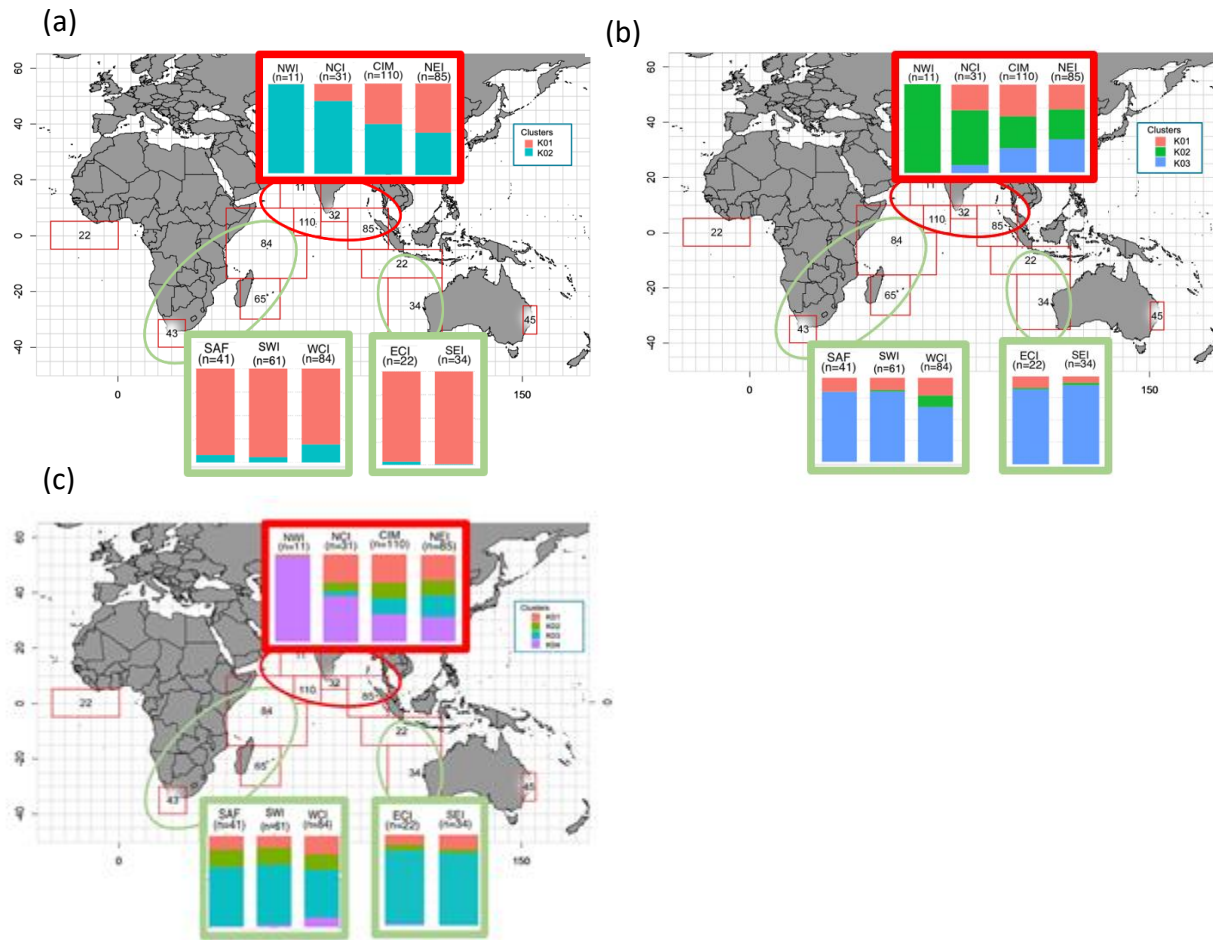
**Append Table 17. Parameters used in the radiator QC filtering of 11 sampling sites (strata) using starting values for radiator of 802 individuals, 39,663 locus fragments, and 54,733 SNP markers.**

<b>Applied FILTER step</b>	<b>Radiator PARAMETERS</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	filter.reproducibility	0.94	802 / 39036 / 53849
Filter monomorphic markers	filter.monomorphic		802 / 36116 / 49652
Filter markers in common	filter.common.markers		802 / 35609 / 49099
Filter individuals based on missingness (with outlier stats or values)	filter.individuals.missing	0.25	776 / 35609 / 49099
Filter monomorphic markers	filter.monomorphic		776 / 35574 / 49050
Filter MAC	filter.mac	6	776 / 27700 / 38584
Filter coverage min / max	filter.coverage	10 / 145	776 / 24495 / 35128
Filter genotyping	filter.genotyping	0.15	776 / 16502 / 24843
Filter SNPs position on the read	filter.snp.position.read	all	776 / 16502 / 24843
Filter markers snp number	filter.snp.number	3	776 / 15767 / 21502
Filter short Id	filter.short.id	mac	776 / 15767 / 15767
detect mixed genomes	ind.heterozygosity.threshold (min/max)	0.135783 / 0.15885	674 / 15767 / 15767
Filter monomorphic markers	filter.monomorphic		674 / 15766 / 15766
detect duplicate genomes	dup.threshold	0.25	546 / 15766 / 15766
Filter monomorphic markers	filter.monomorphic		546 / 15766 / 15766
Filter HWE	hw.pop.threshold / midp.threshold	80.01	546 / 15562 / 15562

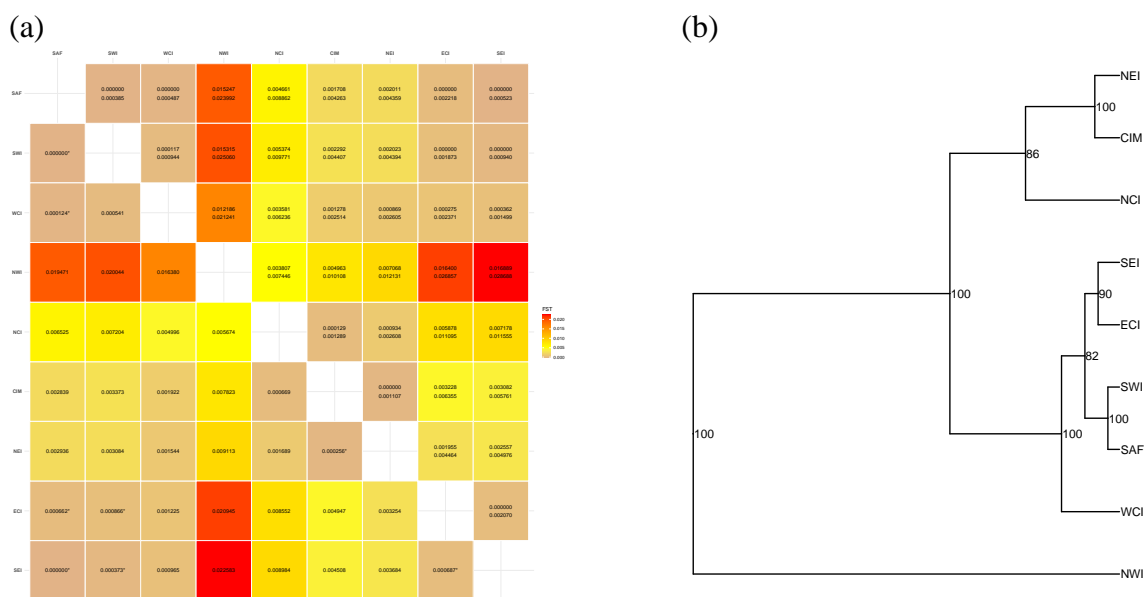




**Append Figure 41: Membership probability of an individual belonging to one of K genetic groupings.** Each vertical bar represents an individual genotypic profile with white profiles representing individuals where there is less than 80% certainty of belonging to a K genetic group. Individuals are sorted by lengths as indicated in the bottom pane. For each K 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.



Append Figure 42. Cumulative probabilities plotted at sample site locations for various models of K genetic groups. (a) K=2, (b) K=3, (c) K=4.



**Append Figure 43. (a)** Heat map and matrix of relative  $F_{st}$  values (below the diagonal) and minimum and maximum ranges (above the diagonal) as calculated among pairs of Indian Ocean sample sites. Colour indicates degree of relative differentiation between pairs of individuals with low (<0.005, beige), medium (0.005 to 0.015, mustard to yellow) and high (>0.015, red). Asterisks indicate minimum ranges of  $F_{st}$  estimates that overlap with zero. **(b)** Distance phenogram tree calculated using values from the  $F_{st}$  matrix as input for relative degree of differences between locations. Numbers at the nodes indicate bootstrap percent confidence of branch configuration (i.e. 100 indicates 100% of calculated trees had that particular branch point).

## Discussion

### *Population subdivision of YFT within the Indian Ocean*

Analysis of single nucleotide polymorphism (SNP) genotype data has indicated presence of temporally stable population heterogeneity within the Indian Ocean. While only based on eleven fish, the predominance (>95%) of a single genetic group in the NWI sample is indicative of a unique reproductively isolated group of fish present within this region. Characterisation of both nuclear (current study) and mtDNA markers within this region also provide further evidence for separate genetically partitioned aggregations (Demmannagoda et al., 2008; and Kunal et al., 2013; Barth et al., 2017). Presence of the predominant genetic group from the Arabian Sea among previously sampled locations (2013 and 2014) demonstrated the occurrence of a stable east-west genetic cline of this genotype among fish sampled north of the equator from the central Indian Ocean (Maldives) through to the East Indian Ocean southward along the coast of Java (Proctor et al., 2019). Further evidence of a temporally stable yet restricted east-west connectivity of yellowfin tuna comes from absence among three eastern Indian Ocean locations of a parasite that was found in the Maldives (Proctor et al., 2019, Moore et al., 2019). This east-west split was not apparent among samples from locations south of the equator but rather the spatial distribution modelled at  $K=2$  genetic types suggests restricted mixing between regions sampled north and south of the equator. These results concur with previous studies of Barth (et al., 2017) and provide evidence of a third genetically partitioned region of Indian Ocean yellowfin tuna. In contrast, the most parsimonious model of  $K=2$  genetic groupings, which is based on the AIC of our stockR results, supports a two-stock hypothesis as suggested by Morita and Koto (1970). However, the existence of at least three separate

morphologically differentiated stocks of Indian Ocean yellowfin tuna postulated by Kurogane and Hiyama (1958) gives rational biological impetus to warrant consideration of analyses that fit models of the data at values of  $K \geq 3$ . Further spatially targeted and temporally stratified sampling is required to further address this latter hypothesis of  $K > 2$  stocks and assess its intra-annual stability.

Localised environmental selection could also explain the observed genetic distribution patterns and provides an alternative mechanism to reproductively isolated spawning aggregations sampled from mixed feeding aggregations. Localised selection resulting in geographical patterning of differentiation of an otherwise reproductively panmictic population has been demonstrated for two species anguillid eels (Gagnaire et al., 2009; Ulrik et al., 2014). For yellowfin tuna, environmental gradients such as seasonal patterns of sea surface temperature, chlorophyll levels, and oxygen gradients are known to correlate with distribution patterns and could potentially act to preclude specific genotypes from certain habitats (Mohri and Nishida, 2000; Rajapaksha et al., 2014). Other variables such as winter and summer monsoonal turnover influence seasonal variability of Arabian Sea phytoplankton biomass and also coincides with anecdotal observations of peak recruitment periods for young tuna (<20cm) at FADS located near the Maldives (R. Jauhary, pers comm.; Marra and Moore, 2009; Kunarso et al., 2018). Interestingly, in the region of the Arabian Gulf high variability of salinity gradients, high sea surface temperatures, and low oxygen levels at shallow depths create unique physiological barriers potentially imposing positive selectivity for individuals genetically capable of surviving such a highly variable environment stresses (e.g. oxygen deficit regions; see Section 3 of this report). Perhaps unsurprisingly we see evidence of a single genotypic group in the Arabian Gulf which is almost absent from southern waters where of sea surface temperatures are cooler and oxygen levels remain high at depth. Strong genetic partitioning of both nuclear (current study) and mtDNA (Demmannagoda et al., 2008; and Kunal et al., 2013) are suggestive that the area where fish were collected from are representative of a potential strong environmental transition zone in the region of the Maldives (CIM), Sri Lanka (NCI), and Pakistan (NWI). Analysis of the available data is currently unable to resolve between competing hypotheses of whether patterns of genetic differentiation represent differential selection or whether the samples represent mixed feeding aggregations of reproductively isolated stocks. Resolving questions focussed on oceanographic determinants of observed geographic differentiation will require targeted analysis of larvae or actively spawning adults representing potential recruitment sources for genetically different groups of yellowfin tuna highlighted by the current study.

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

The pantropical distribution of yellowfin tuna indicates the potential for global connectivity of this species, however, a number of studies have demonstrated evidence of significantly restricted gene flow between the Indian Ocean and both the Atlantic and Pacific Oceans. Connections between major ocean basins are confined to a narrow bands of suitable water mass in the western Indian Ocean around the Cape of Good Hope while in the east, exchange between Indian and Pacific Ocean is restricted to transiting through suitable habitat found solely the Indonesian archipelagic waters (see FAO <http://www.fao.org/figis/geoserver/factsheets/species.html>). In this region, Mullins (et al., 2018) suggested the Benguela current likely provides a sufficient barrier of cold water to maintain isolation of fish and curtail gene flow between the Indian and Atlantic Ocean spawning areas. Indeed, genetic evidence of population subdivision observed by several authors (Pecoraro et

al., 2015; Barth et al., 2017; Mullins et al., 2018; Pecoraro et al., 2018) concurs with our own study and supports the conclusion that yellowfin tuna from the eastern Atlantic Ocean originate from separate source populations to fish collected between South Africa and the Mozambique channel. While the Indian/Atlantic Ocean appears to have obvious population differentiation, which Barth (et al., 2017) were able to detect with less than eight fish per sample location, the genetic divergence observed between Indian and Pacific Oceans is more subtle with much lower  $F_{st}$  values (between WTS (Pacific Ocean) and eastern Indian Ocean locations of ECI and SEI (0.005 versus 0.001 respectively; Append Figure 43). Absence of some parasites among yellowfin tuna sampled from 5 sites within the Indonesian archipelago and 3 nearby locations in the Pacific (Sorong, Jayapura, and Solomon Islands) are indicative of limited movement between Indian and Pacific Ocean for this species (Moore et al., 2019). Consistency among results from several different studies indicate the Indian Ocean effectively comprises a closed system for yellowfin tuna at contemporary time scales that has sufficiently promoted genetic differentiation among the Atlantic, Indian, and Pacific Oceans populations of this species.

#### 8.2.4 Yellowfin tuna (*Thunnus albacares*) – otolith microchemistry

Otolith microchemistry of yellowfin tuna has been used to investigate natal origins and population connectivity in the Pacific (Rooker *et al.* 2016, Wells *et al.* 2012) and discriminate yellowfin from different spawning grounds and nursery areas in the Atlantic Ocean (Kitchens *et al.* 2018, Shuford *et al.* 2007). In the Indian Ocean, Artetxe-Arrate *et al.* (2019) investigated stock structure using stable isotopes and trace elements in young-of-the-year yellowfin tuna.

For this study, samples were collected from young-of-the-year yellowfin in 3 known spawning sites in the northern Indian Ocean during 2 subsequent years, so the aim of the sampling design was to obtain a spawning ground signature from the smaller fish, to determine if the signature was temporally stable and to investigate if the spawning ground signatures matched those in otoliths from older yellowfin from 3 southern locations. For that, our assumptions were that the young-of-the-year fish had not moved significantly from their natal area so that the analysed core and edge otolith material had been deposited while the fish were in their nursery ground.

##### Methods

Otoliths from 197 yellowfin tuna from 6 locations (Append Figure 44) were analysed: 18 from South Africa (SAF), 11 from South west Indian Ocean (SWI), 71 from western central Indian Ocean (WCI), 45 from Maldives (CIM), 43 from north east Indian Ocean (NEI) and 9 from south eastern Indian Ocean (SEI). In this section, we will refer to WCI, CIM and NEI as the three northern locations, and SAF, SWI and SEI as the three southern locations.

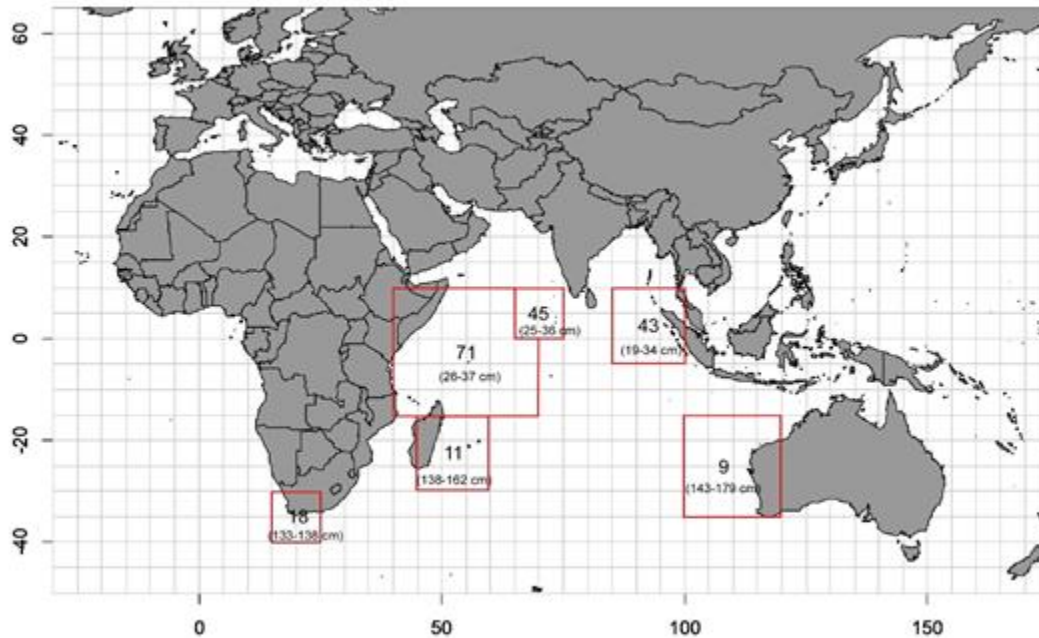
All northern samples were young-of-the-year (YOY, 19.5- 37.5 cm FL) and those from the southern sites were adults (133-179 cm FL) and were obtained in two consequent years, 2018 and 2019 (Append Figure 44, Append Table 18). Although it was not intended in the original design, samples from different locations and years were collected during different periods, due to several reasons (see Section 2).

- WCI: March-April 2018 and April 2019
- CIM: August 2018 and February 2019
- NEI: April 2018 and November 2018
- SAF: March 2018
- SWI: March 2018 and Feb 2019
- SEI: May 2019

YOY samples were analysed separately for each sampling year. In this section, we will refer to the second round of NEI samples (those collected in November 2018) as the 2019 samples from this location.

The otoliths were analysed at the Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux Université de Pau et des Pays de l'Adour/CNRS using high resolution LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Eight trace elements were measured (see Section 4.5.2), 4 were below the limit of detection and therefore discarded for analyses.

The spot 65  $\mu\text{m}$  from the core ( $\pm 13$ -15 days of life) was selected as natal origin signature. The spot at the edge of the otolith was also analysed for YOY 2018, since these data reflect the fish's known capture location, and can be used for validation purposes.



Append Figure 44. Map showing the number of yellowfin tuna otoliths analysed for each of six sampling locations, referred to as South Africa (SAF), South-West Indian Ocean (SWI), Western Central Indian Ocean (WCI), Maldives (CIM), North-East Indian Ocean (NEI), and south east Indian Ocean (SEI); and the size range of fish at each location.

Append Table 18. Number, sampling period and estimated ages of fish for each of the six sampling locations SAF, SWI, WCI, CIM, NEI and SEI.

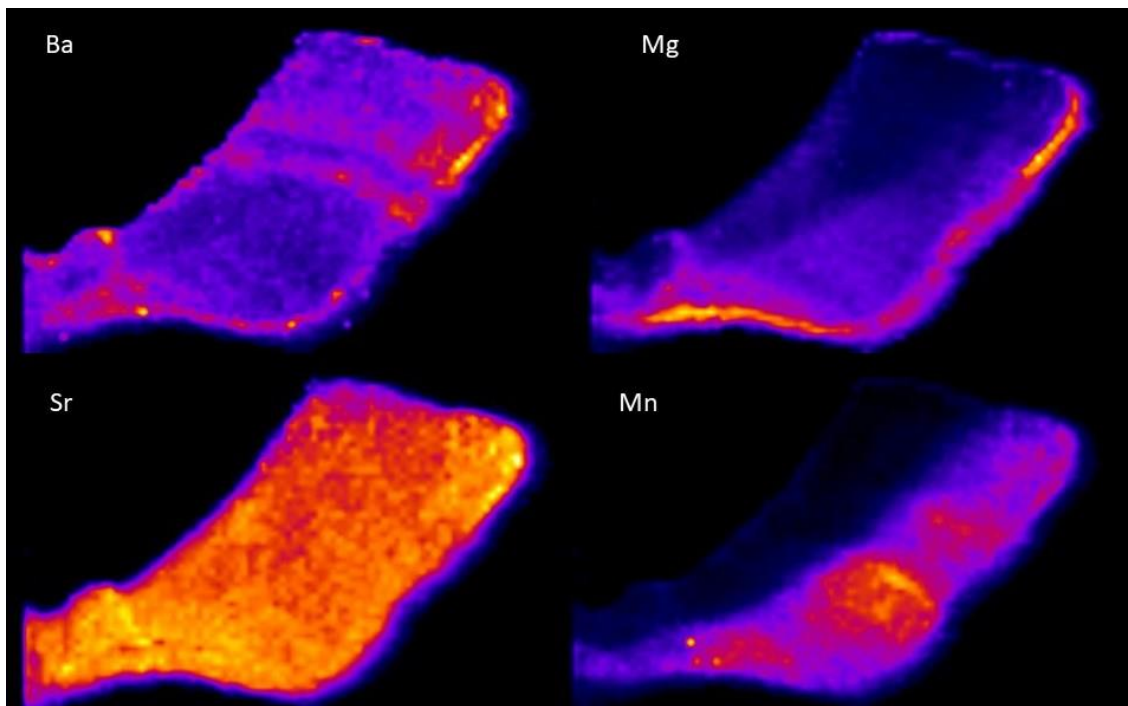
Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
South Africa (SAF)	18	March 2018	133-138	5+
South West Indian Ocean (SWI)	11	March 2018 and Feb 2019	138-162	5+
West Indian Ocean (WCI)	71	March-April 2018 and April 2019	26-37	0+
Maldives (CIM)	45	August 2018 and February 2019	25-36	0+
East Central Indian (ECI)	43	April 2018 and November 2018	19-34	0+
South East Indian (SEI)	9	May 2019	143-179	5+

\* The ranges in ages are for male and females combined (Eveson et al., 2015).

## Results

Core data for YOY 2018 samples included Ba, Mg, Mn and Sr, but YOY 2019 were only analysed for Ba and Sr. The edge signatures data only included Ba and Sr, due to inconsistencies found in the distribution of Mg and Mn along the edge (Append Figure 45).





Append Figure 45. Transverse section showing Element:Ca concentration distribution along the ventral arm of the otolith of a YOY yellowfin tuna (37cm FL).

### CORE results

Boxplots comparing core signatures between locations for YOY yellowfin tuna sampled in 2018 show that Mn, Sr, d18O and d13C varied significantly among locations (Append Figure 46).

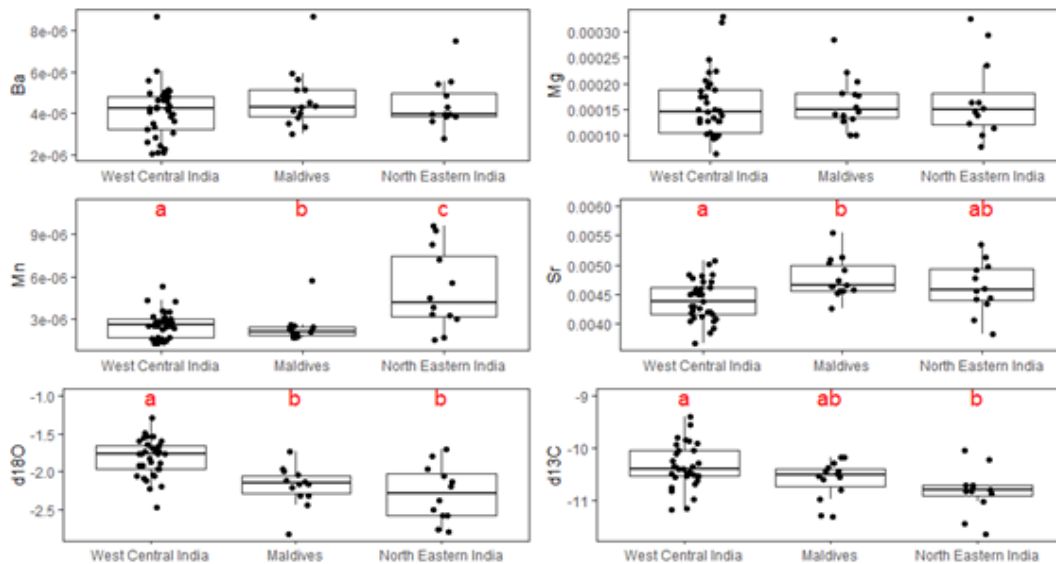
Boxplots comparing core signatures between locations for YOY yellowfin tuna sampled in 2019 show that Mg and Mn varied between locations (Append Figure 47).

Based on PERMANOVA, multielemental core signatures of YOY yellowfin tuna are not equal among all locations both for fish sampled in 2018 ( $p=0.001$ ) and 2019 ( $p=0.001$ ). Based on subsequent pairwise tests between locations:

- All locations differ in core signatures for 2018 samples (Append Table 19)
- All locations differ in core signatures for 2019 samples (Append Table 20)

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data confirms and helps to visualize these findings (Append Figure 48).



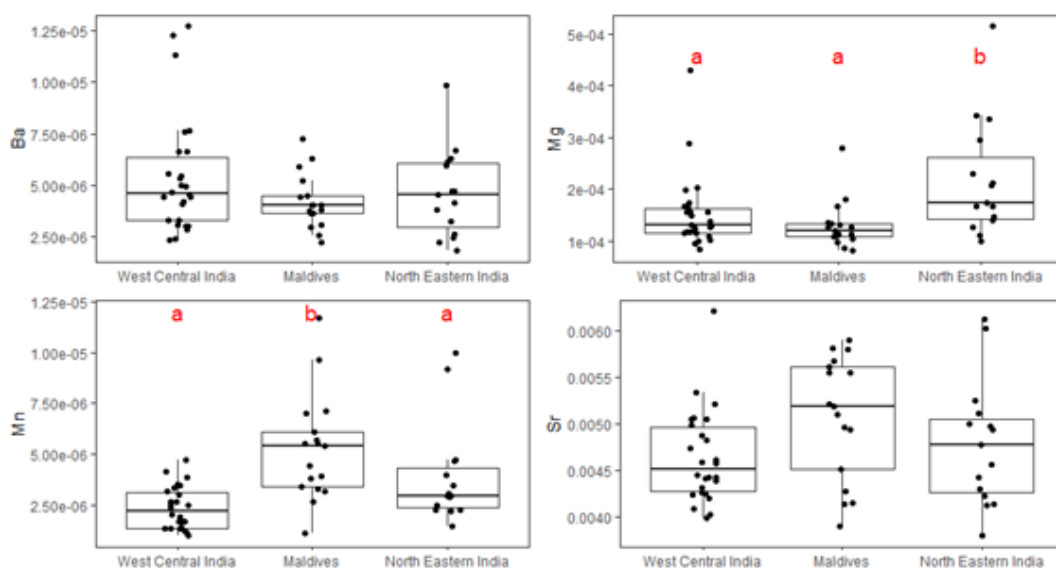


Append Figure 46. Boxplots comparing YOY yellowfin tuna otolith core data between locations for each element for samples collected in 2018. Significant differences ( $p < 0.05$ ) among locations are shown with red letters.

Append Table 19. Results of pairwise comparisons of YOY yellowfin tuna otolith core signatures between locations of samples from 2018. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI	vs	CIM	1	25.351	5.863	0.115	0.001	0.002	**
WCI	vs	NEI	1	45.803	8.478	0.165	0.001	0.002	**
CIM	vs	NEI	1	17.929	3.008	0.111	0.030	0.030	*

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).

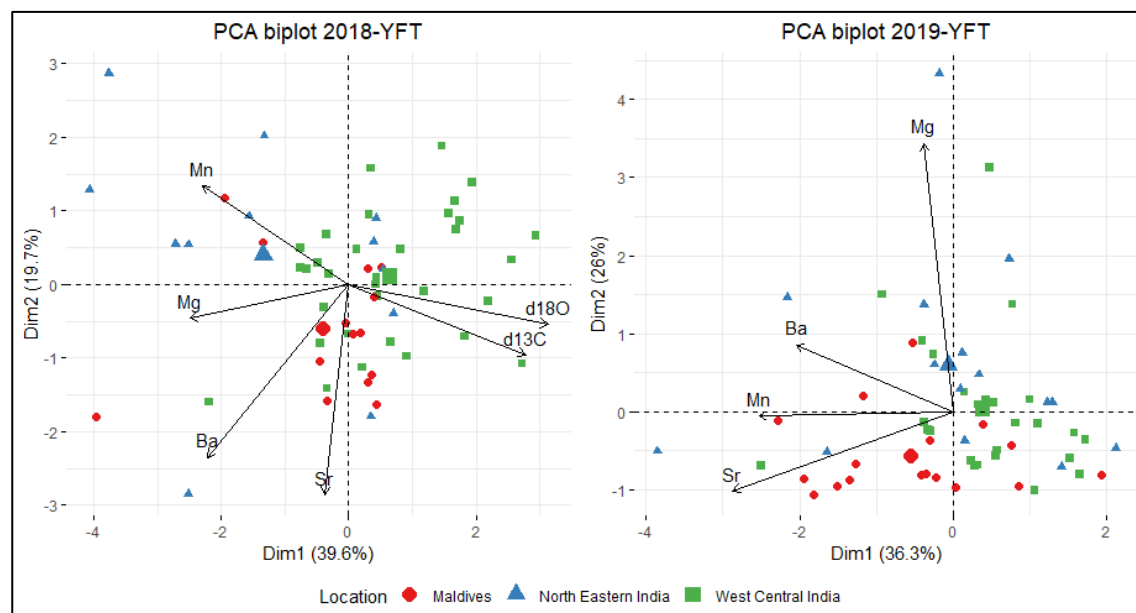


Append Figure 47. Boxplots comparing YOY yellowfin tuna otolith core data between locations for each element for samples collected in 2019. Significant differences ( $p < 0.05$ ) among locations are shown with red letters.

**Append Table 20. Results of pairwise comparisons of YOY yellowfin tuna otolith core signatures between locations of samples from 2018. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian.**

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI	vs	CIM	1	24.804	8.148	0.166	0.001	0.003	**
WCI	vs	NEI	1	11.497	3.081	0.073	0.019	0.019	*
CIM	vs	NEI	1	13.616	3.473	0.104	0.011	0.011	*

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).

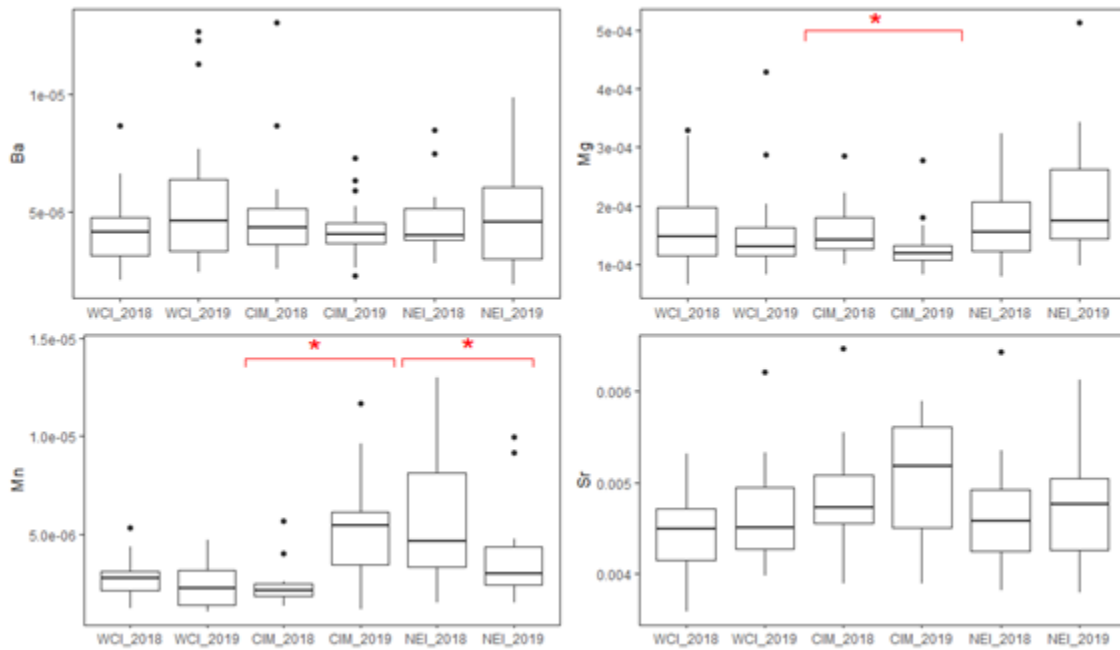


**Append Figure 48. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the yellowfin tuna otolith core signatures for the 2018 and 2019 samples respectively. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.**

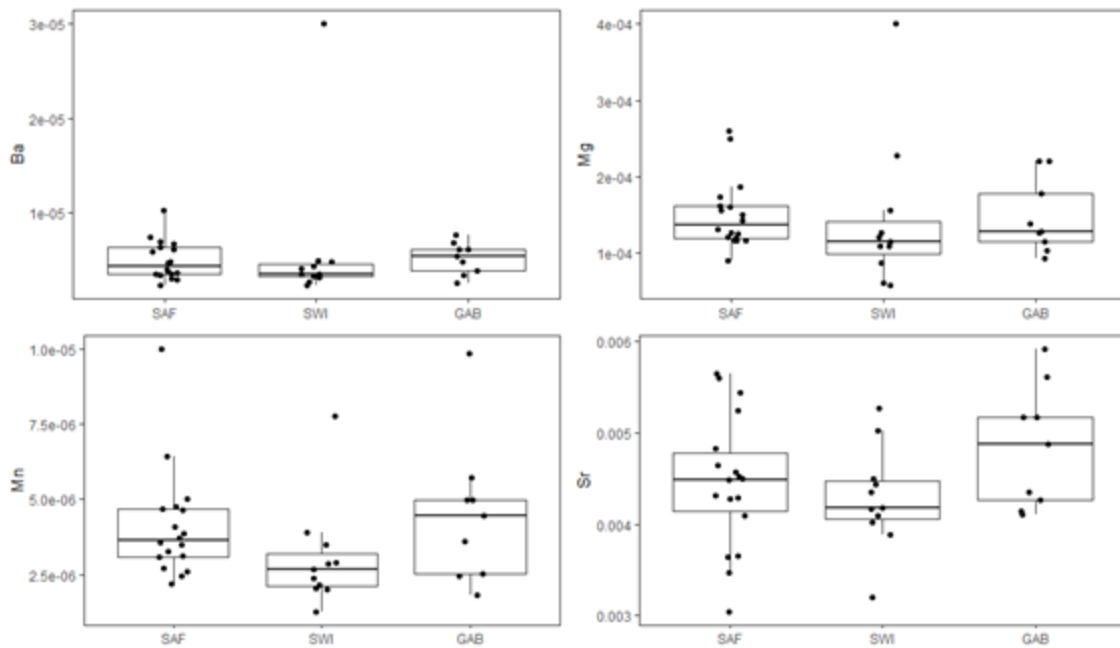
Boxplots comparing core signatures of YOY yellowfin within sampling years show that some elemental signatures were not consistent (Append Figure 49). Mg varied among years in the CIM, whereas Mn varied among years in CIM and NEI. Ba and Sr did not vary among years within any locations.

Core signatures of adult YFT did not vary among sampling locations (Append Figure 50) based on boxplots. Based on PERMANOVA, multielemental core signatures of adult yellowfin tuna did not vary among sampling locations (Append Table 21). A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data confirms and helps to visualize these findings (Append Figure 51).

The optimal number of clusters was estimated to be 1, based on the gap statistics with kmeans clustering.



Append Figure 49. Boxplots comparing yellowfin tuna otolith core data between sampling years for each pair of locations. Significant differences ( $p < 0.05$ ) among years within locations are shown with red \*. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian.

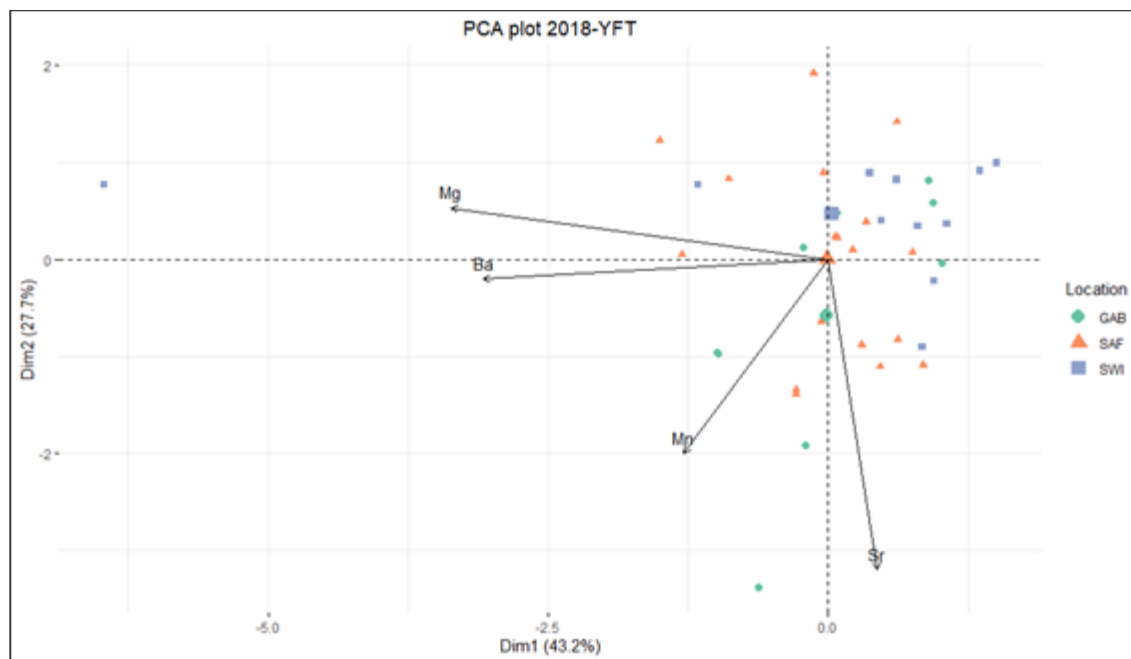


Append Figure 50. Boxplots comparing adult yellowfin tuna otolith core data between locations for each element for samples collected in the three southern locations. Significant differences ( $p < 0.05$ ) among locations are shown with red letters. SAF= South Africa, SWI= South Western Indian and SEI= South Eastern Indian.

Append Table 21. Results of pairwise comparisons of adult yellowfin tuna otolith core signatures between locations. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. SAF= South Africa, SWI= South Western Indian and SEI= South Eastern Indian.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
SAF	vs	SWI	1	2.897	0.676	0.024	0.644	0.644	
SAF	vs	SEI	1	2.193	0.771	0.030	0.514	0.644	
SWI	vs	SEI	1	6.209	1.174	0.061	0.284	0.644	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 51. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the adult yellowfin tuna otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia. SAF= South Africa, SWI= South Western Indian and SEI= South Eastern Indian.

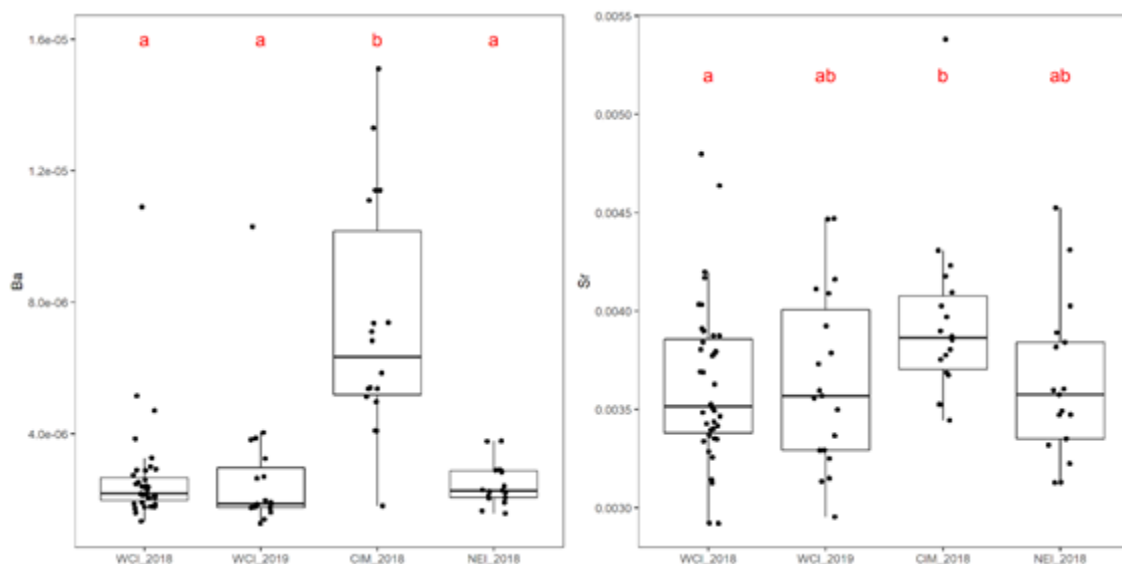
## EDGE results

Boxplots comparing edge signatures between locations show that CIM have higher Ba values than the rest of locations, and higher Sr than WCI samples of 2018 (Append Figure 52).

Based on PERMANOVA, multielemental edge signatures were not equal among all locations ( $p=0.001$ ). Based on subsequent pairwise tests between locations: CIM edge signatures differed from edge signatures from other locations (Append Table 22).

A biplot showing edge Ba and Sr signatures of individuals confirms and helps to visualize these findings (Append Figure 53).

Core and edge signatures were significantly different for all locations (Append Figure 54).

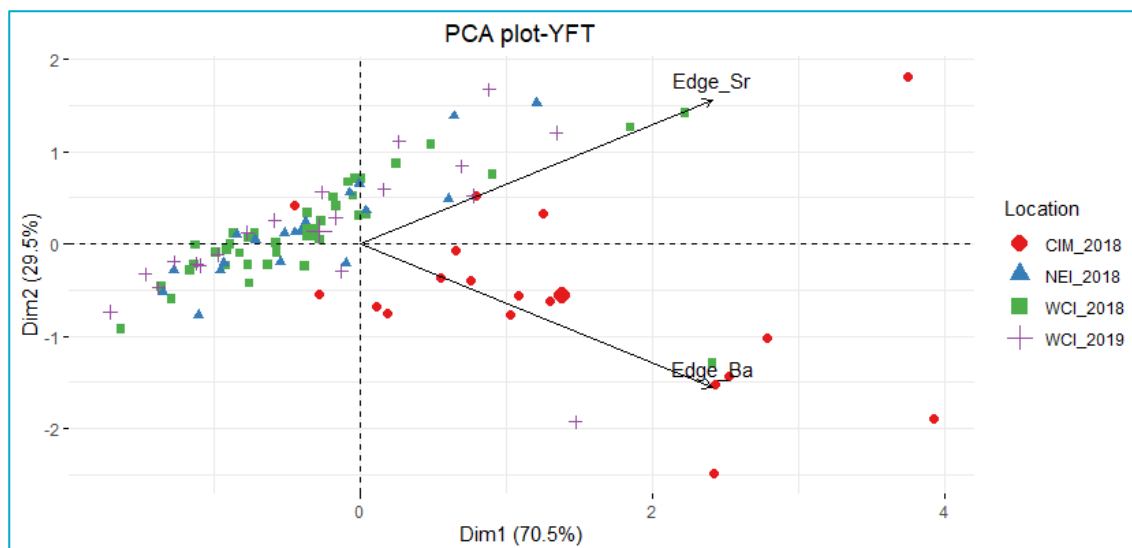


Append Figure 52. Boxplots comparing yellowfin tuna otolith edge data between locations for each element for samples collected. Significant differences ( $p < 0.05$ ) among locations are shown with red letters. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian. For WCI edge was analysed both in 2018 and 2019

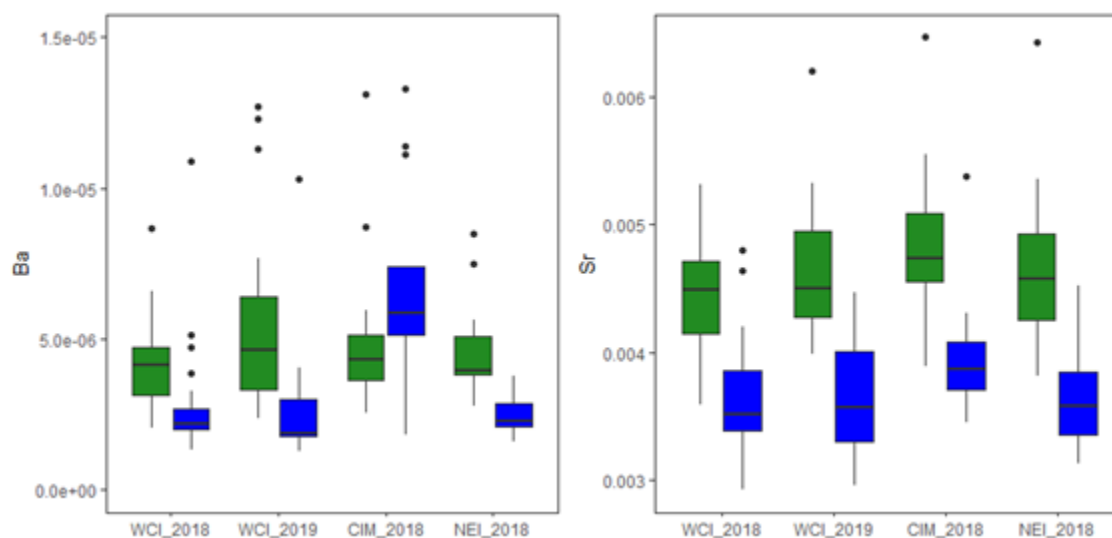
Append Table 22. Results of pairwise comparisons of yellowfin tuna otolith edge signatures between locations. In the Significance column, \* means that the locations differ significantly at the 0.05 level and \*\* at the 0.01 level. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian. For WCI edge was analysed both in 2018 and 2019.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI_2018	vs	CIM_2018	1	42.120	25.570	0.317	0.001	0.002	**
WCI_2018	vs	NEI_2018	1	0.043	0.038	0.000	0.945	0.976	
CIM_2018	vs	NEI_2018	1	31.568	17.454	0.346	0.001	0.002	**
WCI_2018	vs	WCI_2019	1	0.047	0.035	0.000	0.971	0.976	
CIM_2018	vs	WCI_2019	1	30.034	14.089	0.287	0.001	0.002	**
NEI_2018	vs	WCI_2019	1	0.086	0.067	0.002	0.976	0.976	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 53. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the yellowfin tuna otolith edge signatures. Individuals are coded by their sampling location and year. For the variables, the length of the arrow reflects the % of contribution to the total inertia.



Append Figure 54. Boxplots comparing YOY yellowfin tuna core (green) and edge (blue) signatures at the three northern sampling locations. WCI= West Central Indian, CIM= Maldives and NEI= North Eastern Indian. For WCI edge was analysed both in 2018 and 2019.

## Summary

The yellowfin otoliths were collected from three northern locations that are known spawning sites -- WCI, CIM and NEI -- and, by analysing spots near the otolith core, we aimed to obtain an otolith signature that was representative of the nursery locations. By analysing spots on the otolith edge, we aimed to obtain an otolith signature that was representative of capture locations, at time of capture.

For the YOY samples collected in 2018 at the northern locations, core signatures differed between all locations. The NEI and WCI otoliths were from fish of similar lengths that had been caught at the same time, providing evidence that fish from these two locations did, in fact, have distinct spawning

ground otolith signatures. However, the CIM samples were from the same size fish but they had been collected at different times of the year, indicating the CIM fish were spawned in a different season to the NEI and WCI fish. Therefore, the CIM difference in spawning ground otolith signature could be due, at least in part, to seasonal variability in ocean conditions.

Similarly, for the YOY samples collected in 2019 at the northern locations, core signatures differed between all locations. In 2019 the WCI and CIM otoliths were from fish of similar lengths that had been caught at the same time, providing evidence that fish from these two locations did, in fact, have distinct spawning ground otolith signatures. However, the NEI samples were from the same size fish but they had been collected at different times of the year, indicating the NEI fish were spawned in a different season to the WCI and CIM fish. Therefore, the NEI difference in spawning ground otolith signature could be due, at least in part, to seasonal variability in ocean conditions.

Interannual differences were detected in the core signatures of YOY yellowfin in 2 of the 3 northern locations, CIM and NEI, indicating that otolith signatures are not temporally stable between years. However, as fish from CIM and NEI were sampled at different periods in each year, the differences between years could be due to seasonal effects. Fish from WCI were caught at the same time in each year, 2018 and 2019, and the otolith signatures were not significantly different between years.

Adult yellowfin tuna core signatures did not differ among capture locations. Additionally, the posterior cluster analysis (which does not use knowledge of capture location) did not separate the adult core data into groups. This might imply that all the fish had a common origin, or more likely, because adult samples were collected from a wide range of cohorts, differences among years could be larger than differences among locations.

Edge signatures of YOY yellowfin tuna from the WCI and NEI samples were collected at the same time and their edge signatures were not different, possibly because ocean chemistry did not differ significantly between these locations at the time of capture. The WCI and NEI samples were collected in April, when the regional oceanography is more similar in these locations than other time of the year (see Figure 4, Figure 5). Edge signatures of YOY yellowfin from CIM differed from the edge signature of the WCI and NEI yellowfin. This difference could be due to regional differences in ocean chemistry or, as samples of CIM were collected at a different time of the year from WCI and NEI samples, due to seasonal differences in ocean chemistry.

Core and edge signatures were different in all locations, in each of the 2 years. The difference between core and edge could be due to ontogenetic changes during the first months of life strongly influencing the chemical composition of the otolith. However, variation in seasonal oceanography during the first months of life could also have influenced the chemical composition of the otolith.

Overall, the results for yellowfin tuna microchemistry suggest that seasonality can strongly influence otolith signatures and should be considered in any further studies, ideally sampling all fish at the same time. In this study, the results from fish collected at the same indicated that different nursery areas can be discriminated within the Indian Ocean using otolith microchemistry. This might serve as an effective tool to study the connectivity and mixing rates of older individuals within the Indian Ocean. However, due to seasonal and variability, it is crucial that juveniles from different locations are collected at the same time of the year to assess the extent of seasonal and interannual variability, in order to establish a baseline to investigate the natal origin of adults.

### 8.2.5 Bigeye tuna (*Thunnus obesus*) - population genetics

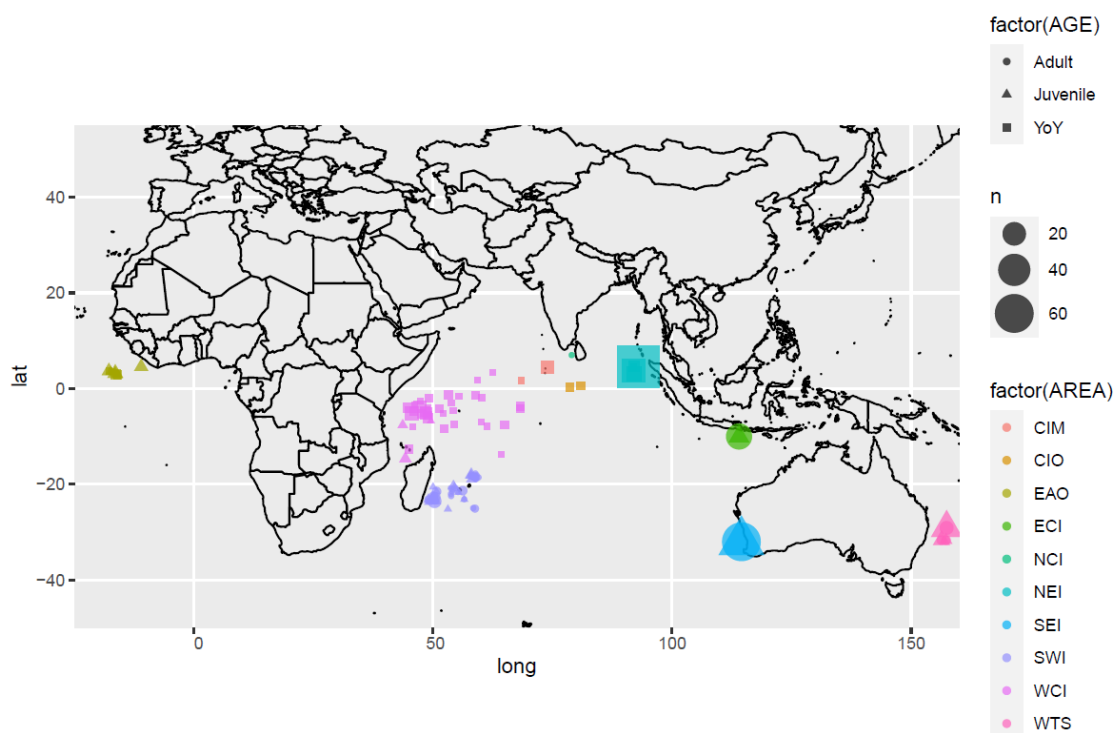
Previous studies on the population structure of the bigeye support inter-oceanic genetic separation (Durand et al. 2005) (Alvarado-Bremer et al. 1998; Chow et al. 2000b; Martinez et al. 2006). Most of the few studies aimed at understanding population structure of this species within the Indian Ocean did not observe signs of heterogeneity, supporting the existence of a single panmictic population (Appleyard, Ward, Grewe 2002; Chiang et al. 2008).

#### Methods

##### *Sampling, DNA extraction and DArTseq library preparation*

Bigeye tuna samples from the Indian, Pacific and Atlantic Ocean were obtained from scientific and commercial fisheries (Append Figure 55). The 496 samples were organized in 10 areas (Maldives, Central Indian Ocean, East Central Indian Ocean, North Central Indian Ocean, North East Indian Ocean, South East Indian Ocean, South West Indian Ocean, West Central Indian Ocean, West Tasman Sea and East Atlantic Ocean). Samples were classified by Straight Fork Length (SFL) as young of the year (YoY) (<45 cm), juveniles (45-120 cm) and adults (>120 cm) according to the age-length classification proposed in the sampling protocol of this project. DNA was extracted from about 15 mg on an Eppendorf 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. DNA sample libraries were created in digestion/ligation reactions using two restriction enzymes, PstI 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. SphI-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” (PstI-SphI) 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.





**Append Figure 55.** Samples collected for this study. Each location is represented by one color (CIM - Maldives, CIO - Central Indian Ocean, ECI - East Central Indian Ocean, NCI - North Central Indian Ocean, NEI - North East Indian Ocean, SEI - South East Indian Ocean, SWI - South West Indian Ocean, WCI - West Central Indian Ocean, WTS - West Tasman Sea and EAO - East Atlantic Ocean) and shapes indicate if samples are young of the year (YoY), juveniles or adults. Size of shapes are proportional to the number of samples collected per point.

### *DART-tag assembly and SNP calling*

Generate Dart-seq reads were analyzed using *Stacks* version 2.4 (Catchen et al. 2013). Using *process\_radtags*, reads were truncated to 69bp so that all reads had the same length after barcode removal and reads with any uncalled base, with total low-quality scores or with quality score below 20 within 10bp size sliding windows were removed. The module *ustacks* was then used to assemble orthologous tags (stacks) per individual, with a minimum coverage depth required to create a stack (parameter -m) of 3, and a maximum nucleotide mismatches allowed between stacks (parameter -M) of 2. Matches to the catalog for each sample were searched using *sstacks* and transposed using *tsv2bam* and the module *gstacks* was used to identifying and genotyping SNPs. Only samples with more than 30,000 and less than 65,000 reads were selected for further analyses. The module *populations* was used to export from the catalog, the SNPs presented in RAD loci found in at least 75% of the individuals. Using *PLINK* version 1.07 (Purcell et al. 2007), SNPs with more than 5% missing data and a minimum allele frequency (MAF) smaller than 0.05 or failing the Hardy-Weinberg equilibrium test ( $p < 0.05$ ) in at least two location (excluding North Central Indian (NCI), Central Indian Ocean (CIO) and Maldives (CIM) locations which contained less than 10 individuals), as well as samples with more than 10% missing data were excluded from downstream analyses. Finally, only the first SNP per loci were kept and resulting genotype table was exported into *Structure* and *Genepop* formats. Related individuals were identified using GCTA (Yang et al. 2011).

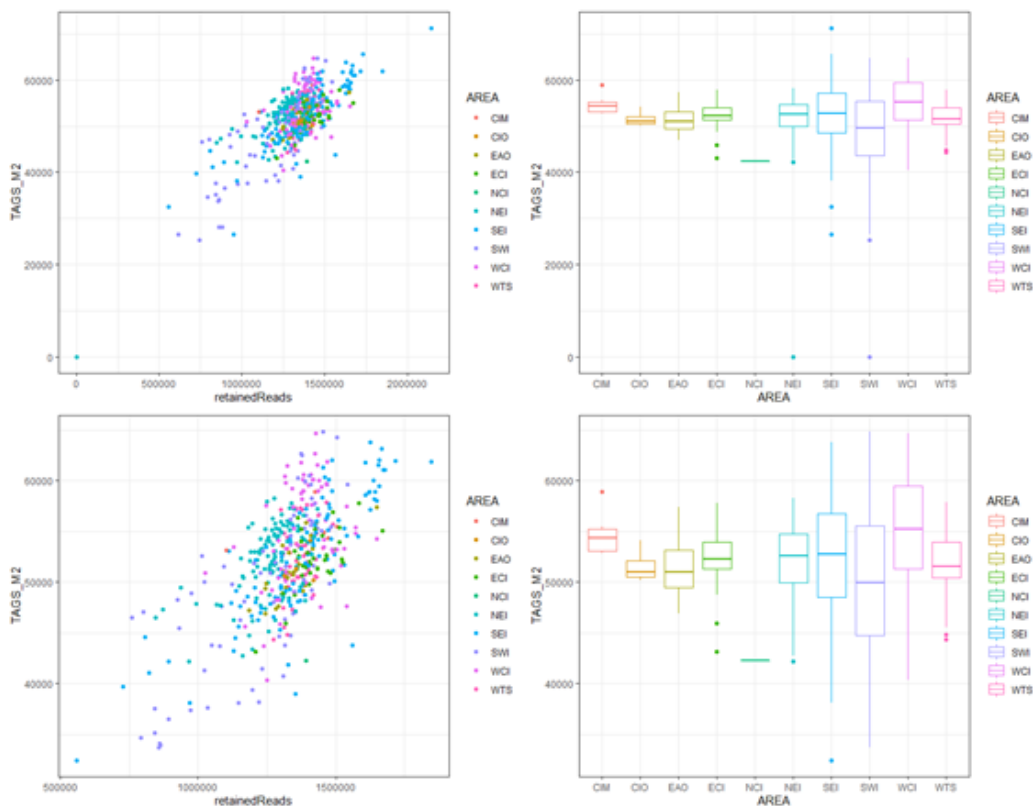
### *Genetic diversity and population structure analyses*

Principal Component Analysis (PCA) was performed using the *ade4* R package (Jombart, Ahmed 2011), and ADMIXTURE (Alexander, Novembre, Lange 2009) was run assuming from 2 to 4 ancestral populations (K) setting default parameters. The value of K with lowest associated error value was identified using ADMIXTURE's cross-validation procedure testing K values from 1 to 6. Groups were sought in the genetic data as implemented in the R package stockR (Foster et al. 2018). 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 (25 times in this initial analysis) the genetic data and see how well the groupings match those from the analysis of the full data. The groupings are displayed using probabilities of individual fish to each genetic group are obtained using bootstrap methods (Foster et al. 2018), using 100 resamples in this initial analysis.

## Results

### Data quality

Considering all 496 individuals, an average of 61.3% of the reads were retained for de novo assembly and average coverage was 24X. Number of tags per individual was variable, and some samples resulted in a two low (due to a low number of reads) or two high (potentially due to cross-contamination) number of tags (Append Figure 56).

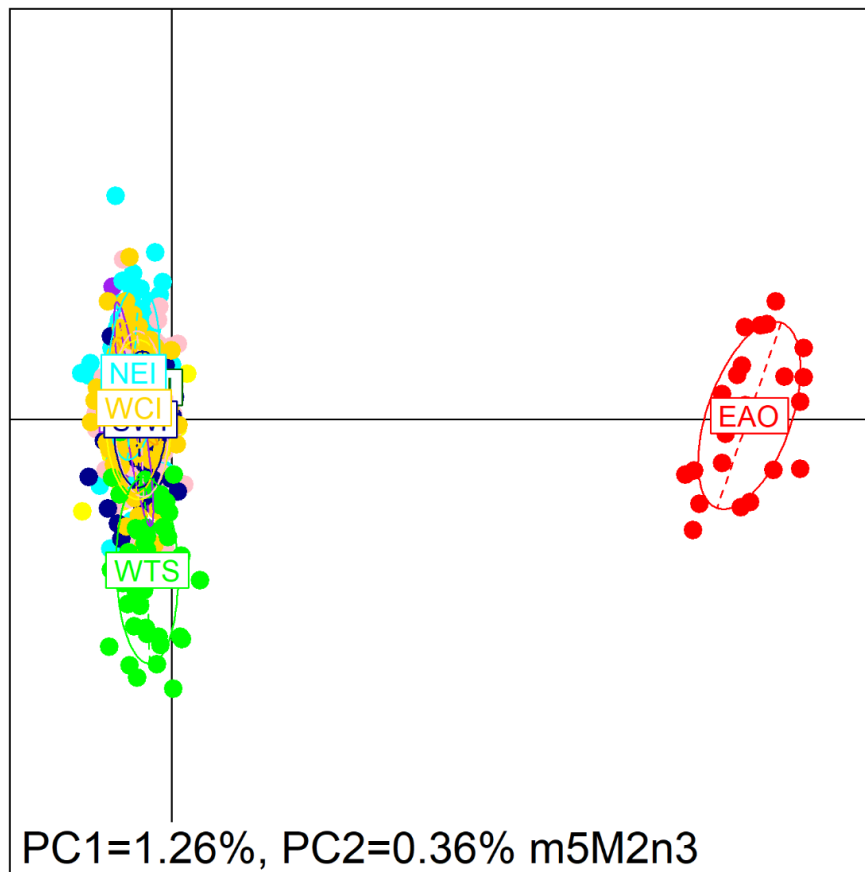


Append Figure 56. Relation between number of tags and number of reads per individual (left) and boxplot of the number of tags per location (right) for the complete dataset.

After removing individuals with too many or too low number of tags, the datasets consisted of 487 individuals. After SNP filtering and related individual removal, the final dataset resulted in 472 and 6,093 SNPs.

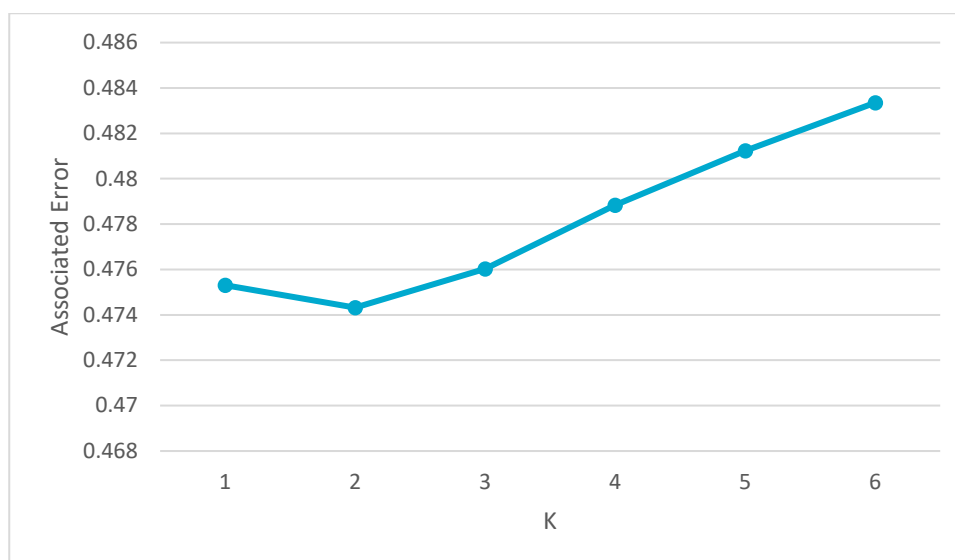
#### *Genetic diversity and population structure*

Principal Component (Append Figure 57) show strong differentiation between samples from the Atlantic (East Atlantic Ocean, EAO) and samples from the Indian and Pacific Oceans, but do not suggest any intra-oceanic structure.

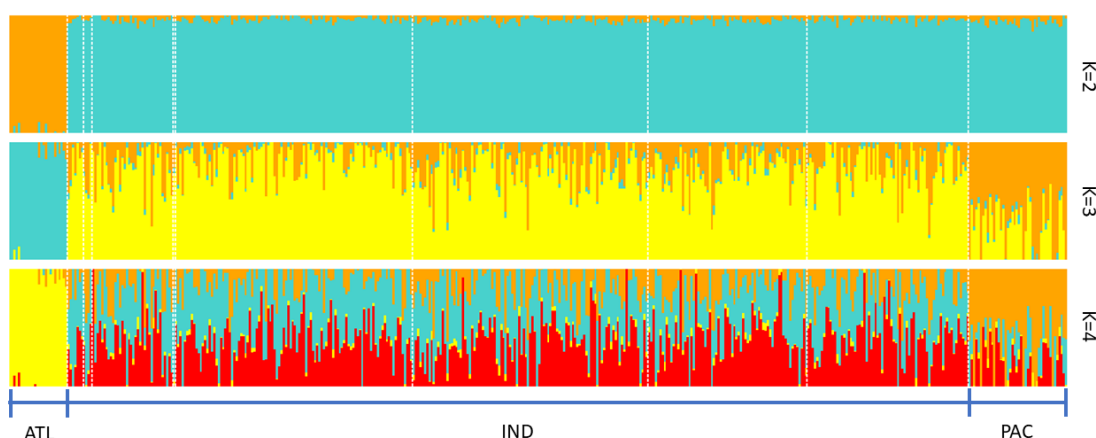


Append Figure 57. Principal Component Analysis (PCA) performed using the final dataset containing 472 samples and 6,093 SNPs. Different colors in the PCA represent samples from the different locations at which samples were captured.

ADMIXTURE analyses show that the number of assumed ancestral populations with the lowest associated error (K) is 2 (Append Figure 58). Individual ancestral proportions (Append Figure 59), show three clusters corresponding to the three oceans.

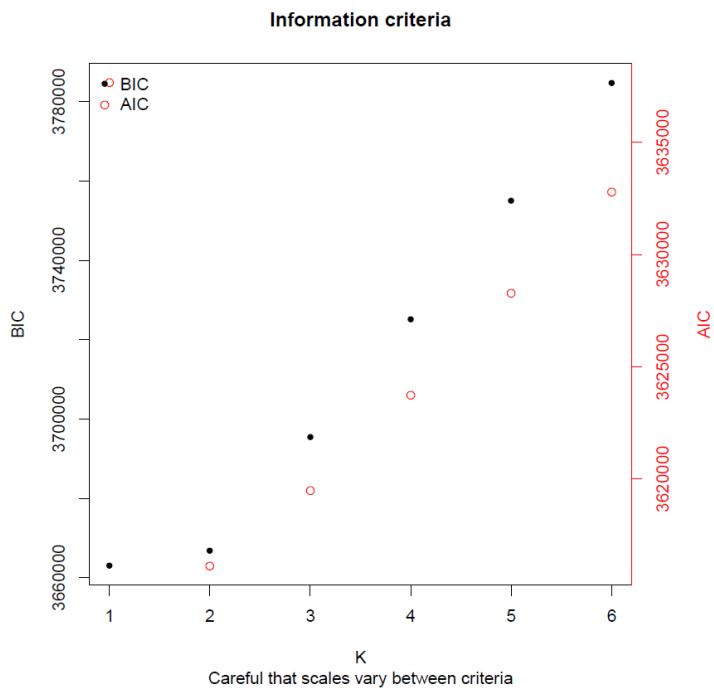


Append Figure 58. Associated error for each value of assumed ancestral population (K) estimated by ADMIXTURE.

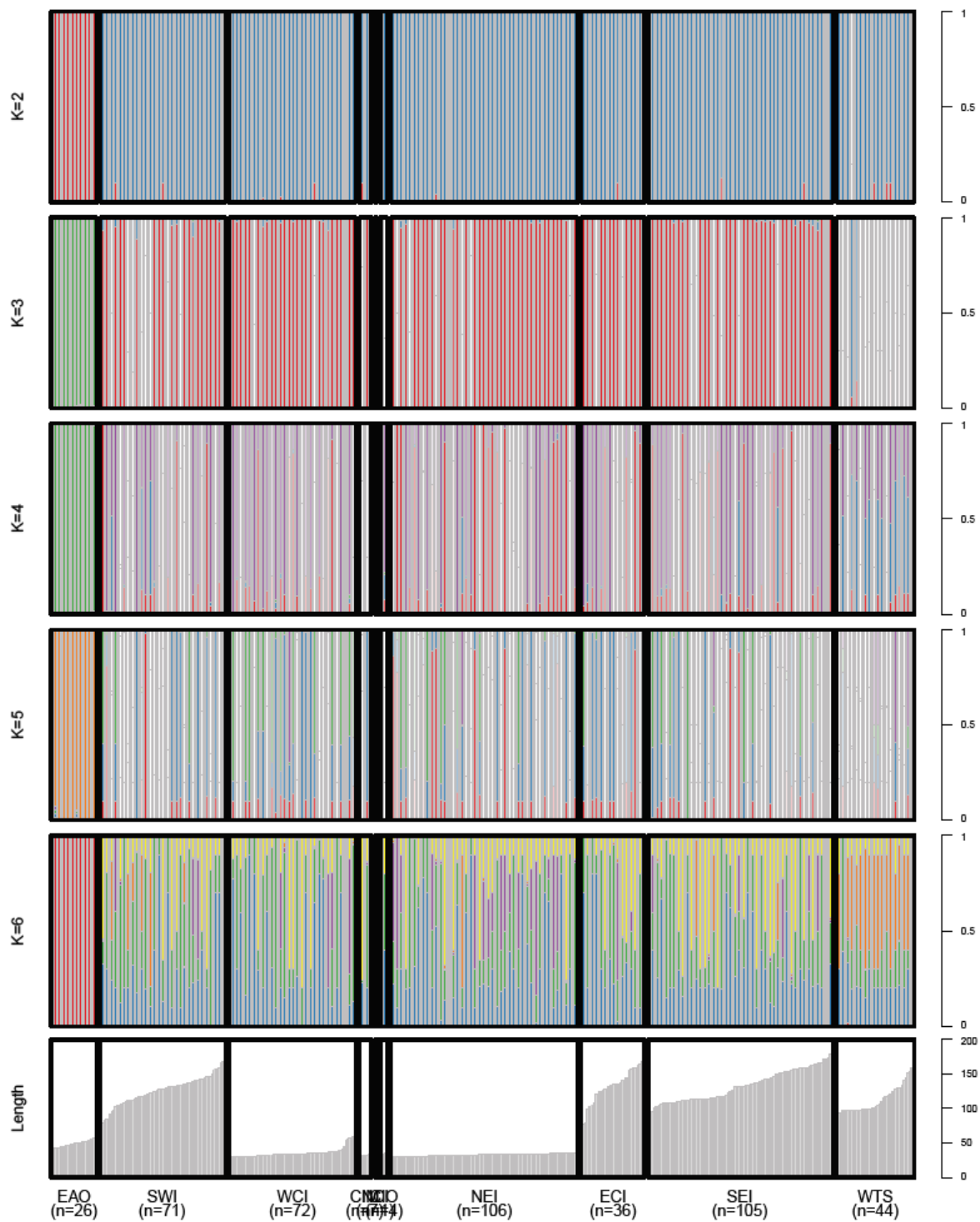


Append Figure 59. Individual ancestry proportions estimated by ADMIXTURE when assuming from two to four ancestral populations (K) of samples from different location in the Atlantic (ATL), Indian (IND) and Pacific (PAC) Oceans.

Similarly, StockR analyses suggest the presence of at least two groups (Append Figure 60) with a clear separation between the Atlantic from the Pacific and Indian Oceans, and a further separation of the Pacific from the Indian Ocean samples (Append Figure 61). No differentiation is observed within the Indian Ocean, suggesting that bigeye in the Indian Ocean belongs to the same genetic population.



**Append Figure 60. Information Criterion plot summarizing the results of model fits (using AIC and BIC) for the most likely number of genetic groups from the distribution of SNP data in the sample.**



Append Figure 61. Assignment of fish to groups (assuming 1 to 6 possible groups) according to StockR.

## Summary

Bigeye from Atlantic, Indian and Pacific oceans form three genetically distinct populations. Within the Indian ocean, no genetic differentiation can be observed suggesting that the bigeye from the Indian ocean forms a single population.

## 8.2.6 Bigeye tuna (*Thunnus obesus*) – otolith microchemistry

The investigation of otolith microchemistry of bigeye tuna (BET) has thus far been limited to a study in the Pacific Ocean (Rooker et al. 2016). They found spatial variation in the otolith elemental and stable isotope signatures from young of the year (YOY) BET that were collected in 4 locations in the western and central Pacific Ocean. These signatures were then used as a baseline to classify age-1 to age 2+ fish to their natal origin by comparing the YOY signatures with otolith material from the older fish ablated near the core.

In this report, we concentrated on analysing elemental signatures from the otolith core, since these data should reflect the fish's spawning origins. However, it was also useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.

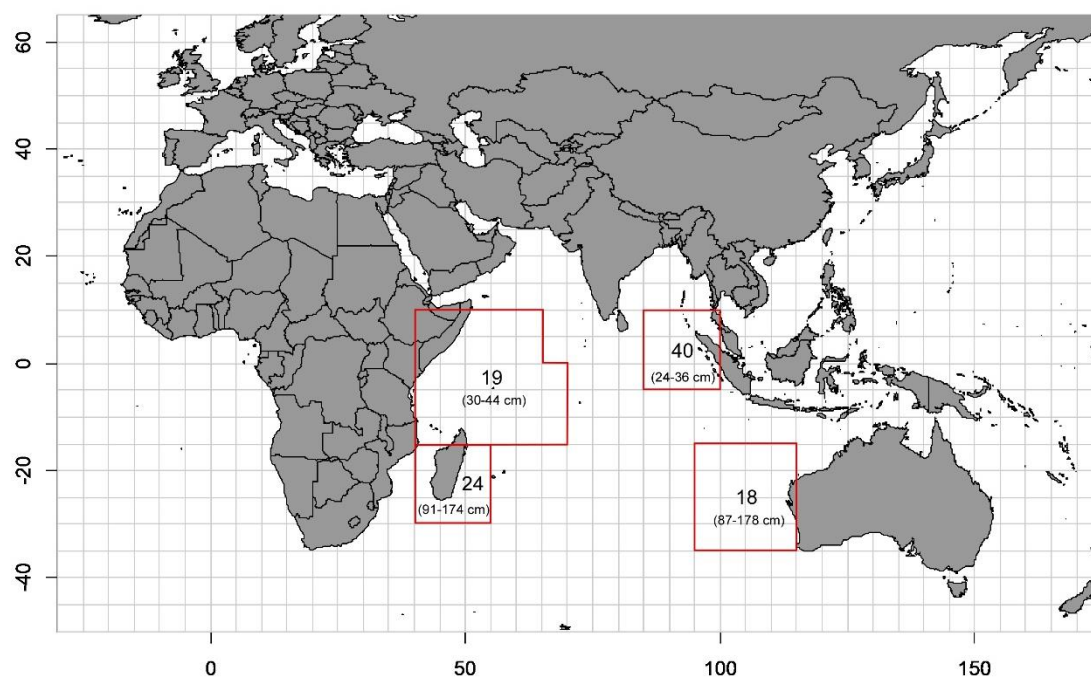
### Methods

The bigeye otolith samples analysed are from four sampling locations (Append Figure 62), referred to as Western Central Indian Ocean (WCI), North-East Indian Ocean (NEI), South-West Indian Ocean (SWI) and South-East Indian Ocean (SEI) (see Figure 2). In this section, we will refer to WCI and NEI as the two northern locations, and SWI and SEI as the two southern locations.

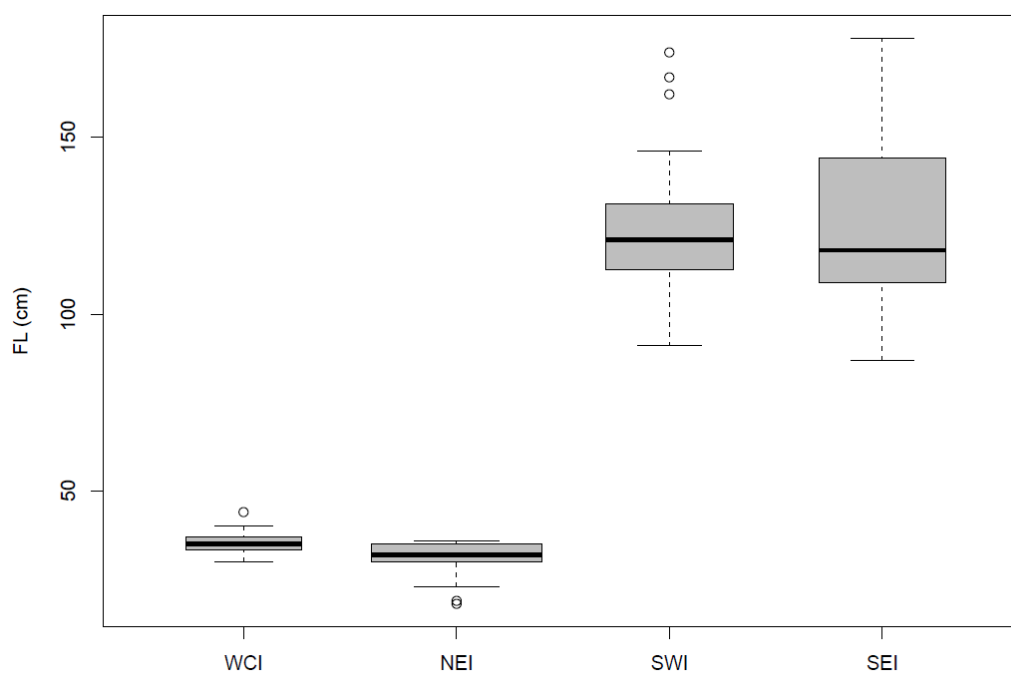
Although it was not intended in the original design, samples from different locations were collected during different periods, due to several reasons (see Section 2). All fish from the two northern sites were juveniles (24-44 cm FL) (Zudaire et al. 2016) while those from the southern sites were larger (87-178 cm FL) (Append Figure 62, Append Figure 63, Append Table 23). The two northern locations are known spawning sites for bigeye (Nishikawa 1985, Stequert and Marsac 1989, Suman et al. 2015), so the aim of the sampling design was to obtain a spawning ground signature for these locations, then see whether the core signatures from the adults sampled at the southern locations corresponded to either of these spawning sites.

The otoliths were analysed at the Centre for Ore Deposits and Earth Sciences (CODES) at the University of Tasmania using LA-ICP-MS. The laser ablated 30 micron spots at 4 positions along the otolith from the core (earliest-deposited material) to the edge (the most recently-deposited material). Thirteen chemical elements were measured (see Section 4.5.2).

The spot near the core was examined to identify the chemical signatures deposited during the first weeks of life, which are most likely to reflect the fish spawning origins. However it was useful to consider signatures from the otolith edge, since these data reflect the fish's known capture location, and can be used for validation purposes.



Append Figure 62. Map showing the number of bigeye otoliths analysed for each of four sampling locations, referred to as Western Central Indian Ocean (WCI), North-East Indian Ocean (NEI), South-West Indian Ocean (SWI) and South-East Indian Ocean (SEI); and the size range of fish at each location.



Append Figure 63. Boxplots of bigeye fork length (FL, cm) by sampling location, including only fish whose otoliths were selected for analysis.



**Append Table 23. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.**

Location	N	Sampling dates	FL (cm)	**Estimated age range (years)
Western Central Indian Ocean (WCI)	19	February-April 2018 (primarily)	30-44	0+
North-East Indian Ocean (NEI)	40	April and November 2018	24-36	0+
South-West Indian Ocean (SWI)	24	August-October 2017 (primarily)	91-174	2-15
South-East Indian Ocean (SEI)	18	May 2019	87-178	2-15

\*\* based on results from Eveson et al. 2015, Farley et al. 2006, Sardenne et al. 2015.

Analyses that were performed on the core or edge data:

- Univariate tests for each element to test for differences among locations. The exact test used depended on whether the data were normally distributed (assessed using shapiro.test in R) and had equal variances between locations (assesses using fligner.test in R), and whether there were two or more locations.
- PERMANOVA (non-parametric version of MANOVA): used to test for differences in the multi-elemental signatures of fish among locations and/or cluster groupings (results obtained using the adonis function from the vegan package in R).
- PCA: used to help visualize the data as this can be difficult with so many elements, and to determine which elements account for most of the variability in the data (results obtained using the dudi.pca function from the ade4 package in R).
- Clustering: in some circumstances, used for identifying the most likely number of separate spawning origins, and to investigate whether spawning origins differ among fish from different sampling locations (results obtained using the hclust function in R with method="ward.D2", and the dissimilarity matrix calculated using Euclidean distance).

Note that prior to any multivariate analyses, the data were standardised (i.e., for each element, the data was centred by subtracting the mean and scaled by dividing by the standard deviation).

## Core Results

Core and edge signatures were significantly different for most elements (Append Figure 64). This is clear in the otolith data for the northern locations, where even though fish are estimated to be only about 3 months old and assumed not to have changed locations between spawning and capture, their core and edge signatures are still significantly different.

Boxplots comparing core signatures between locations show that many elements are similar between locations; however, for those elements that show clear differences (e.g. 39K, 85Rb), the two northern locations (NEI and WCI) appear similar to each other but different than the two southern locations (SEI and SWI) (Append Figure 65).

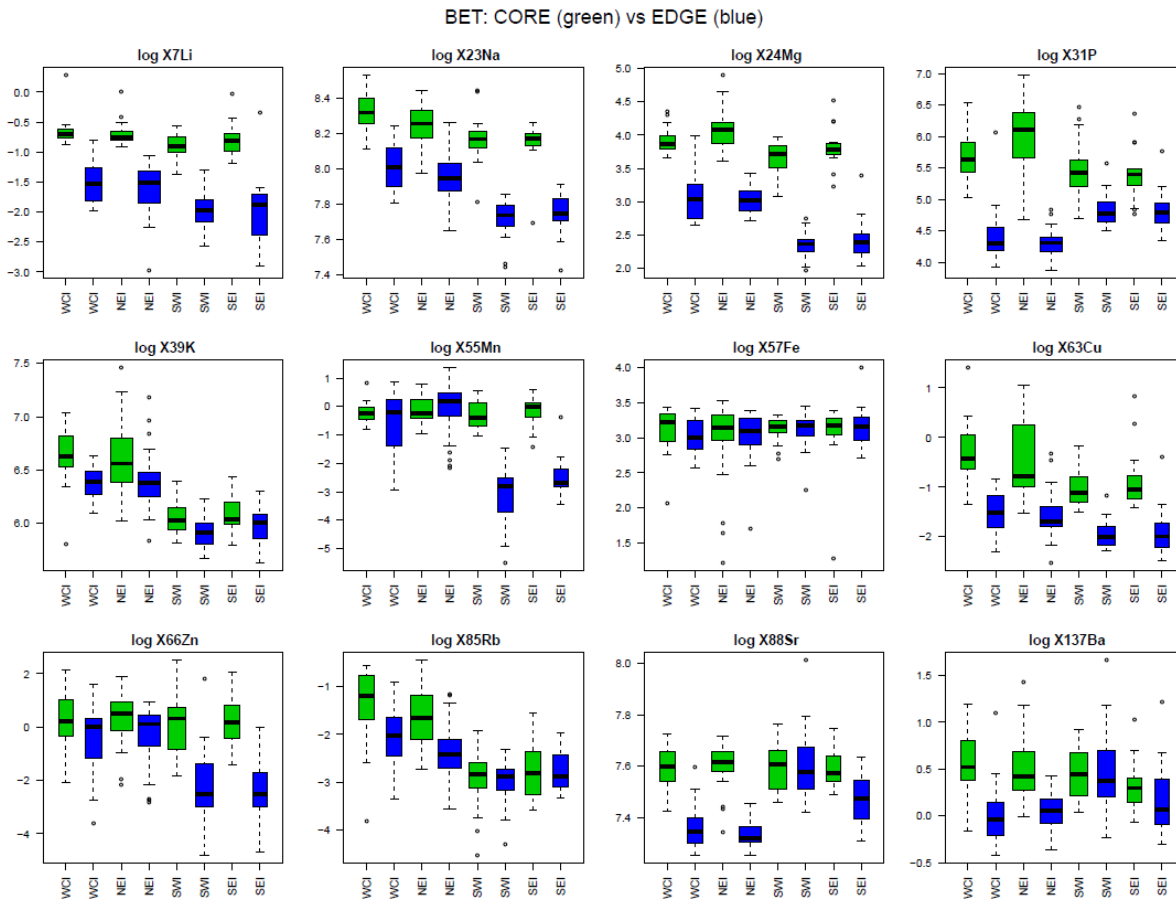
Based on the PERMANOVA, core signatures are not equal among all locations ( $p=0.001$ ).

Based on the subsequent pairwise tests between locations:

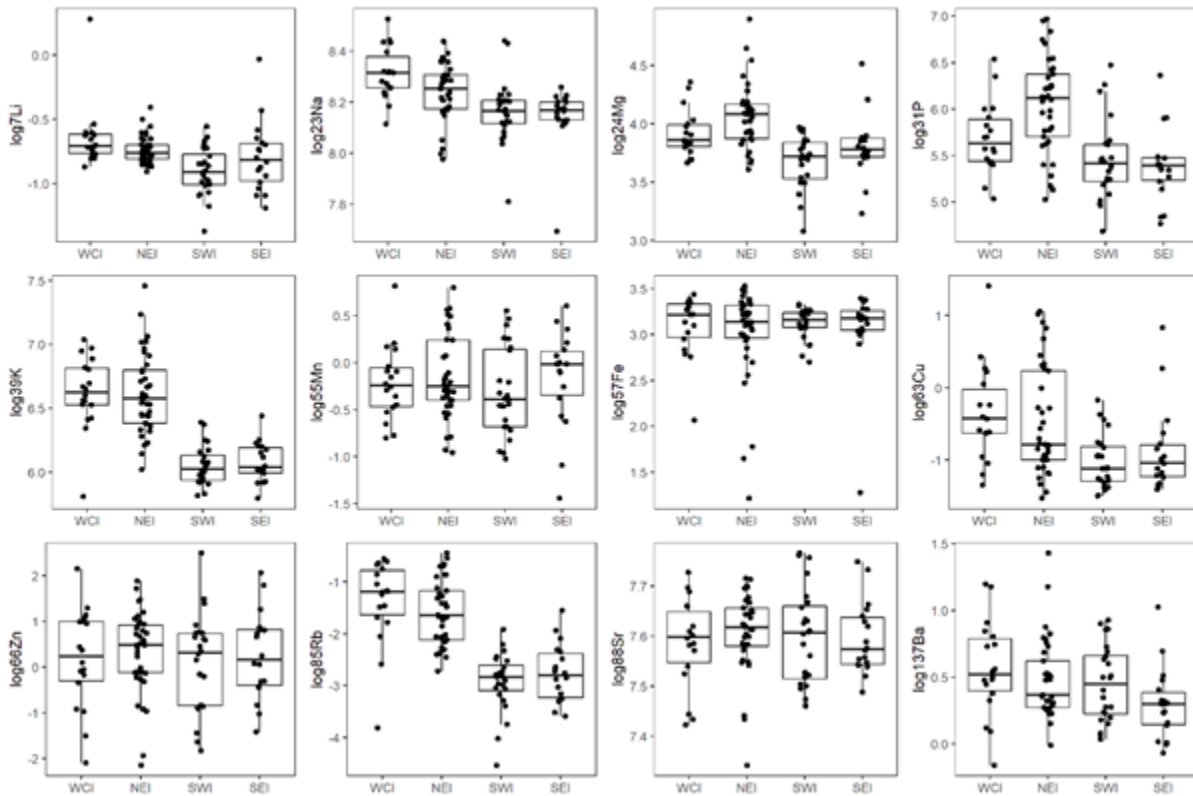
- Core signatures do not differ between the two northern (spawning) locations (Append Table 24);

- Core signatures do not differ between the two southern locations (Append Table 24);
- However, core signatures do differ between the northern locations and the southern locations (Append Table 24).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data confirms and helps to visualize these findings (Append Figure 66).



Append Figure 64. Boxplots comparing bigeye core (green) and edge (blue) signatures at the four sampling locations.

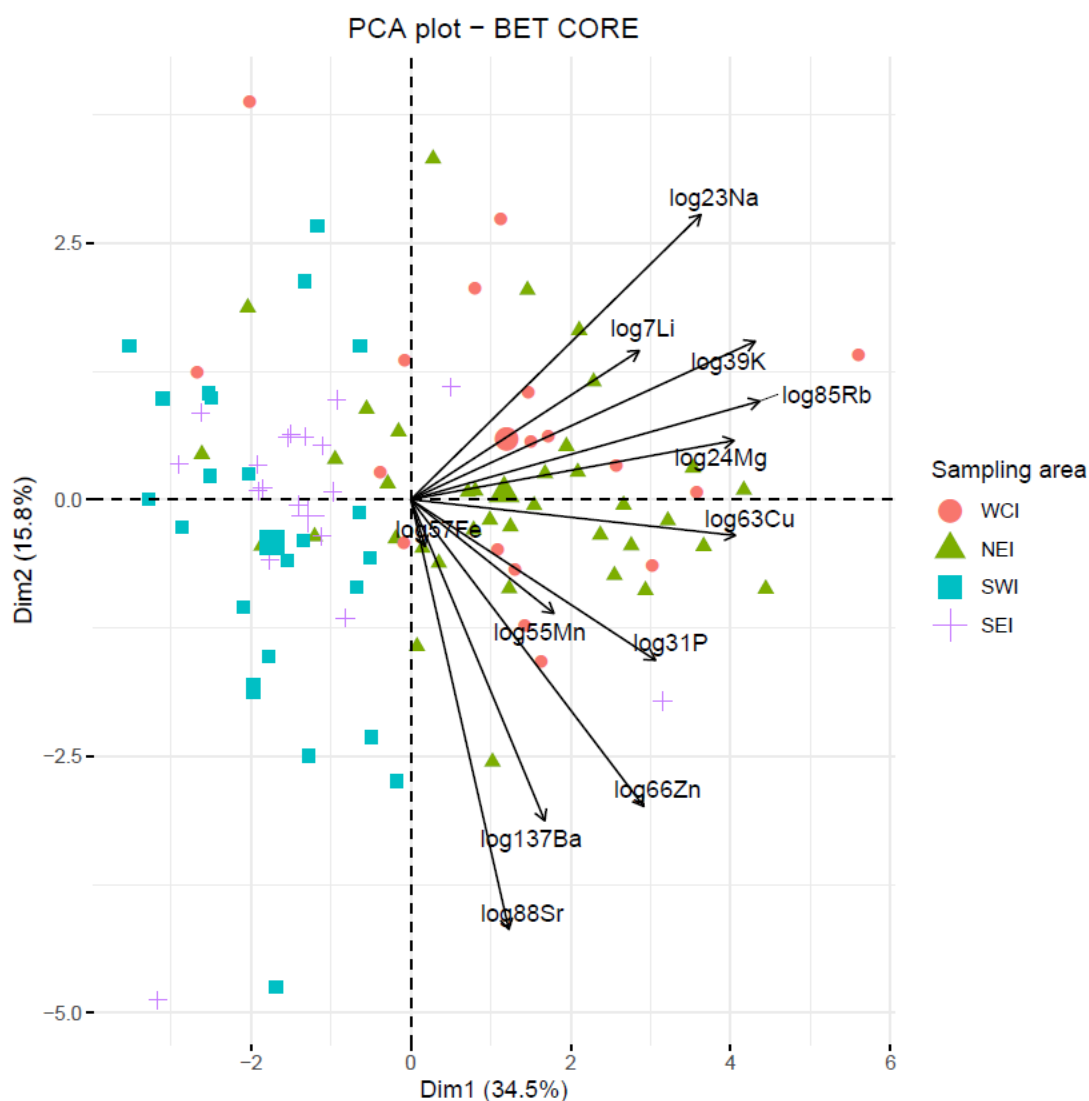


Append Figure 65. Boxplots comparing bigeye otolith core data between locations for each element. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 24. Results of pairwise comparisons of bigeye otolith core signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI	vs	NEI	1	18.729	1.793	0.033	0.106	0.127	
WCI	vs	SEI	1	76.059	7.287	0.177	0.001	0.002	*
WCI	vs	SWI	1	104.763	11.260	0.220	0.001	0.002	*
NEI	vs	SEI	1	90.234	8.654	0.140	0.001	0.002	*
NEI	vs	SWI	1	138.617	14.349	0.196	0.001	0.002	*
SEI	vs	SWI	1	10.700	1.153	0.028	0.324	0.324	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 66. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the bigeye otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Edge results

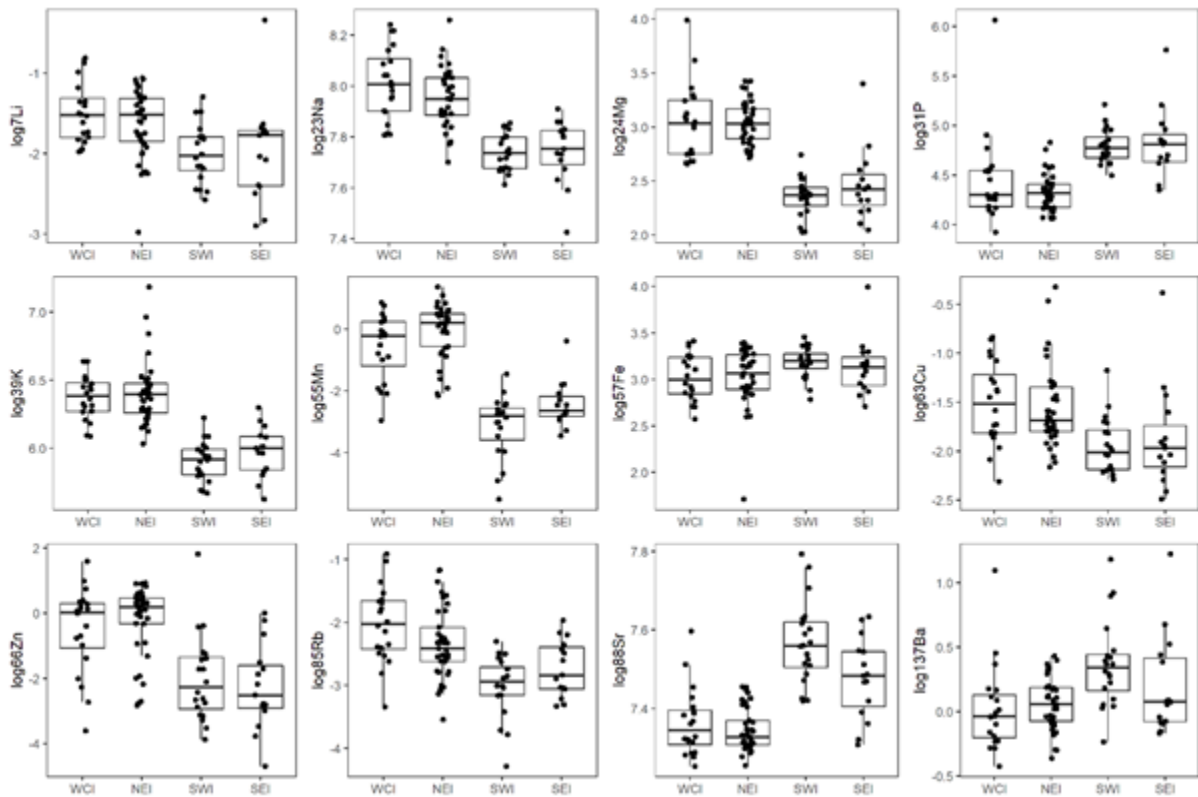
Boxplots comparing edge signatures between locations for each element show that, in general, the data for the two northern locations (NEI and WCI) look similar, and likewise for the two southern locations (SEI and SWI) (Append Figure 67).

Based on the PERMANOVA, edge signatures are not equal among all locations ( $p=0.001$ ).

- Based on the subsequent pairwise tests between locations:
- Edge signatures do not differ between the two northern (spawning) locations (Append Table 25);
- Edge signatures do not differ between the two southern locations (Append Table 25);

However, edge signatures do differ between the northern spawning locations and the southern locations (Append Table 25).

A biplot showing individuals projected onto the first plane (i.e., the first two axes) of a PCA run on the core data confirms and helps to visualize these findings (Append Figure 68).

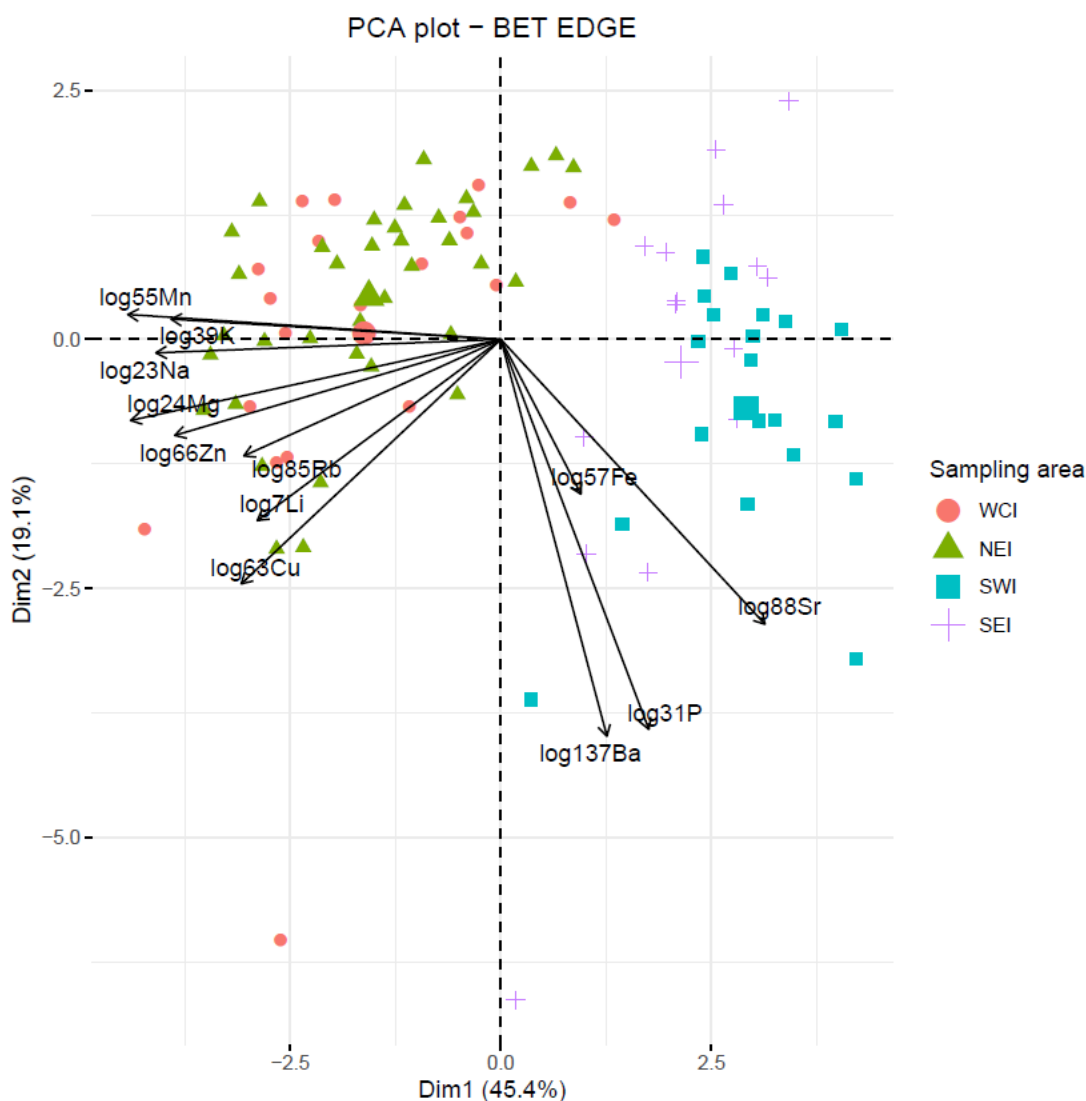


Append Figure 67. Boxplots comparing edge data for each element between locations. Note that the data have been log-transformed to reduce skewness and facilitate comparison.

Append Table 25. Results of pairwise comparisons of multi-elemental edge signatures between locations. In the Significance column, a blank means that the locations do not differ significantly at level 0.05, a dot (.) means they differ at level 0.05, and a star (\*) means they differ at level 0.01.

Pair			Df	Sum-of-Squares	F-statistic	R-squared	P-value	Adjusted P-value	Significance
WCI	vs	NEI	1	10.174	1.323	0.023	0.222	0.222	
WCI	vs	SEI	1	123.875	12.518	0.275	0.001	0.002	*
WCI	vs	SWI	1	213.857	27.962	0.424	0.001	0.002	*
NEI	vs	SEI	1	155.281	19.843	0.284	0.001	0.002	*
NEI	vs	SWI	1	276.356	42.776	0.437	0.001	0.002	*
SEI	vs	SWI	1	12.359	1.575	0.046	0.140	0.168	

Results were obtained using the pairwise.adonis function in R with Euclidean distance to calculate the similarity matrix and the Benjamini and Hochberg (BH) method for calculating the adjusted p-value).



Append Figure 68. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the bigeye otolith edge signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Summary

Because the size/age range of adult fish analysed was wide, meaning fish were spawned in a wide range of years, the results obtained must be interpreted with caution because differences in otolith core signatures among locations might partly be due to cohort effects, i.e. differences among locations might result more from inter-annual shifts in oceanic water masses composition at a given place in the IO than from different geographical spawning origins.

It is well known that how elements are incorporated into otoliths is influenced by age and life stage; this is clear in the otolith data for the northern locations, where even though fish are estimated to be only about 3 months old and assumed not to have changed locations between spawning and capture, their core and edge signatures were still significantly different.

For young-of-year (YOY) fish caught in the two northern locations (WCI and NEI), we assume they have not moved far from where they were spawned, meaning their capture location equals their spawning location. For adult fish caught in the two southern locations (SWI and SEI), we do not know

their spawning location, but we may be able to infer it by comparing their core signatures with the core signatures for the northern spawning locations.

The fact that the core signatures do not differ between the two northern spawning locations suggests either:

- the ocean chemistry does not differ significantly between these locations, or
- fish were spawned in a single location and their larvae were transported within the first few days of life to separate locations.

We were able to investigate the second hypothesis by looking at whether the edge signatures differed between fish from the two locations. We found that they did not (see next section on edge results), suggesting that the ocean chemistry is indeed similar. Unfortunately, this means that these data are not useful for distinguishing which of the two northern spawning locations adult fish sampled in the south originated from. The fact that the core signatures do not differ between the two southern locations suggests that fish from these regions were spawned in waters with similar ocean chemistry.

Interestingly, however, the core signatures for the southern locations differ significantly to the core signatures of the northern spawning locations, indicating either: (i) the fish from the south were not spawned in either of the northern locations; or (ii) they were in fact spawned in one of the northern locations but the ocean chemistry was very different in the years they were spawned (estimated to cover a wide range from the mid-2000s to the mid-2010s), than in the years that the fish from the northern locations were spawned (2017-2018). If the latter hypothesis were true, we might have expected the core signatures of the southern fish to have much greater variability than the northern fish but still to overlap significantly, which is not the case (Append Figure 64). However, we cannot rule out that the ocean chemistry changed significantly between the two periods.

Edge data should be representative of capture locations. Thus, if we find differences in edge signatures between locations, this verifies that otolith chemistry can be useful for classifying fish to these locations. If we don't, this suggests that the ocean chemistry does not differ significantly between locations, and otolith chemistry is not useful for classifying fish to these locations. Note, one caveat is that if fish from the different locations are of very different ages (e.g. YOY vs adult), then ontogenetic influences on their otolith chemistry could blur any existing differences between locations.

The fact that the edge signatures do not differ between the two northern locations, for which all fish are YOY, suggests the ocean chemistry does not differ significantly between these locations. Similarly, the fact that the edge signatures do not differ between the two southern locations, for which all fish are adults, suggests the ocean chemistry does not differ significantly between these locations. The edge signatures do differ significantly between the northern and southern locations; however, this could in part be due to ontogenetic effects.

## 8.3 Temperate tuna

### 8.3.1 Albacore (*Thunnus alalunga*) - population genetics

We have investigated the population genetic structure of albacore tuna (*Thunnus alalunga*) with a focus on population structure and connectivity with the Indian Ocean. For this purpose, 288 individuals caught in the Northeast and Southeast Atlantic, Southwest Pacific and Indian Ocean were sampled, measured, sexed, and genotyped using Diversity Arrays Technology on 103,676 SNPs. Fork length of individuals ranged from 44 cm to 116 cm with an arithmetic mean at 84 cm and a median at 89 cm (CI 74.0-99.8 cm). The smallest individuals were sampled in the Southwest Pacific and the largest in the Southwest Indian Ocean. The females have a mean size smaller than males. Weight ranged from 16.3 to 33.5 kg, with a mean at 23.2 kg and a median at 23.4 kg (CI 20.7-26.2 kg).

The SNPs varied in quality and we retained only the most informative SNPs of the highest quality. Dataset with and without outliers were used. A total of 224 individuals were hence analyzed by genotyping 20,220 SNPs on the filtration selected to classical structure analyzed. Three genetic groups were identified according to the oceans (Atlantic, Indian, and Pacific Oceans). Sampling areas in Indian Ocean (southwest and Indonesia) were genetically undifferentiated, as well as those in the Southwest Pacific (Tasmania and Australia). However, Northeast and Southeast Atlantic (South Africa) were differentiated when including outliers in the data, but this differentiation is unclear when using other methods. Therefore, uncertainty remains as to whether the samples from both North Atlantic and South Africa are differentiated or not. Moreover, our analyses indicate that the genetic assignment in the North Atlantic is insufficient to establish a group signature in this area.

Heterozygosity was low (0.11-0.12 Hobs) with significant deficiency of He in Southwest Indian Ocean and Southwest Pacific.  $F_{st}$  values were also low between main geographic areas, as is to be expected for large populations, ranging from 0.003 to 0.020 with supposed neutral loci (20,038 SNPs) and 0.004 to 0.030 with all loci (20,220 SNPs). Significant differences by  $F_{st}$  were detected between oceans: Northeast Atlantic, South Africa, Indian Ocean, and Southwest Pacific. However, the difference between the Northeast Atlantic and South Africa comes from only 183 loci supposedly under selection.

We inferred evolutionary history under different scenarios to refine our understanding on the separate genetic clusters identified. To do this, we tested different scenarios of divergence among pairwise comparisons of genetic clusters, namely North Atlantic vs. South East Atlantic; North Atlantic vs. Pacific Ocean; South East Atlantic vs. Indian Ocean; South East Atlantic vs. Pacific Ocean and Indian Ocean vs. Pacific Ocean. In all comparisons we found that a secondary (SC2m) contact was the best model to describe the observed genetic data. Moreover, the duration of allopatric phase corresponds in all cases to the impact of the last glaciation event. These results suggest that each sub-clustering observed in DAPC correspond to independent glacial lineages, despite the relatively importance of migrant per generation ( $N_m$ ). The evolutionary history results support the existence of four distinct genetic clusters reflecting independent evolution during the last glaciation. We therefore recommend a stock assessment by oceanic regions (North Atlantic, South Atlantic, Indian Ocean, South Pacific). South Africa (the only samples from the South Atlantic) showed high levels of divergent loci and we discuss necessary investigation of sampling in South Atlantic.



## Methods

### *Samples*

A total number of 288 individuals of albacore tuna was sampled (Figure 41). Sampling consisted of taking pieces of tissue to genotype SNPs along with individual length (cm), sex, geographic location (latitude and longitude) and vessel characteristics and gear. The low number (47) of individuals measured in curved fork length (CFL) were transformed in straight fork length (FL) based on the equation  $FL\text{ (cm)} = 0.9529779 \times CFL - 0.11566164$  (Bonhommeau et al. 2019). For albacore, FL versus CFL does not exist, so this equation concerns tuna (based on albacore, bigeye, and yellowfin tunas in Indian Ocean).

The spatial distribution of samples was mapped using ArcGIS software ([www.arcgis.com](http://www.arcgis.com)). The adults and mature (> 90 cm FL) are mainly caught in the Southwest Indian Ocean. The proportion of male and female is indicated only for the SW Indian Ocean, as sex was not provided at the other sampling sites.

Samples came mainly from longline (LL) catch (Pacific and Indian Ocean) with a total of 157 samples, then from pole and Line (PL) with 68 samples from Atlantic and South Africa, handline (HL) with 57 samples from Pacific and only 6 samples by purse seine (PS) from Indonesia (East Indian Ocean). Regarding the distribution of body length size (cm), the mean length size of individuals caught by LL and PS is larger than individuals caught by HL and PL. Therefore, based on estimates of size at first maturity (50%) at around 85-90 cm FL for albacore (Bard 1981; Farley et al. 2014; Dhurmeea et al. 2016), corresponding to an age of 4-5 years, individuals caught with PL and HL gears are considered as immatures, whereas individuals caught with LL and PS gears are considered as adults.

### *DNA Sequencing process*

Genomic DNA was extracted from the 288 tissue samples and processed for reduced representation library construction, sequenced, and genotyped by Diversity Arrays Technology using the DArTseq™ technique (DarT Pty Ltd). Genome complexity reduction was achieved with a double restriction digest using a *Pst*I and *Sph*I methylation-sensitive restriction enzyme combination and various adaptors (include a barcode to allow disaggregation) added to the terminal sequence fragments to allow Illumina short-read sequencing to proceed. The fragments of DNA selected by this process, Illumina HiSeq 2500 platform, are around 75 bp. More detail on the method can be found in Sansaloni et al. (2011), Kilian et al. (2012), and Georges et al. (2018). For initial assessment of read quality and sequence representation, raw reads obtained were processed using Illumina CASAVA v.1.8.2 software. Then, DArTtoolbox performed filtering and variant calling, and generated final genotypes (Kilian et al. 2012). More details in the sequences process to generate SNP genotyping are in Georges et al. (2018).

The resultant dataset contained SNP genotypes and associated metadata (ex. repAvg, avgPIC, etc.) relevant for our analysis. Because it is extremely rare for a mutation to occur twice at the same site in the genome, the SNP data is biallelic. The one-row data is presented with a zero (0) score to denote a homozygote for the reference allele, one (1) score to denote a homozygote for the SNP allele, and a two (2) a heterozygote (i.e. both SNP and reference alleles are present). A missing genotype (i.e. no allele has been called) is indicated by a dash (-).

### *Additional SNP genotyping process*

The raw data contained 103,676 SNPs for 288 individuals. These SNPs varied in quality, and we retained only the most informative SNPs of the highest quality for our analyses. We created a genotype file from DarTseq one raw file for use in population genetic analyses. We extracted and filtered the data to create a file formatted for use in different software packages. A function of Radiator package developed by Thierry Gosselin (`filter_rad`) was used (Gosselin et al. 2020). This function, designed for RADseq and also DarTseq data in the latest version of the Radiator package, can be found at [https://thierygosselin.github.io/radiator/articles/get\\_started.html](https://thierygosselin.github.io/radiator/articles/get_started.html). Detailed parameters of filtration for the dataset were provided in Appendix Table 26. The final dataset contained 20,220 SNPs for 224 individuals.

Methods for the detection of loci under selection, such as `fdist` and `BayeScan`, have been shown to suffer from large false positive rates as a result of violations of assumption of independence i.e. it does not allow for the non-random correlations among pairs of populations that cause the non-independence of sampling (Meirmans 2012; Bierne et al. 2013; De Mita et al. 2013; Fourcade et al. 2013; Lotterhos and Whitlock 2014). Hence, we searched for candidate loci under selection by testing hypotheses about excess genetic differentiation across the genome, using the `outflank` function provided by Whitlock and Lotterhos (2015).

Analyses of genetic polymorphism offer the opportunity to distinguish locus-specific effects from genome-wide effects at many loci. Identifying and using presumably neutral regions and loci of the genome, that are assumed to be influenced only by genome-wide effects and exclude the regions presumed to be under selection, is therefore essential to reliably infer population demography (Vitalis 2020) and the evolutionary history. In our study, we carried out population differentiation analyses by removing markers showing a deviation from neutral expectation. However, we also explored analyses that included these supposed non-neutral loci in the data and also based on the non-neutral loci markers alone to better understand the impact on population differentiation analyses when the indices are low ( $F_{st}$ ).

**Append Table 26. Radiator filtering steps for the blue shark *Prionace glauca*, including threshold values and the number of individuals, locus and markers after each step (the raw dataset consisted in 20,220 SNPs on 224 samples analysed). Last lines also detail the number of SNPs removed after radiator filters because detected as outliers by applying approach OutFLANK, and removed from the dataset before further analysis.**

PARAMETER S	VALUES	NAME FUNCTION	VALUE OF FILTER APPLY	BEFORE			AFTER		
				Individual s	Locus	Marker s	Individual s	Locus	Marker s
DArT	reproducibility	filter.reproducibility	0.959	288	60520	103676	288	57480	95557
monomorphic	markers	filter.monomorphic		288	57480	95557	288	53326	86728
markers in	common	filter.common.markers		288	53326	86728	288	53003	86265
individuals based on missingness (with outlier stats or values)	based	filter.individuals.missing	0.3	288	53003	86265	251	53003	86265
monomorphic	markers	filter.monomorphic		251	53003	86265	251	52716	85615
MAC (Minor Allele Count)		filter.mac	4	251	52716	85615	251	40344	60576
coverage	min/max	filter.coverage	10/175	251	40344	60576	251	30386	47281
genotyping		filter.genotyping	0.1	251	30386	47281	251	21115	31665
SNPs	position on the read	filter.snp.position.read	all	251	21115	31665	251	21115	31665
markers snp	number	filter.snp.number	3	251	21115	31665	251	20233	27731
short	ld	filter.short.ld	mac	251	20233	27731	251	20233	20233
mixed	genomes	ind.heterozygosity.threshold (min/max)	0.102504/0.135774	251	20233	20233	225	20233	20233
monomorphic	markers	filter.monomorphic		225	20233	20233	225	20233	20233
duplicate	genomes	dup.threshold	0.25	225	20233	20233	224	20233	20233
monomorphic	markers	filter.monomorphic		224	20233	20233	224	20233	20233
HWE	hw.pop.threshold	midp.threshold	30.0001	224	20233	20233	224	20220	20220
Radiator filtration:						20,220 SNPS on 264 individuals			
Post radiator: filtering based on outlier detected using OutFLANK						20,038 SNPS on 264 individuals			

### Population genetic analysis

We calculated pairwise  $F_{st}$  values and average pairwise difference with 1,000 bootstraps using Arlequin software (Excoffier et al. 2005) with our final dataset (20,220 SNPs), without outliers detected (20,038 SNPs) and only with the outliers (182 SNPs). Principal Coordinates Analysis (PCoA) on allelic frequencies was performed on dataset with neutrals loci (without 182 outliers).

Supplement hierarchical genetic structure were estimated by the Bayesian individual clustering assignment performed with STRUCTURE 2.3.4 software (Pritchard et al. 2000) with the admixture model and correlated allele frequencies. The number of clusters,  $k$ , was determined by comparing log-likelihood ratios in 5 runs for different values of  $K$  between 1 and 6 (number of main geographic

sampling + 1) with a burn-in period of 10,000 steps followed by 10,000 MCMC replicates. For obtaining optimal K, results were analysed through methods from Evanno et al. (2005) and Puechmaille (2016) (approach adapted on uneven sampling) from Structure Selector (<http://lmme.qdio.ac.cn/StructureSelector/>). Plots for optimal K was performed with CLUMPAK (Kopelman et al. 2015).

Model for clustering individuals were also tested in the R package stockR (Foster et al. 2018). Information about the number of groups that the data support is obtained using two sources. Information criteria (AIC and BIC) and resampling method similar to cross-validation were used. The resampling method gives an empirical indication of performance. We repeatedly resample (25 times in this initial analysis) the genetic data and see how well the groupings match those from the analysis of the full data. The groupings are displayed using probabilities of individual fish to each genetic group are obtained using bootstrap methods (Foster et al. 2018), using 250 resamples in this initial analysis (Append Figure 69).

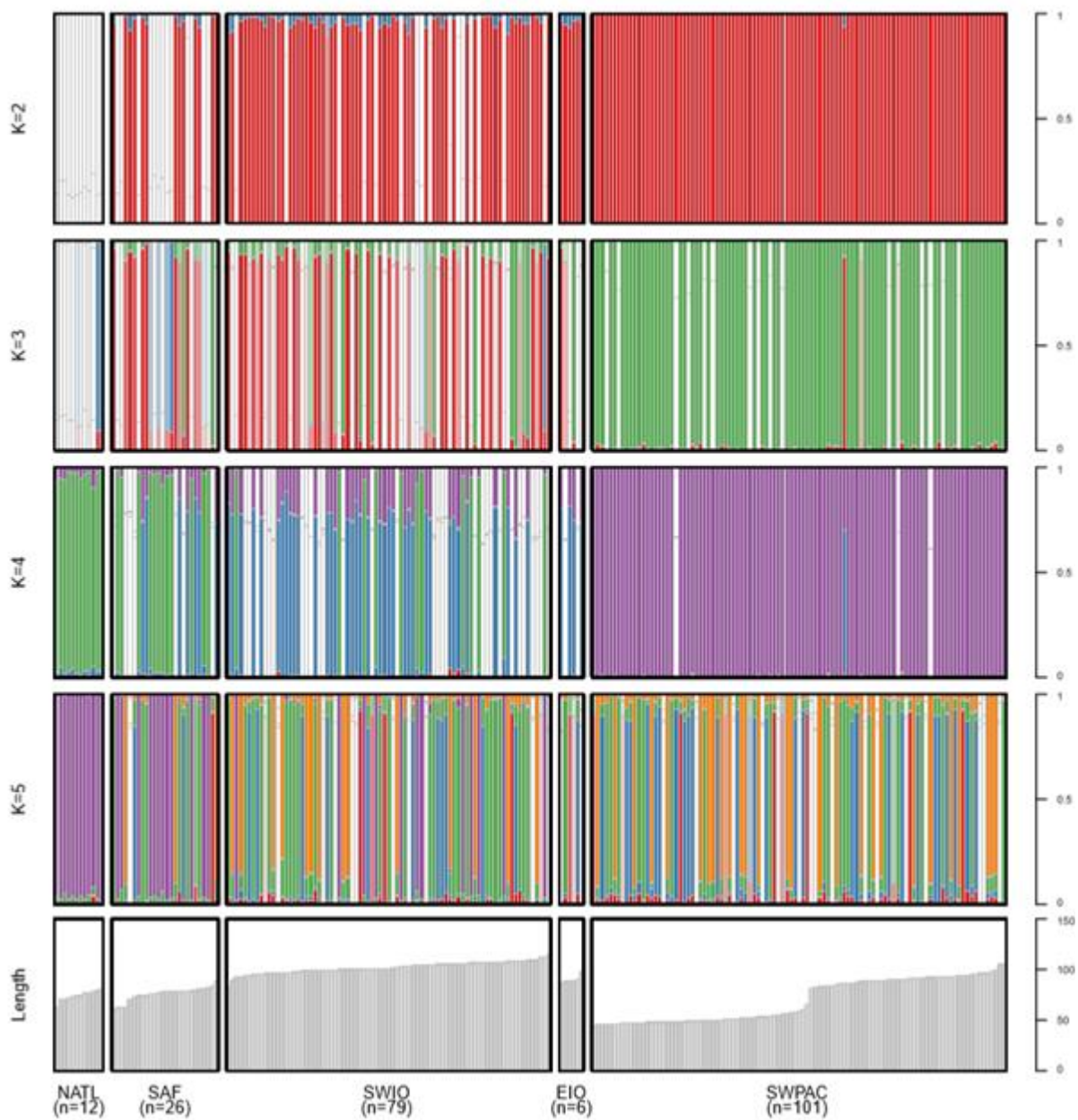
Discriminant Analysis of Principal Components (DAPC) (Pritchard et al. 2000; Jombart et al. 2010; Grünwald and Goss 2011) is a multivariate statistical approach to infer population structure by determining the number of clusters (groups) observed without prior knowledge and without making assumptions of panmixia. In this approach, the variance in the sample is partitioned into a between and within-group component to maximize discrimination between groups. We used adegenet (Jombart 2008; Jombart et al. 2010; Jombart and Amhed 2011) and ade4 (Dray and Dufour 2007; Bougeard and Dray 2018; Chessel D et al. 2004; Dray S et al. 2007) packages to DAPC analysis and plot. The correct number of PCs retained is important to avoid sources of variation. We hence selected the correct number of PCs from a cross-validation (50) in DAPC according the putative origin of individuals and with 1,000 replicates. Based on the retained discriminant functions, it is possible to derive group membership probabilities in order to assess how clear-cut or admixed the clusters are. The trade-off between power of discrimination and over-fitting can be measured by the a-score, which is simply the difference between the proportion of successful reassignment of the analysis (observed discrimination) and values obtained using random groups (random discrimination) (Jombart and Collins 2015).

We used assignPOP R package (Chen et al. 2018) to perform population assignment using a machine-learning framework and employed genetic and non-genetic (fork-length) datasets. Assignments were run with and without the option to remove low variance loci across the dataset. The default of variance threshold is set at 0.95, meaning that a locus will be removed from the dataset if its major allele occurs in over 95% of individuals across the populations. A low variance locus - which has a major allele in most individuals and a minor allele in very few individuals - is unlikely useful because if an allele only occurs in the training or test data, it will not help ascertain population membership of test individuals. We tested the assignment accuracies by removing low variance loci leaving 14,183. The results revealed that this process improved the accuracies for discriminate the populations. When using genetic-morphometric data, the assignment accuracies of populations increased. These results showed the potential of using multiple data types to improve assignment success.

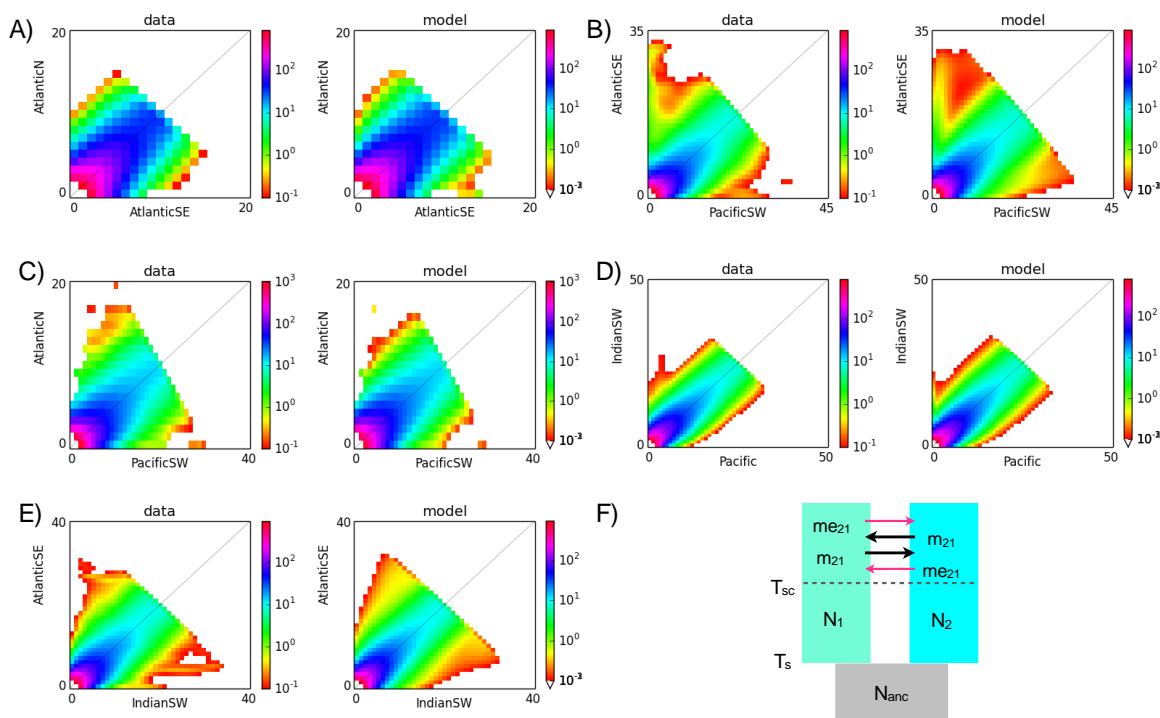
In order to infer the history of the four potential populations (North Atlantic, South Atlantic, Indian Ocean and South Pacific) identified with DAPC, we analyzed their joint allele frequency spectrum (JAfS) using  $\delta a \delta i$  v1.7.0 (Gutenkunst et al. 2009). This program, given an observed JAfS, estimates

the likelihood of an assumed divergence history model for each population and its relevant evolutionary parameters (e.g. population size, duration of isolation, migration rates, etc.). Here, we explored seven models, derived from the four classical models of divergence, capturing important demographic and selective aspects related to populations divergence history. The models implemented in this study were initially built in Rougeux *et al.* (2017). Briefly, all the models were built on a common basis of a simple isolation (SI) model, which corresponds to the split (at split time:  $T_S$ ) of two populations from an ancestral population. Then, we implemented the presence of gene flow occurring at variable duration of divergence between daughter populations - at initial stages only (ancient migration: AM) during  $T_{AM}$ , during the divergence ( $T_{IM}$ ) and therefore allowing contemporary gene flow (isolation with migration: IM) or gene flow after an isolation period (secondary contact: SC) of  $T_{SC}$ . Therefore, we inferred historical divergence history between pairs of populations from North Atlantic, South East Atlantic, Indian Ocean and Pacific. We projected down JAFS from i) Atlantic Ocean and Indian Ocean, ii) Atlantic Ocean and Pacific Ocean, iii) Indian Pacific oceans and iv) Atlantic N and Atlantic SE based on 15,599 genotyped SNPs. We ran each model for 25 optimizations runs in order to check for model convergence, for statistical power purpose and retained the best model based on Akaike Criterion Information (AIC). We defined a conservative threshold to retain models with  $\Delta AIC_i = AIC_i - AIC_{min} < 10$  and computed Akaike weights ( $w_{AIC}$ ) to estimate the relative probability of each model to be the best. We then converted estimated demographic parameters into biologically units in order to compare informative values (e.g. duration and strength of gene flow) among populations, using the same mathematical framework as developed in Rougeux *et al.* (2017).

For each pairwise comparison, seven alternative divergence models of variable complexity were fitted to polymorphism data and compared with each other. This comparative framework enabled us to consider different aspects of gene flow. These include temporal variation in migration rate ( $m$ ) as well as chromosomal variation of the migration rate ( $me$ ) to capture putative selective effects. From the observed JAFS, we observed variable amount of divergence depending on sampling zones (i.e., increasing number of alleles onto orthogonal axis of the diagonal). Namely, JAFS for Atlantic-SE vs. Pacific, Atlantic-N vs. Pacific and Atlantic-SE vs. Indian showed higher levels of divergent loci (Append Figure 70). We performed a model selection based on AIC to penalize model likelihood by the number of parameters to avoid any over fitting. Applying criterion of  $\Delta AIC < 10$  and considering the  $w_{AIC}$ , we retained two models for Atlantic-N/Atlantic-SE (SC2m and SC) and only one model (SC2m) for Atlantic-SE/Pacific-SW, Atlantic-N/Pacific-SW, Indian-SW/Pacific and Atlantic-SE/Pacific as the fittest models to the data centralized their respective inferred parameters. We found asymmetrical effective population size ( $N_e$ ) and gene flow among pairs of comparisons, then relatively similar durations of isolation (allopatric phase) matching with the last glacial event.



Append Figure 69. Individual length frequencies and results of population structure analysis of DArTSeq using StockR for albacore tuna over the 20,038 SNPS assuming 2-5 genetic groups.



**Append Figure 70.** Historical demography of the albacore species pairs. Observed joint allele frequency spectrum (JAFS) obtained by projection of empirical data to 20, 45, 40, 50, 40 diploid individuals for A) Atlantic-N/Atlantic-SE, B) Atlantic-SE/Pacific-SW, C) Atlantic-N/Pacific-SW, D) Indian-SW/Pacific and E) Atlantic-SE/Pacific, respectively. For each JAFS, the color scale indicates the number of SNPs falling in each bin defined by a unique combination of the number of mutations observed in population 1 (Y-axis) compared to population 2 (X-axis). For each comparison, the observed JAFS is described as the data-JAFS while the JAFS of the fittest model is described as model-JAFS. F) Representation of the fittest model for each comparison. Nanc: ancient population size. N: population size.  $m_{12}$ ,  $m_{21}$ : migration rate from the population 1 to the population 2 and vice versa.  $me_{12}$ ,  $me_{21}$ : reduced effective migration rates for loci influenced by selection when semi-permeability was assumed.  $T_{sc}$ : duration of the secondary contact.  $T_s$ : time from the initial split to present.

## Summary

This novel genetic research which incorporated selected SNPs (20,028 SNPs and 224 individuals) highlights the complexity of global population structuring in albacore tuna. The PCA, assignment, and hierarchical Bayesian results revealed a minimum of three genetic groups corresponding to the three oceans: Atlantic, Indian and Pacific Oceans. Alternatively,  $F_{st}$  including potential outliers and Puechmaille analysis on hierarchical clustering indicated 4 potential groups: North Atlantic, South Atlantic (South Africa), Indian Ocean (southwest and Indonesia), and Southwest Pacific (Australia and Tasmania). The low differentiation value (low  $F_{st}$ ), despite significant differentiation between the 4 groups, is not surprising, considering the large inferred effective population sizes ( $N_e$ ).

Differences were present on our k-means clustering (ex. AIC and stockR) and the model-based clustering methods (ex. Structure) because the best number of clusters (K) is only a representation of models and therefore we have to be careful about these results (best K). One aspect in which k-means clustering differs from the other methods is that it does not aim at providing estimates of the amount of admixture of individuals, but as producing a “hard” clustering, where individuals are always assigned to a single population (Stift et al. 2019). A recent study (Stift et al. 2019) compared



the performance of k-means clustering and the model-based clustering methods implemented in STRUCTURE, ADMIXTURE, FASTSTRUCTURE and INSTRUCT. They showed that only STRUCTURE allowed unbiased inference with weak population differentiation. Nevertheless, researchers are cautious in their interpretation of STRUCTURE results (Lawson et al. 2018) and it is important to use several approaches based on different underlying assumption and clustering algorithms, in order to appraise the stability/accuracy of delivered insights.

Inferences of the demo-selective history allow refining the understanding about genetic clusters across Atlantic, Indian and Pacific Oceans. Indeed, we found that the SC2m (model including divergence and secondary contact) model best fitted the different JAFS from our empirical data. According to temporal aspects of isolation and duration of secondary contact in contemporary populations of albacore tuna, the different clusters recognized here may correspond to independent glacial lineages. Therefore, those results would support the existence of four distinct genetic clusters reflecting significant periods of independent evolution (in isolation), possibly during the last glaciation.

Future research should focus on overcoming the lack of sampling in areas with low power of assignment as North Atlantic and that are known spawning areas, such as the southwest Atlantic Ocean. The number of individuals in Indonesia is too low to draw any conclusions (6 individuals) and further sampling is required to infer the status of albacore tuna in this area. The number of individuals genotyped in North Atlantic is also too low to conclude and we encouraged to genotype the 26 supplement samples collected in this project. Moreover, sampling and analyzing distant regions (ex. South Africa) would be needed to gain a holistic understanding of the number and dynamics of albacore stocks. Our results suggest strong interconnectivity between the Southwest Pacific and the Indian Ocean, highlighting the need for further investigation of the Southeast Pacific and the North Pacific.

With regards to the results on South Africa, the distribution and mixture of fish from each stock (Atlantic and Indian Ocean) may be affected by the dynamics of the currents around South Africa and is discussed in the paper of Nikolic et al. (2020). South African individuals were more similar to the samples from the North Atlantic samples than to those from the Indian Ocean. Nevertheless, the results in this study (example with the differentiation pattern obtained using outliers loci) suggested the possibility that individuals from another geographic area could also supply South Africa e.g. Southwest Atlantic cannot be excluded. A clear sampling protocol for future research must be drawn up to address questions that arise from our current study.

### **8.3.2 Albacore (*Thunnus alalunga*) - otolith microchemistry**

The otolith microchemistry of albacore tuna has been analysed to investigate nursery origins and migrations in the North Atlantic Ocean (Fraile et al. 2016), population structure in the eastern North Pacific (Wells et al. 2015) and mixing and movement in South Pacific albacore (McDonald et al. 2013).

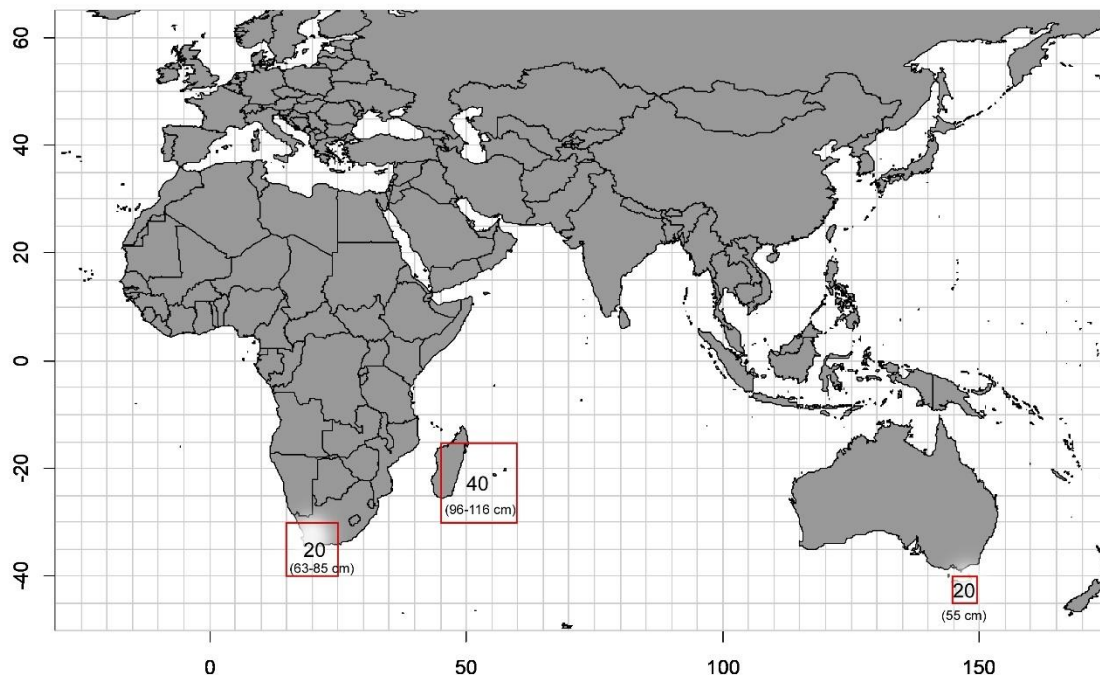
Due to sampling limitations, albacore otoliths for this study were available from only three locations. They spanned multiple age-classes; the South African and Tasman Sea samples (55-85 cm FL) were



juveniles (or sub-adults) and the south west Indian Ocean samples were adults, given the length at 50% maturity for females is 85 cm FL (Dhurmeea et al. 2016).

## Methods

Otoliths from eighty albacore were analysed (55-113 cm FL). Twenty were collected from the south west Tasman Sea, 40 from the south west Indian Ocean and 20 from South Africa in the south west Atlantic Ocean at 2 sites, Hout Bay and Saldana Bay (Append Figure 71). The samples were collected between February 2018 and February 2019 (Append Table 27). The otoliths were grouped by area for analysis.



Append Figure 71. Map showing the number of albacore otoliths analysed for each of the three sampling locations, referred to as south west Indian Ocean (SWI), South Africa (SA) and south west Tasman Sea; and the size range of fish at each location.

Append Table 27. Number, sampling period, size range and estimated ages of fish for each of the sampling locations.

Location	N	Sampling dates	FL (cm)	*Estimated age range (years)
south west Indian Ocean (SWI Feb 18)	12	February 2018	96-104	7-10
south west Indian Ocean (SWI May 18)	8	May 2018	98-113	7-15+
south west Indian Ocean (SWI 2019)	20	December 2018	96-116	7-15+
South Africa – Hout Bay (SAS)	6	March-April 2018	63-85	2-5
South Africa – Saldana Bay (SAN)	14	March-April 2018	63-85	2-5
South west Tasman Sea (SWTS)	20	February 2019	55	1

\* The ranges in ages are for male and females combined (Xu et al. 2014).

The otoliths were analysed at Montpellier University, Plateforme AETE-ISO (France) using LA-ICP-MS. The laser ablated along a transect between the otolith core and the edge, therefore acquiring a chemical signal from material deposited throughout the life of the fish.

The portion of the otolith transect near the core was examined to identify the chemical signatures deposited during the first weeks of life. These are most likely to reflect the fish spawning origins, i.e. the physicochemical characteristics of the water masses in which spawning occurred. The first laser point, on the core, was not included, in order to avoid any maternal influence on otolith composition. The mean of the next 3 points, between 10 and 40 microns after the core and corresponding to the first weeks of life, was used for analysis.

Fifteen chemical elements were measured. Those elements where more than 75% of results were above the limit of detection were retained for further analysis.

## Results

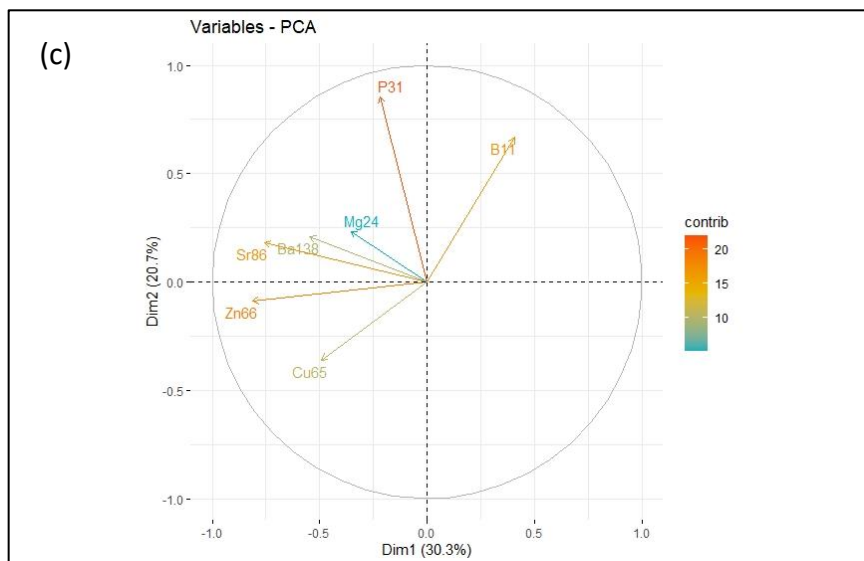
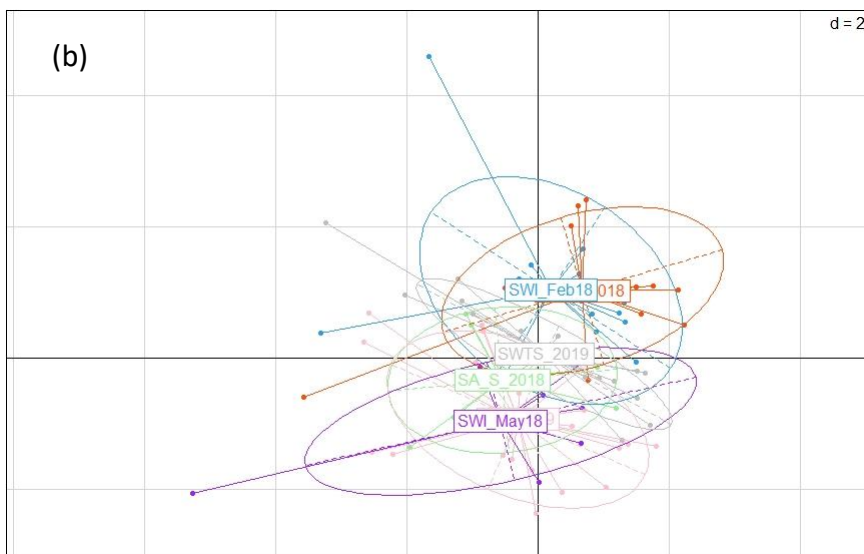
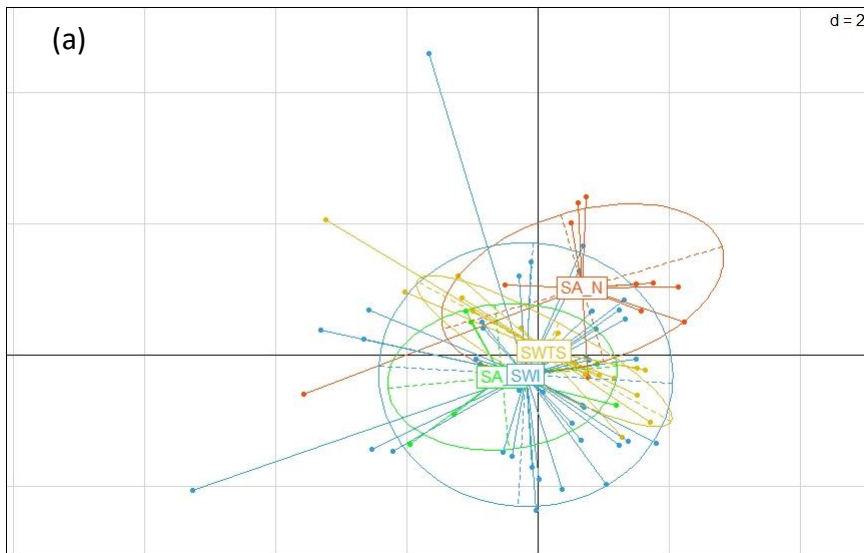
After examination of the levels of detection for the 15 elements analysed by ICPMS, 7 elements were retained for further analysis (B, Mg, P, Cu, Zn, Sr and Ba).

### *Principal component analysis - PCA*

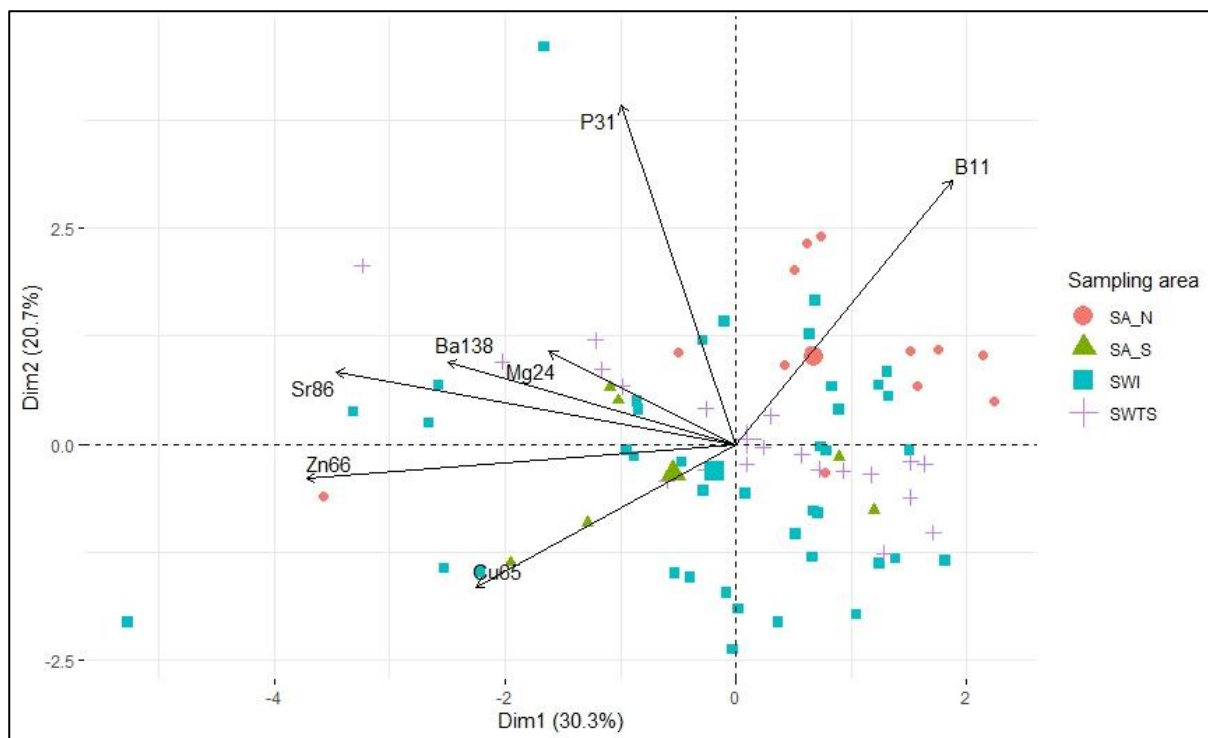
The otolith signatures from all areas largely overlapped but PCA identified some variation in spawning origin according to fish capture location, although this was not significant (PERMANOVA,  $p = 0.057$ ).

Out of the 7 elements above LODs, only P, Zn, Sr, and B significantly contributed to the differences in otolith core signatures among albacore individuals (contrib >20% for dim 1 and/or 2 in the PCA; Append Figure 72 and Append Figure 73). Therefore, only their signatures were used for investigation of fish spawning origin.

Although there is some overlap between SA-North and SA-South they are distinct groups (Append Figure 72a). The core signature of SA-North overlaps that of SWI-Feb 18 and the core signature of SA-South overlaps those from SWI-May18 and SWI 2019 (light pink) (Append Figure 72b).



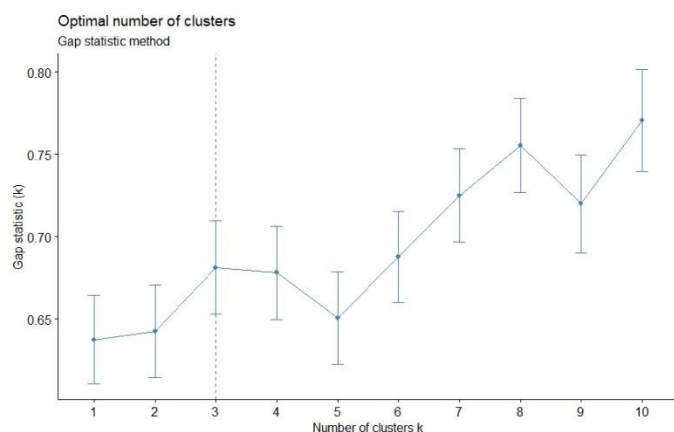
**Append Figure 72.** Graphs of individual (a, b) and variable (c) projections on the first plane of the PCA made with the multi-elemental (B, Mg, P, Cu, Zn, Sr, Ba) signatures of the otolith core of the 80 albacore analysed in this study.



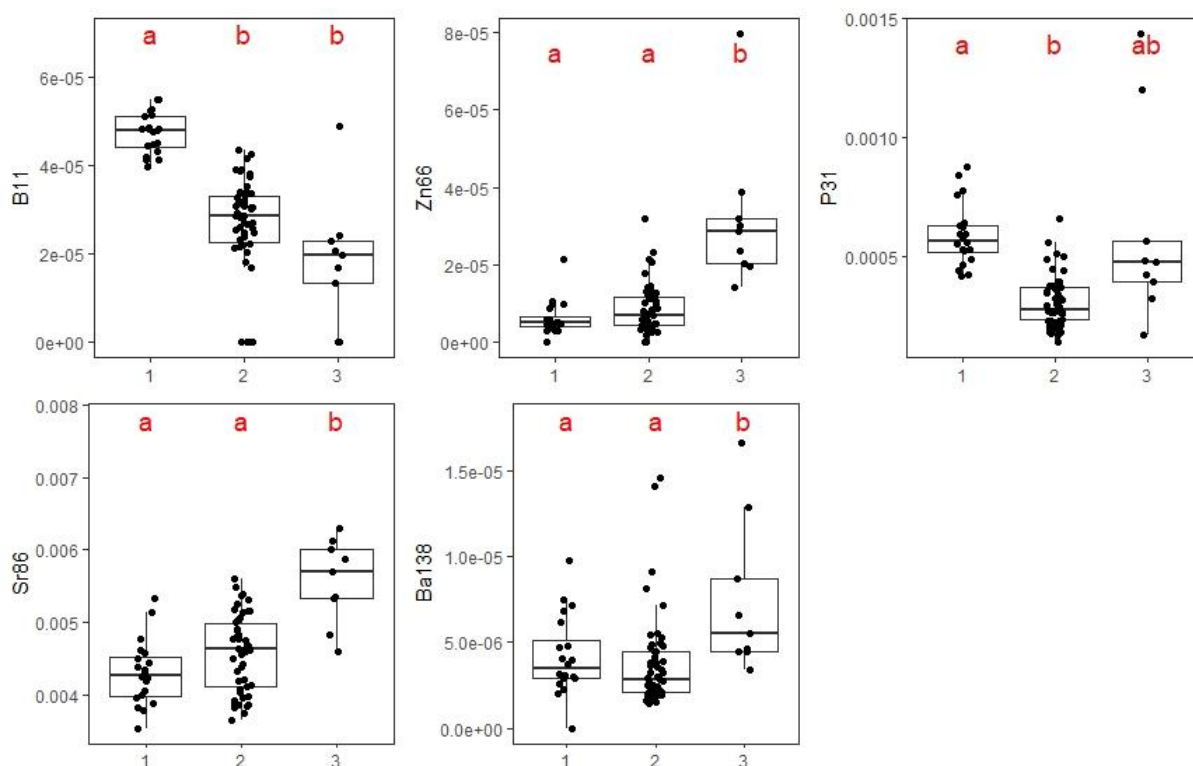
**Append Figure 73.** Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the multi-elemental (B, P, Cu, Zn, Sr, Ba) signatures of the otolith cores of the 80 albacore analysed. Individuals are coded by their final sampling location -- South Africa north (SA-N); South Africa south (SA-S), South West Indian Ocean (SWI) and Southwest Tasman Sea (SWTS). Their size on the graph is proportional to the quality of their representation in this plane. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

### Clustering

Based on the variation in P, Zn, Sr, and B core signatures, the most relevant number of clusters (i.e. distinct spawning origins) was found to be 3 (Append Figure 74). The 3 corresponding putative spawning origins (SpO) had significantly (PERMANOVA,  $p = 0.002$ ) distinct multi-elemental signatures (Append Figure 75). Ba was included in the analyses as it was just near the limit of significance.



**Append Figure 74.** Results of the gap statistic with K means clustering

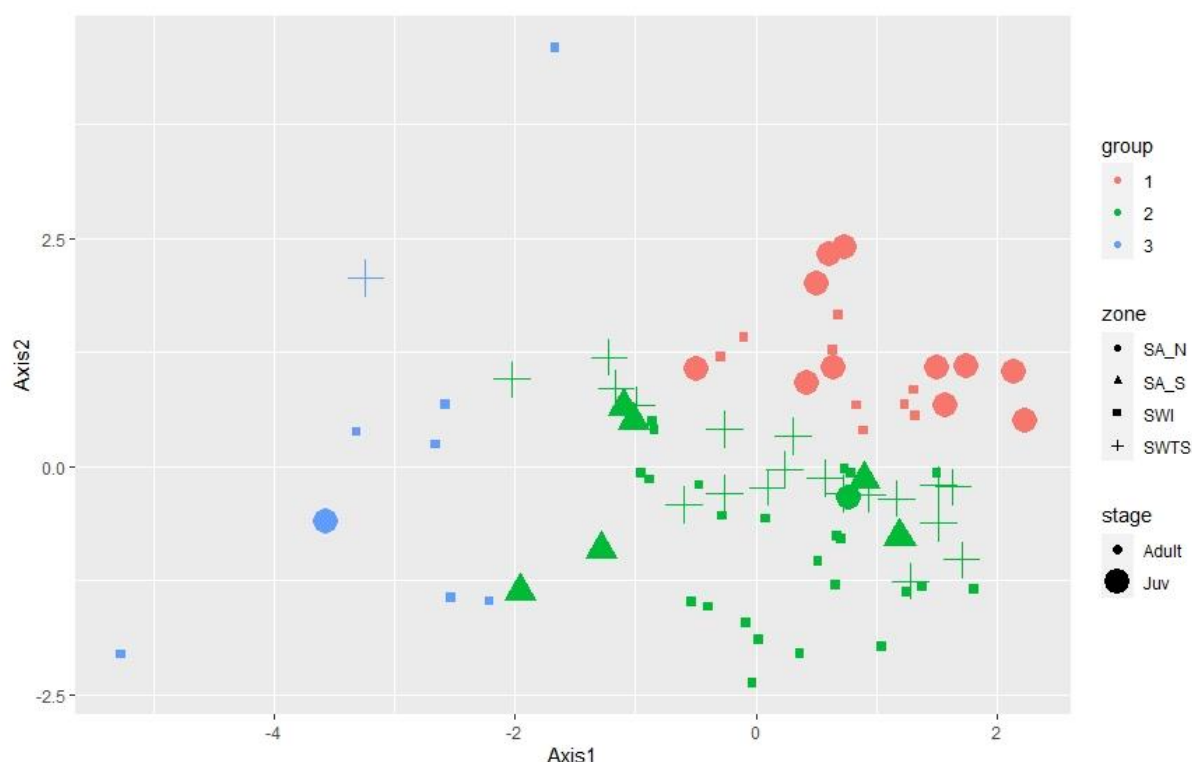


**Append Figure 75.** Projection of the clusters identified as potential discrete spawning origins (SpO 1 to 3) on the first plane of the PCA made with the multi-elemental (B, P, Zn, Sr, Ba) signatures of the otolith core of the 80 albacore analysed and corresponding elemental signatures (black dots) and boxplots (Q1, median, Q3). Letters in red indicate groups with significantly distinct signatures ( $p < 0.05$ ).

In particular, the chemical signature of the otoliths differed significantly among all SpO in terms of B, with maximum values in SpO-1. SpO-3 had significantly higher Zn, Sr and Ba signatures.

All three SpO contributed to the stocks of the 3 areas investigated, but in varied proportions (Append Figure 76) except SpO1 that doesn't contribute to SWI samples. SpO-2 was the main spawning source for the albacore analysed (63%) (Append Table 28) principally in SWTS (95%), SWI (60%) and SA (37%, and all the SA-S samples). SpO-1 provided 25% of the fish sampled, principally in SA (58%, and almost all the SA-N samples) and SWI (22.5%). SpO-3 contributed to only 12% of the total fish analysed.

The differences in fish origin among locations explained the slight variation in otolith core signatures according to sampling location observed in the original PCA.



Append Figure 76. Projection of individuals on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr) signatures of the otolith core of the 80 albacore analysed. Colours on the graph represent the spawning origins (1 to 3) for each fish and symbols represent its final sampling area.

Append Table 28. % of individuals assigned to each spawning origin (SpO 1 to 3) in the albacore analysed (total N of fish tested = 80, FL = 55-116 cm) for each of the 3 sampling locations: South Africa (SA), South-Western (SWI) and Southwest Tasmanian Sea (SWTS).

	SA	SWI	SWTS
SpO-1	58%	22.5%	0%
SpO-2	37%	60%	95%
SpO-3	5%	17.5%	5%
Total	<b>100%</b>	<b>100%</b>	<b>100%</b>

## Summary

Analysis of the multi-elemental signatures of otolith cores identified 3 potential different spawning origins for the 80 juveniles and adult albacore (*T. alalunga*) analysed, captured between February 2018 and December 2018 in South Africa, the south western IO and south western Tasman Sea. All 3 putative spawning origins apparently contribute to each of the stocks fished in the 3 locations sampled, except putative spawning origin 1 (SpO-1) to SWTS samples, but in varying proportions. SpO-1 was a spawning source for albacore from the 2 western IO sampling locations but made no contribution to albacore from the Tasman Sea.

Fish caught in the 2 South African locations had distinct signatures (north and south), the northern signature was found only in adults caught in the SWI in February 2018 and the southern signature was found in adults caught in SWI in May 2018. This could indicate 2 spawning sources for south African fish or two spawning times, with different environmental conditions.

SpO-2 signatures were found in otoliths from all locations except SA North. SpO-2 was the most common signature in albacore from the south west Tasman Sea with 95% of the otolith cores having the SpO-2 signature.

The highest proportion of SpO-3 signatures was seen in otoliths of fish caught in SWI.

The samples collected do not allow us to confirm that these different spawning origins correspond to spatially discrete zones in the Indian Ocean. Ideally, young-of-the-year fish would have been sampled from each location at the same time to minimize temporal variability in otolith signatures. However, the size-range of the fish analysed was 55 to > 116 cm FL and with these ranges in size (and age), the results must be interpreted with caution as differences in otolith core signatures among locations might be due in part to cohort effects, i.e. differences among locations might be the result of inter-annual or seasonal shifts in oceanic water composition at one location rather than swordfish otoliths having different geographical spawning origins.

Before drawing any conclusion about spawning origin in this species and the mixing of its stocks in the IO, these preliminary results must be confirmed by analysing more fish (if possible YOYs from the same cohorts, collected in several successive years -at least 3) from the same sampling areas, but also from others in the Indian Ocean where albacore are present but that could not be sampled for this project. Only this will confirm that the clusters found here do correspond to spatially distinct spawning zones and give a clear pattern of the population structure of the species in this ocean, in combination with genetic data.



## 8.4 Billfish

### 8.4.1 Swordfish (*Xiphias gladius*) - population genetics

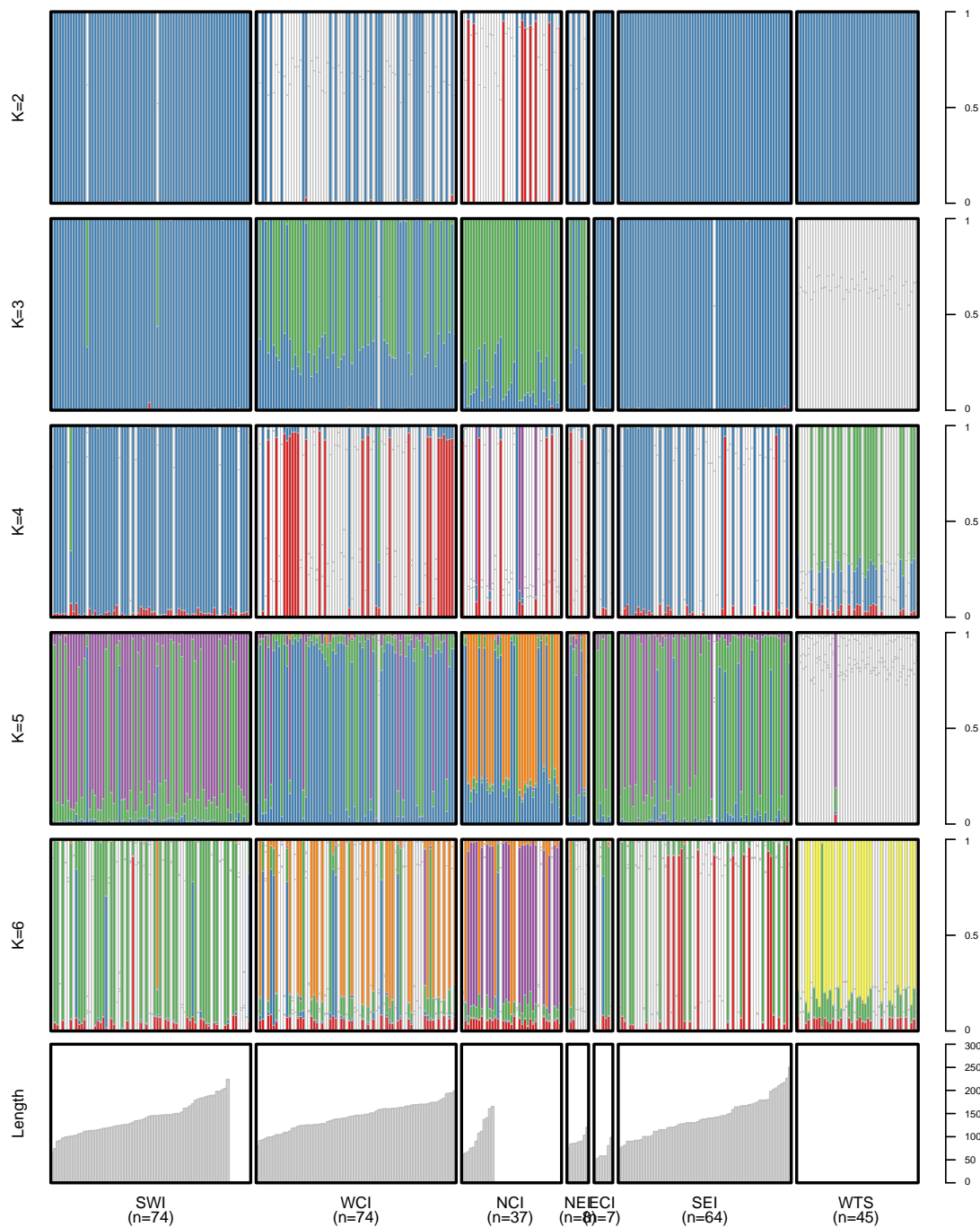
Swordfish (*Xiphias gladius*) is one of the most widely distributed pelagic species found in both tropical and temperate waters of the Atlantic, Indian and Pacific Oceans. In contrast to its broad distribution and occurrence in diverse habitats, tagging data has suggested a high degree of residency and homing behaviour (Sedberry and Loefer, 2001). Expectation of restricted gene flow based on tagging data showing lack of movement across the equator, has been 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). Lu et al., (2006) that have shown differentiation and others that were unable to find population subdivision (Jean et al., 2006; Muths et al., 2009; Muths et al., 2013). 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 the analysis of previous studies to further characterise genetic structure of swordfish population within the Indian Ocean through examination of nuclear DNA variation at SNP markers among fish sampled from broadly distributed locations within the Indian Ocean and south-west Pacific outlier location.

### Results

The population structure of Indo-Pacific 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 46; Append Figure 77). The sizes of sampled fish ranged from sub-adults (80-100cm) to large (>200cm) mature adults (Append Figure 77). Following quality control filtering using Radiator (Gosselin 2019, <https://github.com/thierygosselin/radiator>) the DArT-SEQ SNP sequencing data produced by Diversity Arrays Technology produced genotype information at 15,070 SNP loci (Append Table 29). The genotype data was further assessed using the program StockR to examine geographical distribution for each individual when modelled for K genetic groupings from 1 through 7 (Append Figure 77). Examination of each individual's K probabilities modelled for values  $k > 1$  suggest consideration of a more realistic and value of  $K=2$  or even  $K=3$  based on geographical groupings (Append Figure 77). 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



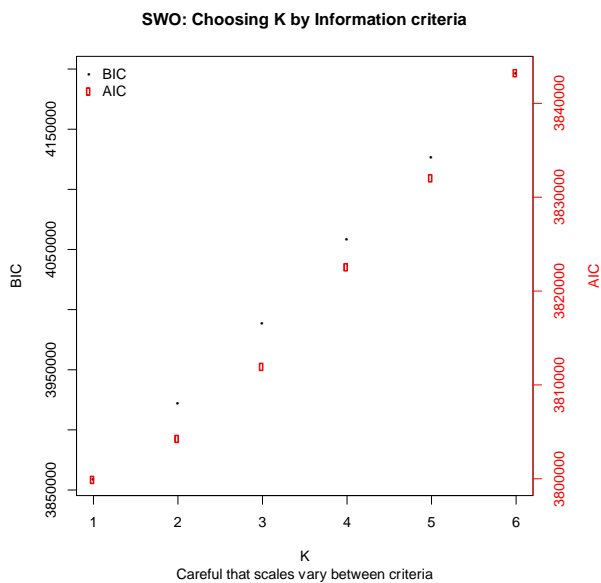
value for K (Append Figure 78). However, when modelled at  $K=2$  and  $k=3$ , there appears to be a strong group of genotypes restricted to samples from the northernmost locations with highest, almost fixed frequency at NCI and present, but at reduced frequency, in the two closest locations, WCI (40%) and NEI (50%). The genotype was rare to absent at the southern sample locations, SWI (3%) and ECI (0%) and SEI (0%), (Append Figure 77). Initial analysis of fish sampled from WTS (western Tasman Sea) failed to be assigned to a single specific K using assignment probabilities greater than 80%. However, a more focussed comparison of the eastern Indian locations (ECI, SEI) and the south-western Pacific (WTS) clearly revealed two distinct groups of swordfish between these sampled regions (Append Figure 79). The separation WTS from Indian Ocean, as well as the close relationship among SWI, SEI, and ECI sample locations was consistent across model results at higher values of K and, in particular at  $K=5$  (Append Figure 77, Append Figure 81). In addition, analysis of  $F_{st}$  indicated restricted gene flow between fish sampled from WTS (Coral Sea, Western Pacific) and those from NCI (Sri Lanka) with NCI also the most consistently divergent from other Indian Ocean sampling locations (Append Figure 79). Interestingly, the southwest Indian and the southeast Indian were the least genetically differentiated samples based on  $F_{st}$  (Append Figure 80), while the SWI and WCI samples included a substantially different mix of genetic groups, indicating they were not sampled from the same genetic distribution. 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. However, further scrutiny of the WCI population demonstrated lack of sex bias contributing to this observation with 56% (28/50) males carrying this genotype. Interestingly, this same genotype is present in the NCI (Sri Lanka) sample at a nearly fixed frequency of 95% (30/32 fish).



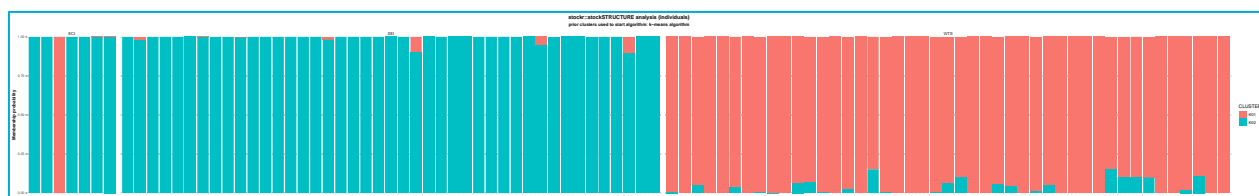
**Append Figure 77. Membership probability barplots for each individual belonging to one of K genetic groupings. For each panel below there are K colours representing K genetic groupings. The probability of each individual is then plotted *proportionally* relative to assignment probability to each of the K groups with white profiles representing individuals where there is less than 95% certainty of belonging to a K genetic group.**

**Append Table 29. Parameters used in the radiator QC filtering of 11 sampling sites (strata) using starting values for radiator of 802 individuals, 39,663 locus fragments, and 54,733 SNP markers.**

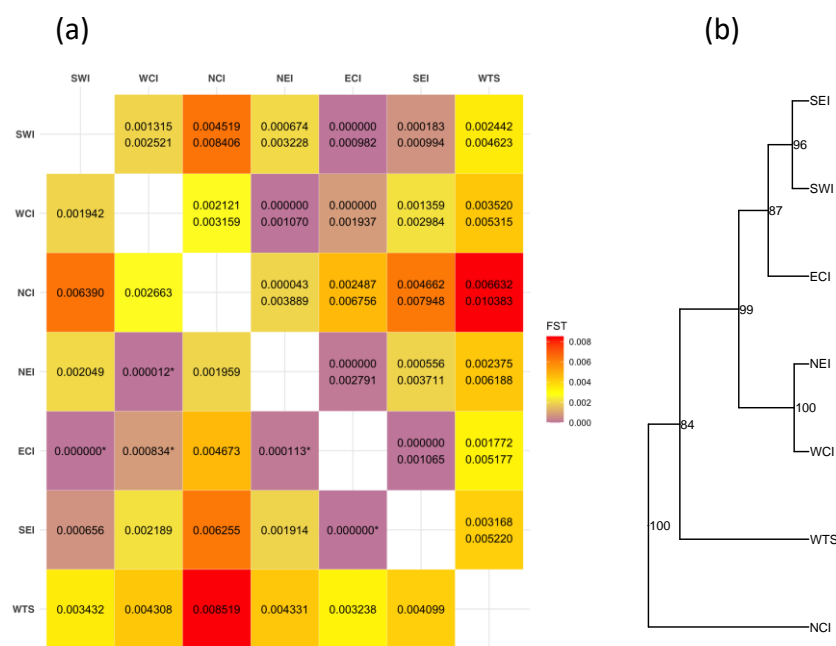
Applied FILTER step	VALUES	Individuals / Locus / Markers
Filter DArT reproducibility	0.93	493 / 71021 / 126843
Filter monomorphic markers		493 / 71021 / 121665
Filter markers in common		493 / 68712 / 121665
Filter individuals based on missingness (with outlier stats or values)	0.25	492 / 68712 / 118575
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 Id	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



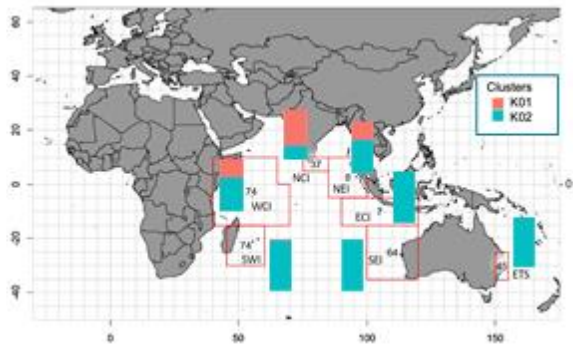
**Append Figure 78. Plot of information criteria for each tested model of from K=1 through K=6.**



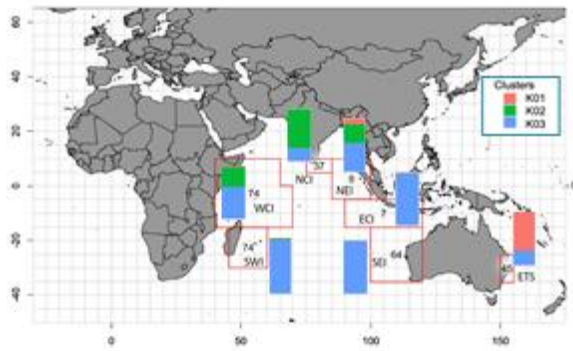
Append Figure 79. Results of focussed comparison of eastern Indian Ocean (ECI and SEI) and south-western Pacific sampling locations (WTS) Individual bar plot probability of assignment to a particular K genetic group when modelled at K=2 (cyan and red). From left to right are ECI (east central Indian), SEI (south east Indian), WTS (west Tasman Sea).



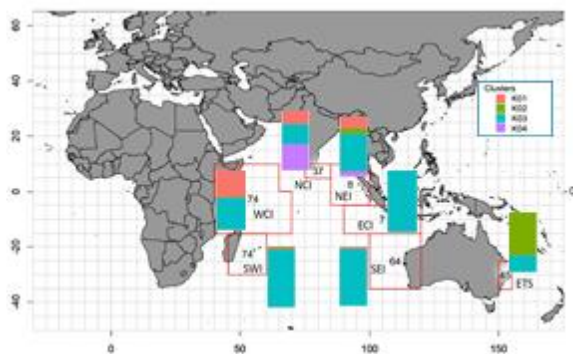
Append Figure 80. (a) Heat map of  $F_{st}$  values between pairs of sampling locations (below diagonal) and 95% confidence range (above diagonal). Colour levels as indicated by the scale bar to highlight relative  $F_{st}$  values calculated between pairs of sample locations with low (<0.003, purple 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.



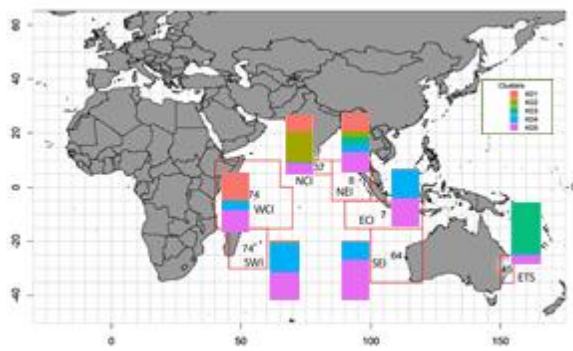
K=2



K = 3



K = 4



k = 5

Append Figure 81. Cumulative probabilities by location modelled at various K genetic groups.

## Summary

### Population structure within the Indian Ocean

Analysis of SNP genotypes has demonstrated presence of population heterogeneity of Swordfish within the Indian Ocean. While evidence of heterogeneity has previously been detected with maternal markers (e.g. Lu et al., 2006; Bradman et al., 2011), results from nuclear markers have been unsuccessful (Muths et al., 2009; Muths et al., 2013). These latter studies, which used both mtDNA and nuclear DNA markers and sampled at similar spatial and temporal scales to the current study, suggested the failure to detect evidence of population structure was largely due to maternal philopatry and subsequent male mediated gene flow. However, evidence of population differentiation revealed by SNP markers using differentiation at DNA microsatellites loci is likely due to homoplasy which results from the rapid rate of strand slippage mutation returning alleles to the population at a similar rate to that at which they are lost due to genetic drift. Homoplasy, identified through sequencing of swordfish mtDNA d-loop versus the ND2, also casts doubt on conclusions drawn from studies relying on this rapidly evolving section of the mitochondrial genome (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 of swordfish within the Indian Ocean. The advent of SNP marker technology has clearly overcome this disadvantage and provide a powerful platform to further investigate fine scale resolution of population structure.

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 indicating 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  $F_{st}$  (Append Figure 80). 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 (Append Figure 77 and Append Figure 80). 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 genotypes 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 genotype 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 (Append Figure 77, Append Figure 81). Analysis of SNP data has demonstrated the potential for resolving stock structure and delivered evidence for both northern and eastern stocks of swordfish in the Indian Ocean. Extension of this approach to include additional targeted

sampling 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 samples 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 SWO 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 would further 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 grounds will also be valuable for developing provenance markers to better inform sustainable swordfish management both within the Indian Ocean and on a globally.



#### 8.4.2 Swordfish (*Xiphias gladius*) - otolith microchemistry

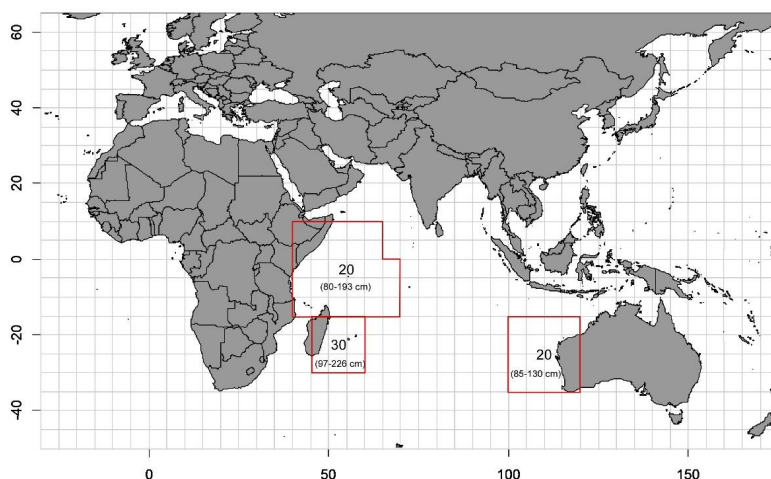
The investigation of otolith microchemistry of broadbill swordfish has thus far been limited to studies in the Pacific Ocean. Humphreys et al. (2005) used solution-based ICP-MS to obtain elemental fingerprints from otoliths collected around the Hawaiian Islands, from which they distinguished nursery areas. In a subsequent study, Humphreys and DeMartini (2008) used laser ablation ICP-MS on swordfish otoliths collected at sites in the northern and central Pacific Ocean but to date the results have not been published.

Otoliths from the Indian Ocean have been analysed using another otolith-based method -- otolith shape (Mahe et al. 2016). They examined swordfish otoliths collected at 6 locations in the western and central Indian Ocean and found no significant differences between otolith shape in the different locations.

Due to sampling limitations, swordfish otoliths for this study were available from only three locations. The numbers of otoliths were uneven among sampling locations and sampling periods so all the individuals from a location were grouped for the analyses. However, for some analyses we differentiate individuals by size: all swordfish were  $\geq 80$  cm LJFL so were considered to be either sub-adults or adults given the estimated LJFL at 50% maturity is 170 cm for females and 120 cm for males (Poisson and Fauvel 2009).

#### Methods

Otoliths from seventy swordfish were analysed (80-226 cm LJFL): 30 collected from the south west Indian Ocean, 20 from the western central Indian Ocean and 20 from the south east Indian Ocean (Append Figure 82). The fish were sampled during 3 periods: Nov-Dec 2017, March-May 2018 and December 2018 (Append Table 30). The otoliths were grouped by area for analysis.



Append Figure 82. Map showing the number of swordfish otoliths analysed for each of the three sampling locations, referred to as south west Indian Ocean (SWI), west central Indian Ocean (WCI) and south east Indian Ocean (SEI); and the size range of fish at each location.



**Append Table 30. Number, sampling period, size range and estimated ages of fish for each of the three sampling locations SWI, WCI and SEI.**

Location	N	Sampling dates	LJFL (cm)	* Estimated age range (years)
south west Indian Ocean (SWI)	20	Nov-Dec 2017	97-226	1-10
south west Indian Ocean (SWI)	10	Dec 2018	111-204	2-8
west central Indian Ocean (WCI)	20	Mar-May 2018	80-193	1-7
south east Indian Ocean (SEI)	20	Mar-May 2018	85-130	1-3

\* The ranges in ages are for male and females combined (Varghese et al. 2013, Wang et al. 2010).

The otoliths were analysed at Montpellier University, Plateforme AETE-ISO (France) using LA-ICP-MS. The laser ablated 20 micron spots along a transect between the otolith core and the edge, therefore acquiring a chemical signal from material deposited throughout the life of the fish.

The portion of the otolith transect near the core was examined to identify the chemical signatures deposited during the first weeks of life. These are most likely to reflect the fish spawning origins, i.e. the physicochemical characteristics of the water masses in which spawning occurred. The first laser point, on the core, was not included, in order to avoid any maternal influence on otolith composition. The mean of the next 3 points, between 10 and 40 microns after the core and corresponding to the first weeks of life, was used for analysis.

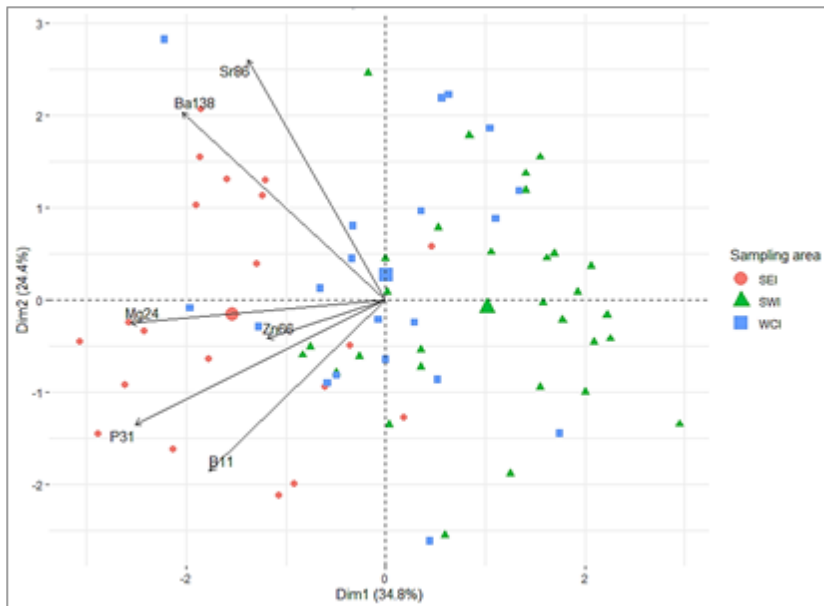
Fifteen chemical elements were measured. Those elements where more than 75% of results were above the limit of detection (LOD) were retained for further analysis.

## Results

After examination of the levels of detection for the 15 elements analysed by ICPMS, 6 elements were retained for further analysis (B, Mg, P, Zn, Sr and Ba).

### ***Principal component analysis - PCA***

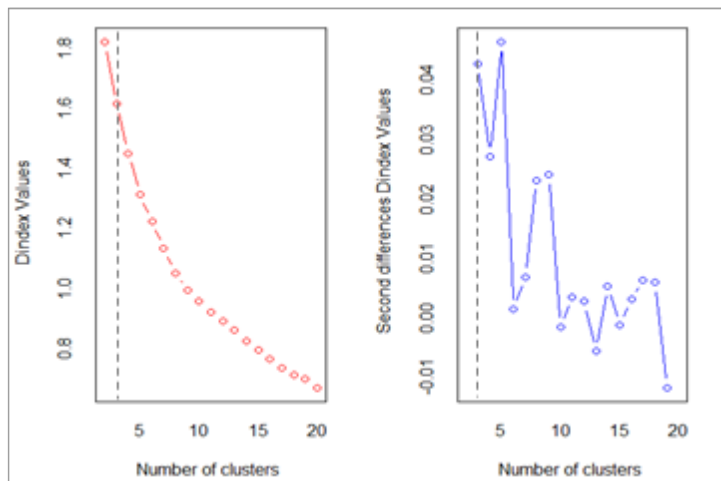
The otolith signatures from all areas largely overlapped but PCA suggested some variation in spawning origin according to fish capture location, although this was only just significant (PERMANOVA,  $p = 0.047$ ). Of the 6 elements above LODs, only Mg, P, Sr, Ba and B significantly contributed to the differences in otolith core signatures among swordfish individuals (>20% for dim 1 and/or 2 in the PCA; Append Figure 83). Therefore, only their signatures were used for investigation of fish spawning origin.



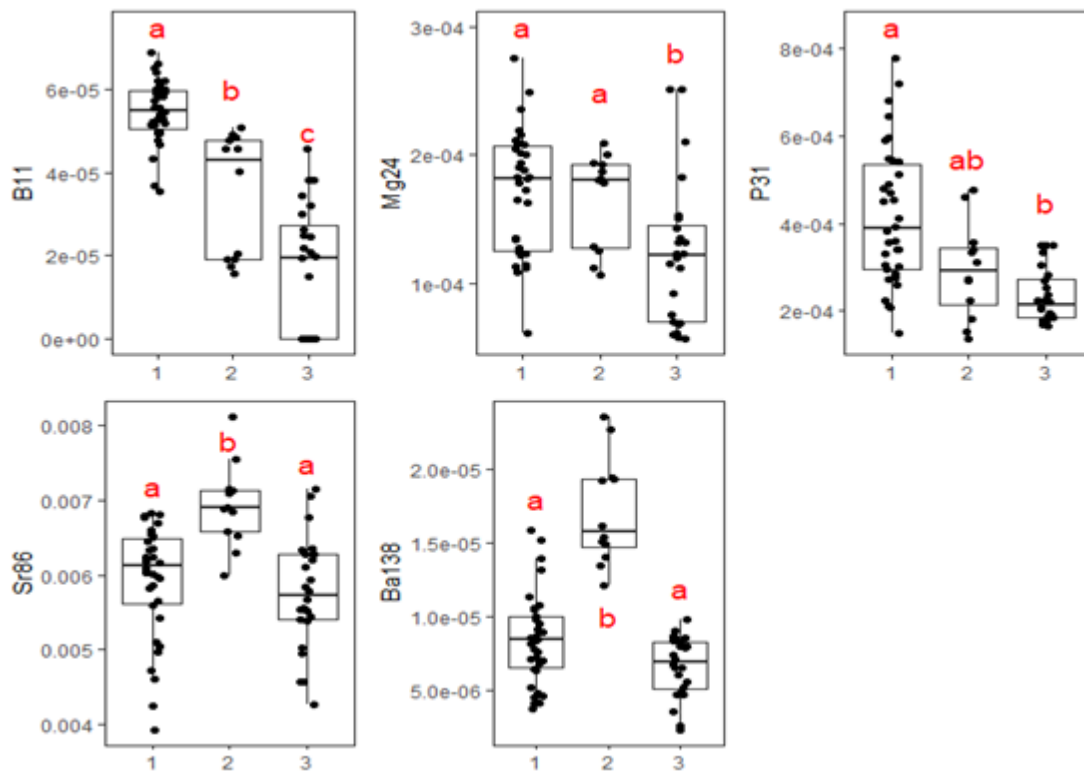
Append Figure 83. Biplot of individual (fish) and variable (chemical elements) projection on the first plane of the PCA made with the swordfish otolith core signatures. Individuals are coded by their sampling location. For the variables, the length of the arrow reflects the % of contribution to the total inertia.

## Clustering

Based on the variation in Mg, P, Sr, Ba and B core signatures, the most likely number of clusters (i.e. distinct spawning origins) was found to be 3 (Append Figure 84). The 3 corresponding putative spawning origins (SpO) had significantly (PERMANOVA,  $p = 0.002$ ) distinct multi-elemental signatures (Append Figure 85).



Append Figure 84. Results of the NbClust function (Charrad et al. 2014) using hierarchical clustering with the Ward method and Euclidian distances). On both graphs, the dotted line shows the most likely number of clusters defined according to the majority rule among all indices.

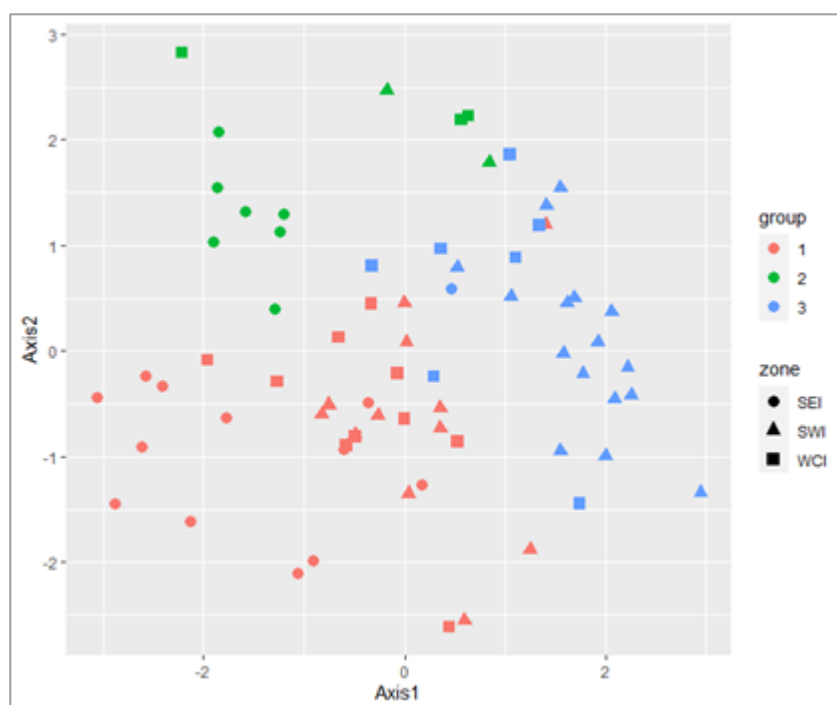


Append Figure 85. Projection of the clusters identified as potential discrete spawning origins (SpO 1 to 3) on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr, Ba) signatures of the otolith core of the 70 swordfish analysed and corresponding elemental signatures (black dots) and boxplots (Q1, median, Q3). Letters in red indicate groups with significantly distinct signatures ( $p < 0.05$ ).

In particular, boron differed significantly among all SpO, with maximum values in SpO-1 and minimum ones in SpO-3. SpO-2 had significantly higher Sr and Ba signatures and SpO3 lower Mg and P signatures than the two other SpO.

All three SpO apparently contributed to the stocks of the 3 areas investigated, but in varied proportions (Append Figure 86). SpO-1 was the main spawning source for the swordfish analysed (49%). It provided 40-60% of the fish analysed in all the areas sampled (Append Table 31). SpO-3 provided 34% of the fish sampled, principally in SWI (53%) and WCI (35%). Its contribution to SEI was very low (5%). SpO-2 contributed to only 17% of the total N of fish analysed but its contribution was twice to 4 times greater for SEI (35%) than for WCI (15%) and SWI (7%).

The differences in fish origin among zones explained the slight variation in otolith core signatures according to sampling location observed in the original PCA. Most differences were observed between SEI and SWI, which were both mainly originating from SpO-1 and SpO-3, respectively (Append Table 31).



Append Figure 86. Projection of individuals on the first plane of the PCA made with the multi-elemental (B, Mg, P, Zn, Sr, Ba) signatures of the otolith core of the 70 swordfish analysed. Colours on the graph represent the spawning origins (1 to 3) for each fish and symbols represent its final sampling area.

Append Table 31. Percent of individuals from each spawning origin (SpO 1 to 3) in the sub-adult and adult swordfish analysed (total N of fish tested = 70, FL = 80-226 cm) for each of the 3 sampling locations in the Indian Ocean: South-East (SEI), South-Western (SWI) and Central-Western (CWI).

	SEI	SWI	WCI
SpO-1	60%	40%	50%
SpO-2	35%	7%	15%
SpO-3	5%	53%	35%
Total	100%	100%	100%

## Summary

From the multi-elemental signatures of otolith cores, 3 potential spawning origins were identified for the 70 sub-adult and adult swordfish (*X. gladius*), which were captured between November 2017 and December 2018 in 3 locations: south-east, south-western and central-western parts of the Indian Ocean. All 3 putative spawning origins apparently contribute to each of the stocks fished in the 3 locations sampled, but in varied proportions.

Ideally, young-of-the-year fish would have been sampled from each location at the same time to minimize temporal variability in otolith signatures. However, the size-range of the fish analysed was 80 to >220 cm LJFL and with these large ranges in size (and age), the results must be interpreted with caution. Indeed, differences in otolith core signatures among locations might be due in part to cohort effects, i.e. differences among locations might be the result of inter-annual or seasonal shifts in oceanic water composition at one location rather than swordfish otoliths having different geographical spawning origins.

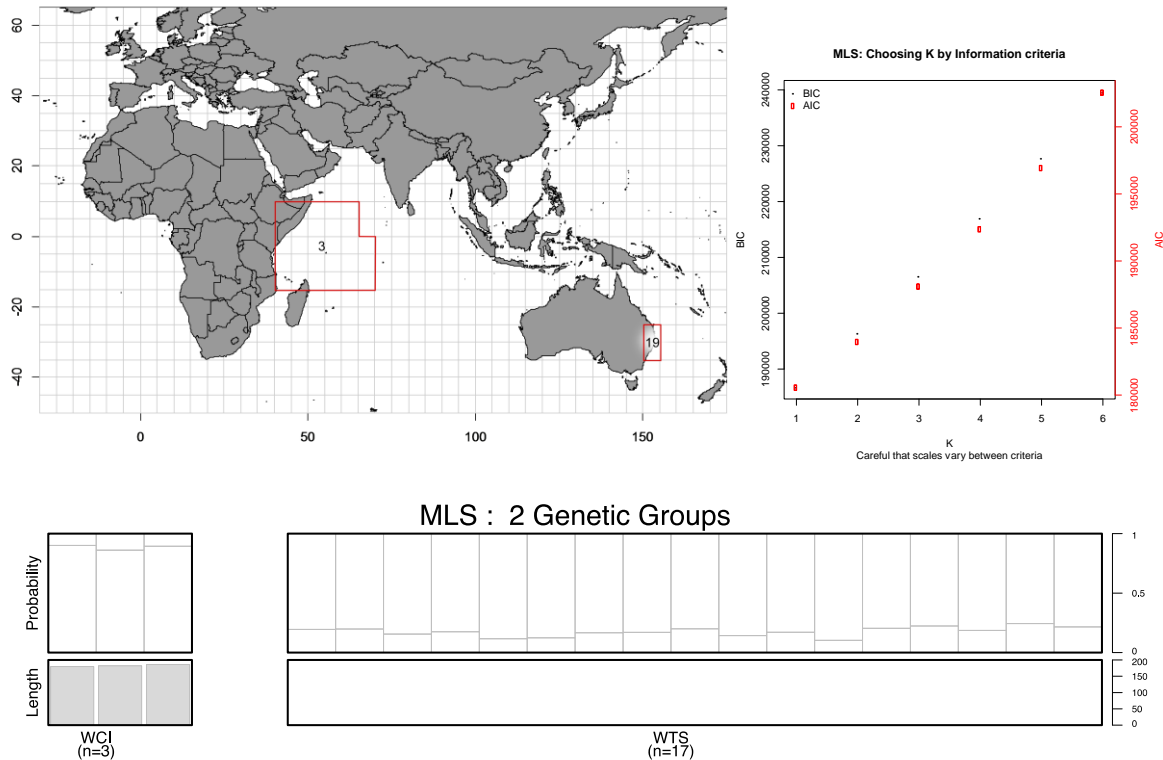
However, based on the available knowledge on swordfish spawning grounds, migrations in the IO (Fu et al. 2017, Nisikawa 1985) and surface circulation (see section 3) we can propose the following hypotheses:

- SpO-1, characterized by high B, Mg and P signatures, but low Sr and Ba values, would appear to be the most common spawning source in the IO with 49% of swordfish otolith cores having the SpO-1 core signatures. In fact, SpO-1 was the most common signature in each of the 3 sampling locations, with a minimum of 40% in SWI and a maximum of 60% in SEI). This suggests that SpO-1 may correspond to the spawning aggregation reported in the Java sea, in the north west Indian Ocean.
- SpO-3, characterized by low signatures in all 5 elements, provided 34% of the total fish analysed: 53% in SWI, 35% in WCI and 5% in SEI. It may therefore correspond to the spawning ground located in the north west Indian Ocean (Fu et al. 2017).
- SpO-2, characterized by high Sr and Ba signatures, contributed to only 17% of the total number of fish analysed. Its contribution was 35% in SEI, 15% WCI and 7% SWI. This could be an indication that this spawning ground is located in the north central Indian Ocean.

Before drawing any conclusions about swordfish spawning origin and mixing of stocks in the Indian Ocean, these preliminary results must be confirmed by analysing more samples. Ideally, samples would be collected from young-of-the-year from the same cohorts, collected in several successive years (at least 3) from the same sampling locations as this study. In addition, we suggest collecting swordfish from other locations in the IO where juveniles are caught, adults are known to spawn or where larvae have been reported. This will provide evidence to test whether the hypothetical spawning ground clusters found in this study correspond to spatially distinct spawning zones. In combination with genetic data, the information from otolith microchemistry could then give a clear pattern of the population structure of the species in this ocean.

### 8.4.3 Striped marlin (*Kajikia audax*) - population genetics

- 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 (Append Figure 87, Figure 51).
- Three samples from the Mozambique channel were genetically different to those collected from the western Tasman Sea indicating very limited to no connectivity between fish from the Western Indian Ocean and Western Pacific Ocean locations.



Append Figure 87. Top Left: 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. Top Right: Information criterion used to assess the likelihood of different numbers of genetic groups (k), lower indicating more likely. Bottom: Results of population structure analysis of DARTSeq using StockR for striped marlin for 2 genetic groups.

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. 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 the high levels of differentiation between these two regions with reasonable statistical precision greater than 95%. However, average individual heterozygosity levels were different between the west Indian Ocean compared to those from the Pacific (27% versus 32% respectively), admittedly based on only three west Indian Ocean versus 17 from the Pacific, are indicative of taking samples from two genetically distinct genetic cohorts of fish. Further temporal and spatial confirmation of previously observed population structure of striped marlin within the Indian Ocean will require focussed engagement and cooperation of local fisheries at broader sampling scales.

#### 8.4.4 Indo-Pacific sailfish (*Istiophorus platypterus*) - population genetics

Little investigation of stock structure exists for Indo-Pacific sailfish 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 SFA (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.

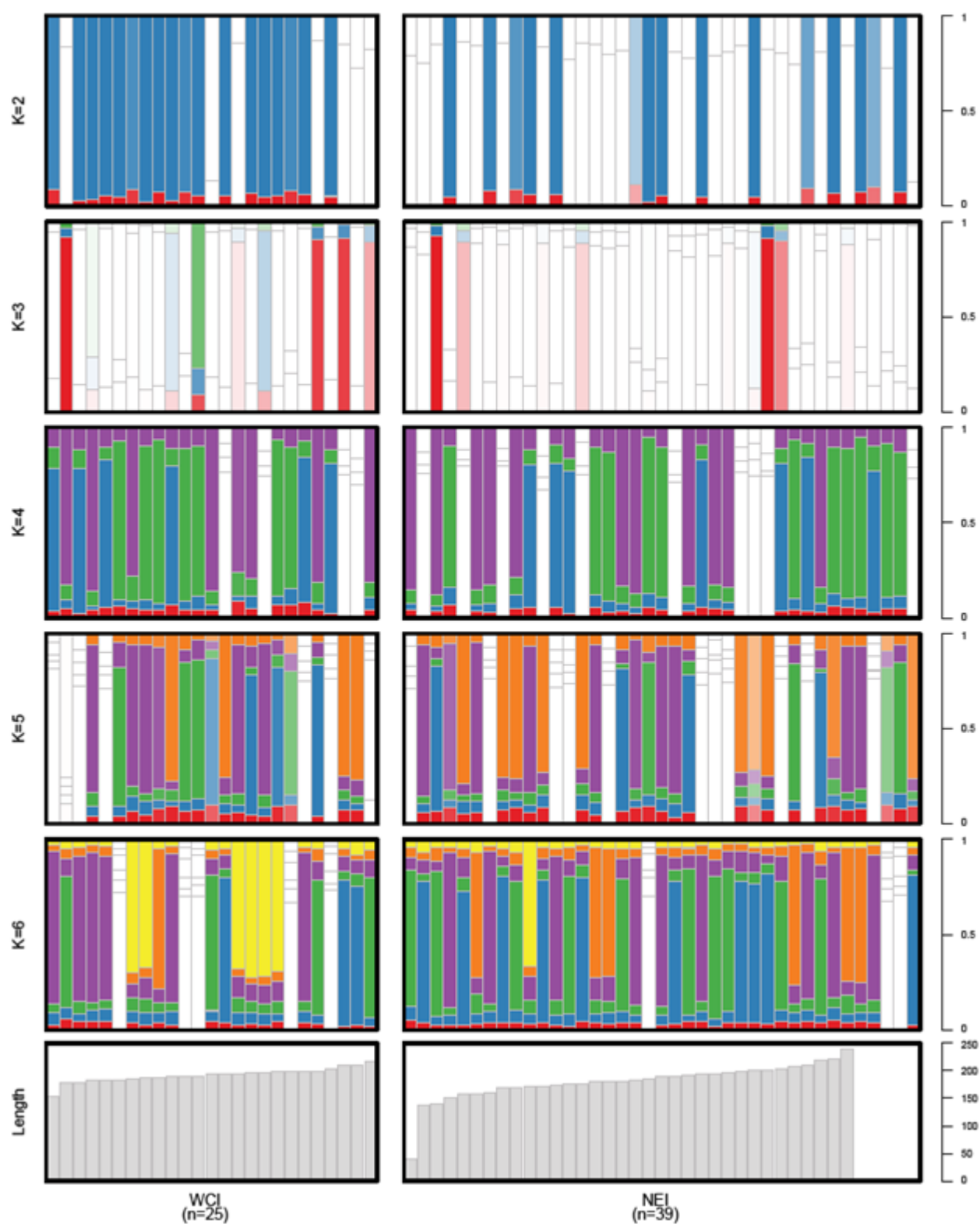
In this project we genotyped 99 Indo-Pacific Sailfish samples (including technical replicates) for a total of 54,827 SNPs. Our Radiator quality control left 65 unique individuals and 49,030 SNPs available for downstream analysis. This approach has been previously described in section 4.2.2 - 4.2.4 and details for each filter are presented in Append Table 32.

Our stockR analysis (Append Figure 88) revealed no genetic structure between sampling locations (Seychelles and Lampulo) in the Indian ocean. Further clarification of stock structure including samples from the Arabian Gulf and possibly the East Coast of Australia would be required in order to implement appropriate management protocols within the Indian Ocean for sailfish.

**Append Table 32. 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 Id	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





Append Figure 88. Individual length frequencies and results of population structure analysis of DARTSeq using StockR for all Indo-Pacific sailfin tuna assuming 2-6 genetic groups.

## 8.5 Sharks

### 8.5.1 Blue Shark (*Prionace glauca*) - population genetics

To the best of our knowledge, all studies performed thus far reported the lack of genetic population structure across the Pacific and the Atlantic and Mediterranean for *Prionace glauca* prior to this study. In fact, neither the use of electronic tags in the North (Queiroz *et al.* 2012; Vandeperre *et al.* 2014) and South Atlantic (Kohler and Turner 2008) and in the north-eastern Pacific (Maxwell *et al.* 2019), nor population genetic studies in both Atlantic and Pacific Oceans (King *et al.* 2015; Taguchi *et al.* 2015; Leon *et al.* 2017; Verissimo *et al.* 2017; Bailleul *et al.* 2018) provide any evidence against the assumption of large scale movements and population homogeneity within and across ocean basins (da Silva *et al.* 2010). Electronic tags demonstrate blue sharks capacity to swim very large distances, including inter ocean migrations (Kohler *et al.* 2002; da Silva *et al.* 2010; Queiroz *et al.* 2012; Vandeperre *et al.* 2014; Maxwell *et al.* 2019). Trans equatorial migrations are suspected to be limited (Kohler and Turner 2008) and reproductive cycles in the Northern and Southern hemispheres are reported to be off-set (Nakano and Seki 2003; Nakano and Stevens 2008), consistent with the current appraisal of two separate reproductive in the Atlantic. In the case of genetic studies, those based on mitochondrial DNA or/and microsatellite markers have not been able to detect any consistent pattern of genetic differentiation, even between ocean basins (King *et al.* 2015; Taguchi *et al.* 2015; Li *et al.* 2017; Verissimo *et al.* 2017; Bitencourt *et al.* 2019), with the exception faint signs of differentiation from the Mediterranean sea (Leon *et al.* 2017; Bailleul *et al.* 2018), and between Western Australia and the eastern South Pacific (Taguchi *et al.* 2015). However, the blue shark is an archetype of those species for which the lack of genetic differentiation may result from a broad range of demographic situations, including the existence of demographically independent populations which size is too large and /or separation too recent to be detected, resulting in a population grey zone effect (Bailleul *et al.*, 2018). For this reason, the authors suggested that the enhanced resolution allowed by genome scan analysis may help screening for signatures of genetic differentiation, if such demographic independence would exist and have left even faint footprint on the genome.

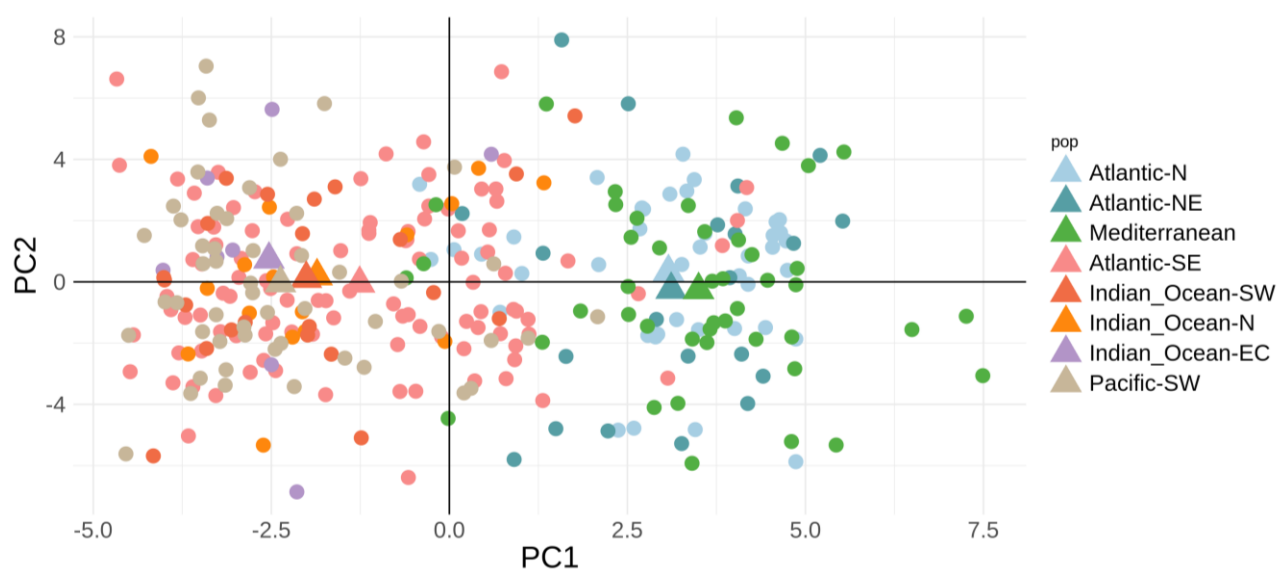
In this project, nearly all samples came we caught by Longline gear (n=348) with a total of 356 samples and only 8 samples caught by Purse seine (Indonesia area – Indian Ocean East). We genotyped 364 blue sharks (with 109 technical replicates) that delivered a total of 172,384 SNPs. Our Radiator quality control filtering left 312 unique individuals and 45,810 SNPs available for downstream analysis (Append Table 33). Further filtering removed 143 sex-linked markers based on presence-absence and heterozygosity patterns between sexes for Y-linked and X-linked markers respectively, as implemented in radiator (Gosselin *et al.*, 2020). Nine  $F_{st}$  outlier loci were identified, based on both results from PCAdapt (Luu *et al.*, 2017) and OutFLANK (Whitlock *et al.*, 2015) methods, and removed. All further analysis were based on the 45,658 SNPs and 312 individuals.

Our  $F_{st}$  and multidimensional analysis (PCA and DAPC) revealed that each Ocean (and basin within the Atlantic Ocean) hosts a distinct genetic group, with the exception of the Indian and South-western Pacific Ocean between which no significant structure was detected (Append Figure 89, Append Table 34). This is the first time that population structure between oceans has been demonstrated for blue sharks. The  $F_{st}$  values were extremely low (an order of  $10^{-3}$  to  $10^{-4}$ ) (Append Table 33, which explains difficulties to identify groups within the Atlantic Ocean using stockR (Foster

et al., 2018). In fact StockR was designed on purpose to discriminate groups with no contemporary mixture based on probability for classification (Foster et al., 2018), as opposed to other clustering algorithms such as Structure (Pritchard et al., 2000) or Tess (Caye et al., 2016), which are designed to detect clusters and infer the level of admixture of each sample possibly due to the history of migration and isolation among clusters. Notwithstanding this, StockR does show the strongest distinction between Atlantic-Mediterranean versus Pacific-Indian Ocean groups. Although genetic structure and clustering analysis confirm the occurrence of independent populations in each ocean the null hypothesis of panmixia could not be rejected at the within-basin scale, except for some indication of weak differentiation between Northern and southern Atlantic Ocean. Further analysis will be performed to explore these results in more detail. The power of the SNP markers to detect genetic differentiation as demonstrated here for the first time, provides strong support to extend sampling and analysis to still poorly represented areas in the species global range, such as the Northern Pacific and the Western and Northern-most parts of the Atlantic.

**Append Table 33. Radiator filtering steps for the blue shark *Prionace glauca*, including threshold values and the number of individuals, locus and markers after each step (the raw dataset consisted in 172384 SNPs on 364 samples analysed). Last lines also detail the number of SNPs removed after radiator filters because located on sexual chromosomes, or detected as outliers by applying two approaches (PCAdapt and OutFLANK), and removed from the dataset before further analysis.**

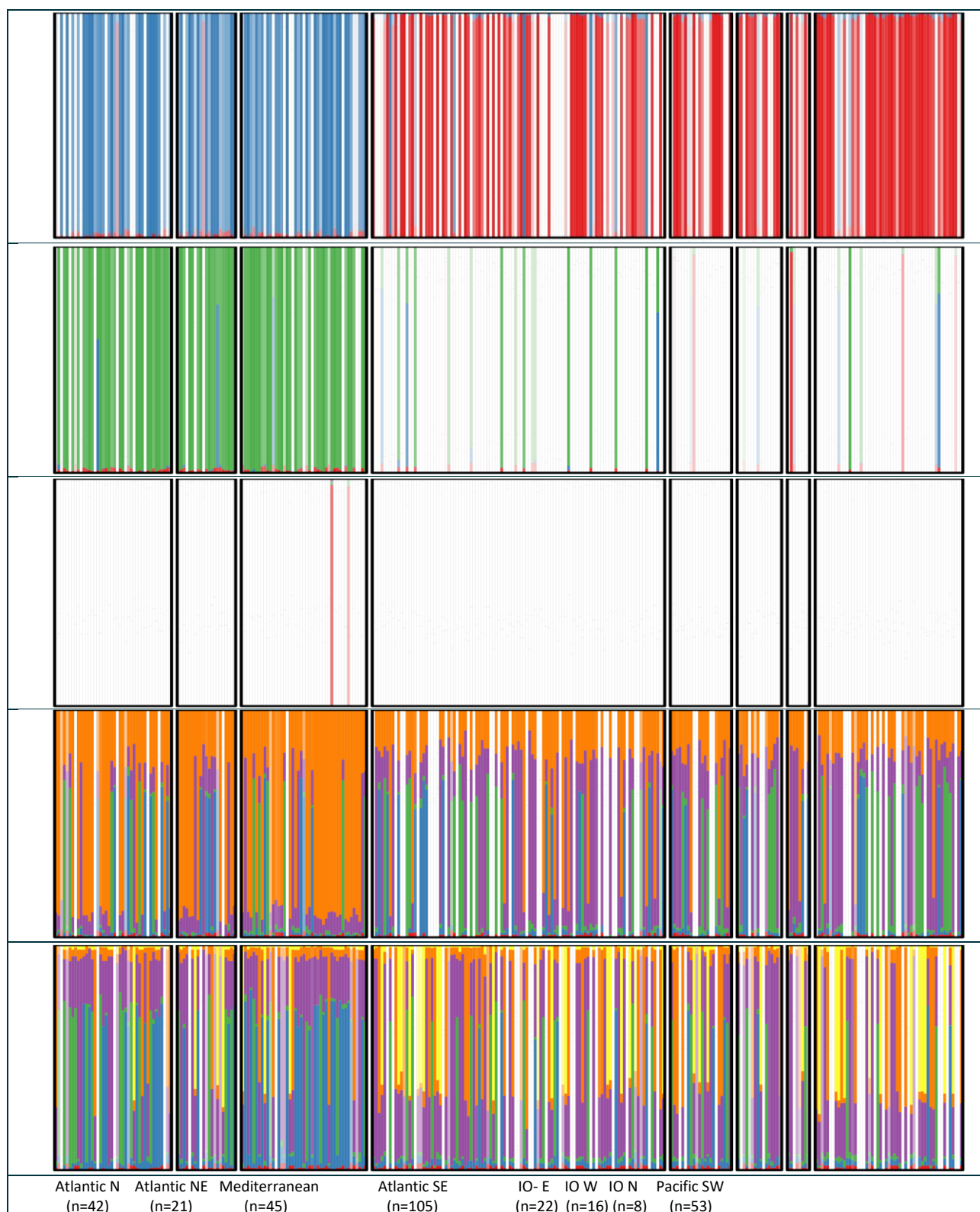
<b>Filters</b>	<b>VALUES</b>	<b>Individuals / Locus / Markers</b>
Filter DArT reproducibility	0.959(outliers)	364 / 95699 / 156195
Filter monomorphic markers		364 / 95699 / 156195
Filter markers in common		364 / 85836 / 142272
Filter individuals based on missingness	0.185(outliers)	332 / 85836 / 142272
Filter individuals based on heterozygosity	0.0601-0.0779 (outliers)	312 / 85836 / 142272
Filter monomorphic markers		312 / 84082 / 136648
Filter minor allele count	4	312 / 71622 / 110261
Filter coverage min / max	7/200	312 / 60859 / 95216
Filter missingness	0.1	312 / 46128 / 68083
Filter SNPs position on the read	8bp (outliers)	312 / 46128 / 68083
Filter markers snp number	4 (outliers)	312 / 45889 / 66837
Filter short linkage disequilibrium	MAC	312 / 45889 / 45889
detect mixed genomes (ind. heterozygosity)	0.117-0.15	312 / 45889 / 45889
detect duplicate genomes	0.1	312 / 45889 / 45889
Filter Hardy-Weinberg Equilibrium	Min 3 pops; p<0.0001	312 / 45810 / 45810
Post radiator: filtering based on sex markers		312 / 45667/ 45667
Post radiator: filtering based on outlier detected using PCAdapt and OutFLANK		312 / 45658/ 45658



Append Figure 89. PCA of blue shark based on 45658 SNPs (after removal of sex-linked and outlier SNPs), showing the segregation of two groups, one including Mediterranean and Northern Atlantic and the other the Pacific, Indian Ocean and Southern Atlantic.

Append Table 34. Pairwise  $F_{st}$  values and their level of significance after correction with q-value (\*  $p < 0.01$ , \*\*  $p < 0.001$ , \*\*\*  $p < 0.0001$ ).

<i>Fst</i>	MED (45)	ATL-N (42)	ATL-NE (21)	ATL-SE (105)	IO-EC (8)	IO-N (16)	IO-SW (22)	PAC-SW (53)
MED		0.0007***	0.0010***	0.0015***	0.0023***	0.0017***	0.0017***	0.0022***
ATL-N			0.0006**	0.0013***	0.0010*	0.0011***	0.0015***	0.0017***
ATL-NE				0.0018***	0.0014**	0.0015***	0.0015***	0.0020***
ATL-SE					0.0001	0.0000	0.0000	0.0000
IO-EC						0.0004	0.0000	0.0000
IO-N							0.0000	0.0000
IO-SW								0.0001
PAC-SW								



**Append Figure 90. Results of population structure analysis of DARTSeq using StockR for all blue sharks assuming 2 to 6 genetic groups.**

## 9 Appendix 2: Project outputs

Bibliography of population structure studies relevant to the Indian Ocean

Standard Operating Procedures for sample collection and distribution

Methods comparison for Radseq and ddRAD

Data and sample management

## 10 Appendix 3: Starting Values for Genetic Analyses

The analytical method employed here is based on Foster et al. (2018), which is a model-based statistical method. In particular, the model parameters (allele frequencies and probabilities) are estimated using maximum likelihood methods. Since there are a large number of parameters from allele frequencies, the parameter space is difficult to search. 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.


In an attempt to find these higher peaks, we search for local maxima that may turn out to be global maxima. We do this by introducing extra sets of starting values, based on random groupings of the individuals (random starts, Givens & Hoeting; 2013). The full set of starting values used are:

1. Completely random starts. Each individual is assigned at random to one of the putative groups.
2. Random assignment of sampling regions. Each sampling region (e.g. the North Central Indian, NCI). 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.
3. 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.

Anecdotal evidence suggests that the original method in Foster et al. (2018) is rarely surpassed by a random start. This is especially so for smaller numbers of genetic groups. Further, the number of principle components to perform the rotation with does not seem to affect the results overly. Again, this finding is weakened as the number of putative genetic groups is increased. It is stressed; however, this is anecdotal evidence only as the source of the maximum was not recorded.







**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](mailto:csiroenquiries@csiro.au)  
[www.csiro.au](http://www.csiro.au)

**For further information**

Oceans and Atmosphere  
Dr Campbell Davies  
+61 3 6232 5222  
[Campbell.davies@csiro.au](mailto:Campbell.davies@csiro.au)  
[csiro.au](http://csiro.au)