Australia's National
Science Agency

Genetic population structure of neritic tunas in the Indian Ocean from the PSTBS-IO Project

Pierre Feutry, Scott Foster, Peter Grewe, Jorden Aulich, Matt Lansdell, Scott Cooper, Grant Johnson, Daniel Fernando, Zulkarnaen Fahmi, Wudianto, Fayakun Satria, Umair Shahid, Syed Meesum Reza Kazmi, Mohamed Ahusan, Pratiwi Lestari, Muhammad Taufik, Asep Priatna, Achmad Zamroni, Craig Proctor, Jessica Farley, Hilario Murua, Francis Marsac and Campbell Davies

1 July 2020

Paper submitted to the 10th meeting of the Working Party of Neritic Tuna

Table of Contents

Abstract

Longtail tuna (*Thunnus tonggol*), kawakawa (*Euthynnus affinis*) and narrow-barred Spanish mackerel (*Scomberomorus commerson*) are important fish resources in the Indian Ocean. These species are currently managed as single stocks. Studies to date have not demonstrated strong evidence of population structure. In this study we report results from the multi-agency, collaborative PSTBS-IO project, which involved large-scale sampling and cutting-edge sequencing technology to investigate the genetic population structure of these three neritic species, along with priority tropical and temperate tuna, billfish and sharks. These analyses provide strong evidence of at least three, two and four different populations in the Indian Ocean for longtail tuna, kawakawa and the narrow-barred Spanish mackerel respectively. These results warrant re-consideration of how these species are monitored and managed in the Indian Ocean and highlight the need for further targeted research to confirm the temporal stability of these results and provide a comprehensive understanding of population boundaries for these species within the Indian Ocean.

Introduction

Longtail tuna, kawakawa and narrow-barred Spanish mackerel are species of substantial commercial and food security value in the Indian Ocean. Like other tuna species in the region, these neritic species straddle multiple coastal EEZs and internationals waters. All three are assumed to be highly migratory, forming a single stock in the Indian Ocean and their management, for which the Indian Ocean Tuna Commission (IOTC) is responsible, requires a multi-national effort. If the scientific stock assessment advice is based on invalid assumptions, management may not achieve stated objectives related to conservation and optimal economic use of the resource. Specifically, if there is more than one population (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. Over the last two decades, several genetic studies have the investigated the population structure of neritic species in the Indian Ocean or parts of it. The conclusions of these studies differ for each species. For longtail tuna (Kunal 2014, Zamroni 2018) and kawakawa (Kumar et al. 2012, Masazurah et al. 2012, Jonhson et al. 2016) the analysis of mitochondrial markers did not reveal any structure and the assumption of a single stock could not be rejected. For the narrow-barred Spanish mackerel, however, there has been contrasting results. Whist some studies did not detect any structure (Hoolihan et al. 2006, Abedi et al. 2012, Vineesh et al. 2016), the analysis of microsatellite and/or mitochondrial markers by Fauvelot & Borsa (2011), Habib & Sulaiman (2017) and van Herwerden et al. (2006) has seriously challenged the single stock hypothesis.

A key component of population genetic studies, is that they are only as good as the markers and the samples they use. The lack of detection of population structure can either be due to true absence of barriers to geneflow or simply due to insufficient resolution with the method employed. In this context, the contrasting results between various genetic studies on the narrow-barred Spanish mackerel can be reconciled due to differences in marker used, samples sizes and

sampling locations. It also shows that the single stock hypothesis for longtail tuna and kawakawa is remains to be resolved for the management of these species.

Recent advances in sequencing technologies and analytic procedures have greatly improved the power to detect population structure through examination of Single Nucleotide Polymorphisms (SNPs) and drastically reduce the risk of failing to identify barriers to geneflow due to a lack of adequate genetic resolution. Complexity reduction genotyping by sequencing methods, in particular, offer inexpensive ways of screening large sample sizes for non-model species such as those of interest here. This highlights the need for a more systematic and unified approach to evaluate the population structure of these species in the Indian Ocean, using the best sequencing technology available.

In 2017, CSIRO (Australia) in collaboration with AZTI Tecnalia (Spain), IRD (France) and CFR (Indonesia) and a number of Indian Ocean coastal state collaborators commenced a 3-year project on population structure of tuna, billfish and sharks of the Indian Ocean (PSTBS-IO) funded by the European Union and the consortium partners through the Indian Ocean Tuna Commission (Davies et al. 2017; Davies et al. 2020).

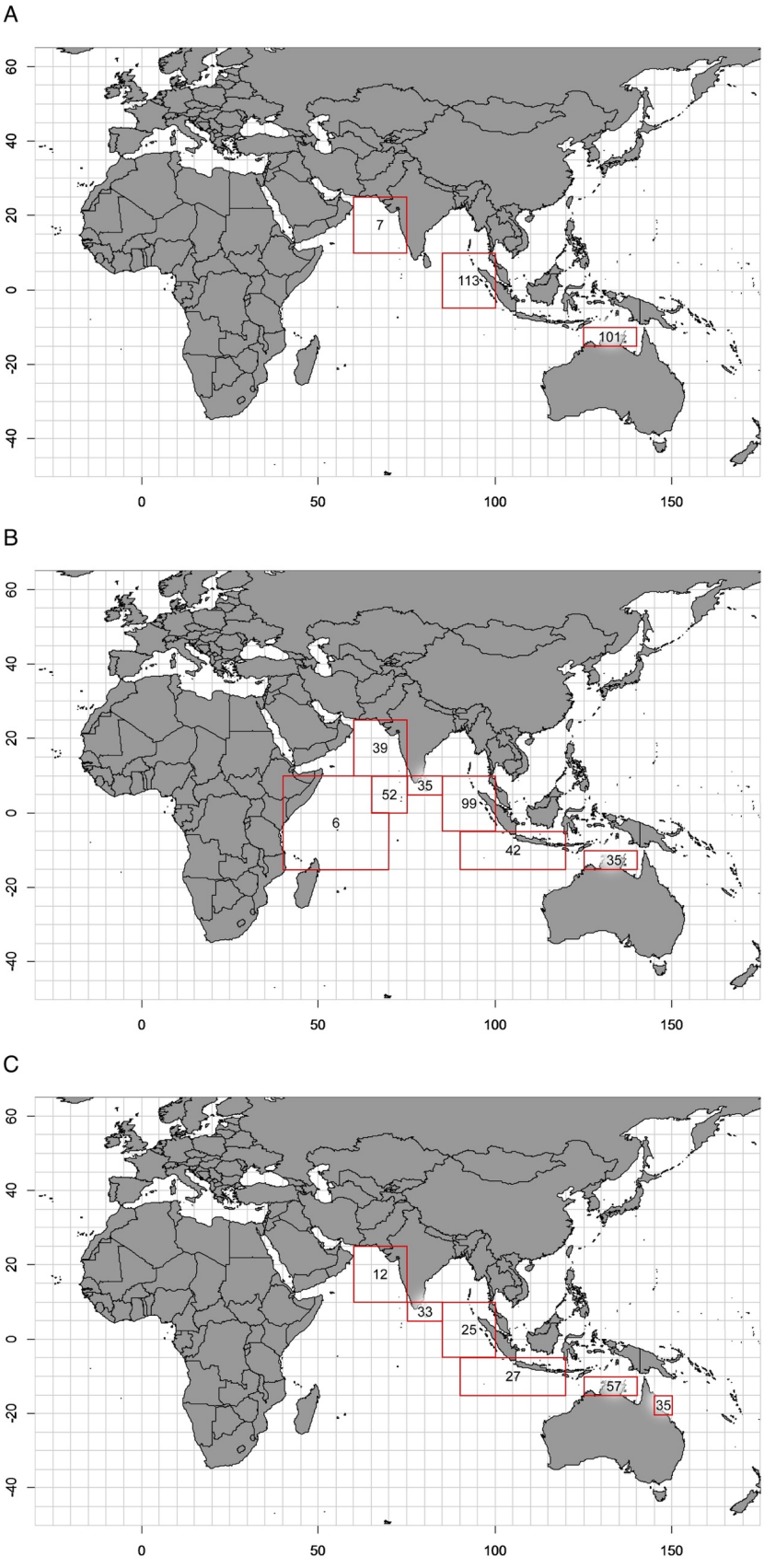
In this paper we report on the genetic population structure of longtail tuna, kawakawa and narrow-barred Spanish mackerel results from the PSTBS-IO project, using a proven sampling strategy combined with a cutting-edge genotyping-by-sequencing protocol. In doing so, we welcome further input from the WPNT on these results and their interpretation. The results for other species (i.e. tuna, billfish, sharks) from the project will be reported to the relevant working parties and Scientific Committee meeting later in 2020.

Methods

Sampling

Tissue samples of longtail tuna, kawakawa and narrow-barred Spanish mackerel were collected from several locations within the Indian Ocean between 2017 and 2019 (Fig. 1; Table 1). Samples from the Pacific Ocean were also obtained for narrow-barred Spanish mackerel. The samples were taken following the protocols developed for the PSTBS-IO project (Anon 2018). The samples were removed using a clean sharp knife and stored in nucleic acid preservation buffer before being transported to CSIRO Marine Labs for analysis. Information on fish length (cm), catch location, sampling date and sex (if possible) of the sampled fish was recorded. Fish were selected for analysis based on capture location in order to obtain equal sample sizes from each region, as far as possible. For regions where the number of samples was low, additional samples were selected from other regions.

Figure 1. Sampling locations of (A) longtail tuna (*Thunnus tonggol*), (B) kawakawa (*Euthynnus affinis*) and (C) narrow-barred Spanish mackerel (*Scomberomorus commerson*).



Species	No. tissue samples	No. selected for sequencing	No. analysed (past quality control)	Size range analysed (mean) (cm)
LOT	316	298	221	43-103 (75)
KAW	546	362	308	21-68 (40)
COM	256	210	189	18-120 (59)

Table 1. Summary of all samples used in this study.

DArTSeq library preparation and sequencing

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

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

Quality control filtering of loci and individuals

A step wise process for data quality control using the package RADIATOR (Gosselin 2020) was performed to filter out both poor-quality DNA markers (SNP loci) and poor-quality individual samples (i.e. low DNA quality/quantity or DNA contamination). The following parameters were evaluated as part of the SNP quality control: reproducibility (proportion of repeatable genotype calls estimated via inclusion of technical replicates, i.e. some DNA samples are prepared and sequenced twice); call rate (proportion of samples genotyped); minor allele count (how many times the allele in lowest frequency was observed); coverage (how many times a locus was observed in average across all individuals); position of the SNP on the fragment being sequenced; linkage disequilibrium, multiple SNP on the same fragment are not independent of each other and we only retained the one with the highest minor allele counts (most informative); Hardy Weinberg

equilibrium within each sampling location; missingness (proportion of times a loci was reported as NA); minor allele frequency per sampling region. For the quality control of individuals, the following parameters were evaluated: missingness (proportion of missing values); heterozygosity (proportion of locus exhibiting two different alleles); total coverage (total number of sequences obtained per individual); genetic distance (measure of how similar a pair of individuals are). Thresholds used during the filtering process foreach of these parameters are provided in appendix 1-3.

Statistical analysis

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

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

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

Results and discussion

Longtail tuna

A total of 298 longtail tuna samples were genotyped and 123,600 SNPs were called by DArT-Soft14. Our Radiator quality control left 221 unique individuals and 18,670 SNPs available for the population grouping analysis (App. 1).

Our stockR approach revealed that each of the three sampling locations in the Indian ocean hosts a distinct genetic group, therefore providing the first evidence of population structure of longtail tunas in this region (Fig. 2).

To the best of our knowledge, four attempts were made at detecting genetic population structure across the Indo-Pacific distribution of *Thunnus tongol* prior to this study. All four 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 2014). Similarly, no structure was detected for this species within the Indonesian archipelago (Zamroni 2018).

The results of the present study 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 across its full range in the Indian Ocean.

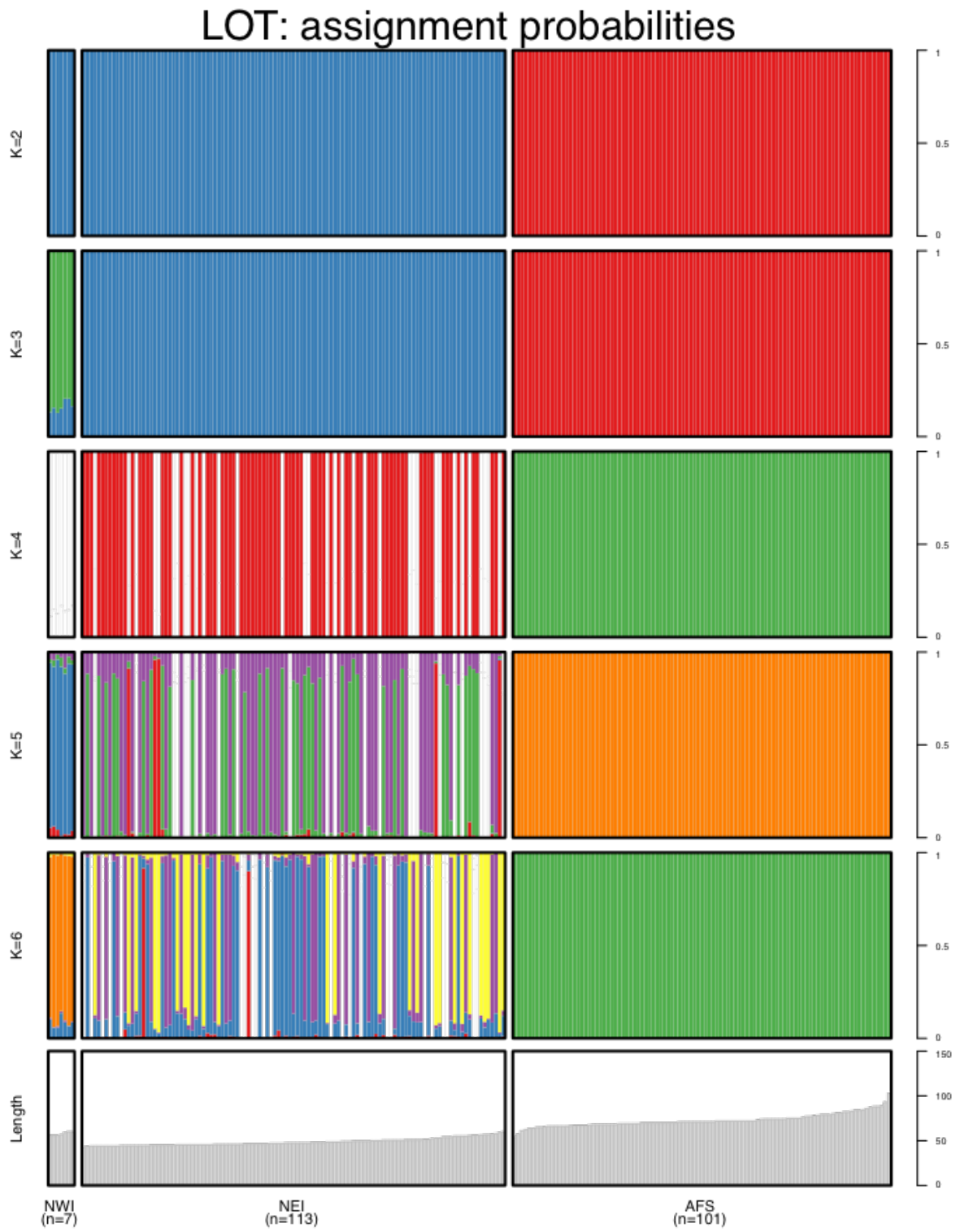


Figure 2. Individual length frequencies (bottom panel) and results of population structure analysis of DArTSeq using StockR for all longtail tuna (*Thunnus tonggol*) assuming 2-6 genetic groups (top panels).

Kawakawa

A total of 362 kawakawa samples were genotyped and the DArTseq analysis yielded 66,910 SNPs. Out of these, 308 unique individuals and 10,334 SNPs passed our quality control filtering (App. 2).

The stockR analysis (Fig. 3) detected the presence of at least two distinct genetic groups amongst the 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.

Prior to this study, the investigation of the genetic population structure of kawakawa *Euthynnus affinis* had 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) investigated population structure within the strait of Malacca. None of these studies detected any sign of genetic structure. Our results provide higher resolution insight into the population structure in kawakawa in the Indian Ocean.

Outside the Indian Ocean previous attempts at detecting population structure had mixed results. In Southeast Asia, an 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 scales in this species.

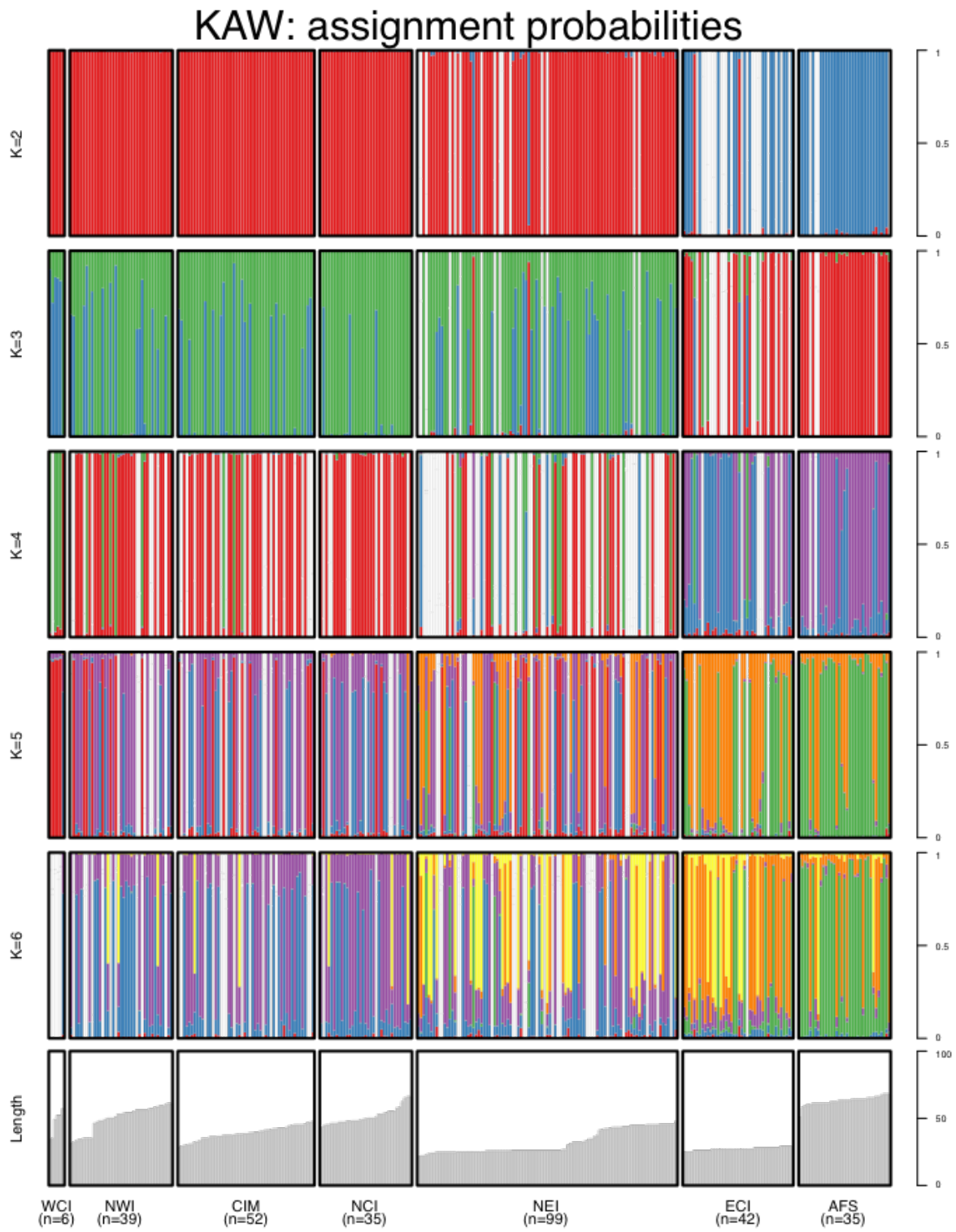


Figure 3. Individual length frequencies (bottom panel) and results of population structure analysis of DArTSeq using StockR for all kawakawa (*Euthynnus affinis*) assuming 2-6 genetic groups (top panels).

Narrow-barred Spanish mackerel

In this study 210 narrow-barred Spanish mackerel samples were genotyped and the DArTseq analysis generated 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 as has been hypothesised by Fauvelot & Borsa (2011). This additional complexity 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. Quality control details for the both rounds of filtering are given in Appendix 3.

Out of all three species analysed here, the narrow-barred 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 (Fig. 4). The first one was made of the NWI and NCI sampling locations (App. 4a), the second and third one corresponded to the NEI and ECI sampling locations respectively (App. 4b) and the fourth group comprised the AFS and WCS sampling locations (App. 4c). The second round of stockR analyses detected further spatial clustering within this fourth group, in which there was a hint of at K = 5 of the first round of analysis (Fig. 4), but not within the other groups (App. 4a, 4b). Having only 32 individuals that past quality control for NWI and NCI combined was possibly a limiting factor to detect further sub-structure in that region.

Investigation of the genetic population structure of the narrow-barred Spanish mackerel over the last two decades, using either mitochondrial DNA or microsatellite nuclear markers, had already showed evidence of limited connectivity. These studies generally found strong genetic population structure at the regional scale. At least three genetic stocks have been identified in the Pacific (Shaklee et al. 1990, Buckworth et al. 2007, Sulaiman & Ovenden 2010). In fact, the genetic differentiation is so high that the possibility of a complex of cryptic species has been proposed (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 sample size available to Habib & Sulaiman (2017) was too low (N=5) to allow definitive 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).

This study detected a total of four genetic groups amongst five sampling locations within the Indian Ocean, taking the understanding of the narrow-barred Spanish mackerel to a new level, from regional to local.

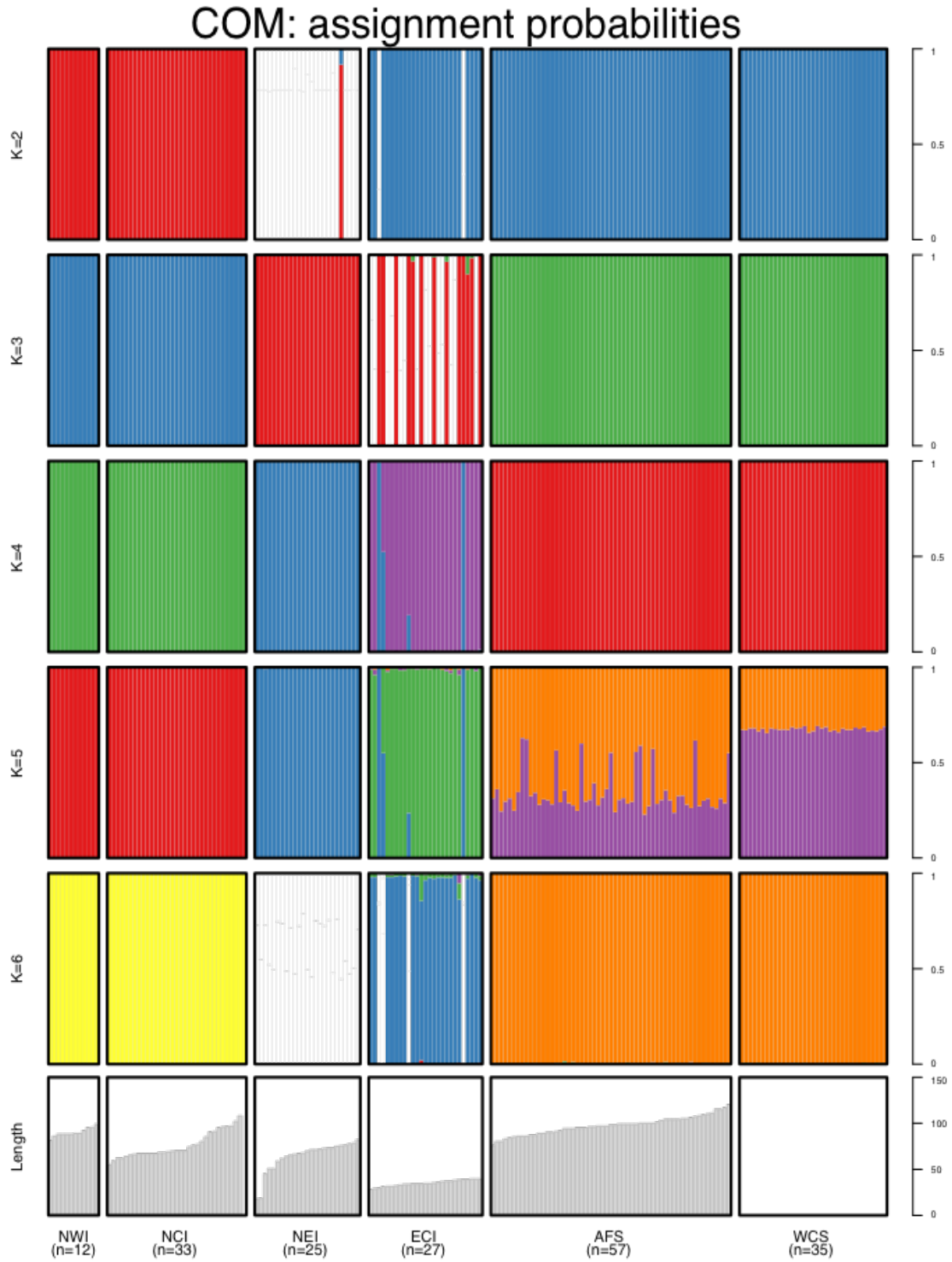


Figure 4. Individual length frequencies (bottom panel) and results of population structure analysis of DArTSeq using StockR for all narrow-barred Spanish mackerel (*Scomberomorus commerson*) assuming 2-6 genetic groups (top panels).

Conclusion

Our study has substantially advanced the understanding of population structure of each of these three neritic species within the Indian Ocean. Barriers to gene flow inside the Indian Ocean were found for the first time for long tail tuna and kawakawa, whereas much stronger, more localised structure, compared to previous studies, was found in the narrow-barred Spanish mackerel.

This research highlights the need for the investigation of the spatial resolution of genetic structure at even finer scale for these species across the full range within the Indian Ocean. For longtail tuna and the narrow-barred Spanish mackerel in particular, the fact that almost each location sampled as part of this study was found to host a distinct genetic group indicate that further sampling with higher spatial resolution is likely to reveal additional distinct genetic groups.

Importantly, the technique deployed here integrates gene flow over evolutionary timescales, whilst in a perfect world stock should be managed on demographic timescales. It is possible that genetic approaches with higher temporal resolution, such as the study of shared fragment of genomic fragments (Ralph & Coop 2013) or Close-Kin Mark-Recapture (Feutry et al 2017), would reveal structure at even smaller scale.

From a stock assessment and management perspective, these results indicate that monitoring, population biology and management programs should be designed to support management at the unit stock scale. While, we recognize this presents considerable challenges, given the data poor nature of many of the fisheries in the region, it is evident from the results of this collaborative study, among others, that advances can be achieved.

Acknowledgements

We wish to thank all those who assisting with the sample collection and analysis, in particular, 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; Ahmed Haikaland Shafiya Naeem, Marine Research Centre, and Shiham Adam, Ministry of Fisheries, Marine Resources and Agriculture, Maldives; Northern Territory Department of Primary Industry and Resources, Fisheries Division; Joanne Langstreth, John Cavallaro, Queensland Department of Primary Industry and Fisheries, Australia; Anais Medieu, Marianne Pernak, Pascal Bach, IRD; Seychelles Fishing Authority; Iñigo Krug, AZTI; Dale, Kolody, CSIRO; the IOTC secretariat; the many vessel owners, skippers, observers, crew and processors who provided fish for sampling and access to facilities. The project is supported by funding from CSIRO Oceans and Atmosphere, AZTI Tecnalia, Institut de Recherche pour le Développement (IRD), and Indonesia's Center for Fisheries Research (CFR) and financial assistance of the European Union (GCP/INT/233/EC –Population structure of IOTC species in the Indian Ocean). The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

Bibliography

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.

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

Buckworth, R. C., Newman, S. J., Ovenden, J. R., Lester, R. J. G., 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.

Davies, C.R., Murua, H., Marsac, F., Fahmi, Z. (2017). Population Structure of IOTC species and sharks of interest in the Indian Ocean. Presentation to the IOTC Scientific Committee. Seychelles, December 2017.

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., Zamroni, A. (2020). Study of population structure of IOTC species and sharks of interest in the Indian Ocean using genetics and microchemistry: Final Report to IOTC. May 2020.

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.

Feutry, P., Berry, O., Kyne, P. M., Pillans, R. D., Hillary, R. M., Grewe, P. M., Marthick, J. R., Johnson, G., Gunasekera, R. M, Bax, N. J., Bravington, M. (2017). Inferring contemporary and historical genetic connectivity from juveniles. *Molecular ecology 26* (2): 444-456.

Foster, S. D., Feutry, P., Grewe, P. M., Berry, O., Hui, F. K. C., Davies, C. R. (2018). Reliably discriminating stock structure with genetic markers: Mixture models with robust and fast computation. *Molecular Ecology Resources* 18: 1310-1325.

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

Grewe, P. M., Grueger C., Aquadro, C. F., Berminghmam, E. (1993). Mitochondrial DNA Variation among Lake Trout (*Salvelinus namaycush*) Strains Stocked into Lake Ontario. Canadian Journal of Fisheries and Aquatic Sciences 50: 2397–2403.

Grewe, P. M., Feutry, P., Hill, P. L., Gunasekera, R. M., Schaefer, K. M., Itano, D. G., Fuller, D. W., Foster, S. D., Davies, C. R. (2015). Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. Scientific reports, 5, 16916.

Habib, A., Zohrah, H., Sulaiman, Z. (2012). Population genetic structure of longtail tuna (*Thunnus tonggol*) along Wallace's Line. Department of Biology, Faculty of Science, Universiti Brunei Darussalam.

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.

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.

Jackson, A. M., Erdmann, M. V., Toha, A. H. A., Stevens, L. A., Barber, P. H. (2014). Phylogeography of commercial tuna and mackerel in the Indonesian Archipelago. *Bulletin of Marine Science 90* (1): 471-492.

Johnson, M. G., Mgaya, Y. D., Shaghude, Y. W. (2016). Genetic Stock Structure and Phylogenetic Relationship of Kawakawa *Euthynnus affinis*–Cantor (1849) in the Northern Coastal Waters of Tanzania Using Mitochondrial DNA Control Region. *IOTC WPNT-06-22*.

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.

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. (2014). Population genetic structure of longtail tuna *Thunnus tonggol* Bleeker 1811 and yellowfin tuna *Thunnus albacares* bonnaterre 1788 from the Indian region. PhD Dissertation. Acharya Nagarjuna University. 118p.

Masazurah, A. R., Azizah, S., Samsudin, B. (2012). A preliminary study of population structure of kawakawa, *Euthynnus affinis* (Cantor 1849) in the straits of Malacca. *IOTC WPNT-02-23*.

Ralph, P., Coop, G. (2013). The geography of recent genetic ancestry across Europe. *PLoS biology*, *11*(5).

Santos, M. D., Lopez, G. V., Barut, N. C. (2010). A pilot study on the genetic variation of eastern little tuna (*Euthynnus affinis*) in Southeast Asia. *Philippine Journal of Science* 139: 43–50.

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.

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.

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.

Vineesh, N., Kathirvelpandian, A., Divya, P. R., Mohitha, C., Basheer, V. S., Gopalakrishnan, A., 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.

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.

Zamroni, A., Suwarso, S., 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.

Appendix 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.

Filters	VALUES	Individuals / Locus / Markers
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 ld	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

Appendix 2

Radiator filtering steps for kawakawa *Euthynnus affinis*, 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	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 ld	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

Appendix 3

Radiator filtering steps for the narrow-barred 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.

A - Filters	Values	Individuals / Locus / Markers
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 ld	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

B1 - Filters	Values	Individuals / Locus / Markers
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 ld	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
Filter HWE		32 / 7570 / 7570

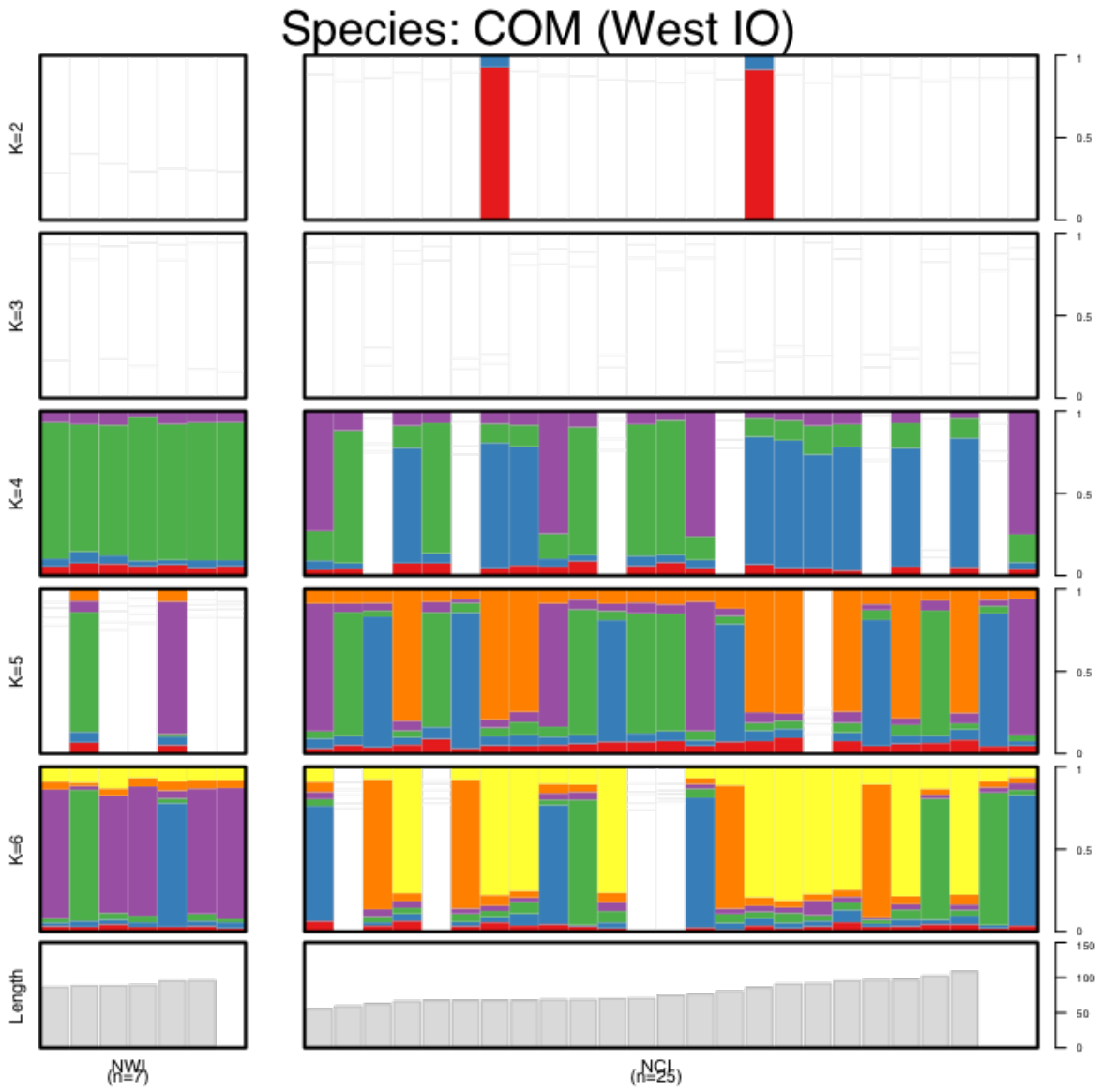
B2 - Filters	Values	Individuals / Locus / Markers
Filter DArT reproducibility	0.95	77 / 106487 / 194591
Filter monomorphic markers		77 / 103039 / 187335
Filter markers in common		77 / 68227 / 101404
Filter MAC	10	77 / 67999 / 101126
Filter coverage min / max	10 / 100	77 / 30519 / 36140
Filter genotyping	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 ld	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

B3 - Filters	Values	Individuals / Locus / Markers
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 ld	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
Filter HWE	20.01	73 / 12010 / 12010

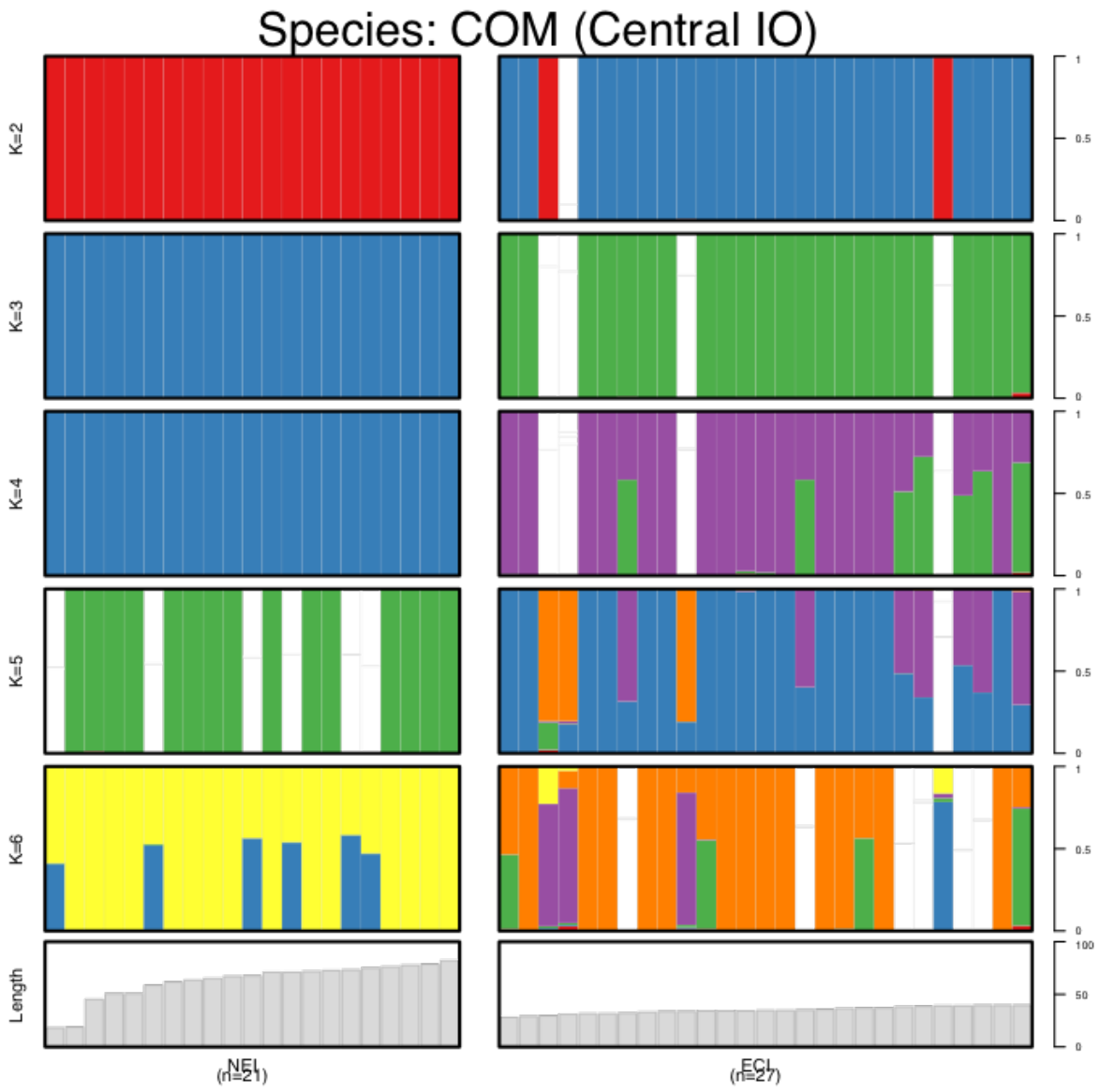
Appendix 4a

Individual length frequencies (bottom panel) and results of population structure analysis of DArTSeq using StockR for the narrow-barred Spanish mackerel *Scomberomorus commerson* from the NWI and NCI subset assuming 2-6 genetic groups (top panels).



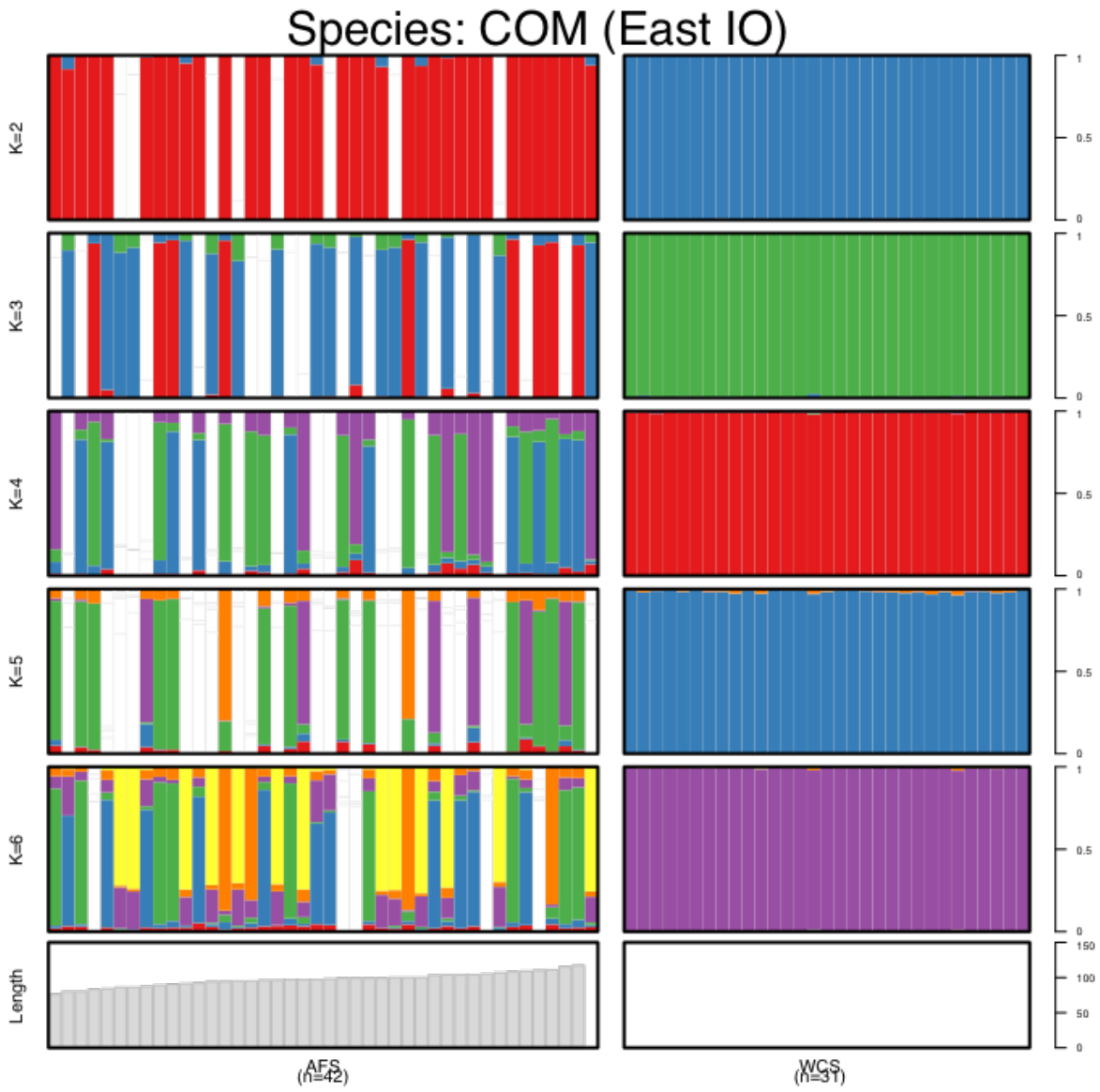
Appendix 4b

Individual length frequencies (bottom panel) and results of population structure analysis of DArTSeq using StockR for the narrow-barred Spanish mackerel *Scomberomorus commerson* from the NEI and ECI subset assuming 2-6 genetic groups (top panels).



Appendix 4c

Individual length frequencies (bottom panel) and results of population structure analysis of DArTSeq using StockR for the narrow-barred Spanish mackerel *Scomberomorus commerson* from the AFS and WCS subset assuming 2-6 genetic groups (top panels).



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

CSIRO. Unlocking a better future for everyone.

Contact us

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

For further information

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