CONTRACT FOR THE ICCAT ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA (GBYP) 02/2023- BIOLOGICAL STUDIES (ICCAT GBYP – PHASE 13)

Final Report for ICCAT

July 31st, 2024

Scientific coordinators: Dr. Barbara Block[1] and Chloe Mikles[1]

[1]Departments of Biology and Oceans of Stanford University

Hopkins Marine Station of Stanford University, Pacific Grove, CA, 93950, USA

This project is co-funded by the European Union

INDEX:

1. Context

The proposal "Tender for Investigating Adaptive Divergence in Atlantic Bluefin Tuna using Whole Genome Sequencing" to the ICCAT GBYP CIRCULAR 02/2023 Biological Studies was awarded to Stanford University in response to the call for tenders on July 26th, 2023.

The agreement of Stanford to conduct the work for ICCAT finalized on February 6th, 2024.

This report contains the required information for **Deliverable 3,** which consist in a first version of the final scientific report, containing a full description of the research carried out, a detailed description of methodology and protocols, maps of areas in which research was carried out, tables showing results of research activities, and detailed description of the analyses carried out. A version including including further data analyses of larvae provided by ICCAT, will be presented as a SCRS paper within the SCRS BFT species group meeting that will be held in September 2024.

2. Executive summary

Determining biologically meaningful populations is critical for preserving biodiversity and directing management efforts, though can be highly complex among highly migratory species such as the Atlantic bluefin tuna *Thunnus thynnus*. We summarize our progress thus far performing whole genome resequencing on over 300 samples representing Atlantic bluefin tuna populations in the western, eastern Atlantic, and Mediterranean Sea, with the goal of understanding the population structure at a genome-wide level. We detail preliminary results on comparisons of Atlantic bluefin tuna samples from the Gulf of Mexico and Mediterranean Sea populations. To date, we have isolated DNA and sequenced the genomes of over 100 samples including adults and larvae. We report significant genomic variation that can be explained by markers at both the neutral and adaptive genome. We additionally discovered outlier markers that may be advantageous in the assignment of individuals of unknown origin in the future. We also have obtained the raw sequencing data for 230 additional samples, including those directly provided to us through the ICCAT GBYP 02/2023 Biological Studies program, for which we are currently processing and analyzing the genomic data.

3. Scientific report

Introduction

Evidence from various techniques, including electronic tagging, otolith microchemistry, and genomics, suggest the population structure of Atlantic bluefin tuna *Thunnus thynnus* (ABT) is more complex than implied by the current ICCAT two-stock management strategy (Fromentin et al. 2014, Rodriguez-Ezpeleta et al. 2019, Brophy et al. 2020). There has been significant effort to understanding the genetic basis of Atlantic bluefin tuna population structure, in an attempt to clarify and expand upon existing hypotheses for more effective management. It is well established that ABT have two primary spawning grounds in the Gulf of Mexico (GOM) and Mediterranean Sea (Med), and the two stocks have been differentiated using mitochondrial DNA, allozymes, microsatellite loci, and SNP loci (summarized in Table 1). However, studies have found varying levels of differentiation, as well as evidence for sub-populations within these spawning grounds (ie., Carlsson et al. 2004, 2007;), and a population in the Slope Sea (SS) region (Rodriguez-Ezpeleta et al. 2019; Díaz Arce et al. 2024). Studies thus far have yet to reach a consensus on overall ABT population structure, though there is broad agreement that the use of genetics is necessary to characterize biological diversity and can help inform fisheries management efforts.

Rapid advances in modern genetic techniques, paired with decreasing sequencing costs, have made possible studies of population genetic structure at the genome-wide level. Single nucleotide polymorphisms (SNPs) serve as valuable markers for population studies due to their broad and abundant distribution across the entire genome and low mutation rate, while also facilitating the potential to investigate the roles of selection and adaptation. Recent studies have utilized SNP markers to characterize population structure between the GOM and Med, finding significant differentiation between populations (Puncher et al. 2018, Rodriguez-Ezpeleta et al. 2019), with evidence of interbreeding occurring in the SS (Díaz-Arce et al. 2024).

In this report, we summarize our utilization of whole genome resequencing (WGS) to understand the drivers of genomic divergence between GOM and Med populations. The use of WGS allows for the identification of adaptive genetic diversity, qualifying it as a powerful tool for improved management of vulnerable, locally adapted populations (Funk et al., 2012). This work is facilitated by an existing high-resolution chromosome-level reference genome for Atlantic bluefin tuna, assembled through PacBio HiFi and Dovetail sequencing (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_963924715.1/). The increased resolution of WGS data will provide much more power behind conclusions of population structure than studies with fewer markers. Recognizing that there may be additional populations, in this report we focus on the two internationally managed stocks (GOM and Med), which are most likely the major contributors to the overall spawning stock biomass.

Authors	GOM and Med distinct	Evidence of additional populations	Marker
Edmunds and Sammons 1973	no	no	Allozyme
Broughton and Gold 1997	yes	no	Microsatellite
Takagi et al. 1999	no	no	Microsatellite
Alvarado Bremer 1999, 2003	yes	no	mtDNA
Ely et al. 2002	no	no	mtDNA
Carlsson et al. 2004, 2007	yes	within Med	Microsatellite & mtDNA
Vella et al. 2006, 2009	–	within Med	Microsatellite
Boustany et al. 2008	yes	no	Microsatellite & mtDNA
Riccioni et al. 2010, 2013	–	within Med	Microsatellite
Viñas et al. 2011	–	no	Microsatellite
Vella et al. 2016	–	no	Microsatellite
Antoniou et al. 2017	–	no	SNPs
Puncher et al. 2018	yes	no	SNPs
Rodriguez-Ezpeleta et al. 2019	yes	Slope Sea	SNPs
Johnstone et al. 2021	yes	no	Microsatellite
McDowell et al. 2022	–	–	SNPs
Puncher et al. 2022	yes	no	SNPs
Díaz-Arce et al. 2024	yes	Slope Sea	SNPs

Table 1. Summary of population genetic studies on Atlantic bluefin tuna from 1973 – 2024.

Sampling design and methodology

Larval samples from Sicily arrived on February 14th, 2024, and larval samples from the Balearics arrived on March 1st, 2024. We performed DNA isolations of the larval samples preserved in ethanol using the Qiagen DNeasy blood and tissue kit from larvae and adults and obtained higher-than-optimal DNA concentrations for whole genome sequencing. We submitted 177 samples for whole genome sequencing (WGS) on May 7th, 2024 (Table S1). These data are being combined with two prior sequencing runs including an additional 154 samples (funded through other sources), yielding a total of 331 whole genome sequences for analysis of genomic structure of individuals of GOM, Med, and SS populations. We have an additional 70 samples to prepare for sequencing from the first Sicily shipment that require different laboratory techniques, as the tissues are older (collected between 2010-2016) and degraded, which makes it more challenging to get high enough quality DNA for full genomic sequencing.

For analyses, all larvae were assigned to a population (GOM or Med) based on being collected in the region, and not being old enough to have moved significantly from the spawning location. Adults were assigned to GOM or Med based on entry to the respective spawning grounds (e.g.,

Figure 1). Adult fish were assigned to the SS by Aalto et al. (2023) based on spatial analyses of position in the spawning season and behavioral data from electronic tags.

DNA concentrations of all samples were quantified using the Qubit dsDNA High Sensitivity Assay Kit (Life Technologies) and DNA. Library prep and sequencing was conducted with Texas A&M AgriLife Genomics and Bioinformatics Service. Libraries were sequenced on six Illumina NextSeq 6000 lanes in total. FASTQC was performed on all raw sequence data for quality control, reports for each sample can be made available on request.

Data from 97 samples, funded by the Block Lab at Stanford University, have been fully processed and analyzed, and the remaining 234 samples are currently being processed through a custom bioinformatics pipeline at Stanford that yields variant call format files for each individual, resulting in single nucleotide polymorphisms on the scale of several million. The analyses that have already been completed (and that will be done for the remaining samples) are detailed as follows:

Basic descriptive statistics, including nucleotide diversity (pi), individual heterozygosity, and Tajima's D, were estimated for each population in 25 kb windows using VCFTOOLS (Danecek et al., 2011). F_{ST} was calculated for each SNP, for each population in 25 kb windows, and pairwise among populations. Private alleles were counted within populations using BCFTOOLS and SAMTOOLS (Danecek et al. 2021).

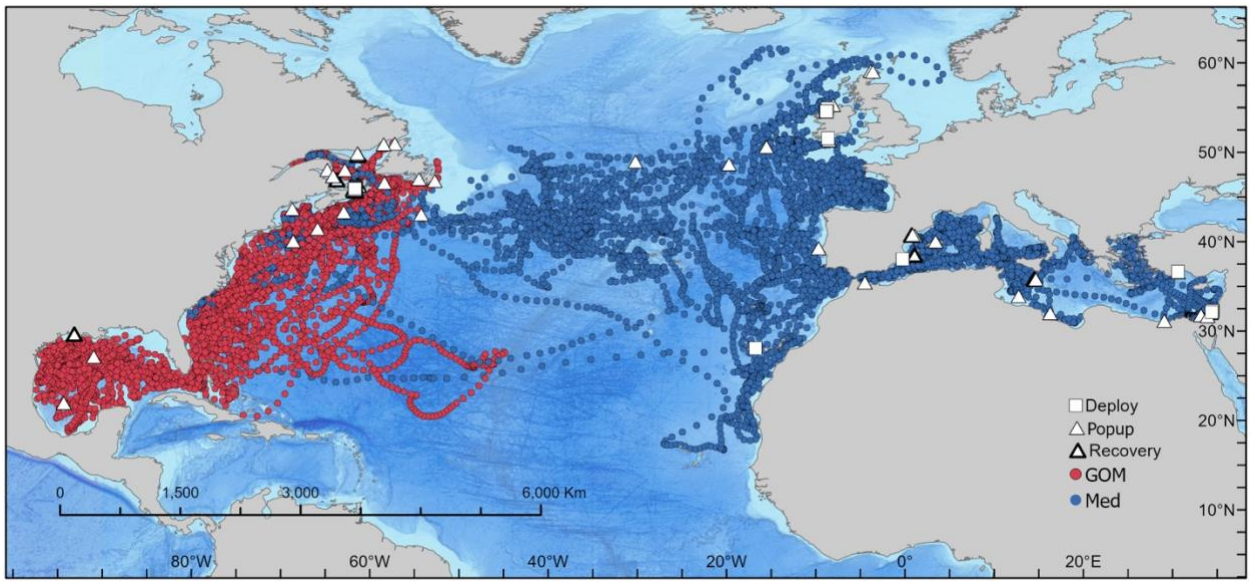

Figure 1. Electronic-tagged adult Atlantic bluefin tuna that visit one of the two recognized spawning grounds. Samples from these adults are included in this preliminary report.

Population genomic differentiation

Whole genome sequences allowed us to characterize the contributions of both adaptive and neutral markers in shaping differences between populations. For analyses of neutral population structure, we used a filtered dataset that was pruned to avoid linkage and excluded SNPs that were mapped to exons or within 10kb of exons. For both the full and the putatively neutral dataset, we estimated basic descriptive statistics, including nucleotide diversity (pi), individual heterozygosity, and Tajima's D, for each population in 25 kb windows using VCFTOOLS (Danecek et al., 2011). We also counted private alleles within populations using BCFTOOLS and SAMTOOLS (Danecek et al. 2021). To evaluate population structure, we performed a principal component analysis (PCA) using the R package "SNPRelate" (Zheng et al. 2012). We ran ADMIXTURE (Alexander et al. 2009) for only the neutral dataset. To characterize genome-wide patterns of divergence between GOM and Med populations, we calculated pairwise F_{ST} values. We also calculated F_{ST} for individual SNPs and in 25 kb windows.

Patterns of selection

Peaks in F_{ST} were visualized using Manhattan plots constructed using the R package QQMAN (Turner, 2018). We defined outlier SNPs as those that exhibited elevated divergence between GOM and Med and, specifically, those that had F_{ST} estimates greater than four standard deviations above the weighted mean. For each outlier SNP, we extracted regions within 40 kb of each SNP and used the R package GENOMICRANGES (Lawrence et al., 2013) to find overlapping annotations within the Atlantic bluefin tuna reference genome from our lab (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_963924715.1/). We obtained information on the function of each gene using the UniProt database (http://www.uniprot.org/), and inspected individual outlier regions by plotting F_{ST} with nucleotide diversity and Tajima's D. Finally, we performed a PCA using just the outlier SNPs.

For adult fish with electronic tagging data, we also performed genome-wide associations to investigate whether differences in spawning time, age-at-maturity, and environment between the GOM and Med explain the observed genomic differentiation. To do so we used the Bayesian sparse linear mixed model (BSLMM) implemented in the package GEMMA (Zhou et al., 2013) to scan for SNPs associated with date and age of entry to the spawning grounds, as well as environmental variables (sea surface temperature, salinity, and eddy kinetic energy). Associations were plotted across chromosomes using the qqman package in R (Turner, 2018).

Preliminary Results

We report on results from the first round of sequencing of GOM and Med samples funded by Stanford University. The analyses detailed here will be repeated for the samples provided by ICCAT, and with the bolstered sample size we can conduct finer scale investigations of population structure within the eastern population, combined with our Slope Sea samples.

The mean percentage of reads mapped to the Atlantic bluefin tuna reference genome was 99.44% and the mean genome coverage across the 92 retained individuals was 10.24x. After filtering, we obtained 13,241,151 single nucleotide polymorphisms (SNPs) for subsequent analyses.

Principal component analysis (PCA) revealed subtle, but significant genome-wide differentiation between GOM and Med samples for both the full and neutral dataset (Figure 2). Results from ADMIXTURE on the neutral dataset supported K=1 as the best supported cluster by cross-validation errors, and we did not see any population structuring in the visualization of ancestry

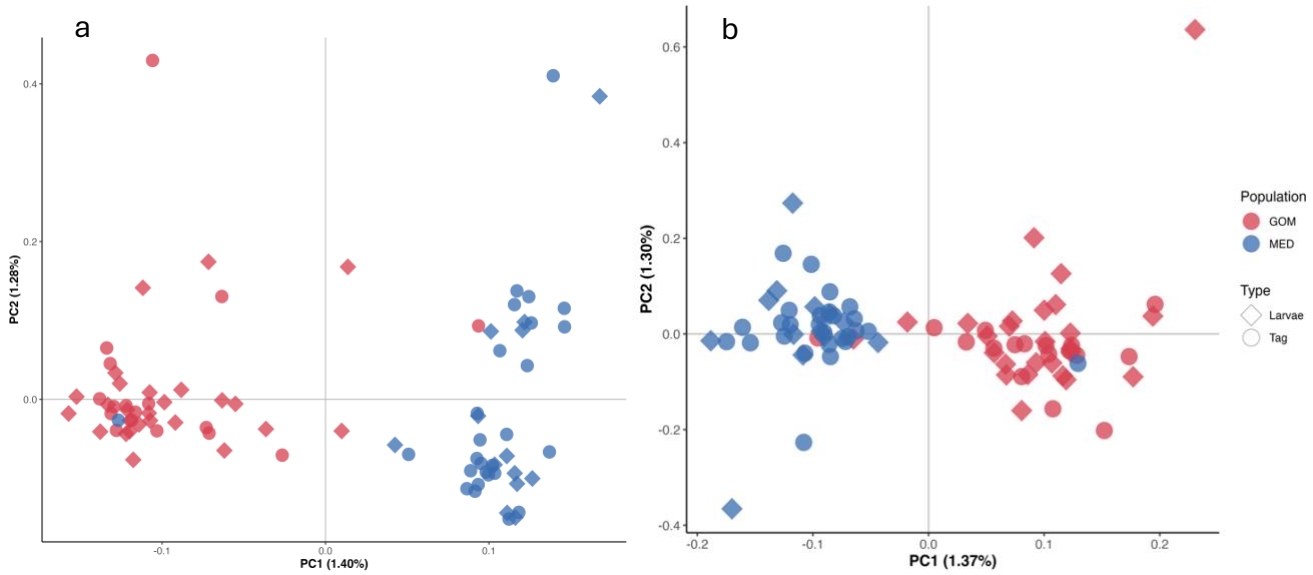

Figure 2. Principal component analysis (PCA) using the full dataset (a) and neutral dataset (b). Points are colored by population, GOM (red) and Med (blue), and shaped by sample type, either larvae (diamonds) or tagged adults (circles).

proportions for higher values of K. There were no differences in heterozygosity between GOM (mean ± SD = 0.21± 0.03) and Med (0.20 ± 0.03).

Genome-wide F_{ST} estimates between GOM and Med suggest moderate levels of divergence (mean F_{ST} = 0.0032), with 158 SNPs having elevated F_{ST} values greater than four standard deviations above the mean (Figure 3). The 158 outlier SNPs overlapped with 468 elevated windows, which mapped to 305 unique gene IDs. Several of these had functions related to development of the vasculature (angiogenesis) and calcium signaling in ion channels (Figure 4).

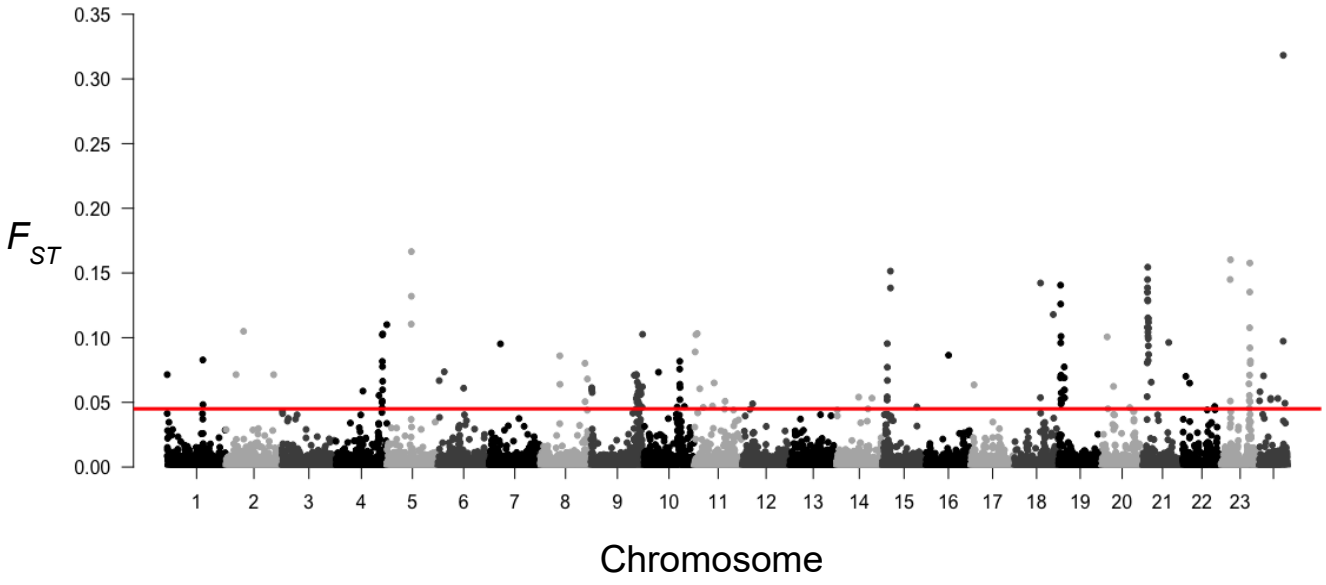

Figure 3. Manhattan plot of genome-wide pairwise F_{ST} between GOM and Med samples. Peaks in F_{ST} are considered significant above the red line, which denotes 4SD above the mean F_{ST}.

Chrom	Gene name	Description
21	a2ml	Liver development
21	plekhg5 b	Angiogenesis, blood vessel endothelial cell migration
21	rcbtb1	Retinal vasculature and blood vessel development
21	MLNR	Calcium-mediated signaling
21	pgr	Ovulation; LH signaling
21	trpc6a	Calcium channel activity

Figure 4. Selection of candidate genes and their putative roles from F_{ST} outlier peak on Chromosome 21.

Conclusion

We report preliminary results from GOM – Med whole genome comparisons that describe subtle, but significant genomic variation using both adaptive and neutral markers. We have identified the most differentiating markers across the entire genome that may aid future efforts in assigning individuals of unknown origin to the western or eastern stock. The analysis of the additional Med larvae from different sites, as well as the adults tagged in Norway, will allow us to investigate the possibility of population structure within the Mediterranean Sea. The addition of Slope Sea spawners will help to better understand their origin and contribution to overall bluefin tuna biodiversity.

The timeline and plan for completion of this project extends in several phases over the next several months. The first phase involves sequencing of the older samples from Sicily, which we expect to be completed in November 2024. Meanwhile, we will complete data processing for the second set of sequenced samples, expected to be finalized September 1, 2024, and the preliminary analyses for them by November 2024. Once the Sicily samples are obtained and analyzed, we will begin preparation of the entire dataset for publication in a peer-reviewed journal.

Acknowledgements

This work was carried out under the provision of the ICCAT. The contents of this document do not necessarily reflect the point of view of ICCAT, which has no responsibility over them, and in no ways anticipate the Commission's future policy in this area.

This work was conducted within the ICCAT GBYP Phase 13 and partially funded by the European Union through the EU Grant Agreement No. 101133291.

References:

Aalto, E. A., S. Dedman, M. J. W. Stokesbury, R. J. Schallert, M. Castleton, and B. A. Block. 2023. Evidence of bluefin tuna (*Thunnus thynnus*) spawning in the Slope Sea region of the Northwest Atlantic from electronic tags. ICES Journal of Marine Science 80(4):861-877.

Alvarado Bremer, J., I. Naseria, and B. Ely. 1999. Heterogeneity of northern bluefin tuna populations. ICCAT Collective Volume of Scientific Papers 49:127-29.

Antoniou A, Kasapidis P, Kotoulas G, Mylonas CC, Magoulas A. 2017. Genetic diversity of Atlantic bluefin tuna in the Mediterranean Sea: insights from genome-wide SNPs and microsatellites. Journal of Biological Research (Greece) 24(1):1–14DOI 10.1186/s40709-017-0062-2.

Boustany, A.M., Reeb, C.A., and Block, B.A. 2008. Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (Thunnus thynnus). Mar. Biol. **156**(1): 13–24. Springer.

Brophy, D., N. Rodríguez-Ezpeleta, I. Fraile, and H. Arrizabalaga. 2020. Combining genetic markers with stable isotopes in otoliths reveals complexity in the stock structure of Atlantic bluefin tuna (*Thunnus thynnus*). Nature Scientific Reports 10:14675.

Carlsson J, McDowell JR, Diaz-Jaimes P, Carlsson JEL, Boles SB, Gold JR, et al. 2004. Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (Thunnus thynnus thynnus) population structure in the Mediterranean Sea. Molecular Ecology 13:3345-56.

Carlsson, J., McDowell, J.R., Carlsson, J.E.L., and Graves, J.E. 2006. Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. J. Hered. 98(1): 23–28. Oxford University Press.

Danecek, P., A., et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156-2158.

Díaz-Arce,, N. et al. 2024. Unidirectional trans-Atlantic gene flow and a mixed spawning area shape the genetic connectivity of Atlantic bluefin tuna. Mol Ecol 33:1.

Fromentin, J., Reygondeau, G., Bonhommeau, S., and Beaugrand, G. 2014. Oceanographic changes and exploitation drive the spatio-temporal dynamics of Atlantic bluefin tuna (Thunnus thynnus). Fish. Oceanogr. 23(2): 147–156. Wiley Online Library.

Funk, C. W. J. K. McKay, P. A. Hohenlohe, F. W. Allendorf. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology and Evolution 27:9:489-496.

Hauser L, Carvalho GR. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries. 2008;9:333-62.

Johnstone, C. M. Pérez, E. Malca, J. M. Quintanilla, T. Gerard, D. Lozano-Peral, F. Alemany, J. Lamkin, A. García, and R. Laiz-Carrión. 2021. Genetic connectivity between Atlantic bluefin tuna larvae spawned in the Gulf of Mexico and in the Mediterranean Sea. PeerJ 9:e11568

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Puncher, G.N., Cariani, A., Maes, G.E., Van Houdt, J., Herten, K., Cannas, R., Rodriguez-Ezpeleta, N., Albaina, A., Estonba, A., and Lutcavage, M. 2018. Spatial dynamics and

mixing of bluefin tuna in the Atlantic Ocean and Mediterranean Sea revealed using next-generation sequencing. Mol. Ecol. Resour. **18**(3): 620–638. Wiley Online Library.

Puncher, G. N., A. Hanke, D. Busawon, E.V.A. Sylvester, W. Golet, L.C. Hamilton, and S.A. Pavey. 2022. Individual assignment of Atlantic bluefin tuna in the northwestern Atlantic Ocean using single nucleotide polymorphisms reveals an increasing proportion of migrants from the eastern Atlantic Ocean. Canadian Journal of Fisheries and Aquatic Sciences.

Riccioni G, Landi M, Ferrara G, Milano I, Cariani A, Zane L, et al. 2010. Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic Bluefin tuna of the Mediterranean Sea. Proceedings of the National Academy of Sciences. 2010;107:2102-7.

Riccioni, G. M. Stagioni, M. Landi, G. Ferrara, G. Barbujani, F. Tinti. 2013. Genetic structure of bluefin tuna in the Mediterranean Sea correlates with environmental variables. PLOS One.

Rodríguez-Ezpeleta, N., Díaz-Arce, N., Walter, J.F., Richardson, D.E., Rooker, J.R., Nøttestad, L., Hanke, A.R., Franks, J.S., Deguara, S., and Lauretta, M. V. 2019. Determining natal origin for improved management of Atlantic bluefin tuna. Front. Ecol. Environ. Wiley Online Library.

Turner S (2018). "qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots." The Journal of Open Source Software.

Viñas, J., A. Gordoa, R. Fernández-Cebrián, C. Pla, Ü. Vahdet, and R. M. Araguas. 2011. Facts and uncertainties about the genetic population structure of Atlantic bluefin tuna (Thunnus thynnus) in the Mediterranean: Implications for fisheries management. Reviews in Fish Biology and Fisheries 21:527-541.

Zhou X, Carbonetto P, Stephens M (2013) Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genet 9(2): e1003264.