SCIENTIFIC COMMITTEE
FIFTEENTH REGULAR SESSION

Pohnpei, Federated States of Micronesia
12-20 August 2019

Population Structure and Connectivity of Tropical Tuna Species across the Indo Pacific Ocean Region

WCPFC-SC15-2019/SA-IP-15

P.M. Grewe, Wudianto, C.H. Proctor, M.S. Adam, A.R. Jauhary, K. Schafer, D. Itano, K. Evans, A. Killian, S. Foster, T. Gosselin, P. Feutry, J. Aulich, R. Gunasekera, M. Lansdell and C.R. Davies.

Population Structure and Connectivity of Tropical Tuna Species across the Indo Pacific Ocean Region

Authors: P.M. Grewe[1], Wudianto[2], C.H. Proctor[1], M.S. Adam[6], A.R. Jauhary[6], K. Schafer[3], D. Itano[4], K. Evans[1], A. Killian[5], S. Foster[7], T. Gosselin[8], P. Feutry[1], J. Aulich[1], R. Gunasekera[1], M. Lansdell[1], C.R. Davies[1].

[1]CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart Tasmania 7001, Australia
[2]Center for Fisheries Research, Jakarta, Indonesia,
[3]Inter-American Tropical Tuna Commission, La Jolla, California, USA.
[4]689 Kaumakani Street, Honolulu Hawaii. 96825, USA.
[5]Diversity Arrays Technology, University of Canberra, Bruce, ACT, Australia
[6]Marine Research Centre, Ministry of Fisheries and Agriculture, Male, Maldives.
[7]Data 61, CSIRO, GPO Box 1538, Hobart Tasmania 7001, Australia
[8]Department of Biology, Laval University, Quebec, Canada.

Abstract

Current stock assessments and management arrangements for tropical tuna (bigeye tuna, skipjack tuna, and yellowfin tuna) in the Indian and Pacific Oceans assume large panmictic stocks confined to boundaries within regional fishery management organisations (RFMOs). More specifically, they assume a single, well mixed spawning stock generates the recruitment for subsequent generations across the region of the assessment. Recent advances in Next-Generation-Sequencing methods have been used to provide greater genetic resolution to address questions of stock structure and provenance. Results from early studies using these methods for yellowfin tuna (e.g. Grewe et al., 2015) suggested there may be more structure in populations of tropical tuna than previously thought. In this paper we summarise results from recently completed research that provides new insights into genetic structure in bigeye tuna, skipjack tuna, and yellowfin tuna across the Indo Pacific region. Comparison of data among these three tuna species suggests genetic population differentiation of skipjack tuna appears to be weaker among current sampling sites collected from the central Indian Ocean through to the Eastern Pacific Ocean suggesting more restricted connectivity exists for both BET and YFT species across the same region. None the less, presence of reduced geneflow across the Indo Pacific and Tropical Eastern Pacific regions demonstrate that the population structure of these species is more complex than previously thought. Additional population genetic analyses and more comprehensive sampling are required to provide full spatial coverage and confirm (or not) the temporal stability of these results. This additional analysis should shed more light on competing categories of processes that could be responsible for observed differentiation such as: i) isolation by distance; and ii) stronger bio-geographic barriers or loci under differential selection among areas in a way that is consistent with global patterns in biogeography. Improved management for these species should benefit from improved

understanding of stock structure and connectivity in these species but this will require further investigation with careful consideration regarding extension of sampling coverage (both temporal and spatial), quality control of tissue collection, and consistent application of genetic methods.

Introduction

Examination of stock connectivity in the Indo Pacific region for bigeye tuna (BET), *Thunnus obesus*, skipjack tuna (SKJ), *Katsuwonus pelamis,* and yellowfin tuna (YFT), *T. albacares*, through analyses of genetic population structuring has often led to reporting of ambiguous results. Researchers have either concluded lack of genetic differentiation for these species on large geographic scales (thousands of kilometres) or the finding of population subdivision in quite limited geographical range on the scale of hundreds of kilometres. Past studies, which have used a variety of genetic marker techniques (including protein electrophoresis, mitochondrial DNA, DNA microsatellites), have predominantly cited a combination of three general reasons to explain inconclusive results (i.e. lack of genetic differentiation): i) inadequate numbers of individuals collected from each sample site; ii) lack of resolution due to the type of DNA markers used in the analysis; or iii) presence of sufficient gene flow among sampling locations that prevents development of significant population differentiation. The approaches used in this project aimed to overcome the shortcomings of previous studies by initially addressing the first of these two points that appear to be mainly technical issues as a result of financial constraints, genetic marker resolution, or insufficient geographic and temporal sampling coverage. We chose to use single nucleotide polymorphism (SNP) markers that have demonstrated excellent potential for revealing subtle population differentiation when present, where previous techniques of mitochondrial DNA and DNA microsatellite had failed to do so. Through examination of widely separated outlier sample sites, we also hoped to shed light on whether sufficient gene flow was preventing population differentiation between regions where the potential for connectivity (e.g. between ocean basins) was suspected to be much lower than that expected across shorter distance such as within the Indonesian archipelago and nearby areas of the Indian and Pacific Oceans.

Methods

Sample collection, extraction, and DNA sequencing

Tissue samples of BET, SKJ, and YFT were collected from sites across their nominal species ranges with a focus on locations covering the Indo Pacific and Tropical Pacific biogeographic regions (Table 1; Figure 1). For each of the species, a broad size range of individuals was sampled in order to represent multiple year-classes at each location. White muscle was biopsied next to the main dorsal fin and transported to the lab preserved in RNAlater® (Life Technologies). Approximately 15mg of tissue was subsampled from each biopsy and used for individual DNA extractions. Total genomic DNA was isolated using one of two protocols; either a Machery Nagel Nucleo-Mag bead-based DNA isolation kit or a CTAB protocol, a

Phenol- Chloroform based method described by Grewe *et al*. (1993). The magnetic bead-based extractions were performed on an Eppendorf EP-Motion-5075 robotic liquid handling station while CTAB preps were manually processed. DNA quality of each extract was visually inspected using 0.8% agarose electrophoresis gels as a first-pass qualitative check. Samples assessed as containing sufficient amounts of high molecular weight DNA were then checked to confirm taxonomic species assignments had been correctly identified during field sampling. Genetic species identification was done using a testing procedure that examined restriction digests of a mitochondrial DNA PCR amplicon (PCR-RFLP) following protocols described by Chow and Inoue (1993) with further modifications described by Takayama et al. (2001). Samples passing all QC steps were shipped to Diversity Array Technologies in Canberra for genomic sequencing using their proprietary DArTseq™ methodology (for detailed description see Grewe et al., 2015). Approximately 2,000,000 sequences per barcode/sample were identified and fed into a proprietary algorithm (DArTsoft14) to call single nucleotide polymorphism (SNP) genotypes.

Table 1. Location, site codes, and numbers of bigeye tuna (BET), skipjack tuna (SKJ), and yellowfin tuna (YFT) chosen for the genetic analysis.

Location	Site Code	BET	SKJ	YFT
Gulf of Mexico	ATL	-	-	48
Maldives	MAL	92	76	96
Padang	PAD	96	-	96
Palabuhanratu	PAL	96	-	96
Prigi	PRI	84	-	109
Ambon	AMBB	96	-	96
Kendari	KEN	96	-	96
Gorontalo	GOR	49	-	134
Bitung	BIT	27	-	161
Sorong	SOR	79	-	109
Jayapura	JAY	50	-	138
Bismarck Sea	BIS	-	24	-
Solomon Islands	SOL	96	-	92
Coral Sea	MOO	-	37	116
Tasmania	TAS	-	19	-
Tokelau	TOK	64	-	64
Hawaii	HAW	-	-	66
HCP	-	24	-	-
CPO	-	24	-	-
Baja California	BAJ	-	-	64
Eastern Pacific	EPO	48	79	-

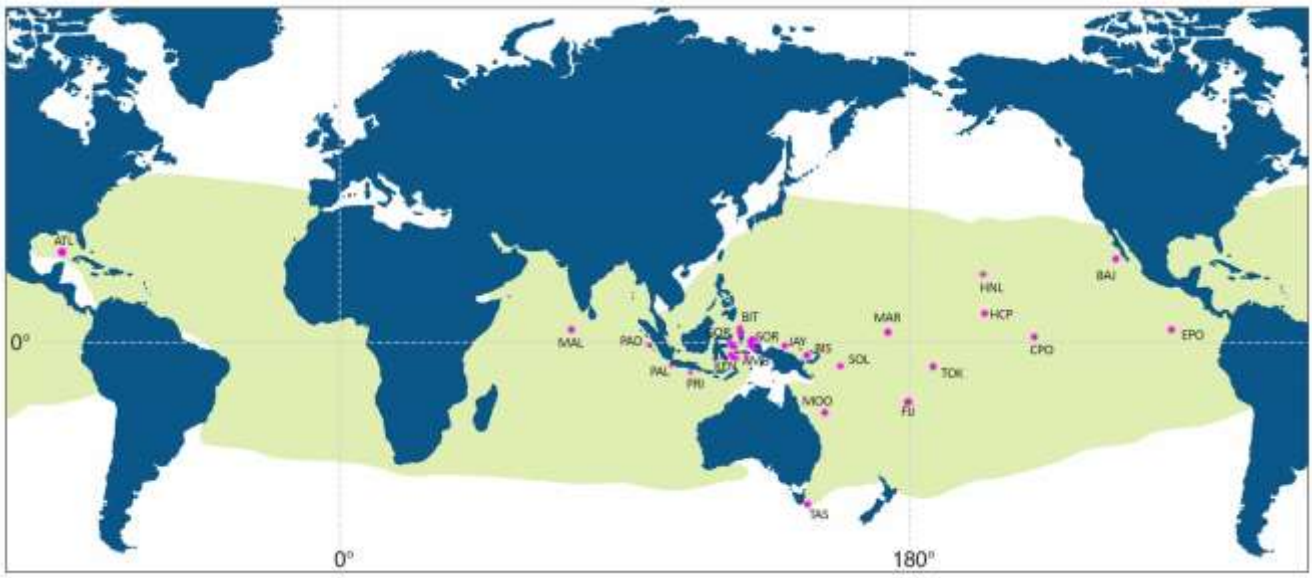

Figure 1. Sampling locations (pink dots) for BET, SKJ, and YFT with sampling site codes listed in Table 1. Green shading highlights approximate species range.

Genotyping quality control and population analysis

A step wise process for data quality control of the DArTseq™ data was performed at two levels using the package RADIATOR to retain: i) good quality DNA loci (i.e. SNP markers deviating from expected metrics); and ii) individuals with good quality DNA (Gosselin 2017). Filters criteria used to check DNA loci included: i) marker reproducibility (using comparison of technical replicates); ii) identification of common markers (these are markers that are present among all individuals); iii) minor allele counts (elimination of sequencing artefacts); iv) minimum and maximum read depth (a reliability index of DNA quality and also identification repetitive DNA genes – for example junk DNA in the genome); v) the number of counts at a locus (ensures SNPs are only from unique chromosome locations); and vi) and whether loci comply with assumption of Hardy Weinberg equilibrium (Andrews 2010). Individuals were assessed against three genetic criteria: i) missing data; ii) average heterozygosity; iii) highly similar/duplicate genotypes. Individuals with higher than 10% missing data or those falling outside specific 95% confidence range of average heterozygosity were removed from further population analysis. Individuals with less than 10% sequence divergence were deemed to be technical replicates or repeated tissue samples. Only one individual from each of the replicate pairs was retained for subsequent population analysis.

The resulting genotype data set was analysed using two different methods. The first examined F_{ST} genetic relatedness to estimate and define genetically related distance between pairs of sites. The second method used an R-package program (stockR), that uses a robust strategy for identifying genetically distinct groups or putative stocks within the data set (Foster et al. 2018). The methods implemented in stockR are directly targeted at finding breeding groups, unlike some commonly-used statistical methods. Genetic relatedness among all individuals was used to cluster them into “K” different groups via a particular

statistical mixture model, using a latent variable for the group membership of each individual fish. The most pertinent output of the modelling was an assignment of each fish to one of "K" genetically related groups over a range of seven different levels (K=2 to 8 groups). The proportions of the different K-groups at each sampling location were compared to overall spatial distribution of the different defined genetic groups using a cross-validation approach that estimated percentage of likely correct assignment to each "K" cluster group. The higher the percentage of correct assignments, the more support the data has for that level of grouping. The cross-validation was performed by examining the genetic group that each individual was assigned using the full data set, and then to hold a random subset of individuals out and re-performing the analysis (we held out a fifth of the data at each location). A total of 100 hold-out sets were evaluated for each number of groups (K). The cross-validation statistic was defined to be the number of hold-out fish that were re-assigned to the same group. The number of correctly reassigned fish is measured by the sum of the posterior probability of group membership. This process was performed for each of K=2 to K=8 putative genetic groups.

Results and Discussion

Sampling coverage achieved for the three species (BET, SKJ, YFT) examined in this study have provided a broad perspective on the connectivity and genetic differentiation across a significant proportion of their distributions from the Maldives in the central Indian through to sites in the Eastern Pacific Ocean (Figure 1). Analysis of SNP data performed well and higlighted a level of population structure not resolved by previously used techniques including protein electrophoresis, mitochondrial DNA, and DNA microsatellites. While some techniques may still prove to be useful in certain contexts (e.g. mtDNA for species ID and DNA microsatellites for individual ID), the throughput capacity (tens of individuals per sampling location) and indeed resolving power of these latter techniques (dozens of markers) for revealing population discrimination has now been clearly outflanked by new genomic sequencing techniques such DArTseq™ (hundreds of individuals and several thousands of markers). Development of statistical approaches specifically tailored to take advantage of this genomics-based fishery independent data can now address a number of fishery management related needs such as population census size via Close-kin Mark-Recapture studies, connectivity assessment through population discrimination, and year class strength though gene-tagging (Bravington et al, 2016; Grewe et al., 2015; Preece et al., 2018).

Sample sizes analysed for YFT ranged from 48 to 161 fish per site (average 91 per site) across 18 sampling locations. Achieving target samples sizes for the BET analysis proved to be more challenging due to limited occurrence of this species in the overall catches at some locations. Despite this, overall sampling effort was considered sufficient over 12 locations with 27 to 96 fish per location (average of 77 per location). Sampling effort for skipjack tuna across the Indo Pacific and Tropical Pacific regions was less intensive between 19 to 79 per site (average of 47 fish per site) at five locations, but provided sufficient coverage to permit broad scale analysis.

While significant differentiation was present for skipjack between Indo Pacific (Maldives, Bismarck Sea, Coral Sea, Tasmania) and the Tropical Pacific (EPO) sampling locations, it was not detected among samples specifically within the Indo Pacific region. In contrast, a more complex picture was present for YFT and BET that indicated strong differentiation between Indo and Tropical Pacific regions as well as evidence of weak population structure among the Indo Pacific sample locations (Maldives through to Central Pacific). Strong population differentiation was also observed at the ocean basin scale for both BET and YFT samples representing Atlantic, Indian, and Pacific Oceans.

The CSIRO O&A group studying genomic applications for managing highly migratory pelagic tuna and billfish species have demonstrated the utility of SNP markers to reveal evidence for both spatially and temporally stable population structure (Grewe et al., 2015; Proctor et al., 2019; Evans et al., 2019). Population differentiation was more pronounced among populations of BET and YFT than that observed for SKJ and sample sites in close proximity to each other appeared to be more genetically similar than those separated at inter ocean scale distances. The reasons for the levels of observed differentiation require further investigation with appropriate sampling designs that achieve finer scale coverage through an increased number of sampling locations as well as through targeted sampling fish from representative spawning aggregations. In addition to the broad scale population structure that has emerged for BET, SKJ, and YFT, our study has also provided a number of valuable lessons that can help guide design of future investigations for these species:

- Temporally stratified sampling design **is essential**
  (e.g. ripe/running adults, YOY, or larvae if possible)
- Cross-contamination of tissue samples can be common. Rigorous sampling protocols, training and monitoring is essential.
- Rigorous QC of large data sets to identify & filter out:
  - poor data resulting from bad DNA/loci,
  - problematic cross contaminated samples,
  - sequencing platform and batch effects.
  - retain only one individual from a group of siblings (FSP/HSP)
- Use of appropriate analysis approaches & packages
  examples – RADIATOR (Gosselin 2019); ADEGENET (Jombard and Ahmed, 2011); STOCK-R (Foster et al., 2019)

Future Directions

Substantial progress has been made, which has highlighted complex population structure in BET, SKJ and YFT using SNP markers, but, further sampling is still required to fill gaps in distribution of each species. In collaboration with other institutes from Spain (AZTI), and France (IRD), and Indonesia (CFR), as well as a number of coastal states, we are currently completing a major sampling and analysis effort across several other tropical and neritic

tuna species (including BET, SKJ and YFT) within the Indian Ocean along with additional outlier populations from the Atlantic and Pacific Oceans. The key to progress has been a systematic sampling effort to obtain ocean basin scale coverage across the range of each of these species. Delivering the outcome for this goal has required global partnerships and collaboration from a number of member countries involved that have historically and continued to fish these species.

Some of the essential issues faced for completing basin scale studies of stock structure include:

- Identification of specific questions relevant to management needs
- Focus on rigorous sampling design and based on current experience:
  - targeted and well-coordinated sampling strategy
  - address spatial and temporal stability using 2 or more years of sampling effort
- Standard operating procedures and attention to QC of sample collection is essential
  -targeted sampling of life-history stages (e.g. YoY and Spawning adults)
- Robust inferences for stock assessment and management purposes requires broad scale sampling coverage
- Capacity building, during project development and implementation.

Several studies that have been recently completed (by us and others), have demonstrated the power of these new genomic sequencing methods. In any future application of this technology, we strongly emphasize the need for carefully designed and implemented sampling designs to provide high quality robust information on population structure, mixed stock catch compositions required to inform stock assessments and implement more rigorous catch and market monitoring systems.

References

Appleyard S., Grewe P., Innes B., and Ward R. (2001). Population structure of yellowfin tuna (*Thunnus albacares*) in the western Pacific Ocean, inferred from microsatellite loci. Marine Biology 139, 383- 393.

Bravington, M. V., Grewe, P. M., & Davies, C. R. (2016). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat Commun, 7*, 13162. doi:10.1038/ncomms13162

Chow, S. & Inoue, S. (1993). Intra- and interspecific restriction fragment length poly-morphism in mitochondrial genes of Thunnus tuna species. Bulletin of National Research Institute of Far Seas Fisheries 30, 207 – 225.

Evans, KE, Grewe, P.M., Scott Foster, S., Gosselin, T., Gunasekera, R., Matt Lansdell, M. 2019. Connectivity of tuna and billfish species targeted by the Australian Eastern Tuna

and Billfish Fishery with the broader Western Pacific Ocean. WCPFC-SC15-2019/SA-IP-13

Foster S.D., Feutry P., Grewe P., Berry O., Hui F. K. C., and Davies C. (2018). Reliably discriminating stock structure with genetic markers: Mixture models with robust and fast computation. . Mol Ecol Resour. 2018;00:1–16. https://doi.org/10.1111/1755-0998.12920

Gosselin T (2019). *radiator: RADseq Data Exploration, Manipulation and Visualization using R*. doi: 10.5281/zenodo.1475182, https://thierrygosselin.github.io/radiator/.

Grewe, Peter, Feutry, Pierre; Hill, Peta; Gunasekera, Rasanthi; Foster, Scott; Schaefer, Kurt; et al. Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports*. 2015; 5(11):9.

Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics.<doi:10.1093/bioinformatics/btr521>

Preece, A. and Braford, R.W. 2018. Gene-tagging recruitment monitoring in 2018. Progress report and RMA request for gene-tagging in 2019. CCSBT-ESC/1809/07, 10pp.

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.

Takeyama, H., Chow, S., Tsuzuki, H. & Matsunaga, T. (2001). Mitochondrial DNA sequence variation within and between Thunnus tuna species and its application to species identification. Journal of Fish Biology 58, 1646 – 1657. doi: 10.1006/ jfbi.2001.1575