Australia's National Science Agency

Genetic population connectivity of yellowfin tuna in the Indian Ocean from the PSTBS-IO Project

Peter Grewe, Pierre Feutry, Scott Foster, Jorden Aulich, Matt Lansdell, Scott Cooper, Naomi Clear, Jessica Farley, Natacha Nikolic, Iñigo Krug, Iñaki Mendibil, Mohamed Ahusan, Denham Parker, Wudianto, Toni Ruchimat, Fayakun Satria, Pratiwi Lestari, Muhammad Taufik, Daniel Fernando, Asep Priatna, Achmad Zamroni, Naiara Rodríguez-Ezpelet, Iraide Artetxe-Arrate, Zulkarnaen Fahmi, Hilario Murua, Francis Marsac, Campbell Davies

28 Sept 2020

Paper submitted to the 10th meeting of the Working Party on Tropical Tuna

Abstract

Yellowfin tuna are a high value pantropically distributed tuna species managed as a single stock within the Indian Ocean. While studies to date have not provided evidence that a revision to this single stock assumption is warranted, further exploring and understanding the level of population heterogeneity is a priority for sustainable management of these fisheries. This paper presents results from a recent investigation of population structure of yellowfin tuna using cutting-edge sequencing technology as part of a larger collaborative project "Population Structure of IOTC species and sharks of interest in the Indian Ocean (PSTBS-IO)". A total of 1206 individuals from 9 Indian Ocean areas and two outlier locations (east Atlantic Ocean and south-west Pacific Ocean) were collected. 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, matched to the intended sampling design of the study, were chosen to be sequenced using DArTSeq and included in the analysis of population structure and examination of population connectivity. 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. Samples collected for this project, which are representative of major fishing areas within the Indian Ocean, were 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 may 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. In addition, 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.

Introduction

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 within the Indian Ocean (western and two eastern) 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 genetic studies, which have examined a variety of 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 within the Indian Ocean and supported conclusions of these earlier studies that were consistent with presence of a single stock of yellowfin tuna (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; Proctor 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 basin-scale sampling design to further investigate temporal and spatial population differentiation within the Indian Ocean.

Methods

Sampling

Tissues from individual fish were collected following a sampling design that broadly covered opposing corner reqions within the Indian Ocean taking two sets of temporally space samples at each location between 2017 and 2019 (Figure 1). Biopsied tissue samples were immediately preserved in RNA-Later following protocols developed for the project (Anon 2018). All samples were shipped to the CSIRO labs in Hobart Australia.

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 *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 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 Table 1.

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; Anon., 2018). This statistical method aims to find the groups of individuals within which the genetic profiles are more similar than those between groups. 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

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. The samples consisted of a mix of young of year (YoY) fish and mature adults collected from locations that closely matched the intended sampling design requirements of our study, which was to collect two temporally separated sets of sampled individuals from strategic locations roughly located at four opposing corners of the Indian Ocean (Figure 1). DNA extractions from a subset of the total collected individuals (n=664) were sent to

Diversity Arrays Technology (DArT) and sequenced using their proprietary DArTSeq™ technology (for details see Davies et al., 2020).

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 showing evidence of DNA cross contamination (for details see Davies et al., 2020), 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 (Table 1; Figure 1).

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 was the most parsimonious solution based on its AIC value (Figure 2). Furthermore, the distribution of individuals belonging to each of K genetic groups (K2, K3, K4) appeared to have a non-random geographic partitioning (**Error! Reference source not found.** 5). 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-non-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 (Figure 5). Notably, the group membership plots for the K = 2, 3 and 4 consistently show that NWI fish are relatively genetically homogeneous with all 12 fish having higher than 95% assignment probability to single divergent group (Figure 5). 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://rdrr.io/github/thierrygosselin/assigner/man/fst_WC84.html), also supports the uniqueness of the fish from NWI (Figure 4**Error! Reference source not found.**, a). The resulting F_{st} dendrogram demonstrates substantial differentiation between NWI and the rest of Indian Ocean sampling locations (**Error! Reference source not found.**gure 4, 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 SEI) 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 (Figure 5). 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 500 km 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 SWI and SAF southern adult locations (1 out of 67 or 1.6%).

Examination of the Atlantic Ocean and Pacific Ocean outgroup populations indicated high relative levels of differentiation between Ocean basins. Notably, Fst values indicated the lowest degree of differentiation was detected between the SAF and SWI locations but these were found to be substantially different from the Atlantic Ocean sample (Figure 4). Interestingly the NWI (sampled from Pakistan) showed the highest degree of intra-Indian Ocean differentiation (Figure 4,a).

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 could potentially indicate a unique reproductively isolated group of fish present within this region. The characterisation of both nuclear (current study) and mtDNA markers within this region has previously provided 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 observed in the current study and among previously sampled locations (2013 and 2014) as reported by Proctor (et al., 2019) demonstrated the occurrence of a temporally 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 and southward along the coast of Java. Further evidence of this 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 cline was not apparent among samples from locations south of the equator but rather the spatial distribution for K=2 genetic groups suggests very restricted mixing between regions sampled north and south of the equator. These results are consistent with previous studies of Barth (et al., 2017) and provide evidence of a third genetically partitioned region of Indian Ocean yellowfin tuna. 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 a plausible biological rationale to warrant consideration of results for models of K≥3 groups. Further spatially targeted and temporally stratified sampling of reproductively active fish is required to test this latter hypothesis of K>2 stocks and assess its intra-annual stability.

Localised environmental selection is one explanation for 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 differentiation of an otherwise reproductively panmictic population as has been demonstrated for two species anguillid eels (Gagnaire et al., 2009; Ulrik et al., 2014). In the case of 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 as a barrier to specific genotypes for 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, which 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 potential physiological barriers which may exert positive selectivity for individuals genetically capable of surviving such a highly variable environment stresses (e.g. oxygen deficit regions; Davies et al., 2020). 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 may be 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 the competing hypotheses of whether the observed patterns of genetic differentiation represent a) differential selection or b) 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 the genetically differentiated 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 suggests 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). 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 (**Error! Reference source not found.**). 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 for contemporary time scales and that this has sufficiently promoted genetic differentiation among the Atlantic, Indian, and Pacific Oceans populations of this species.

Acknowledgements

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;
- 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;
- 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;

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

Bibliography

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

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, Floriaan Devloo-Delva, Clément Rougeux, Denham Parker, Natalia Diaz-Arce, Wudianto, Toni Ruchimat, Fayakun Satria, Pratiwi Lestari, Muhammad Taufik, Asep Priatna, Achmad Zamroni. 2020. Population Structure of IOTC species and sharks of interest in the Indian Ocean: Estimation with Next Generation Sequencing Technologies and Otolith Micro-chemistry. Final Report to IOTC Secretariat. May 2020. 228 pp.

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.

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

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

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

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

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

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.

Kunarso, Ario, R., Munandar, B., & Prayogi, H. (2018). Impact of Monsoon to Aquatic Productivity and Fish Landing at Pesawaran Regency Waters. *IOP Conference Series: Earth and Environmental Science*, *116*, 012021. doi:10.1088/1755-1315/116/1/012021

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.

Marra, J., & Moore, T. (2009). Monsoons, Islands, and Eddies: Their Effects on Phytoplankton in the Indian Ocean. *Washington DC American Geophysical Union Geophysical Monograph Series*, *185*, 57-70. doi:10.1029/2008GM000701

Moore, Bradley & Lesteri, Pratiwi & Cutmore, Scott & Proctor, Craig & Lester, Robert. (2019). Movement of juvenile tuna deduced from parasite data. ICES Journal of Marine Science. 10.1093/icesjms/fsz022.

Mohri, Masahiko & Nishida, Tom. (2000). Consideration on distribution of adult yellowfin tuna (Thunnus albacares) in the Indian Ocean based on Japanese tuna longline fisheries and survey information. IOTC Proc.. 3.

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

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

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.

Rajapaksha, Jagath & Samarakoon, L & Gunatilake, Jagath. (2013). Environmental Preferences of Yellowfin Tuna in the North East Indian Ocean: An Application of Satellite Data to Longline Catches. International Journal of Fisheries and Aquatic Sciences. 2. 72-80.

Sharp, G. D. (1978). Behavioral 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

Ulrik, M.G., Pujolar, J.M., Ferchaud, A. *et al.* Do North Atlantic eels show parallel patterns of spatially varying selection?. *BMC Evol Biol* **14,** 138 (2014). https://doi.org/10.1186/1471-2148-14-138

Tables:

Table 1. 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.

			Numbers remaining post filter step		
Applied FILTER step	Radiator PARAMETERS	Chosen VALUES	individuals	loci	markers
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 ld	filter.short.ld	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

Figures:

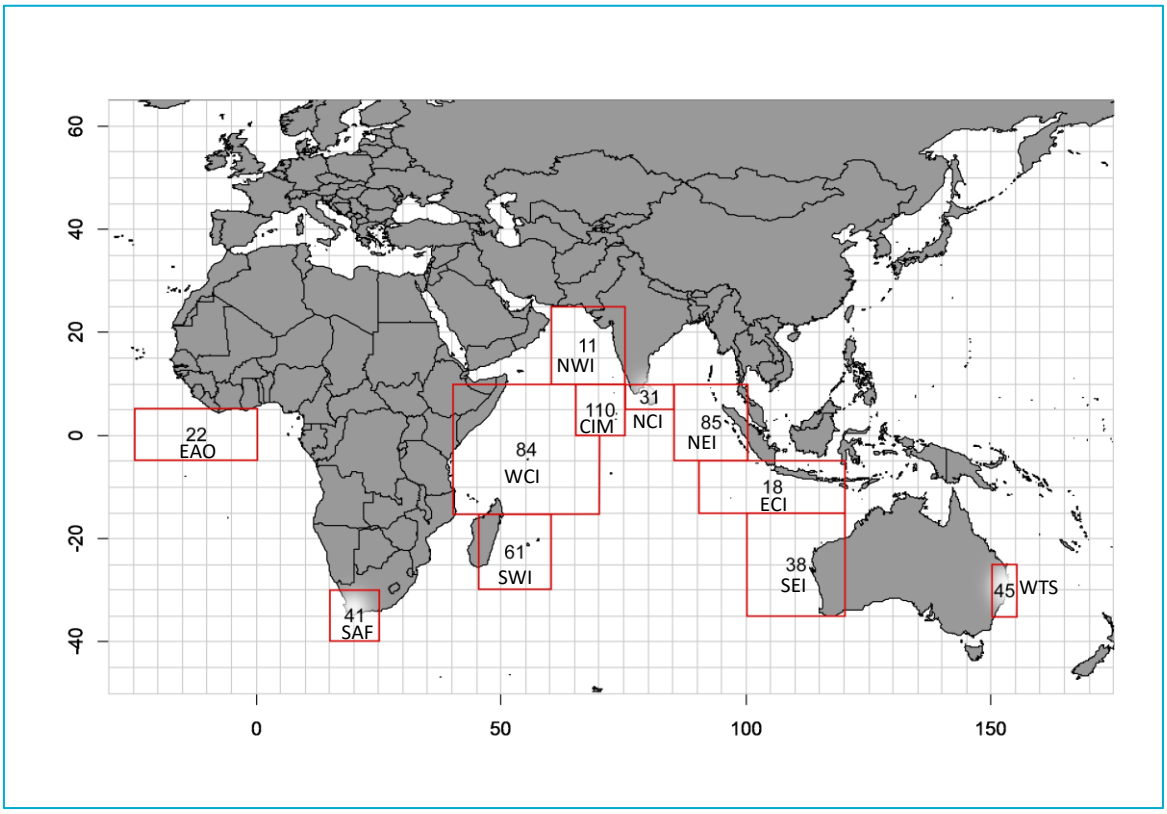


Figure 1. Sample locations and numbers of yellowfin tuna passing initial DNA quality checks and analysed for population genetics in this study. East Atlantic Ocean (EAO,n=22), South Africa (SAF, n=41), Southwest Indian (SWI, n=61), West Central Indian (WCI, n=84), North West Indian (NWI, n=11), Maldives (CIM, n=110), Sri Lanka (NCI, n=31), Lampulo (NEI, n=85), east central Indian (ECI, n=18), Perth (SEI, n=38), east Tasman Sea (ETS, n=45) and numbers for yellowfin tuna.

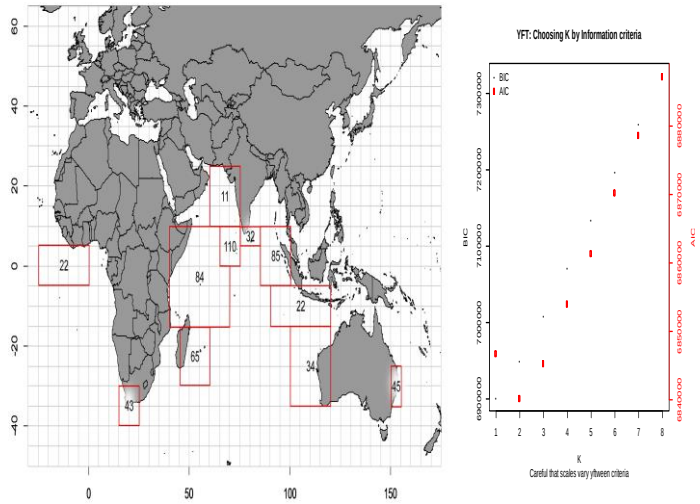


Figure 2 1. Left: Distribution of samples (N= 664) of yellowfin tuna (*Thunnus albacares*) for both rounds of sampling sequenced using DArTSeq by sampling region for PSTBS-IO project. Right: Information criterion used to assess the likelihood of different numbers of genetic groups (K).

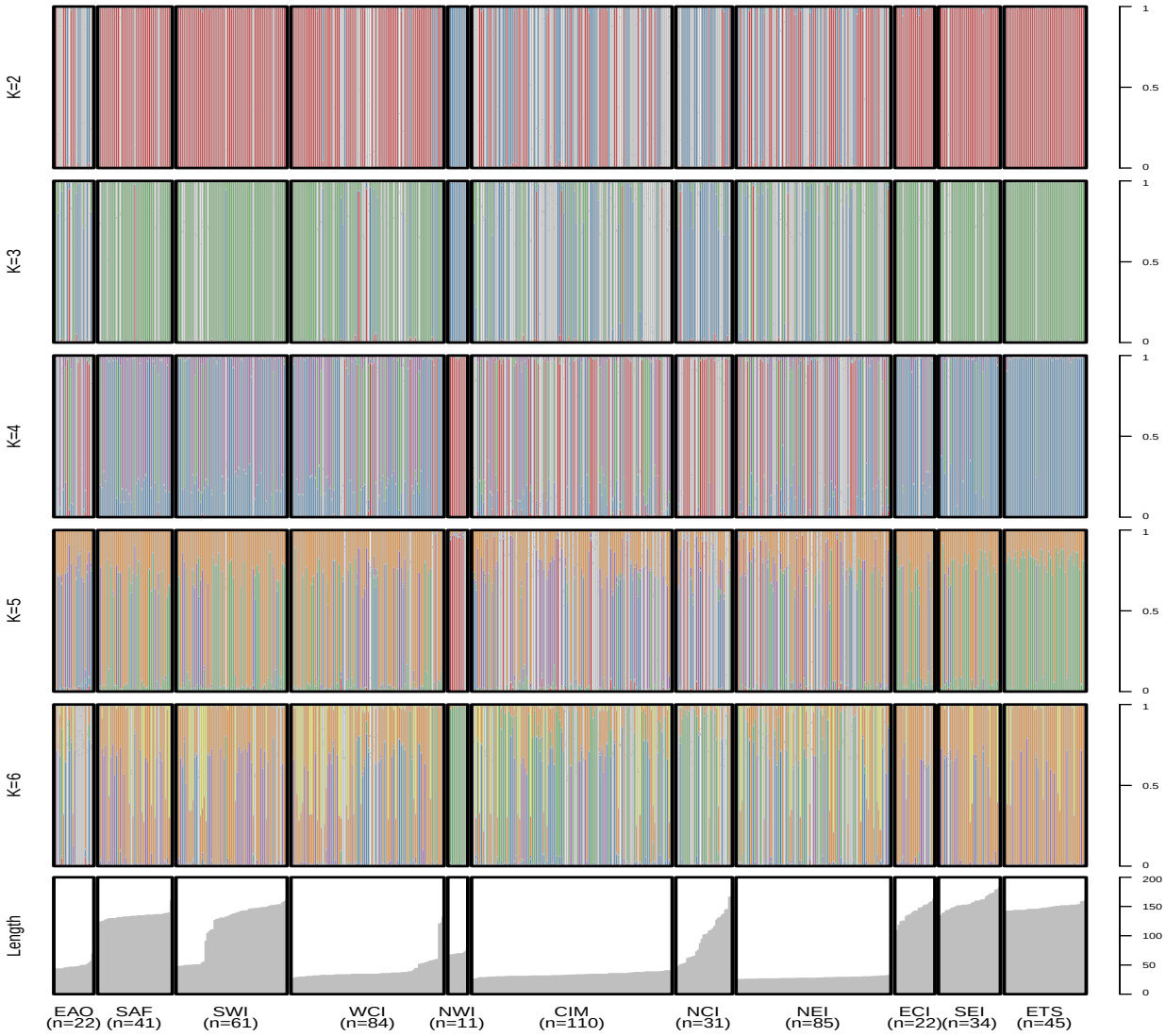


Figure 3. Membership probability of an individual belonging to one of K genetic groupings. Each vertical bar represents an individual genotypic profile with white profiles representing individuals where there is less than 80% certainty of belonging to a K genetic group. Individuals are sorted by lengths 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.

(a) (b)

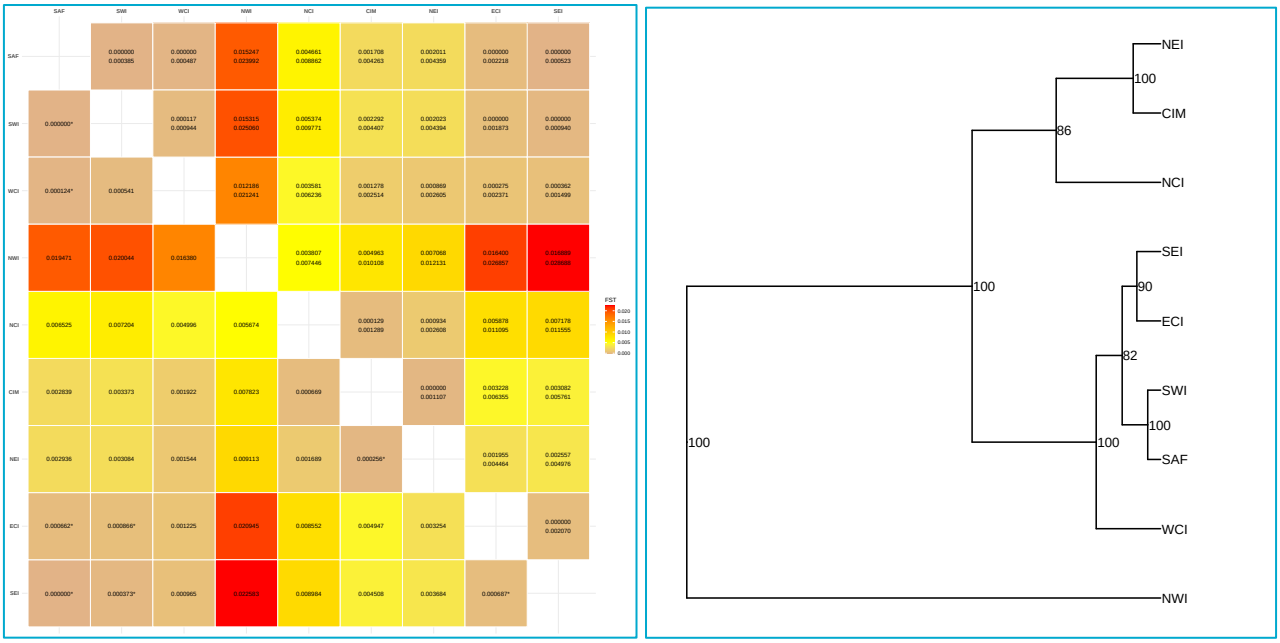
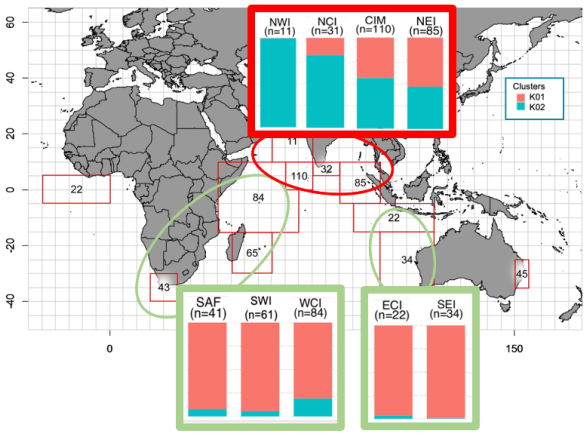
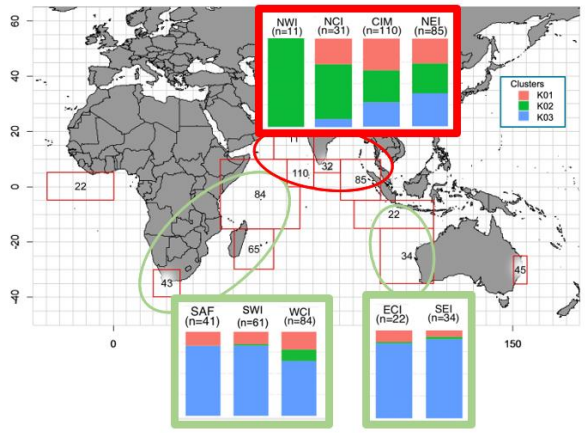


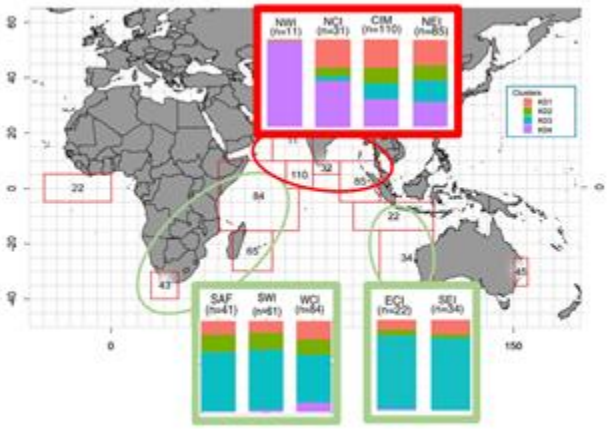
Figure 4 (a) Heat map and matrix of relative F_{st} values (below the diagonal) and minimum and maximum ranges (above the diagonal) as calculated among pairs of Indian Ocean sample sites. Colour indicates degree of relative differentiation between pairs of individuals with low (<0.005, beige), medium (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).



K=2



K = 3



K = 4

Figure 5. Cumulative probabilities plotted at sample site locations for various models of K genetic groups. (a) K=2, (b) K=3, (c) K=4.

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