SHORT TERM CONTRACT FOR THE BIOLOGICAL STUDIES (ICCAT GBYP 02/2023) OF THE ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA (GBYP Phase 13)

Final Report
(Deliverable # 4) for:

ICCAT

Scientific coordinator:

Dra. Igaratza Fraile (AZTI-Member of Basque Research & Technology Alliance)

Pasaia, May 31st, 2024

This project is co-funded by the European Union

PARTNERS:

Fundación AZTI – AZTI Fundazioa,

(AZTI)

Instituto Español de Oceanografía,

(IEO-CSIC)

Galway-Mayo Institute of Technology,

(GMIT)

Fisheries Resources Institute, Japan Fisheries Research and Education Agency

(FRI)

Universidad de Cádiz,

(UCA)

Institut Français de Recherches pour l'Exploitation de la Mer

(IFREMER)

Texas A&M University

(TAMU)

National Oceanic and Atmospheric Administration

(NOAA)

SUBCONTRACTORS and COLLABORATORS:

Institute of Marine Research

(IMR)

University of Arizona

Xenetica Fontao

Data Value Managment

CEAB-CSIC

Scientific Angler

TABLE OF CONTENTS

TABLE OF CONTENTS 4
CONTEXT 5
EXECUTIVE SUMMARY 6
1. SAMPLING 8
    1.1 - Fish sampling 8
    1.2 – Larvae sampling 14
2. MAINTENANCE AND MANAGEMENT OF THE ICCAT GBYP TISSUE BANK 17
3. GENETIC ANALYSES 19
    3.1 Exploring the existence of genetic mechanisms regulating connectivity between spawning grounds 19
    3.2 Characterizing genomic mechanisms affecting adaptation at the earliest development stages 26
    3.3 Exploring the genetic origin of previously unassigned individuals 32
    3.4 Evaluate the sex assignment power of genetic markers included in the SNP array 41
4. INFORMATION SYSTEM 46
5. ANNEX 52

CONTEXT

On July 26th 2023, the consortium coordinated by Fundación AZTI-AZTI Fundazioa, formed by partners Fundación AZTI-AZTI Fundazioa (AZTI), Instituto Español de Oceanografía (IEO-CSIC), Galway-Mayo Institute of Technology (GMIT), Japan Fisheries Research and Education Agency (FRI), University of Cádiz (UCA), Texas A&M University (TAMU), and the National Oceanic and Atmospheric Administration (NOAA), presented a proposal to the call for tenders on biological and genetic sampling and analysis (ICCAT-GBYP 02/2023).

This proposal was awarded and the final contract between ICCAT and the consortium represented by Fundación AZTI-AZTI Fundazioa was signed on September 13th, 2023.

The present report corresponds to the revised final report (**Deliverable #4**) to be submitted to ICCAT in the framework of this contract.

EXECUTIVE SUMMARY

The main objective of this project is to enhance our knowledge about Atlantic bluefin tuna (ABFT) population structure, mixing, and growth, as well as to develop methodologies that integrate the current biological and ecological knowledge for an effective stock management.

During Phase 13, the Consortium sampled a total of 573 Atlantic bluefin tuna (135 YOY, 1 juvenile fish, 24 medium sized fish and 413 large fish) from different regions (148 from the Balearic Sea, 30 from Canary Islands, 85 from the Bay of Biscay, 131 from the Norwegian Sea and 149 from the Central Atlantic). In total, 986 biological samples (286 otolith samples, 127 fin spines and 573 genetic samples) were collected by the Consortium and incorporated into the tissue bank. The Consortium also received samples apart from those agreed in the contract. In total, the Consortium handled 1384 biological samples (405 otolith samples, 256 fin spines and 723 genetic samples) from 732 individuals. All these samples have been catalogued and stored together within the biological tissue bank. The information from this and previous phases is being reviewed and uploaded to the BioTuna application developed in AZTI. The distribution of samples within the ICCAT-GBYP biobank and associated metadata is available in BioTuna application, which is a data repository and visualization tool that enables interactive exploration of data (http://aztidata.local/BioTuna).

Additionally, ABFT larvae from surveys conducted in the Balearic Sea spawning ground were sorted and identified for potential close-kin analyses. In total, 2923 individuals from 25 samples collected during 2023 were identified. Bluefin tuna larvae were found in 21 out of the 25 samples analysed. The sorted individuals were preserved in 100% ethanol in different 4 ml jars and kept in the freezer for a perfect conservation.

On the genetic analysis, based on whole genome sequencing analysis we have identified a set of candidate gene variants potentially affecting Atlantic bluefin tuna fitness originated from a past introgression event from the albacore tuna, which was

confirmed to have occurred in the Mediterranean Sea. We showed that stock origin assignment based on the 96 SNP panel could overestimate the presence of ABFT of Gulf of Mexico origin and that the ABFT SNP array could be used for a more comprehensive monitoring of the species populations dynamics and mixing. Using this tool, we identified a set of candidate genomic variants potentially affecting survivorship during earliest life-stages that require further analysis, and performed kinship analysis which suggested the participation of the same individual in multiple spawning events at different locations within the same spawning season, expanding our knowledge on the demographic connectivity of the species. The genomic markers for sex determination included in the ABFT SNP array showed an assignment rate of the 92.6%.

Overall, most of the objectives of the project were met. These analyses continue to provide relevant information for a better understanding of the biology of Atlantic bluefin tuna, which in turn improves the stock assessment and management advice of this valuable species.

1. SAMPLING

Task Leader: Iraide Artetxe-Arrate (AZTI), Igaratza Fraile (AZTI) & Patricia Reglero (IEO)

Participants:
AZTI: Iker Zudaire, Natalia Diaz, Patricia Lastra-Luque, Martin Cabello de los Cobos, Gorka Markalain
UCA: Jose Luis Varela, Antonio Medina, Esther Asensio
FRI: Yohei Tsukahara
IEO-CSIC: Enrique Rodriguez Marín, Patricia Reglero, Rosa Delgado de Molina, Pablo Quelle
IMR: Ørjan Sørensen, Leif Nøttestad
CEAB-CSIC: Ana Gordoa
IFREMER: Tristan Rouyer, Oliver Derridj

The biological sampling conducted under Phase 13 follows a specific design, aimed to cover key geographic areas to allow to better understand mixing dynamics with implications for the MSE, key geographic areas and/or life stages that could be used in the design of a pilot CKMR study and/or areas where important changes have been detected in spatial dynamics of bluefin tuna in recent years, which may have resulted from interactions between biological factors and a gradually changing environment. As such, the sampling conducted under this project is independent from other routine sampling activities for fisheries and fishery resources monitoring (e.g., the Data Collection Framework).

1.1 - Fish sampling

Fish sampling has been completed in nearly all the designated areas outlined in the proposal (Table 1.1.1). In total 986 samples from 573 ABFT individuals have been achieved by the Consortium, consisting of 286 otolith pairs, 127 dorsal fin spines and 573 muscle tissue and/or fin clips for genetics (Table 1.1.2).

100 young-of-year (YOY) samples were planned to be collected in the **Balearic Sea** by CEAB-CSIC in collaboration of the recreational fishermen association ©Scientificangler.es. During the sampling season in September, the planned sampling was impeded due to the atypical vertical distribution of this age class. As a result, only 12 YOY specimens were obtained, including 12 otolith pairs, 12 spines, and 12 muscle tissues. However, the sampling of this age class and strata was successfully accomplished using tissue samples from 123 YOY individuals captured between September 2019 and October 2021 in the Balearic Sea and generously donated by the UCA (University of Cadiz). Besides, 13 muscle samples from large individuals were obtained from a tagging event carried out within a collaboration agreement between AZTI and IMEDEA-CSIC.

The UCA conducted sampling of large individuals in the **Strait of Gibraltar**. Specifically, they collected muscle tissue from 30 large individuals caught in the trap of Barbate in May 2024.

In February, the fishing season opened in the **Canary Islands**, but only a small number of fish were caught initially. During spring, the IEO collected 30 pairs of otoliths and tissue samples for genetic analysis

Adult sampling was accomplished in **Norwegian waters.** In total, 95 otoliths, 115 spines and 131 muscle tissue for genetics from 131 large individuals were sampled by the IMR. These individuals were caught mainly from purse seine landings from M/S Spjæringen (115), while 14 individuals were collected from national electronic tagging programs performed by IMR on board of recreational fishing vessels (by rod and reel). Two additional individuals from farms have also been sampled from the Hjartholmosen and Troland farms.

As regards the **Central Atlantic,** a sampling program conducted by FRI collects otolith and muscle tissue samples from ABFT caught by Japanese longline fishery. This ongoing survey has yielded approximately 100 samples from ABFT captured in 2022. In addition, due to resumption of the regular CPC observer activity, FRI also conducted voluntary onboard samplings. Over 200 samples were collected, and after assessing the condition of each sample, 149 individuals' samples were chosen to contribute to the ICCAT-GBYP biological biobank. More specifically, samples collected by FRI included 99 otoliths and 99 muscle tissue from individuals caught in the eastern Central Atlantic and 50 otoliths and 50 muscle tissue from individuals caught in the western Central Atlantic.

During this Phase 13 sampling in the **Bay of Biscay** was also resumed. Muscle and/or fin clip tissue of 31 large individuals were sampled by AZTI, taking advantage of the mid

water trollers' unloading in the fish markets of Pasaia. These fish were mainly sampled during the winter season (December-March) 2023-2024. In addition, IFREMER collected muscle tissue from 54 ABFT individuals, including 1 juvenile, 24 medium-sized individuals, and 29 large individuals. The samples were obtained from recreational fishing vessels using methods such as rod and reel, mid-water trolling, and longline fisheries. Most of the sampling occurred during the summer months of 2023 (July to October).

In addition to the committed samples, 189 ABFT individuals (40 YOYs, 7 medium-sized individuals and 142 large individuals) were sampled by other partners/contracts (Table 1.1.3). In the Mediterranean Sea, the Regional Observer Programme (ROP) provided 18 muscle tissue samples from 1 medium and 17 large ABFT individuals caught in the **Balearic Sea**, and additional 5 spines of 5 large individuals caught in the **Strait of Sicily**. Muscle samples from 40 YOY individuals caught in the **Adriatic Sea** were also provided by IZOR. Besides, the Consortium received a contribution from Oceanis SL., including 118 otolith pairs, 124 spines and 122 tissue samples of bluefin tuna captured in the **Tyrrhenian Sea.** Finally, an otolith pair from a large ABFT stranded at **Skagerrak** was included during Phase 13 activity of GBYP biological sampling. In total, the Consortium handled 428 samples, including 119 otolith pairs, 129 spines, and 180 tissue samples from external partners or contracts (Table 1.1.4). These samples were integrated into the GBYP biobank.

Altogether, considering the samples collected by the consortium and those arrived from other contracts/programs, the Consortium handled 1414 samples from 762 individuals, consisting of 405 otoliths (Figure 1.1.1), 256 spines (Figure 1.1.2), and 753 muscle tissue or fin clips for genetics (Figure 1.1.3) from several areas covering the ABFT geographic distribution.

Table 1.1.1: Individuals sampled within the Consortium, in each area and per each age stratum.

ICCAT MSE Region	Area	Responsible	Size class sampled				TOTAL	TARGET	%
			Age 0	J	M	L			
			<3 kg	3-25 kg	25-100kg	>100 kg			
MED	Balearic Sea	UCA	123				**123**	100	148
		AZTI (CEAB-CSIC)	12			13	**25**		
SATL	Str. Gibraltar	UCA				30	**30**	30	100
	Canary Islands	IEO-CSIC				30	**30**	35	86
EATL	Bay of Biscay	AZTI				31	**31**	30	283
		IFREMER		1	24	29	**54**		
	C. Atlantic East	FRI					**99**	50	198
NATL	Norway	AZTI (IMR)				131	**131**	30	437
WATL	C. Atlantic West	FRI				50	**50**	50	100
TOTAL			**135**	**1**	**24**	**413**	**573**	**325**	**176**

Table 1.1.2: Detailed number of otoliths, dorsal fin spines and muscle/fin tissue samples achieved in the framework of the Consortium, in each area.

ICCAT MSE Region	Area	Responsible	Tissue sampled			Total
			otoliths	spine	muscle/fin	
MED	Balearic Sea	UCA			123	**123**
		AZTI (CEAB-CSIC)	12	12	25	**49**
SATL	Gibraltar Strait	UCA			30	**30**
	Canary Islands	IEO-CSIC	30		30	**60**
EATL	Bay of Biscay	AZTI			31	**31**
		IFREMER			54	**54**
NATL	Norway	AZTI (IMR)	95	115	131	**341**
WATL	Central and North Atlantic	FRI	149		149	**298**
TOTAL			**286**	**127**	**573**	**986**

Table 1.1.3: Individuals sampled under other contracts and stored by the Consortium, in each area and per each age stratum.

ICCAT MSE Region	Area	Responsible	Size class sampled				Total
			Age 0	Juvenile	Medium	Large	
			<3 kg	3-25 kg	25-100 kg	>100 kg	
MED	**Adriatic Sea**	IZOR	40				**40**
	Balearic Sea	ROP			1	17	**18**
	Tyrrhenian Sea	OCE			6	119	**125**
	Strait of Sicily	ROP				5	**5**
NATL	**Skagerrak**	SVA				1	**1**
TOTAL			**40**	**0**	**7**	**142**	**189**

Table 1.1.4: Detailed number of otoliths, dorsal fin spines and muscle/fin tissue samples received from other partners/contracts outside the Consortium in each area.

ICCAT MSE Region	Area	Responsible	Material sampled			Total
			otoliths	spine	muscle/fin	
MED	**Balearic Sea**	ROP			18	18
	Strait of Sicily	ROP		5		5
	Tyrrhenian Sea	OCE	118	124	122	364
	Adriatic Sea	IZOR			40	40
NATL	**Skagerrak**	AZTI (SVA)	1			1
TOTAL			**119**	**129**	**180**	**428**

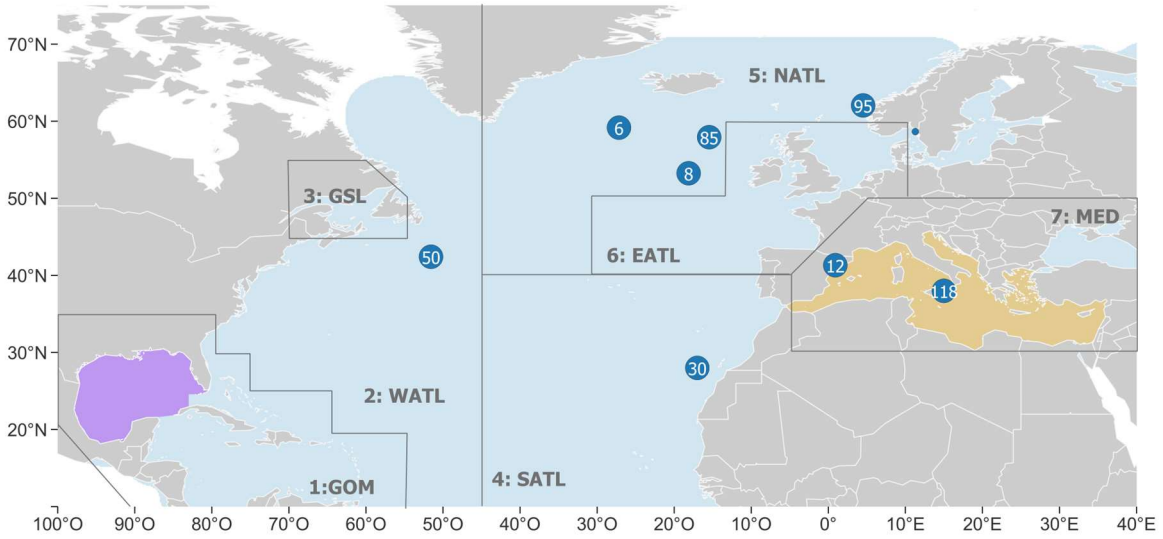

Figure 1.1.1: Geographical distribution of otolith samples collected during GBYP Phase 13 by the Consortium and other contract/partners. Locations are approximate, not actual latitudes and longitudes.

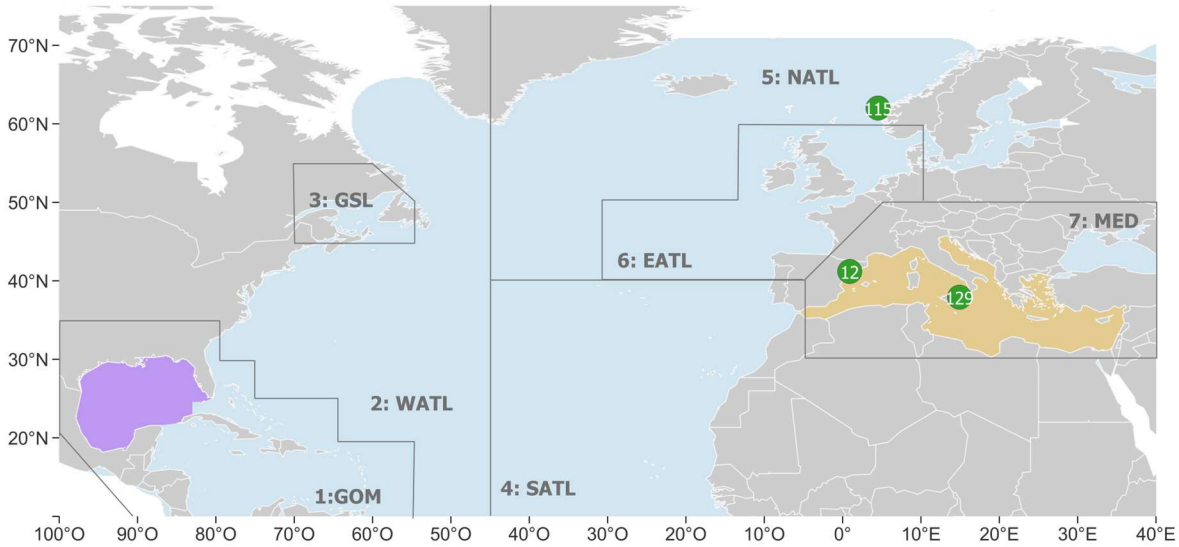

Figure 1.1.2: Geographical distribution of spine samples collected during GBYP Phase 13 by the Consortium and other contract/partners. Locations are approximate, not actual latitudes and longitudes.

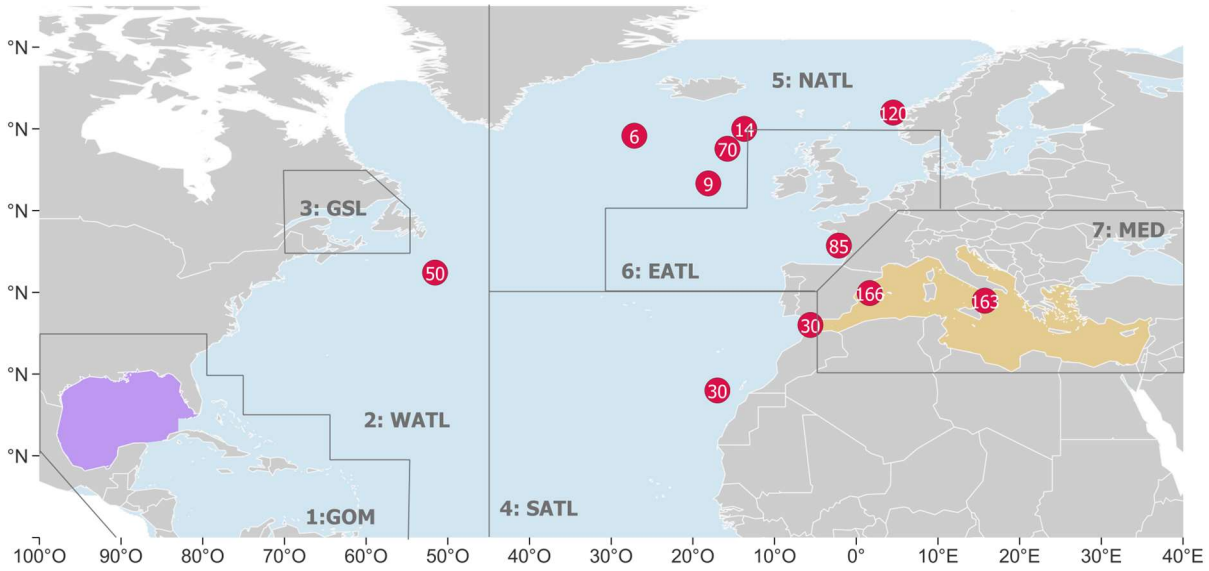

Figure 1.1.3: Geographical distribution of muscle tissue and fin clip samples collected during GBYP Phase 13 by the Consortium and other contract/partners. Locations are approximate, not actual latitudes and longitudes.

1.2 – Larvae sampling

The collection of Atlantic bluefin tuna larvae in the primary spawning area of the western Mediterranean represents a unique opportunity to contribute early life stage samples to the biological sample bank. While adults and juveniles have long been sampled within the GBYP framework for various studies, the inclusion of larvae adds valuable insights to our understanding of this remarkable species. National programs play a crucial role in collecting tuna larvae during the summer in key spawning grounds for Bluefin tuna, utilizing Bongo nets. Among the collected samples, one collector is preserved in formalin. These routine samples provide essential data for calculating a larval index, which serves as a proxy for monitoring the evolution of the spawning stock biomass. This is because formalin preservation method allows to maintain the shape of the larvae, enabling precise length measurements necessary for estimating the larval index. However, it's essential to recognize that formalin preservation is not suitable for other purposes, such as genetics or growth studies based on otoliths. Since 2019, one of the collectors used in the sampling has been preserved in ethanol. This preservation method ensures that larvae can be utilized for purposes beyond species identification and measurement.

Considering this framework, GBYP larval sampling within Phase 13 has consisted in identifying and counting Atlantic bluefin tuna larvae from ethanol-preserved jars. The goal is to obtain sufficient larvae for conducting the necessary studies to develop a

feasibility assessment regarding the potential implementation of the CKMR approach for the Atlantic bluefin tuna Eastern stock.

1.2.1 Field sampling and laboratory processing

In June-July 2023, tuna fish larvae were collected from 25 stations around the Balearic Islands, in the western Mediterranean Sea. The collection process involved sorting and identifying the larvae from one of the replicates of a Bongo net (90-cm diameter and 500-µm mesh size). The net was towed obliquely down to a depth of 30 meters at a cruising speed of 2 knots. The collected larvae were preserved directly in 100% ethanol for further processing. Identification was done using a dissection microscope, specifically identifying bluefin tuna larvae from the total plankton sample. These sorted individuals were then stored in different 4 ml jars and kept in the freezer for optimal conservation.

1.2.2 Results

Among the 25 samples analyzed, 21 contained bluefin tuna larvae, (see table below). A total of 2923 individuals from these 21 samples were identified as BFT larvae.

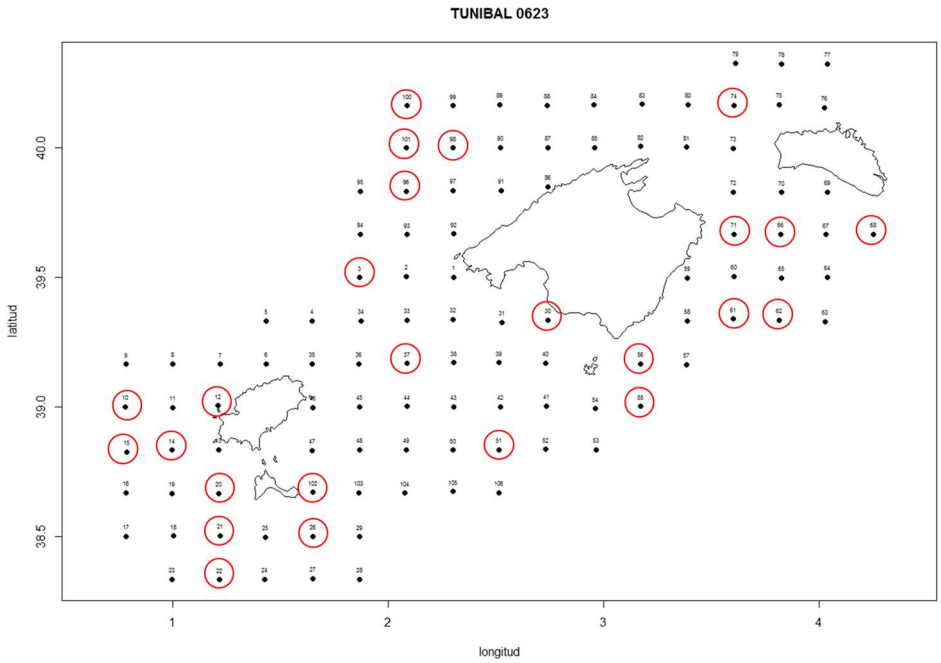

Figure 1.2.1: Geographical distribution of larvae collection in the Balearic Sea. Red circles denote the stations selected for bluefin tuna larvae identification.

Table 1.2.1: Detailed number of otoliths, dorsal fin spines and muscle/fin tissue samples received from other partners/contracts outside the Consortium in each area.

Year	Station	order	Sampler	Number larvae
2023	699	100	Bongo 90-500	0
2023	713	74	Bongo 90-500	7
2023	790	101	Bongo 90-500	0
2023	792	98	Bongo 90-500	15
2023	878	96	Bongo 90-500	0
2023	969	71	Bongo 90-500	1
2023	971	66	Bongo 90-500	292
2023	975	68	Bongo 90-500	2
2023	1042	3	Bongo 90-500	1031
2023	1139	30	Bongo 90-500	369
2023	1146	61	Bongo 90-500	2
2023	1148	62	Bongo 90-500	13
2023	1230	37	Bongo 90-500	8
2023	1239	56	Bongo 90-500	0
2023	1312	10	Bongo 90-500	22
2023	1316	12	Bongo 90-500	12
2023	1331	55	Bongo 90-500	7
2023	1400	15	Bongo 90-500	9
2023	1402	14	Bongo 90-500	85
2023	1416	51	Bongo 90-500	158
2023	1494	20	Bongo 90-500	518
2023	1497	102	Bongo 90-500	19
2023	1589	21	Bongo 90-500	147
2023	1593	26	Bongo 90-500	38
2023	1690	22	Bongo 90-500	168

2. MAINTENANCE AND MANAGEMENT OF THE ICCAT GBYP TISSUE BANK

Task Leader: Igaratza Fraile (AZTI) & Iraide Artetxe-Arrate

Participants:
AZTI: Naiara Serrano, Goretti Garcia, Ainhoa Arevalo, Patricia Lastra

The sampling protocols, together with instructions, have been distributed within the Consortium as well as to ICCAT, so that they are distributed to other institutions conducting biological sampling (e.g., as part of tagging activities, Regional Observer Programs, farms, etc.). The sampling protocol can be found as **Annex 1** of this report.

The Consortium has updated and refined, following ICCAT GBYP guidelines, the currently available database on the samples included in the GBYP Tissue Bank. The Consortium has provided appropriate storage for all the biological samples and hard parts already collected and shipped to AZTI within the current and previous GBYP phases. When needed, samples have been relabelled accordingly. In the context of genetics, muscle samples have been replicated. Additionally, management has ensured the delivery of the required samples to the entities responsible for the analyses within this Consortium.

The Consortium has continued to provide appropriate storage for biological samples collected so far within previous phases of GBYP biological sampling program. This implies storing samples that constitute the GBYP tissue bank, including otoliths, spines, gonads, muscle, and fin tissues. These samples facilitate microchemical, genetic, histological, and morphological analyses based on recommendations from the SCRS and the GBYP Steering Committee. The goal is to enhance our understanding of crucial biological and ecological processes affecting Atlantic bluefin tuna (ABFT), particularly by stockpiling samples for population-level genomics studies.

All the information received has been properly recorded in the tissue bank related information system. Data has been standardized according to the criteria agreed in the protocol to follow a common pattern. Metadata is provided following the accorded format of modules: (1) Fish identification data, (2) Sample availability, (3) Storage at AZTI, (4) Sampling information, (5) Biological data, (6) Analytical tasks and (7) Results. A full

detailed and updated catalogue of samples stored in the GBYP Tissue Bank, detailed by stratum and size classes, is so far provided separately as **Annex 2**.

In the current phase, otolith-derived age estimates have been updated following the calendar year adjustment considering the criteria of Rodriguez-Marín et al. (2022).

Bibliography

Rodriguez-Marin, E.; Busawon, D.; Luque, P.L.; Castillo, I.; Stewart, N.; Krusic-Golub, K.; Parejo, A.; Hanke, A. Timing of Increment Formation in Atlantic Bluefin Tuna (*Thunnus thynnus*) Otoliths. Fishes 2022, 7, 227. https://doi.org/10.3390/fishes7050227

3. GENETIC ANALYSES

Task Leader: Natalia Díaz-Arce (AZTI)

Participants:
AZTI: Naiara Rodriguez-Ezpeleta, Iñaki Mendibil, Natalia Gutierrez

3.1 Exploring the existence of genetic mechanisms regulating connectivity between spawning grounds

3.1.1. Introduction

Previous analyses supported signal of introgression from albacore tuna (*Thunnus alalunga*) present in the Atlantic bluefin tuna genome (Díaz-Arce et al. 2024). This signal is stronger among individuals of Mediterranean origin, leading to the hypothesis that the inter-specific introgression event occurred here. Understanding the origin of this introgression event and deciphering past and present connectivity patterns between spawning areas are critical for advancing our knowledge of Atlantic bluefin tuna. These insights can shed light on the species' evolutionary dynamics and ecological resilience. Likewise, inferring the genomic basis of local adaptation is crucial for fisheries management as it provides valuable insights into the specific genetic traits that allow certain populations of Atlantic bluefin tuna to thrive in their respective habitats. This knowledge will help to identify key adaptation traits contributing to populations resilience, informing targeted conservation efforts, sustainable fishing practices, and the implementation of adaptive management strategies tailored to different regions. Using whole genome sequencing to identify genomic regions under adaptive selection allows for a comprehensive examination of the entire genome, enabling the detection of adaptive genetic variants beyond those targeted by reduced representation techniques, enhancing the full spectrum of adaptive variation in the Atlantic bluefin tuna populations. This information will help to uncover the specific genetic mechanisms involved in the species' response to environmental pressures. With this aim, we explored different sources of genomic variation and identified such genomic regions under adaptive selection using whole genome sequencing data.

The objective of this subtask is to understand Atlantic bluefin tuna **population dynamics and its adaptability** to different environmental conditions.

3.1.2. Materials and methods

3.1.2.1. DNA extraction, library preparation and whole genome sequencing of three albacore tuna samples captured in the Mediterranean Sea

The following procedure was applied to 25 Atlantic bluefin tuna and 5 albacore tuna samples. Among those, 27 had been already processed and three albacore tuna individuals captured around the Balearic Islands in June 2023 were newly processed during the current GBYP Phase-13. From each sample, a ~1 cm^3 of muscle tissue was excised and stored in RNA-later or 96% molecular grade ethanol at −20°C until DNA extraction. Genomic DNA extraction was performed using the Wizard® Genomic DNA Purification Kit (Promega), starting from 20 mg of tissue and following the manufacturer's instructions. Extracted DNA was eluted in sterile Milli-Q water and its concentration was determined with Quant-iT dsDNA HS assay kit using a Qubit® 2.0 Fluorometer (Life Technologies). DNA integrity was assessed by electrophoresis, migrating about 100 ng of GelRed™-stained DNA on a 1.0% (w/v) agarose gel. Individual libraries were prepared for each individual sample and sent for sequencing at a NovaSeq 6000 Illumina platform at ~15X depth coverage.

3.1.2.2. Obtention of genotype dataset from whole genome sequencing data

Whole genome sequencing data of 25 Atlantic bluefin tuna and 5 albacore tuna was analyzed together with whole genome sequencing data of 5 Pacific bluefin tuna (*Thunnus orientalis*) and 4 yellowfin tuna (*Thunnus albacares*) downloaded from public repositories (Table 3.1.1) were filtered using TRIMMOMATIC (Bolger et al. 2014) to trim low quality reads (reads were screened using sliding window of 3 nucleotide size and quality threshold of 28), discard reads of less than 50 nucleotides length and those containing adapter sequences.

Filtered reads were mapped the most updated version available of the reference genome of the Pacific bluefin tuna (GCA_021601225.1) using BWA-MEM algorithm (Li 2013) and only correctly mapped read pairs and primary alignments were kept. Duplicated read pairs, likely deriving from PCR DNA fragment duplication performed during NA preparation, were marked using the MarkDuplicates module from picard tools

(https://github.com/broadinstitute/picard) and removed using *SAMtools* (Li et al. 2009). Genomic variants were called using *freebayes* (Garrison and Marth 2012) and exported to vcf format. Only biallelic SNP variants located within the 260 biggest scaffolds of the reference genome (which cover >90% of the total reference genome) with minimum quality of 20, minimum read coverage of 10 and a minimum allele count of 3 were kept using *VCFtools* (Danecek et al. 2011). Only individuals with less than 3% missing data were included in the datasets. Two datasets were generated: one including filtered samples from all species and a second one including ABFT individuals only. From the dataset including all species only SNP variants without missing data were kept. For the ADMIXTURE (Alexander et al. 2009) and *TreeMix* (Pickrell and Pritchard 2012) analysis, SNPs under Linkage Disequilibrium were pruned using *Plink1.9* (Purcell et al. 2007) by removing SNPs showing a pairwise r^2 value higher than 0.2 estimated within windows including 50 variants sliding 20 variants each time. Mediterranean-like Gulf of Mexico ABFT individuals already identified in (Díaz-Arce et al. 2024) and confirmed in the Report of the Phase 12, were removed from the dataset used for the *TreeMix* analysis to avoid bias introduced by mixed groups. Alleles frequencies within each group was estimated using *Plink1.9* (Purcell et al. 2007) and genotype tables were exported into *Treemix* format using a custom script available at https://github.com/barbatom/plink2treemix. For the dataset including ABFT individual samples only SNPs showing a minimum allele count of 3 among the ABFT individuals were kept.

Table 3.1.1. Number of samples (N) of each species included in the whole genome sequencing dataset. Capture location and source of the data is indicated. (*)The 9 samples captured in the Gulf of Mexico included in this dataset include 3 Gulf of Mexico or confirmed Mediterranean genetic origin (Díaz-Arce et al. 2024).

Species	Location	N	Source
ABFT	Mediterranean Sea	7	Data available in AZTI
ABFT	Gulf of Mexico	9 *	Data available in AZTI
ABFT	Slope Sea	9	Data available in AZTI
ALB	Atlantic Ocean	2	Data available in AZTI
ALB	Mediterranean Sea	3	Data newly generated in this GBYP Phase
PBFT	Pacific Ocean	5	Downloaded (Acc. Numbers: DRR177383-87)
YFT	Atlantic Ocean	4	Downloaded (Acc. Numbers: ERR1462407-12, ERR1462433-35, ERR1462489-94)

3.1.2.3. Analysis of genomic variation of ABFT populations originating from inter-specific introgression

To confirm the genetic diversity between the different species included in the dataset, individuals ancestry proportions were estimated using ADMIXTURE (Alexander et al. 2009) using default parameters and assuming 4 ancestral populations (K value).

To determine whether the observed introgression in ABFT originated from either the Mediterranean or the Atlantic albacore population, we employed two statistical methods. First, we used TreeMix, which constructs a phylogenetic tree considering migration events (Pickrell and Pritchard 2012). Second, we estimated Patterson's D-statistic using the Dsuite software (Durand et al. 2011, Malinsky et al. 2021). Deviations from the expected allele patterns can indicate introgression. These analyses provide insights into historical gene flow between populations. TreeMix was used to estimate historical relationships among populations and species by estimating the maximum likelihood tree for a set of populations allowing historical gene flow events. TreeMix was run allowing from 0 to 10 migration events, obtaining an increasing number of possible gene flow events and associated likelihood values. The most probable number of migration events was selected by stopping adding additional migration events as long as the results remained interpretable and selecting the number showing best-associated likelihood value. Dsuite was used to estimate the Pattersons' D to estimates the excess of shared alleles between Mediterranean ABFT and Atlantic and Mediterranean albacore tuna respectively, compared to Gulf of Mexico ABFT and considering yellowfin tuna as an outgroup. To measure admixture along the genome and identify regions with stronger signal of introgression, f_d (Martin et al. 2014) was estimated within sliding windows covering 50 SNPs size at a distance of 25 SNPs using Dsuite and negative values were converted to zero. Reference genome of *Thunnus orientalis* used to map the whole genome sequences (GCA_021601225) was compared against the reference annotated genome of *Thunnus albacares* (GCA_914725855) using the online software D-genies (Cabanettes and Klopp 2018). SNP window positions showing a f_d value above 1 were explored for corresponding position in the reference genome of yellowfin tuna and matched against the available annotation file to search for coincidence within described genes.

3.1.3. Results and discussion

3.1.3.1. Confirmed the Mediterranean origin of introgression from albacore tuna into the Atlantic bluefin tuna populations

The final genotype table containing individuals from all the species and allowing no missing data included 11,371,709 SNPs and 40 individuals. The individual ancestry analysis revealed genetic differentiation between the four species included in the dataset (Figure 3.1.1)

Figure 3.1.1. Individual ancestry proportions when assuming four ancestral populations represented in red, blue, yellow and green, where it can be observed predominant respective proportions of each of the ancestral populations in the individuals of the different species included in the analysis: albacore tuna (ALB), Atlantic bluefin tuna (ABFT), Pacific bluefin tuna (PBFT) and yellowfin tuna (YFT).

The phylogenetic tree estimated by Treemix showed the expected topology, while the first migration event estimated when allowing for migration events shows gene-flow between the Mediterranean albacore tuna and the Mediterranean ABFT (Figure 3.1.2). While the arrow shows that the migration event occurred from the Albacore tuna to the Mediterranean ABFT, incorrect direction of the arrow is one of the major types of errors produced in TreeMix inferences (Pickrell and Pritchard 2012). While increasing the number of allowed migration events increased the associated likelihood (Figure 3.1.3), the results were hardly interpretable.

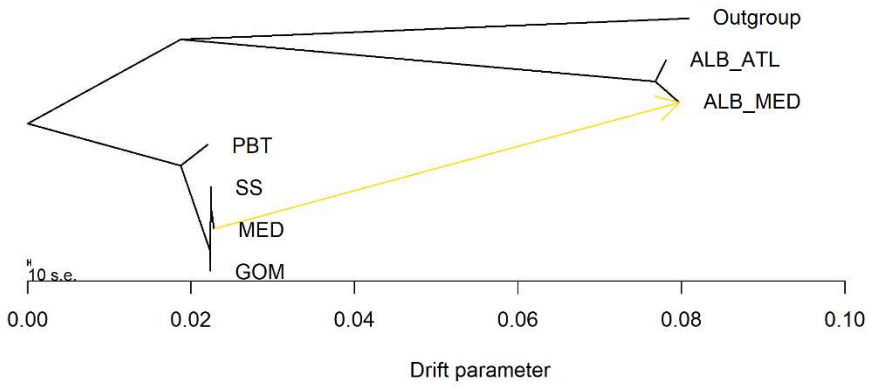

Figure 3.1.2. Phylogenetic tree estimated by TreeMix allowing one migration event (the arrow indicates migration direction and rate) based on the dataset which included samples of 22 ABFT from the Slope Sea (SS), Mediterranean Sea (MED) and Gulf of Mexico (GOM), 6 Pacific bluefin tuna (PBT), 3 and 2 albacore tuna from the Atlantic (ALB_ATL) and the Mediterranean Sea (ALB_MED) respectively and 4 yellowfin tuna.

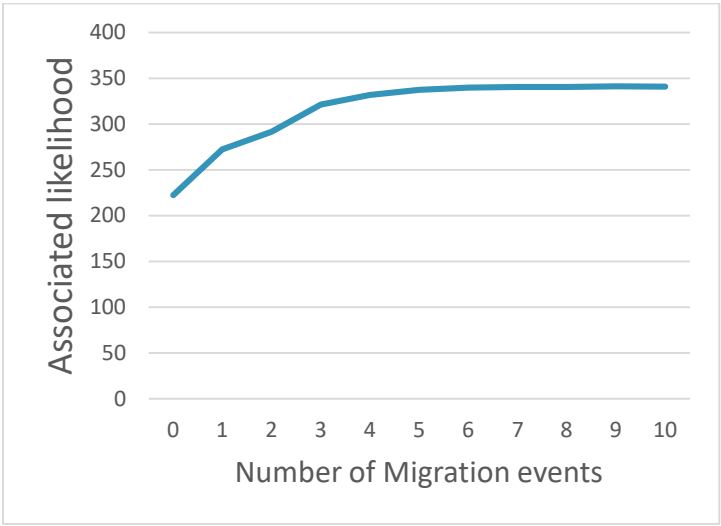

Figure 3.1.3. Associated likelihood (y axis) for each TreeMix analysis performed allowing from 0 to 10 migration events (x axis).

The Patterson D statistic estimated using Dsuite software confirmed the excess of shared alleles between albacore tuna and the Mediterranean bluefin tuna using the Gulf of Mexico ABFT individuals as reference. Interestingly, the signal of introgression was stronger between the Mediterranean albacore (Pattersons D = 0.0116) and Mediterranean ABFT than between the Atlantic albacore tuna and the Mediterranean ABFT (Pattersons' D = 0.0107). In sum, both the TreeMix and Pattersons' D statistic indicate that the

introgression from albacore tuna into the ABFT occurred between the Mediterranean populations.

3.1.3.2. Identification of introgressed genomic regions

The f_d statistic values estimated within window along the whole genome studying introgression of albacore alleles into de Mediterranean population of Atlantic bluefin tuna revealed the existence of a few genomic regions with increased signal (Figure 3.1.4).

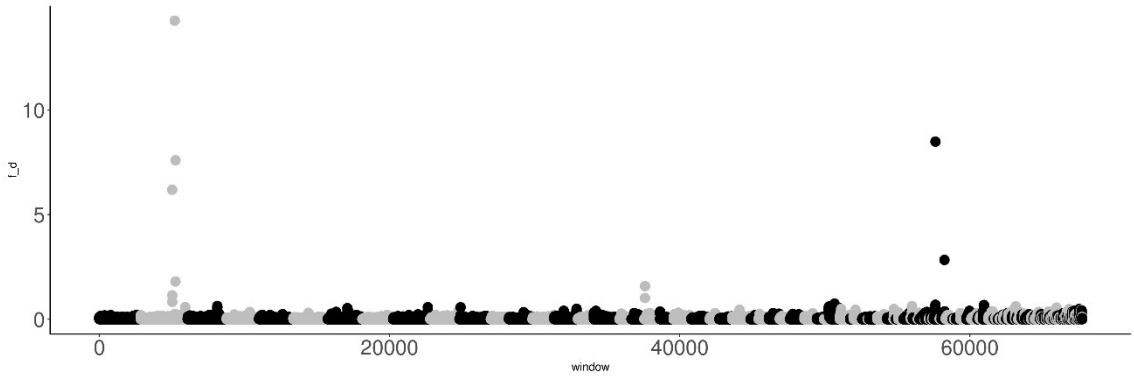

Figure 3.1.4. Genome wide fd values estimated within windows containing 50 SNPs which measures the level of introgression from the Mediterranean albacore into the Mediterranean Atlantic bluefin tuna along the genome. The different scaffolds of the reference genome of Thunnus orientalis are coloured using alternate grey and black colors.

Windows showing f_d values >1 were located in four different scaffolds of the reference genome, which matched against regions located in four reference chromosomes of the yellowfin tuna genome coinciding with the location of four different described protein coding genes. These genes are *plekha7a*, LOC122986472, *rnf213a* and *zmp:0000001020*. *Plekha* family genes are predicted to enable delta-catenin binding activity and act upstream of or within cardiac muscle cell differentiation and regulation of heart contraction. LOC122986472 gene of yellowfin tuna contains protein coding sequence for the emp24 protein, whose expression has been related with fish recovery from heat shocks in previous studies (Buckley et al. 2006, Buckley and Somero 2009). Polymorphisms in the *rnf213a* gene have been associated with growth, fat, muscle and bone formation in other fish species (Zhou et al. 2022).

3.1.4. Conclusions

Past connectivity of ABFT populations:

- The observed introgression signal in ABFT, stronger from Mediterranean albacore than from Atlantic albacore, confirms that the introgression of albacore alleles into ABFT originated from individuals of both species co-occurring in the Mediterranean.
- The confirmed origin of albacore alleles introgressing into Atlantic bluefin tuna from the Mediterranean, along with the observed introgression among Slope Sea larvae and young-of-the-year (YOY), validates the connectivity between the Mediterranean and Slope Sea spawning areas.

Adaptive potential of albacore origin:

- The preservation of specific genomic regions with strong signal of introgression in the genome and the location of protein coding genes within them, suggests that a past introgression event of albacore tuna alleles could confer adaptive potential to the Atlantic bluefin tuna.

3.2 Characterizing genomic mechanisms affecting adaptation at the earliest development stages

3.2.1. Introduction

Only a small proportion of ABFT survives to the early developmental stages, where strong natural selection is presumed to occur. The objective of this task is to identify the genomic basis of selection occurring at early development stages of Mediterranean ABFT. With that aim differences in allele frequencies of already characterized SNPs (included in the ABFT SNP array) between larvae and young of the year captured in the Mediterranean Sea were explored, to identify candidate genomic regions affecting individuals' fitness at these stages. During GBYP Phase 13, newly genotyped larvae and YOY captured in the Mediterranean Sea during the last 6 years were merged with the already existent dataset generated during previous GBYP Phases to compare allele frequencies between both developmental stages and identify genomic regions potentially affecting fitness at the larval stage.

Furthermore, the collected dataset underwent screening to identify kin pairs. These findings not only highlight the spatial and temporal connectivity among spawning sites within the Western Mediterranean spawning area but also inform the assessment of Mediterranean Sea larvae and young-of-the-year (YOY) suitability for future implementation of CKMR models in estimating stock biomass for Eastern Atlantic bluefin tuna.

The objectives of this subtask were:

- **Identify genomic variants** that undergo adaptive selection during the early developmental stages of Atlantic bluefin tuna.

- **Determine spatial and temporal connectivity** between sampled stations within the western Mediterranean spawning area.

3.2.2. Materials and Methods

In total, DNA of 188 larval samples from the years 2020 to 2023 collected in the Balearic Sea, and 112 YOY individuals from the years 2018, and 2020 to 2022 from the different locations within the Mediterranean Sea, was extracted. When dealing with the larvae, DNA extraction was performed from the entire specimen due to the limited tissue available. In addition to expert taxonomic identification of each larva, factors such as length, development stage, yolk-sac presence, and whether it fell into the preflexion or flexion category, were considered, and only those preflexion or flecion larvae were selected for genetic analyses. Importantly, these larvae were not expected to have preyed upon other larvae of the same species, thus minimizing the risk of sample contamination. DNA was extracted using the Wizard® Genomic DNA Purification kit (Promega, WI, USA) following manufacturer´s instructions for "Isolating Genomic DNA from Tissue Culture Cells and Animal Tissue". The starting material was approximately 20 mg of tissue or whole larvae and after extraction all samples were suspended in equal volumes of Milli-Q water. DNA quantity (ng/µl) was evaluated on the Qubit® 2.0 Fluorometer (Life Technologies) and DNA integrity was assessed by electrophoresis.

Extracted DNAs for this task, together with those prepared for Subtasks 3.3 and 3.4 were sent for genotyping and analyzed together following the same procedure. The genotypes of the newly processed samples were obtained by processing the obtained individual CEL files using the Axiom Suite Analysis software and removing individuals with genotyping rates below 0.97 and SNPs with low quality genotype calling discrimination. Obtained

genotype tables were merged to the previously existing dataset generated during the GBYP phases 10, 11 and 12, and SNPs were filtered to include only neutral SNPs genotyped in at least the 90% of the samples.

One genotype table was generated containing information for all larvae and YOY samples captured in the Mediterranean Sea during the years 2018 and 2020 to 2023 that passed the quality filtering criteria. Only SNPs present in the 90% of the individuals and showing a minimum allele frequency of 0.05 were kept.

To detect potential genetic variants affecting individuals fitness at the earlies life-stages, an association study was performed to investigate shifts in allele frequencies between larvae and YOYs captured in the Mediterranean Sea during the years 2018 to 2023 using the *assoc* function in the PLINK1.9 software (Purcell et al. 2007). Due to the reduced number of YOY individuals available from each year, all years were pooled together in the analysis. Those genetic variants with an association p-value <0.01 were identified as potential candidates. To visually confirm the genetic differentiation between larvae and YOY at those SNPs, a PCA was performed on the selected samples based on the obtained list of candidate variants using the R package *adegenet* (Jombart and Ahmed 2011).

The filtered genotype table was converted to VCF format file and relatedness was also estimated using the relatedness function implemented in VCFtools (Danecek et al. 2011). Then, the R package *kinference* (Bravington, in prep.), based on the distribution of pairwise PLOD scores and their fit into the predicted values for unrelated and related pairs was also used to identified pairs of half and full sibling pairs. Those pairs of samples with Genetic Relatedness values above 0.1 (this value is lower than what expected for first cousin pairs, which should be 12.5%, to make sure no kin pairs are missed) and with PLOD-values proximate to those expected for half and full sibling pairs were considered.

3.2.3. Results and Discussion:

3.2.3.1. Genetic variants affecting fitness at earliest life stages of Atlantic bluefin tuna

The filtered genotype table contained 624 larvae (n=490) and YOY (n=134) captured in the Mediterranean Sea during the years 2018 to 2023 and 5975 neutral SNPs. The association study resulted in 63 candidate markers included in the ABFT SNP Array that would affect fitness at earliest life-stages. The PCA showed weak genetic differentiation between larvae and YOY based on these 63 SNPs (Figure 3.2.1).

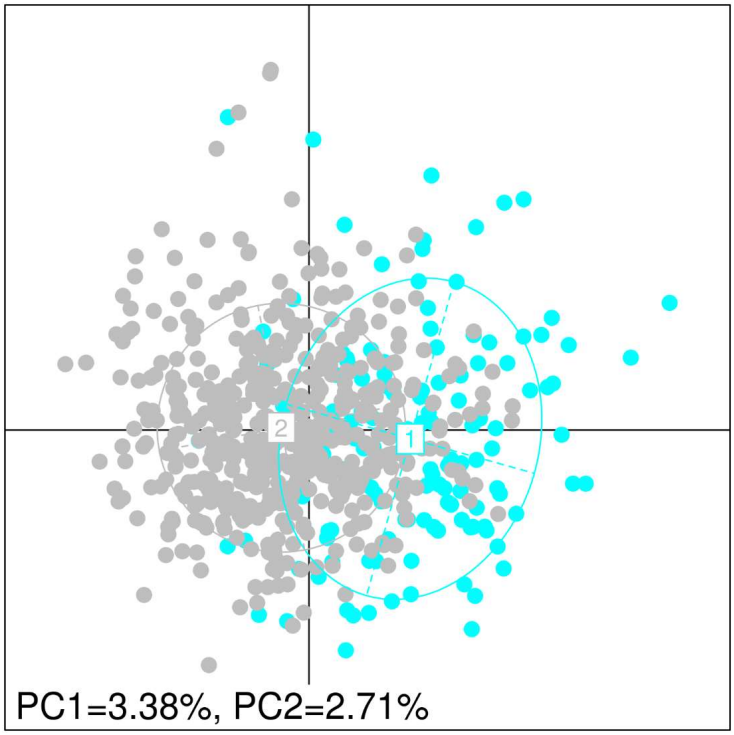

Figure 3.2.1. Principal component Analysis performed based on the 63 candidate SNPs affecting fitness in the earliest life-stages of Mediterranean Atlantic bluefin tuna. The analysis included larvae (grey) and YOY (blue) captured in the Mediterranean Sea during the years 2018 to 2023.

Weak genetic differentiation with big overlap between larvae and YOYs was expected, since no drastic changes are expected within a single generation. Despite this differentiation being low, previous studies have shown the cumulative adaptive potential and the effect on fitness-related traits of many variants of small effect (Yeaman 2015, Rey et al. 2020). The availability of annotated reference genomes for Thunnus species now allows for in-depth exploration of the potential locations of the identified candidate variants within genes. Understanding the function of these variants can shed light on critical threats to the species within its largest known spawning area, the Mediterranean Sea.

3.2.3.2. Kinship analysis among five different cohorts

The PLOD values obtained using the *kinference* package for kinship analysis showed clearly separated predicted values to distinguish between unrelated pairs, half-sibling pairs and full-sibling pairs (Figure 3.2.2).

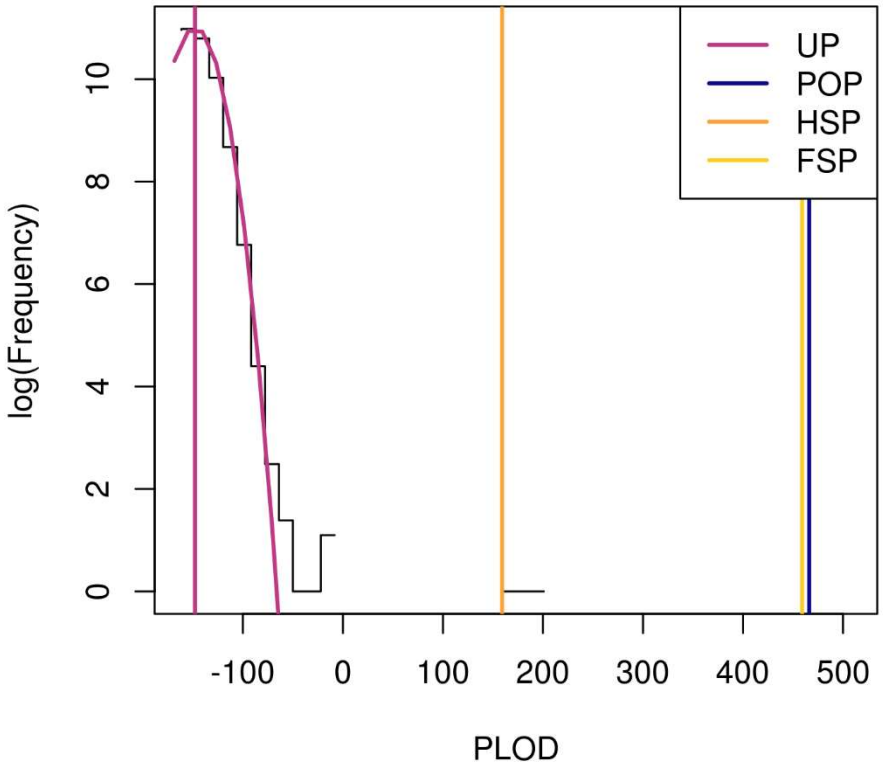

Figure 3.2.2. Histogram showing the distribution of the PLOD values estimated for each individual pair included in the analysis. The magenta line represents the expected distribution of unrelated pairs (UP), the orange, yellow and blue vertical lines show the expected PLOD value for half sibling pairs (HSP), full sibling pairs (FSP) and parent-offspring pairs (POP).

Based on the calculated PLOD values, a total of eight possible kin pairs were identified among the analyzed samples:

1. Full Sibling Pair (FSP): This pair consists of two larvae from the same sampling station in 2020.
2. Half Sibling Larvae Pairs (HSP): There are six pairs corresponding to half-siblings from the same sampling station and year.
3. Different Station Pair: The last pair involves two larvae sampled from two different stations in the same year. Although for this last pair the PLOD value falls much closer to the expected value for a HSP than for an UP, it was the lowest calculated PLOD value for all kin pairs found in the dataset (Table 3.2.1).

Table 3.2.1. Estimated PLOD and Genetic Relatedness (GR) values for each of the eight kin pairs detected among 624 analyzed samples of larvae and YOY individuals captured in the Mediterranean Sea during the years 2018 and 2020 to 2023. For each sample the GBYP database ID, sampled Year and Station are indicated. Finally, the type of each inferred kin-pair (PairType) is indicated based on PLOD and GR values as Half Sibling Pairs (HSP) or Full Sibling Paris (FSP). Pairs indicated with a star (*) are those which involve samples that been analyzed during GBYP-Phase12 and for which all detected pairs were consistently detected in this new analysis involving newly genotyped individuals.

Individual-1	Year	Station	Individual-2	Year	Station	PLOD	GR	PairType
IEO-BA-V-1041	2020	1331	IEO-BA-V-896	2020	882	123.3	0.23	HSP
IEO-BA-V-725	2020	794	IEO-BA-V-645	2020	794	128.5	0.24	HSP*
IEO-BA-V-1082	2022	792	IEO-BA-V-1074	2022	792	158.3	0.25	HSP
IEO-BA-V-692	2020	794	IEO-BA-V-582	2020	794	170.3	0.25	HSP*
IEO-BA-V-1119	2022	792	IEO-BA-V-1078	2022	792	182.2	0.28	HSP
IEO-BA-V-882	2020	794	IEO-BA-V-718	2020	794	195.7	0.29	HSP*
IEO-BA-V-822	2020	794	IEO-BA-V-797	2020	794	282.8	0.36	HSP*
IEO-BA-V-770	2020	794	IEO-BA-V-658	2020	794	510.3	0.56	FSP*

No kin pairs were found among individuals from different sampling stations, years and age classes, except for two larvae sampled at stations located near the Nort-west and South of the Mallorca Island in 2020 with three days of difference. May this pair be a real Half-Sibling pair, would mean that the same adult individual spawned at least at this both stations during the same sampling period. Nevertheless, the relative high amount of kin pairs detected within the same, compared to the amount if kins found between sampling events suggest that this mobility between spawning spots within the same spawning period, even if not necessarily rare, does not result in complete mixing between spawning adults in the Western Mediterranean within the same spawning season.

3.2.4. Conclusions

Understanding the adaptive potential of Atlantic bluefin tuna to changing environments:

- The finding of candidate genomic variants for affecting survivorship during the earliest life-stages of the Atlantic bluefin tuna in the Mediterranean Sea means an advance on the understanding of the genomic basis of adaptive capacity of the species to changing environmental conditions.

Suitability of the larvae for kinship analysis required for CKMR model implementation in the Eastern Atlantic bluefin tuna:

- The obtained results suggest that adult individuals may spawn at different locations in the Westen Mediterranean during the same spawning season. Spatial and temporal connectivity between spawning sites in the Mediterranean Sea needs further study and increased sampling size to determine Atlantic bluefin tuna spawning site fidelity.

- Kinship analysis can be used to explore spawning site fidelity at a fine scale within the Mediterranean Sea.

3.3 Exploring the genetic origin of previously unassigned individuals

During previous GBYP phases (from Phases 6 to 12), more than 3,500 individuals have been assigned to their genetic origin using a 96 SNP panel. While validation of the 96 SNP panel on individuals captured at the spawning areas showed high percentages of genetic origin assignment (see phase 6 Report and Rodríguez-Ezpeleta et al. 2019), approximately 10-25% of individuals could not be assigned to either origin. This could be explained by the presence of genetically intermediate individuals across the North Atlantic as supported by previous results (see GBYP phase 10 Report). Recent studies have shown that these intermediate genetic profiles are more commonly found in the recently discovered spawning area in the Slope Sea (Díaz-Arce et al. 2024). Understanding the origin of previously unassigned individuals will contribute to characterize the mixing of the different genetic groups at feeding aggregates along the North Atlantic and the identification of different migratory patterns.

On the top of this, different studies have reported evidence for changes in the ABFT migratory behavior (Aalto et al. 2021, Jansen et al. 2021), which highlights the need for temporal monitoring to anticipate to these changes. In this context, we have focused on characterizing the genetic origin of individuals captured at anomalous geographic regions or seasons during the last decade analyzing 100 individuals captured during winter and summer seasons in the Bay of Biscay during the last three years using the ABFT SNP array to assess their genetic origin. These findings are useful to evaluate the suitability of these samples for a future implementation of the CKMR model for the eastern Atlantic bluefin tuna stock.

The objectives of this subtask were as follows:

- Understanding bluefin tuna **mixing patterns**: Analyzing the distribution of genetic profiles within the species to gain insights into bluefin tuna mixing patterns.

- Monitoring **migratory behavior**: Investigating Atlantic bluefin tuna migratory behavior to anticipate potential changes in stock distribution.

3.3.2. Materials and Methods

3.3.2.1 Origin assignment based on the 96 SNP panel and sample selection

The genotypes obtained from the 96 SNP panel, initially designed for assigning the genetic origin of Atlantic bluefin tuna (ABFT) individuals (Rodríguez-Ezpeleta et al., 2019), were accessible for a total of 3853 individuals analyzed during previous GBYP Phases. Among these, 769 were larvae, young-of-the-year, and spawning adults, serving as reference individuals in the baseline. To determine the genetic origin of the remaining 3084 samples, we employed the geneclass2 software (Piry et al., 2004). Assignments were made based on the recommendation by Rodríguez-Ezpeleta et al. (2019), considering percentage scores below 80% as unassigned. These assessments utilized the 86 SNPs common to both the first (Rodríguez-Ezpeleta et al., 2019) and second (developed in GBYP – Phase 11) versions of the 96 SNP panel. In the assignment process, we compared eachindividual's genetic origin with the expected origin based on their catch location as recorded in the GBYP sample database. Specifically, ABFT individuals captured west of the 45º meridian were assumed to originate from the Gulf of Mexico, while those captured east of the 45º meridian were associated with the Mediterranean Sea. Among the analyzed individuals, 284 individuals were selected and genotyped using the ABFT SNP array during GBYP Phase 13. The selection included 118 individuals with unexpected origin assignment, 15 individuals assigned to their expected origin (in common with those selected for task 3.4) and 151 unassigned individuals.

To explore the genetic origin of catches in the Bay of Biscay during different seasons, 132 individuals captured in this area during the years 2021 to 2024 and covering different seasons were newly genotyped using the ABFT SNP Array.

3.3.2.2. Analysis of genetic profile based on the ABFT SNP array:

The genotypes of the newly processed samples were obtained by processing the obtained individual CEL files using the Axiom Suite Analysis software and removing individuals with genotyping rates below 0.97 and SNPs with low quality genotype calling discrimination. Obtained genotype tables were merged to the previously existing dataset generated during the GBYP phases 10, 11 and 12, and SNPs were filtered to include only neutral SNPs genotyped in at least the 90% of the samples. Individual ancestry proportions assuming two ancestral populations were estimated using the software ADMIXTURE (Alexander et al. 2009).

3.3.3. Results

3.3.3.1. Origin of previously unassigned or assigned to the unexpected origin using the 96 SNP panel

The final dataset including samples processed with the ABFT SNP Array included 1,763 samples (of which 675 have been newly genotyped in this phase for subtasks 3.2, 3.3 and 3.4) genotyped at 6,069 neutral SNPs. Among these, we observed the following assignments using the 96 SNP panel: 214 individuals were previously unassigned, 66 individuals were unexpectedly assigned to Gulf of Mexico origin, and 172 individuals were unexpectedly assigned to Mediterranean origin. The ABFT SNP Array dataset also included 102 and 134 individuals that had been expectedly assigned to Gulf of Mexico and the Mediterranean origin, respectively. Among the remaining 1,075 individuals not assigned using the 96 SNP panel, 169 were reference individuals (larvae, young-of-the-year, or spawning adults) from the Gulf of Mexico, while 651 were from the Mediterranean Sea.

The individual ancestry values were consistent with previous results, finding genetic differentiation between the Mediterranean and Gulf of Mexico reference individuals with some overlap (Figure 3.3.1). The proportion of Mediterranean ancestral population ranged from 0 to 0.64 and from 0.48 to 1 among the reference Gulf of Mexico and Mediterranean Sea individuals. The origin of those samples that had expectedly assigned to Mediterranean and Gulf of Mexico origin using the 96 SNP panel were confirmed using the ABFT SNP Array, with the exception of 7 individuals (out of 102) that had been expectedly assigned as Gulf of Mexico but genotyping with the ABFT SNP array revealed

proportion of the Mediterranean ancestral population above the range of reference Gulf of Mexico the which was 0.63 (Figure 3.3.1).

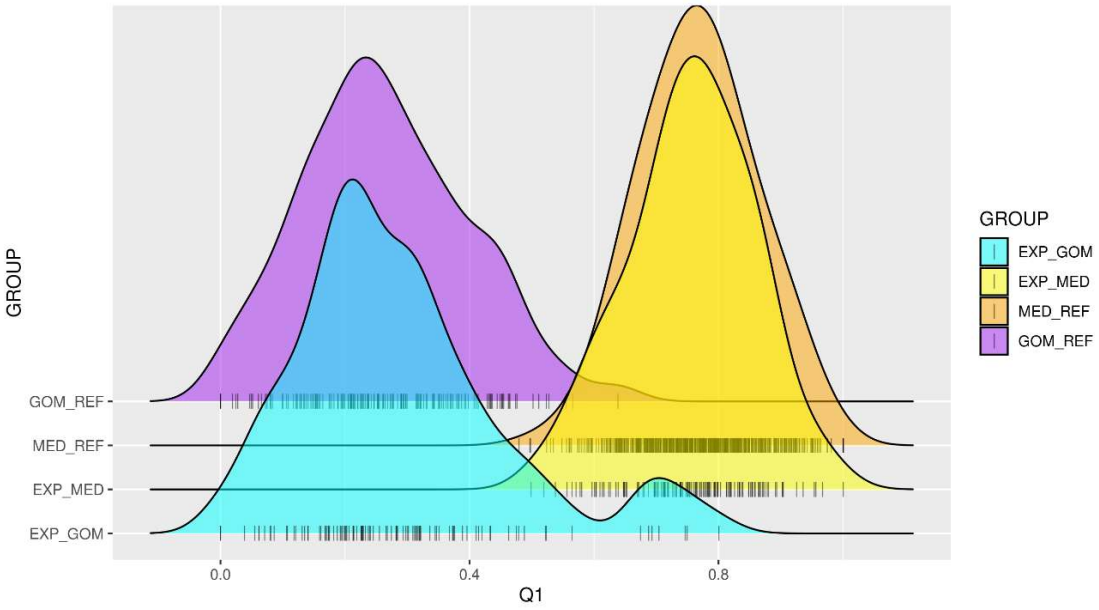

Figure 3.3.1. Distributions of individual ancestry proportion (Q1) of the ancestral population which corresponds to the Mediterranean Sea of the reference individuals from the Gulf of Mexico (GOM_REF, in purple) and the Mediterranean Sea (MED_REF, in orange) and individuals captured to the West and East of the 45º meridian which were assigned to the Gulf of Mexico (EXP_GOM, in blue) and the Mediterranean Sea (EXP_MED, in yellow) respectively with the 96 SNP panel. Individual ancestry proportions were estimated based on the genotypes at > 6000 neutral SNPs obtained with the ABFT SNP Array.

Those samples that were unassigned using the 96 SNP panel covered the whole spectrum of genetic profiles, from pure Mediterranean-like to pure Gulf of Mexico-like (Figure 3.3.2). Moreover, the distribution of the ancestry values of the previously unassigned samples shows two peaks reflecting the distribution of the reference Gulf of Mexico and Mediterranean individuals, and very few showed intermediate genetic profiles, suggesting that the presence of unassigned individuals in the different feeding aggregates using the 96 SNP panel may be explained by limitations of the 96 SNP panel for origin assignment.

The genetic origin of most samples that had been unexpectedly assigned to Mediterranean origin using the 96 SNP panel was confirmed using the ABFT SNP Array, while 8 of these individuals (out of 172) showed MED-like ancestry values below the range of the Mediterranean reference samples. Instead, most individuals that were unexpectedly assigned to the Gulf of Mexico were confirmed to have either Mediterranean-like, intermediate or Gulf of Mexico-like genetic profile (Figure 3.3.3).

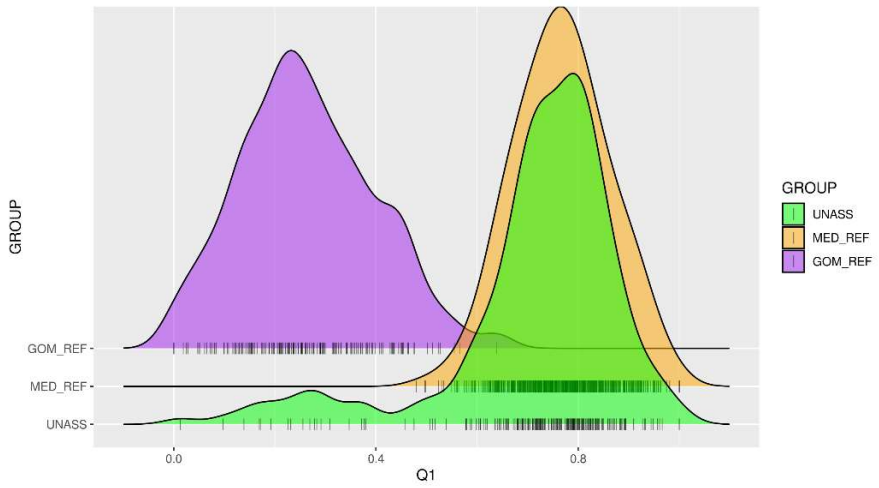

Figure 3.3.2. Distributions of individual ancestry proportion (Q1) of the ancestral population which corresponds to the Mediterranean Sea of the reference individuals from the Gulf of Mexico (GOM_REF, in purple) and the Mediterranean Sea (MED_REF, in orange) and individuals that were unassigned based on results obtained with the 96 SNP panel. Individual ancestry proportions were estimated based on the genotypes at > 6000 neutral SNPs obtained with the ABFT SNP Array.

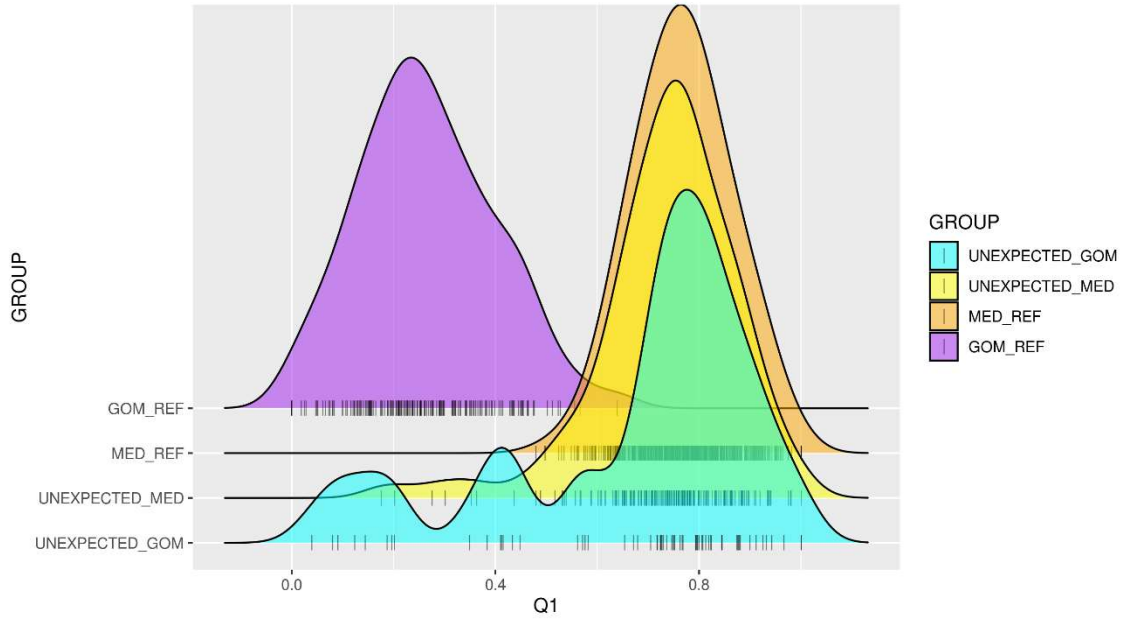

Figure 3.3.3. Distributions of individual ancestry proportion (Q1) of the ancestral population which corresponds to the Mediterranean Sea of the reference individuals from the Gulf of Mexico (GOM_REF, in purple) and the Mediterranean Sea (MED_REF, in orange) and individuals that were unexpectedly assigned to the Gulf of Mexico (UNEXPECTED_GOM, in blue) or to the Mediterranean Sea (UNEXPECTED_MED, in yellow) based on results obtained with the 96 SNP panel. Individual ancestry proportions were estimated based on the genotypes at > 6000 neutral SNPs obtained with the ABFT SNP Array.

These results support that the assignments based on the 96 SNP panel leads individuals erroneously assigned to both Mediterranean and Gulf of Mexico. The mismatch between origin assignment based on the 96 SNP panel and the genetic profile of the individuals was higher among the individuals that were assigned to Gulf of Mexico origin, and it was also higher among the samples that had been assigned to an unexpected origin, meaning that the assignments based on the 96 SNP panel could be overestimating the mixing of both components at both sides of the North Atlantic Ocean, and particularly overestimating the proportion of individuals originating from the Gulf of Mexico.

3.3.3.2. Genetic profile of individuals captured in the Bay of Biscay in different seasons

In total, 190 individuals captured in the Bay of Biscay were successfully genotyped with the ABFT SNP Array. Among those, 101 were captured during the years 2021 to 2024 (Table 3.3.1).

Table 3.3.1. Number of individuals from each catch year and month captured in the Bay of Biscay that were included in the final genotype table.

	YEAR						
Month	**2011**	**2012**	**2021**	**2022**	**2023**	**2024**	**Total**
1			1			5	6
2					6	5	11
3				4	26		30
4					1		1
5					3		3
6	8		3	1			12
7	35	12	4	3	4		58
8	12	8	4	2	6		32
9	12	1	2	2	2		19
10		1	2	5			8
11					5		5
12					5		5
TOTAL	**67**	**22**	**16**	**17**	**58**	**10**	**190**

The 95,8% (182 out of 190) individuals captured in the Bay of Biscay showed individual ancestry values that fall within the distribution of the Mediterranean reference individuals (Q1>0.48) (Figure 3.3.4). However, 8 individuals out of 190 showed a genetic profile compatible with the Gulf of Mexico. These individuals were captured during the months of June of 2021 and from July to September of 2011 and 2012 (Figure 3.3.4 and 3.3.5).

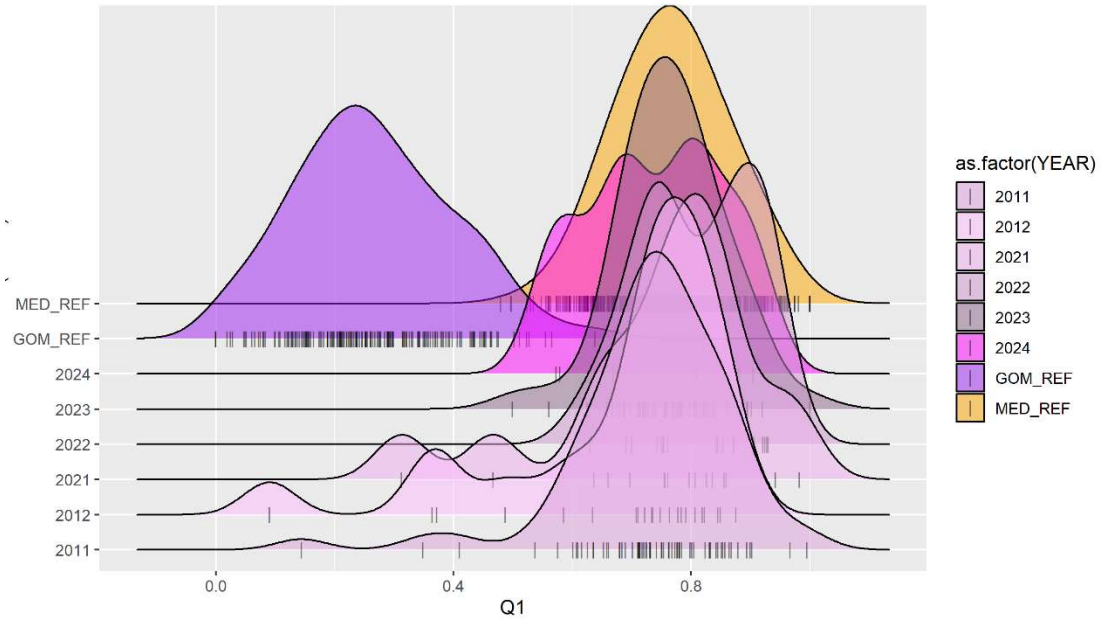

Figure 3.3.4. Distributions of individual ancestry proportion (Q1) of the ancestral population which corresponds to the Mediterranean Sea of the reference individuals from the Gulf of Mexico (GOM_REF, in purple) and the Mediterranean Sea (MED_REF, in orange) and individuals that were captured in the Bay of Biscay in the years 2011, 2012 and from 2021 to 2024. Individual ancestry proportions were estimated based on the genotypes at > 6000 neutral SNPs obtained with the ABFT SNP Array.

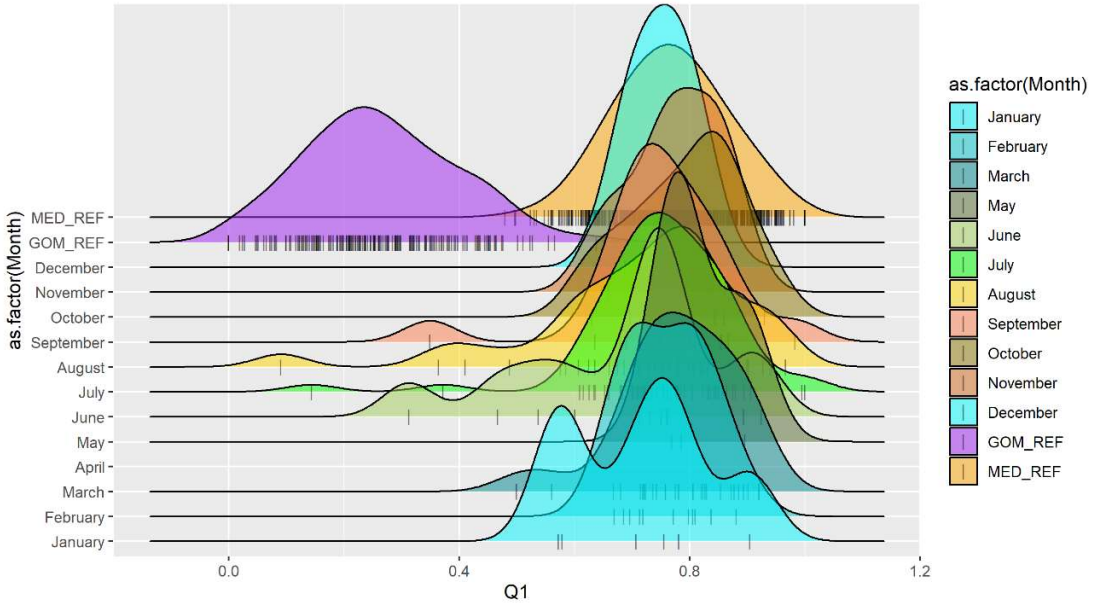

Figure 3.3.5. Distributions of individual ancestry proportion (Q1) of the ancestral population which corresponds to the Mediterranean Sea of the reference individuals from the Gulf of Mexico (GOM_REF, in purple) and the Mediterranean Sea (MED_REF, in orange) and individuals that were captured in the Bay of Biscay during different months of the years 2011, 2012 and 2021 to 2024. Individual ancestry proportions were estimated based on the genotypes at > 6000 neutral SNPs obtained with the ABFT SNP Array.

Interestingly, among those 8 individuals showing Gulf of Mexico-like individual ancestries not compatible with Mediterranean-like genetic profile (out from the overlap area), 5 were of Juveniles age class, 2 were of Medium age class, and only one was Large (captured in the 2011) (Figure 3.3.6). Sex identification based on genetic markers analyzed as described in Subtasks 3.4. revealed that these individuals were of both sexes.

While the sampling size is different across years, the different proportion of Gulf of Mexico-like individuals in different years suggests that migration from western spawning areas into the Bay of Biscay could be variable. Regarding the seasonality, all Gulf of Mexico individuals were captured during the summer months. However, during years 2011 and 2012, which are those with higher proportion of Gulf of Mexico-like individuals, no winter samples were collected (Table 3.3.1). No Gulf of Mexico-like individuals were detected in years 2022 to 2024, although years 2022 and 2023 were sampled during the summer. All the 56 individuals sampled from November to May, which were collected from years 2021 to 2024, showed Mediterranean-like genetic profiles.

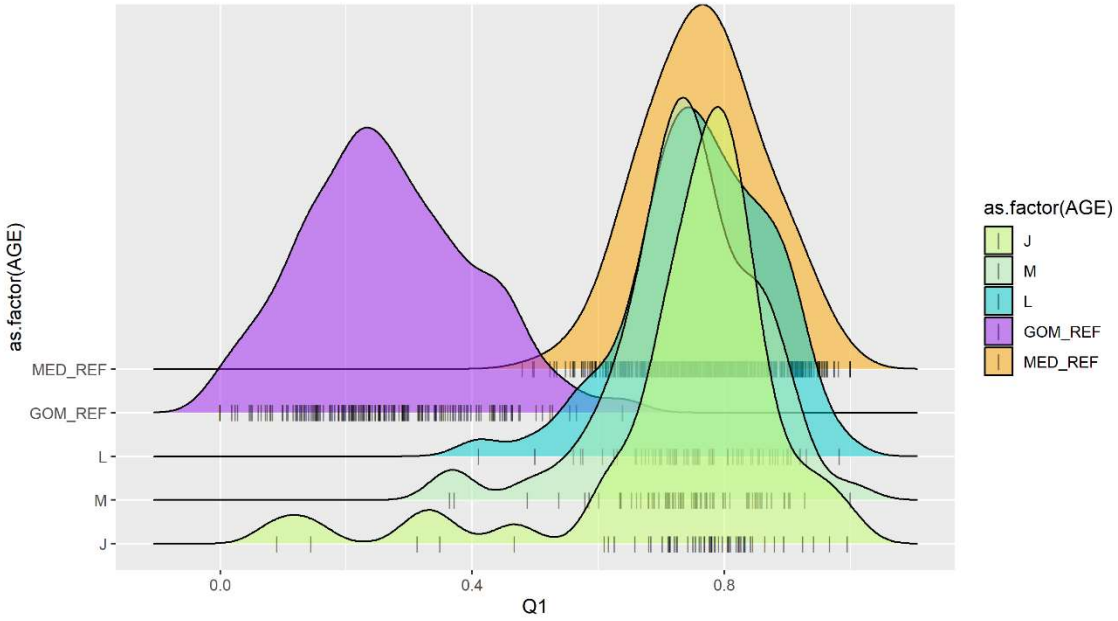

Figure 3.3.6. Distributions of individual ancestry proportion (Q1) of the ancestral population which corresponds to the Mediterranean Sea of the reference individuals from the Gulf of Mexico (GOM_REF, in purple) and the Mediterranean Sea (MED_REF, in orange) and individuals of different age classes (Juveniles = J, Medium = M and Large = L as described for the GBYP database). Individual ancestry proportions were estimated based on the genotypes at > 6000 neutral SNPs obtained with the ABFT SNP Array.

3.3.4. Conclusions

Genetic origin assignments based on the 96 SNP panel could overestimate the proportion of Gulf of Mexico individuals and the mixing of both components at both sides of the North Atlantic Ocean, highlighting the need for a more comprehensive and powerful tool, such as the ABFT SNP Array, for Atlantic bluefin tuna monitoring to provide with more accurate information about the genetic origin of Atlantic bluefin tuna, mixing dynamics and behavior.

Atlantic bluefin tuna individuals of Mediterranean and Gulf of Mexico origin can be found in the Bay of Biscay at different proportions across different years, seasons and age classes, suggesting dynamic migratory behavior of the species.

Unusual increased catches of Atlantic bluefin tuna in the Bay of Biscay during the winter over the last three years are composed by individuals of Mediterranean origin, although higher samples sizes are needed to infer more robust conclusions.

3.4 Evaluate the sex assignment power of genetic markers included in the SNP array

3.4.1. Introduction

To help understanding migratory behavior of ABFT, we have evaluated the success of genetic sex identification based on the genetic markers included in the ABFT SNP array which will allow to pose hypothesis of sex-biased migration. The better characterization of the genetic origin and the evaluation of genetic sex identification will improve the interpretation of the results obtained with the ABFT SNP array for the implementation of the CKMR model in the eastern ABFT.

The objective of this subtask was to **evaluate the assignment accuracy of genetic markers for sex determination** in Atlantic bluefin tuna individuals.

3.4.2. Materials and Methods

Genotypes obtained for 5 genetic sex markers adapted from (Suda et al. 2019) obtained from the result of fluorescent signal of 7 probe-pairs (for two markers two probe-pairs were included) included in the ABFT SNP array were compared with identified sex of 68 individual samples sexed by experts on the reproductive biology of Atlantic bluefin tuna (either from the NOAA and the UCA). Among these 68 individuals, 21 had already been genotyped in previous GBYP Phases and 47 were newly genotyped in this Phase, and 29 had been identified as females and 39 as males. Sex identification rate for each genetic marker was assessed based on concordance with the expected genotypes according to (Suda et al. 2019) among these 68 individuals.

Frequency of different genotype combinations at sex markers with sex identification power > 90% were then explored at 325 individual samples genotyped using the ABFT SNP array for which sex data is available in the GBYP sample database, among which 136 had been identified as females and 189 as males.

3.4.3. Results and discussion

From the 7 probe-pairs corresponding to 5 different genetic markers for sex identification, genotypes obtained from 4 probe-pairs corresponding to 4 different genetic markers provided with a sex identification rate >90% when compared with the expected genotypes

for each sex (Table 3.4.1). Among the 26 female individuals, 24 and 2 consistently showed the expected female and male-characteristic genotypes. Among the 39 analyzed male individuals, 36 and 3 showed consistent expected genotypes for male and female individuals respectively, except for one female-characteristic sample which showed the male expected genotype for one of the markers.

Table 3.4.1. Number (n) and percentage (%) of samples assigned as females (F) and males (M) by visual inspection by experts in reproductive biology of Atlantic bluefin tuna that showed the expected genotypes according to (Suda et al. 2019) for each probe-pair corresponding to the sex genetic markers included in the ABFT SNP array. Markers showing sex identification rate of 90% or higher for both sexes are shaded in green.

	n		%	
MarkerName	F	M	F	M
SexIA	27/29	36/39	93.1	92.3
SexIB	27/29	37/39	93.1	94.9
SexIIA - ProbePair1	0/24	12/39	0.0	30.8
SexIIA - ProbePair2	18/28	27/38	64.3	71.1
SexIIB - ProbePair1	26/28	30/35	92.9	85.7
SexIIB - ProbePair2	27/29	36/39	93.1	92.3
SexIII	27/29	36/39	93.1	92.3

The sex of the 86% (117 out of 136) and the 76.7 % (145 out of 189) of the analyzed ABFT individuals identified as females and males respectively in the GBYP database was confirmed using the genetic markers included in the ABFT SNP array (Figure 3.4.1). Four samples showed genotypes characteristic of female individuals, except for the marker SexIB, which showed the male-characteristic genotype. Among these four samples, three were identified as female and one as male in the GBYP database.

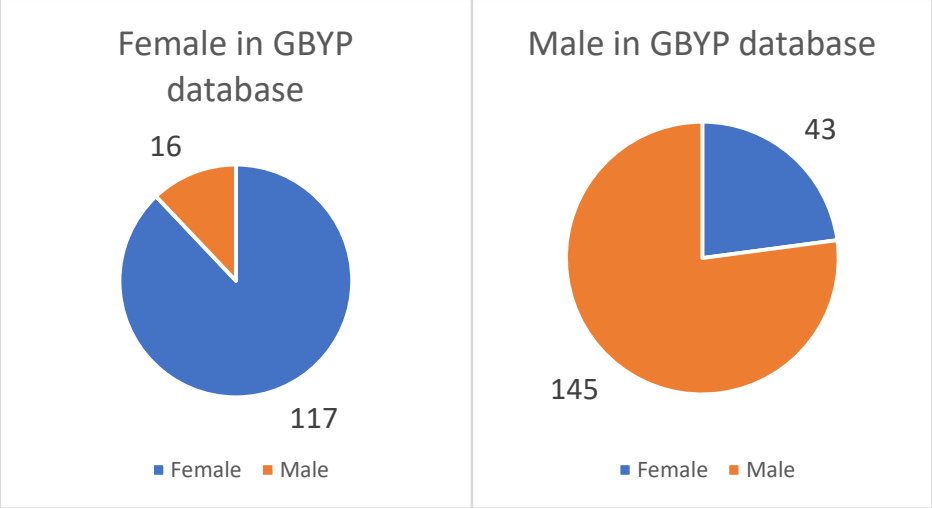

Figure 3.4.1. Pie charts showing the proportion and number of individuals that are registered in the GBYP database as female (left) and male (right) which were confirmed as female (blue) and male (orange) based on best discriminating genetic sex markers.

3.4.4. Conclusions

Genetic sex identification using the ABFT SNP array:

The **correct sex assignation** of the ABFT samples using the genetic markers for sex determination included in the ABFT SNP array will allow the reliable automated sex identification of ABFT individuals genotyped using the ABFT SNP array required for kinship analysis needed for the implementation of CKMR models without increasing costs.

Bibliography

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

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Buckley, B. A., A. Y. Gracey, and G. N. Somero. 2006. The cellular response to heat stress in the goby Gillichthys mirabilis: a cDNA microarray and protein-level analysis. Journal of Experimental Biology 209:2660-2677.

Buckley, B. A., and G. N. Somero. 2009. cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish Trematomus bernacchii to alter gene expression in response to heat stress. Polar Biology 32:403-415.

Cabanettes, F., and C. Klopp. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6:e4958.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and G. P. A. Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Díaz-Arce, N., P.-A. Gagnaire, D. E. Richardson, J. F. Walter III, S. Arnaud-Haond, J.-M. Fromentin, D. Brophy, M. Lutcavage, P. Addis, F. Alemany, R. Allman, S. Deguara, I. Fraile, N. Goñi, A. R. Hanke, F. S. Karakulak, A. Pacicco, J. M. Quattro, J. R. Rooker, H. Arrizabalaga, and N. Rodríguez-Ezpeleta. 2024. Unidirectional trans-Atlantic gene flow and a mixed spawning area shape the genetic connectivity of Atlantic bluefin tuna. Molecular Ecology 33:e17188.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for Ancient Admixture between Closely Related Populations. Molecular Biology and Evolution 28:2239-2252.

Garrison, E. P., and G. T. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv: Genomics.

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

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and G. P. D. P. Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Malinsky, M., M. Matschiner, and H. Svardal. 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. Molecular Ecology Resources 21:584-595.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2014. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. Molecular Biology and Evolution 32:244-257.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLOS Genetics 8:e1002967.

Piry, S., A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. Journal of Heredity 95:536-539.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559-575.

Rey, C., A. Darnaude, F. Ferraton, B. Guinand, F. Bonhomme, N. Bierne, and P.-A. Gagnaire. 2020. Within-Generation Polygenic Selection Shapes Fitness-Related Traits across Environments in Juvenile Sea Bream. Genes 11:398.

Rodríguez-Ezpeleta, N., N. Díaz-Arce, J. F. Walter III, D. E. Richardson, J. R. Rooker, L. Nøttestad, A. R. Hanke, J. S. Franks, S. Deguara, M. V. Lauretta, P. Addis, J. L. Varela, I. Fraile, N. Goñi, N. Abid, F. Alemany, I. K. Oray, J. M. Quattro, F. N. Sow, T. Itoh, F. S. Karakulak, P. J. Pascual-Alayón, M. N. Santos, Y. Tsukahara, M. Lutcavage, J.-M. Fromentin, and H. Arrizabalaga. 2019. Determining natal origin for improved management of Atlantic bluefin tuna. Frontiers in Ecology and the Environment 17:439-444.

Suda, A., I. Nishiki, Y. Iwasaki, A. Matsuura, T. Akita, N. Suzuki, and A. Fujiwara. 2019. Improvement of the Pacific bluefin tuna (Thunnus orientalis) reference genome and development of male-specific DNA markers. Scientific Reports 9:14450.

Yeaman, S. 2015. Local Adaptation by Alleles of Small Effect. Am Nat 186 Suppl 1:S74-89.

Zhou, Y., H.-C. Fu, Y.-Y. Wang, and H.-Z. Huang. 2022. Genome-wide association study reveals growth-related SNPs and candidate genes in mandarin fish (Siniperca chuatsi). Aquaculture 550:737879.

4. INFORMATION SYSTEM

Task Leader: Igaratza Fraile (AZTI) & Iraide Artetxe-Arrate (AZTI)

Participants:
AZTI: Ainhoa Orbe

Following the Commitment to the EU to increase dissemination of the data, in this phase we have developed a detailed and updated catalogue of samples stored in the GBYP Tissue Bank. The catalogue is available through a link to a persistent and reliable public web repository. This user-friendly interface has been developed within a Shiny app and offers an integrated and interactive data visualization tool. This tool enables to compile data from multiple databases and data sources providing a superior overview. Results are visualized with maps, graphs and diagrams, and provides estimates of key biological parameters for stock management. The online version of the application is currently hosted in https://aztidata.local/BioTuna, and is planned to be accessible for external users by the end of the current contract.

BioTuna information system is designed to store and organize sample distribution and associated data for tuna species. It offers the possibility to search for samples based on taxonomic, biometric or tissue type criteria, which allows conducting specific searches for a variety of research initiatives (Figure 4.1). The Map, Graphs and Data tabs are sensitive to the filters used to do the targeted search.

The filters currently available are:

- Species: Multiple choice selector for species by scientific name.
- FAO-Area: Multiple choice selector for areas as defined by FAO major fishing areas for statistical purposes.
- Sub-Area: Multiple choice selector for smaller areas defined inside the FAO areas and defined as sampling strata used for GBYP Biological Studies
- Catch year: Multiple choice selector for the year when the fish was caught from the wild
- Birthyear: Multiple choice selector specific cohorts (estimated from age-length relationship)

- Size class: Multiple choice selector for the class assigned according to fish weight. In the case of ABFT categories are Larvae, Age-0 (<3kg), Juvenile (2-35 kg), Medium (25-100 kg) and Large (>100 kg).
- Straight fork Length: The bar allows selecting the desired size range.
- Total Weight: The bar allows selecting the desired weight range.
- Length based age: The bar allows selecting the desired age range.
- Sex: Multiple choice selector for sex.
- Sample type: Multiple choice selector for each sample type (otolith, muscle, fin, spine, gonad and/or stomach)

In multiple choice selectors, all the options are selected by default. For species-specific searches, filters must be applied to the targeted search.

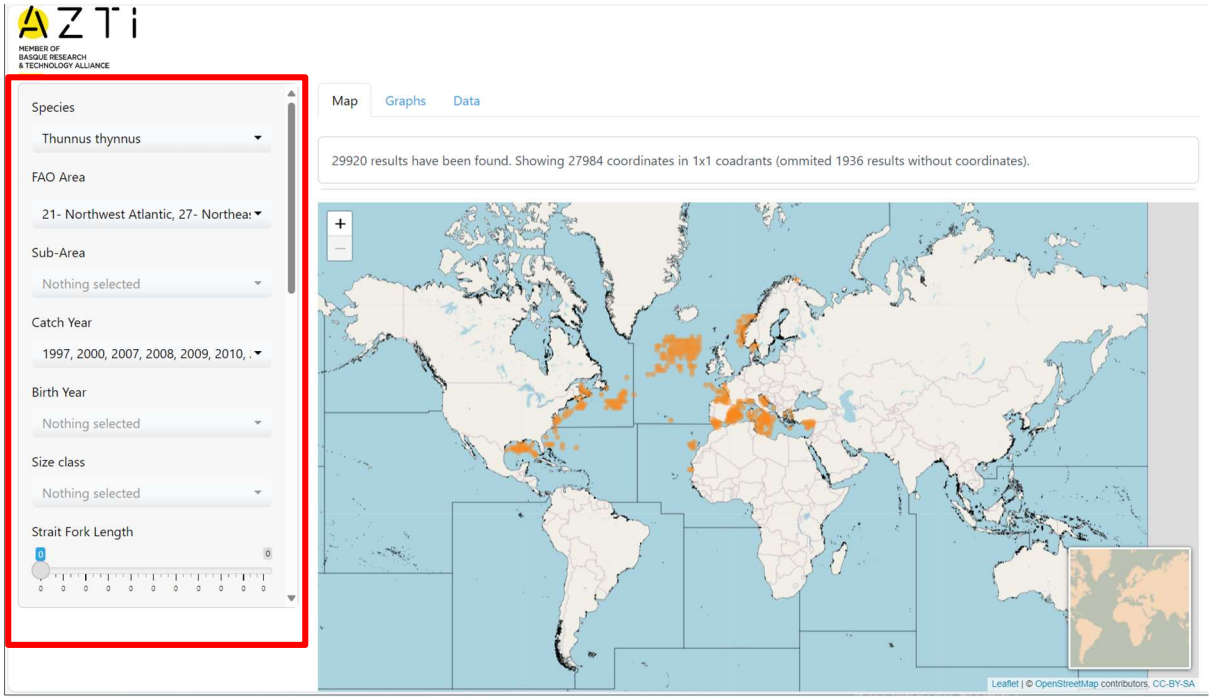

Figure 4.1: Filter selector and corresponding geographic distribution of data. Note that to protect the privacy of individual local data, geographic positions have been masked and transformed into a 1x1 degree grid.

After selecting filters, the user should scroll down and click the "Apply" button at the bottom of the application. This action will update the interactive elements accordingly. If the user needs to start over, they can simply click the "Clean" button to reset the search.

The **Map** tab displays the spatial distribution of data on a 1x1 degree grid (Figure 4.1), with the number of individuals that meet the selected criteria shown on top of it. The positions displayed on the map have been masked to preserve the confidentiality of individual locations. When zooming at the map, the number of individuals represented by each of the dots will become visible. (Figure 4.2).

The **Graphs** tab offers enhanced data visualization through distribution charts. These charts provide insights into various fish characteristics, such as straight fork length, weight, estimated age, year of capture, and estimated year of birth. Users can adjust the bin size to improve visualization. Additionally, the charts offer a convenient way to explore data distribution based on sex or FAO area (Figure 4.3).

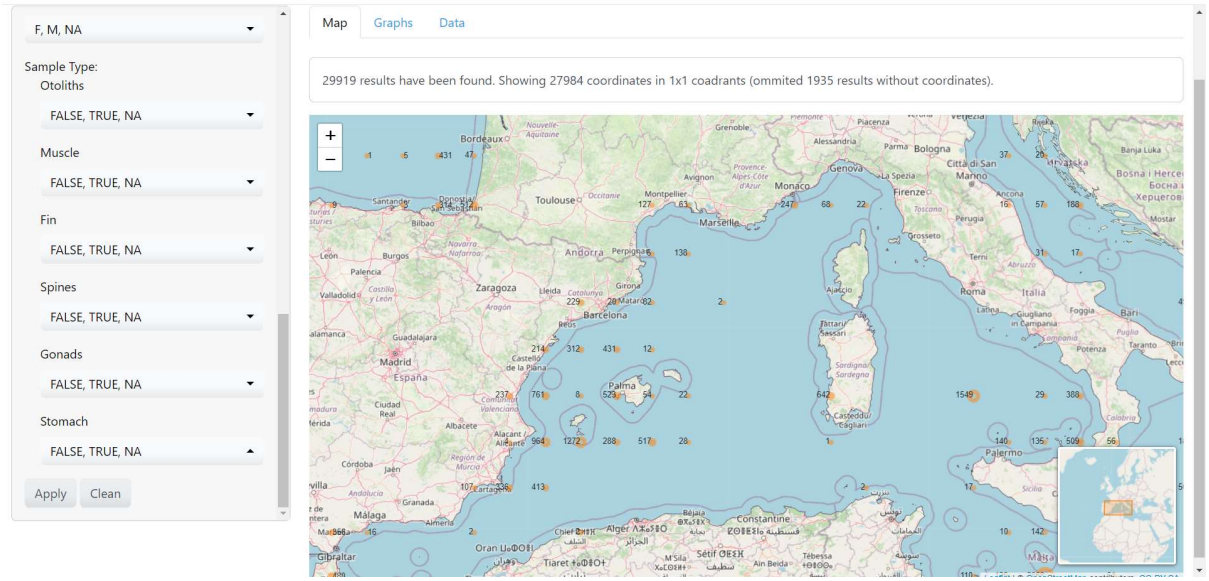

Figure 4.2: Zoomed map showing the number of samples found at each 1x1 degree cell.

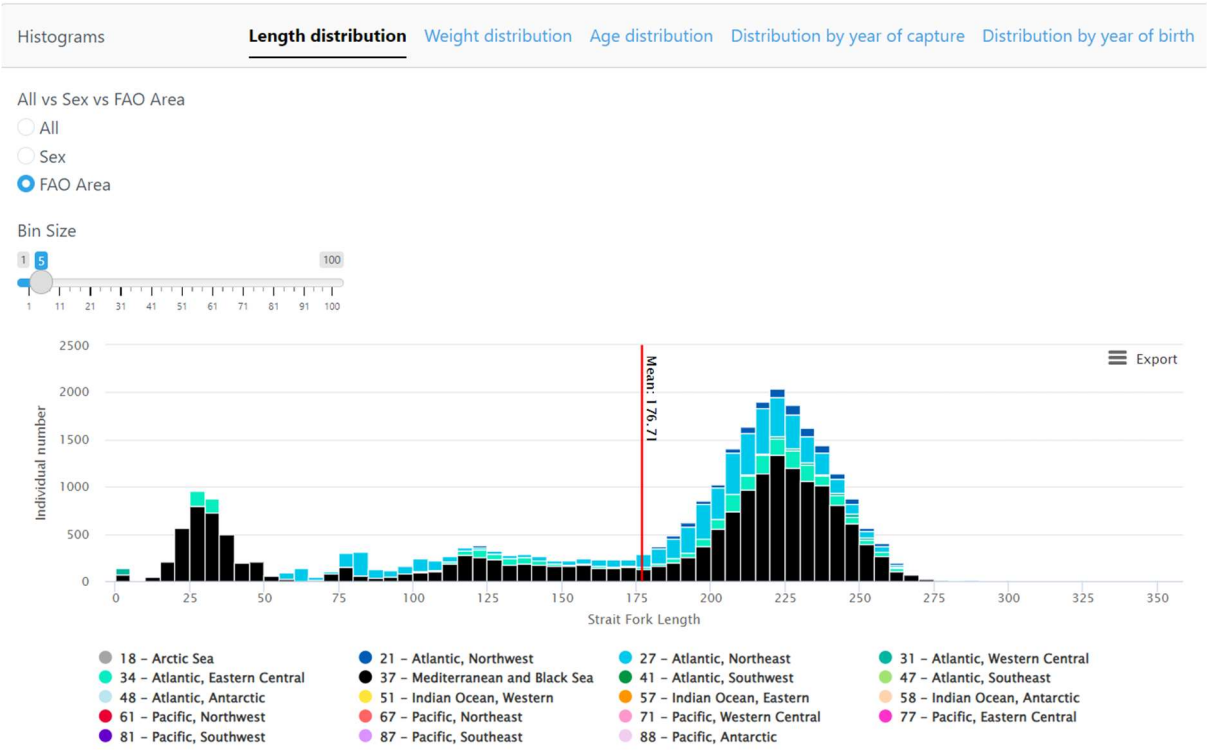

Figure 4.3: Histogram example for length-distribution data by FAO area at a 5-cm bin size. Mean value is indicated by the vertical red line.

Additionally, the species-specific weight-length relationship can be visualized interactively. The resulting relationship is computed for the selected dataset (Figure 4.4). Similarly, male-to-female proportions and maturity curves by length are calculated for each chosen criterion (Figure 4.5). Within the maturity curve, L_{50} value is computed and displayed on screen. All charts generated within the application can be exported in png, jpg, pdf or svg format. The outputs of the application are automatically updated based on user's search criterion. The associated metadata can be downloaded from the "Data" tab with a login name and a password. To protect data privacy, only account owners have the option to access and download this metadata. These credentials will be shared with ICCAT GBYP Biological Program coordinators and can be provided upon request.

In addition to the features mentioned earlier, the application provides several other valuable functionalities:

- Researchers can explore population dynamics by analyzing length data and size distribution across different regions.

- By examining how length changes over time, scientists gain insights into the health and sustainability of fish populations.
- Users can compare weight-length relationships between different areas, sex of years. This interactive feature allows for quick visual comparisons.
- The application allows users to define specific criteria for analysis. Whether it's a particular region or time-period, you can tailor the data selection.
- Customization ensures that the results align with the research focus or management goals.

The BioTuna platform plays a crucial role in streamlining data tracking and reporting. By offering real-time information on essential biological parameters, it facilitates stock evaluation and effective species management. In essence, BioTuna serves as a digital window into the ICCAT-GBYP biological database, providing researchers, policymakers, and stakeholders with valuable insights for sustainable fisheries practices.

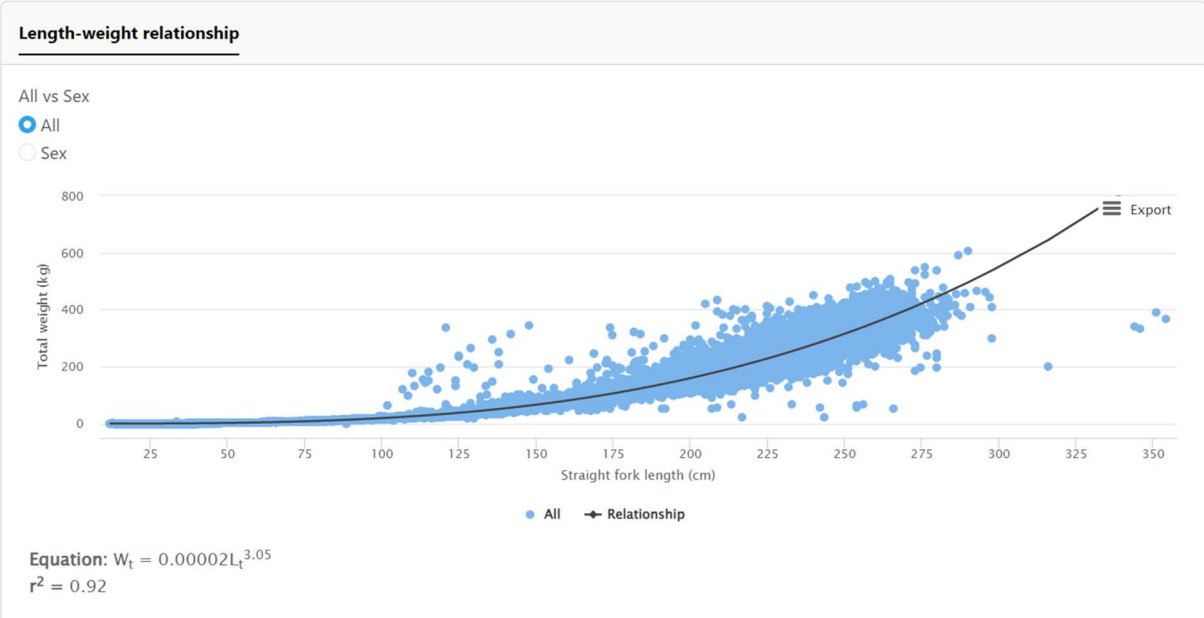

Figure 4.4: Scatterplot of Weight-Length data and associated relationship of the form W_t=aL_t^b for Atlantic bluefin tuna from all areas sampled, where W_t represents total weight (in kg) and L_t and strait fork length (cm). The degree of correlation is denoted by r^2.*

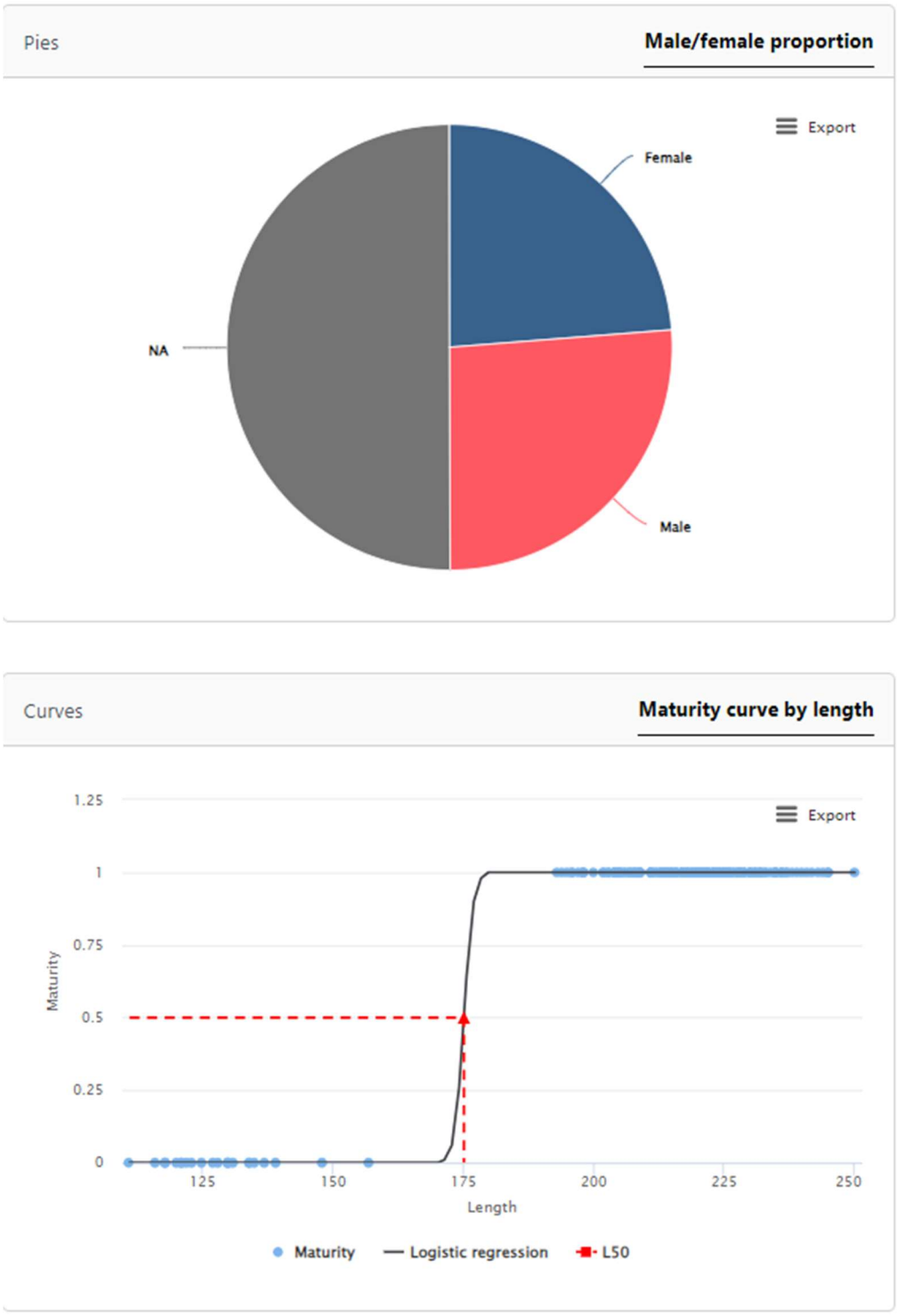

Figure 4.5 Examples of interactive charts for male/female proportion and maturity curve for bluefin tuna. The estimated length of 50 % maturity (L_{50}) is shown in red.

5. ANNEX

- ANNEX I: Sampling Protocol
- ANNEX II: Detailed and updated catalogue of samples stored in the GBYP Tissue Bank
- ANNEX III: Power point presentation of the main results