# Evidence of connectivity of bigeye tuna (*Thunnus obesus*) throughout the Indian Ocean inferred from genome-wide genetic markers

Natalia Díaz-Arce[1], Peter Grewe[2], Iñigo Krug[1], Iraide Artetxe[1], Jon Ruiz[1], Natacha Nikolic[3,#], Anais Medieu[3], Marianne Pernak[3], Matt Lansdell[2], Jorden Aulich[2], Naomi Clear[2], Craig Proctor[2], Wudianto[4], Toni Ruchimat[4], Zulkarnaen Fahmi[4], Fayakun Satria[4], Pratiwi Lestari[4], Muhammad Taufik[4], Asep Priatna[4], Achmad Zamroni[4], Jessica Farley[3], Campbell R. Davies[2], Francis Marsac[3], Igaratza Fraile[1], Hilario Murua[5], Naiara Rodríguez-Ezpeleta[1,*]

[1]Marine Research, AZTI BRTA, Sukarrieta, Bizkaia, Spain

[2] CSIRO Oceans and Atmosphere, Hobart; Tasmania, Australia Marbec, Univ

[3] Montpellier, CNRS, Ifremer, IRD, Montpellier, France

[4] Agency for Marine and Fisheries Research and Human Resources

[5] International Seafood Sustainability Foundation, Washington DC, USA.

[#] Present address: INRAE, Saint-Pée-sur-Nivelle, France

*Correspondence: nrodriguez@azti.es

**Abstract**

In order to resolve the population connectivity of bigeye tuna (*Thunnus obesus*) within the Indian Ocean, we analyzed thousands of genome-wide markers of individuals from a broad geographic area of the Indian Ocean as well as from locations in the Pacific and Atlantic Oceans. Our results support a single panmictic population of bigeye tuna within the Indian Ocean isolated from the Atlantic and Pacific Oceans.

## Introduction

Previous studies on the population structure of bigeye tuna (*Thunnus obesus*) support inter-oceanic genetic separation (Alvarado Bremer et al. 1998; Chow et al. 2000; Durand et al. 2005; Martínez et al. 2006). However, the few studies that applied genetic methods to understanding population structure of this species within the Indian Ocean did not observe signs of heterogeneity, supporting the existence of a single panmictic

population (Appleyard, Ward, Grewe 2002; Chiang et al. 2008). In order to resolve population structure of bigeye tuna within the Indian Ocean, we have analyzed thousands of Single Nucleotide Polymorphisms (SNP) markers from individuals collected throughout the distribution of the species.

## Materials and methods

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

Bigeye tuna muscle tissue samples were obtained from a broad geographic area of the Indian Ocean, together with additional samples from the Pacific and Atlantic Oceans were obtained, by scientific observers on-board purse seiner vessels or by sampling fish during landings at port (Figure 1). In total, 496 fish were sampled from 10 areas (Maldives, Central Indian Ocean, East Central Indian Ocean, North Central Indian Ocean, North East Indian Ocean, South East Indian Ocean, South West Indian Ocean, West Central Indian Ocean, West Tasman Sea and East Atlantic Ocean). Samples were from fish classified by Straight Fork Length (SFT) as young of the year (YoY) (<45 cm), juveniles (45-120 cm) and adults (>120 cm) according to according to previously described age-length relationship (Eveson et al. 2015; Sardenne et al. 2015) and maturity thresholds (Zudaire et al. 2016). DNA was extracted from about 15 mg of tissue 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. DNA sample libraries were created in digestion/ligation reactions using two restriction enzymes, PstI and SphI. The PstI site was compatible with a forward adapter that included an Illumina flow cell attachment sequence and a sequencing primer sequence incorporating a "staggered", varying length barcode region. SphI- generated a compatible overhang sequence that was ligated to a reverse adapter containing a flow cell attachment region and reverse priming sequence. Only "mixed fragments" (PstI-SphI) were effectively amplified by PCR. PCR conditions consisted of an initial denaturation at 94°C for 1 min followed by 30 cycles of 94°C for 20 sec, 58°C for 30 sec and 72°C for 45 sec, with a final extension step at 72°C for 7 min. After PCR, equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to

cBot (Illumina) bridge PCR, followed by sequencing on an Illumina Hiseq2000. The sequencing (single read) was run for 77 cycles.

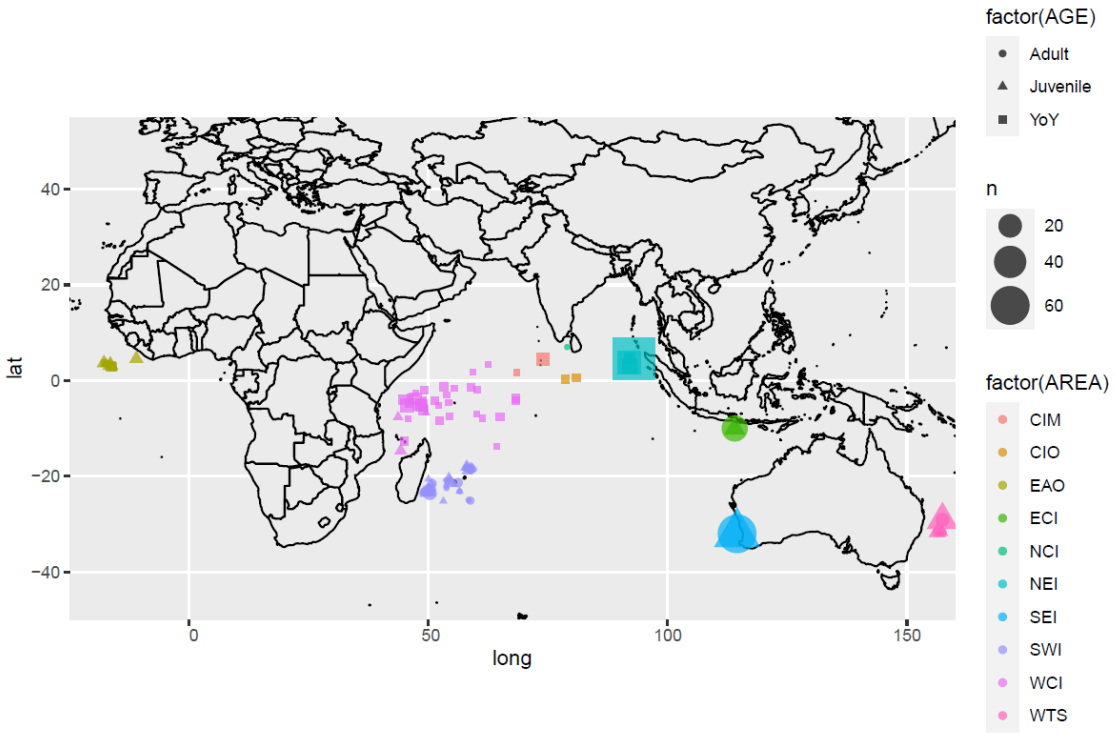


Fig. 1. Samples collected for this study. Each location is represented by one color (CIM - Maldives, CIO - Central Indian Ocean, ECI - East Central Indian Ocean, NCI - North Central Indian Ocean, NEI - North East Indian Ocean, SEI - South East Indian Ocean, SWI - South West Indian Ocean, WCI - West Central Indian Ocean, WTS - West Tasman Sea and EAO - East Atlantic Ocean) and shapes indicate whether samples are young of the year (YoY), juveniles or adults. Size of shapes are proportional to the number of samples collected per area/age.

*RAD-tag assembly and SNP calling*

Generate Dart-seq reads were analyzed using *Stacks* version 2.4 (Catchen et al. 2013). Using *process_radtags*, reads were truncated to 69bp so that all reads had the same length after barcode removal and reads with any uncalled base, with total low-quality scores or with quality score below 20 within 10bp size sliding windows were removed. The module *ustacks* was then used to assemble orthologous tags (stacks) per individual, with a minimum coverage depth required to create a stack (parameter -m) of 3, and a maximum nucleotide mismatches allowed between stacks (parameter -M) of 2. Matches to the catalog for each sample were searched using *sstacks* and transposed using

*tsv2bam* and the module *gstacks* was used to identifying and genotyping SNPs. Only samples with more than 30,000 and less than 65,000 reads were selected for further analyses. The module *populations* was used to export from the catalog, the SNPs presented in RAD loci found in at least 75% of the individuals. Using *PLINK* version 1.07 (Purcell et al. 2007), SNPs with more than 5% missing data and a minimum allele frequency (MAF) smaller than 0.05 or failing the Hardy-Weinberg equilibrium test ($p < 0.05$) in at least two location (excluding North Central Indian (NCI), Central Indian Ocean (CIO) and Maldives (CIM) locations which contained less than 10 individuals), as well as samples with more than 10% missing data were excluded from downstream analyses. Finally, only the first SNP per loci were kept and resulting genotype table was exported into *Structure* and *Genepop* formats. Related individuals were identified using GCTA (Yang et al. 2011).

*Genetic diversity and population structure analyses*

Related individuals were identified using GCTA (Yang et al. 2011). Principal Component Analysis (PCA) were performed using the *adegenet* R package (Jombart, Ahmed 2011), and ADMIXTURE (Alexander, Novembre, Lange 2009) was run assuming from 1 to 6 ancestral populations (K) setting default parameters. The value of K with lowest associated error value was identified using ADMIXTURE's cross-validation procedure.

**Results**

Principal component analyses (Figure 2) based on the filtered dataset consisting of 472 individuals and 6,093 SNPs show strong differentiation between samples from the three oceans, but do not suggest any intra-oceanic structure within the Indian Ocean, even when including only samples from the Indian Ocean. ADMIXTURE analyses show that the number of assumed ancestral populations with the lowest associated error (K) is 2 (Figure 3) and show three clusters corresponding to the three oceans.

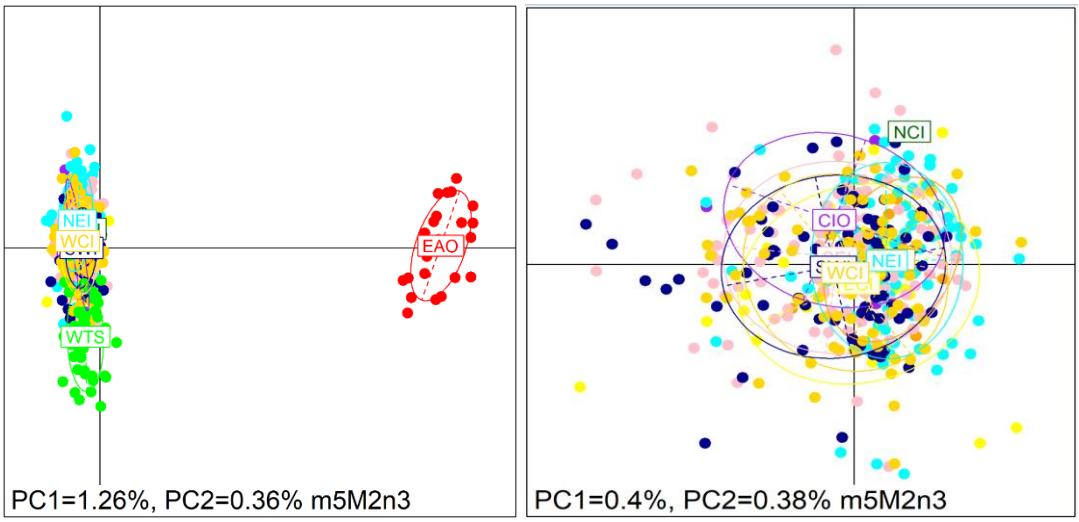


Fig. 2. Principal Component Analysis (PCA) performed using the final dataset containing all samples (left) or only the Indian Ocean samples (right). Different colors in the PCA represent samples from the different locations at which samples were captured.


Fig. 3. Individual ancestry proportions estimated by ADMIXTURE when assuming from two to four ancestral populations (K) when including all samples (top) or only Indian Ocean samples (bottom).

## Main conclusions

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

## Acknowledgements

This work is part of the PSTBS project supported by funding from CSIRO Oceans and Atmosphere, AZTI Basque Research and Technology Alliance, Institut de recherche pour le développement (IRD), and Research Institute for Tuna Fisheries (RITF) and financial assistance of the European Union (GCP/INT/233/EC – Population structure of IOTC species in the Indian Ocean). The views expressed herein can in no way be taken to reflect the official opinion of the European Union. We would like to thank the many vessel owners, skippers, observers, crews and processors who provided fish for sampling and access to facilities.

## References

Alexander, DH, J Novembre, K Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research 19:1655-1664.

Alvarado Bremer, JR, B Stequert, NW Robertson, B Ely. 1998. Genetic evidence for inter-oceanic subdivision of bigeye tuna (Thunnus obesus) populations. Marine Biology 132:547-557.

Appleyard, SA, RD Ward, PM Grewe. 2002. Genetic stock structure of bigeye tuna in the Indian Ocean using mitochondrial DNA and microsatellites. Journal of Fish Biology 60:767-770.

Catchen, J, PA Hohenlohe, S Bassham, A Amores, WA Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124-3140.

Chiang, H-C, C-C Hsu, GC-C Wu, S-K Chang, H-Y Yang. 2008. Population structure of bigeye tuna (Thunnus obesus) in the Indian Ocean inferred from mitochondrial DNA. Fisheries Research 90:305-312.

Chow, S, H Okamoto, N Miyabe, K Hiramatsu, N Barut. 2000. Genetic divergence between Atlantic and Indo-Pacific stocks of bigeye tuna (Thunnus obesus) and admixture around South Africa. Molecular Ecology 9:221-227.

Durand, JD, A Collet, S Chow, B Guinand, P Borsa. 2005. Nuclear and mitochondrial DNA markers indicate unidirectional gene flow of Indo-Pacific to Atlantic bigeye tuna (Thunnus obesus) populations, and their admixture off southern Africa. Marine Biology 147:313-322.

Eveson, JP, J Million, F Sardenne, G Le Croizier. 2015. Estimating growth of tropical tunas in the Indian Ocean using tag-recapture data and otolith-based age estimates. Fisheries Research 163:58-68.

Jombart, T, I Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070-3071.

Martínez, P, EG González, R Castilho, R Zardoya. 2006. Genetic diversity and historical demography of Atlantic bigeye tuna (Thunnus obesus). Molecular Phylogenetics and Evolution 39:404-416.

Purcell, S, B Neale, K Todd-Brown, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559-575.

Sardenne, F, E Dortel, G Le Croizier, J Million, M Labonne, B Leroy, N Bodin, E Chassot. 2015. Determining the age of tropical tunas in the Indian Ocean from otolith microstructures. Fisheries Research 163:44-57.

Yang, J, SH Lee, ME Goddard, PM Visscher. 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics 88:76-82.

Zudaire, I, E Chassot, H Murua, Z Dhurmeea, M Cedras, N Bodin. 2016. Sex-ratio, size at maturity, spawning period and fecundity of bigeye tuna (Thunnus obesus) in the western Indian Ocean. IOTC-2016-WPTT18-37.