SHORT TERM CONTRACT FOR BIOLOGICAL STUDIES - ICCAT ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA (ICCAT GBYP – PHASE 14)

Final Report
(Deliverable # 3) for:

ICCAT

Scientific coordinator:

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

Pasaia, December 20th, 2024

This project is co-funded by the European Union

PARTNERS:

Fundación AZTI – AZTI Fundazioa,

(AZTI)

Instituto Español de Oceanografía,

(IEO-CSIC)

SUBCONTRACTORS and COLLABORATORS:

Xenetica Fontao

If proceed cite this document as Fraile, I., Artetxe-Arrate, I., Diaz-Arce, N., Torres, A., Reglero, P., Rodriguez-Ezpeleta, N., Etxebarria, S., Gutierrez, N., Mendibil, I., Orbe, A., Garcia, G., and Serrano, N. (2024). Draft final report on short term contract for biological studies - ICCAT Atlantic-Wide Research Programme for bluefin tuna (ICCAT-GBYP Phase 14).

TABLE OF CONTENTS

CONTEXT

On September 4th, 2024, the consortium coordinated by Fundación AZTI-AZTI Fundazioa, comprising partners Fundación AZTI-AZTI Fundazioa (AZTI) and Instituto Español de Oceanografía (IEO-CSIC), submitted a proposal in response to the call for tenders for biological and genetic sampling and analysis (ICCAT-GBYP Phase 14).

This proposal was awarded and the final contract between ICCAT and the consortium represented by Fundación AZTI-AZTI Fundazioa was signed on September 23rd, 2024.

The present report corresponds to the revised final report (**Deliverable #3**) to be submitted to ICCAT in the framework of this contract.

EXECUTIVE SUMMARY

The main objective of the GBYP Biological and sampling program is to deepen our understanding of the population structure, mixing patterns, and growth dynamics of Atlantic bluefin tuna (ABFT). By doing so, we aim to develop and refine methodologies that effectively integrate existing biological and ecological knowledge. The efforts have been directed towards gathering comprehensive data that can inform sustainable management practices that support effective stock management and conservation strategies, ensuring the long-term viability of this important species.

The research carried out in 2024 focuses on advancing in the feasibility study for the application of Close-Kin Mark-Recapture (CKMR) to the eastern Atlantic bluefin tuna stock, with a particular emphasis on evaluating the possibility of using larvae in these studies as source of information about the juvenile fraction of the population. The CKMR approach aims to provide a fishery-independent method for monitoring key demographic parameters such as spawning stock biomass and exploitation rates.

Additionally, it has been prioritized the maintenance of the ICCAT-GBYP tissue bank, collected along all the successive Phases of the program, ensuring its continued utility for population-level genomics, age and growth, reproduction, and/or population structure studies.

During Phase 14, a total of 3998 larvae from the Balearic Sea spawning ground were sorted and properly stored. These larvae were collected during the TUNIBAL oceanographic survey conducted in 2022. Of these, 3,822 samples were genotyped, along with 136 larval samples received in Phase 13, resulting in a total of 3,840 larvae genotyped. The remaining larvae were stored in the ICCAT-GBYP tissue bank for future analyses. Additionally, the consortium received finlet, caudal, and keel tissue samples from six tuna captured by Portuguese traps to evaluate the DNA quality of the different sampling strategies.

The distribution of all samples within the ICCAT-GBYP tissue bank 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). Appropriate storage for all biological samples collected during previous phases, now available in the Biobank, has been guaranteed.

This report provides a detailed description of the available datasets prepared for the analysis of mitochondrial genome sequencing data in ABFT, focusing on preparing specific datasets from three main data sources: pooled-sequencing data and individual whole genome sequencing data produced by AZTI, and sequences downloaded from GenBank. In pooled-sequencing data, the coverage depth varied between pools, with an average depth ranging from 233x to 1512x per pool, each pool containing DNA from 10 to 50 individuals. This high average coverage enables the identification of alternative nucleotides and their frequency estimation across different groups of pooled individuals. The average coverage depth for the individual whole genome sequencing data ranged from 2.8x to 66.5x, ensuring high-confidence recovery of whole individual mitochondrial haplotypes. These data will be crucial for individual accurate mitochondrial haplotype reconstruction. Ten complete mitochondrial genome sequences of *Thunnus thynnus* were obtained from GenBank and successfully mapped against the reference genome of the ABFT. Designing specific, cost-effective assays targeting suitable mitochondrial genomic regions is necessary to determine mitochondrial haplotypes. This enables the assessment of whether half-sibling bluefin tuna share the same mother or father with an associated probability. The dataset prepared here will enable the analysis of mitochondrial haplotypes in the Mediterranean Sea, which is the first step in determining the feasibility of such assays.

The genotyping results obtained for the 3,822 larvae collected in 2022 were merged with previously produced data for larvae and young-of-the-year collected in the Mediterranean Sea between the years 2018 and 2023. This dataset was analyzed for the detection of kin-pairs. As expected, no parent-offspring pairs were found, but 229 full-sibling and 1,019 half-sibling pairs were identified. The analysis revealed that in the Balearic Sea the number of within-station kin-pairs increases with sampling effort and confirmed that one adult can spawn across multiple stations in a season.

Detection of half-sibling pairs involving larvae from different years supports spawning fidelity to the same area.

Overall, 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, SAMPLE PREPARATION AND MAINTAINANCE OF THE GBYP TISSUE BANK AND ASSOCIATED CATALOGUE

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

Participants:

AZTI: Natalia Diaz-Arce, Iñaki Mendibil, Naiara Serrano, Goretti Garcia, Ainhoa Orbe, Natalia Gutierrez

IEO-CSIC: Asvin Perez-Torres

The biological sampling conducted under Phase 14 follows a specific design aimed at addressing key questions necessary for the implementation of CKMR for the eastern Atlantic bluefin tuna. Consequently, the sampling in this project included the collection of larvae from the Balearic Sea in the western Mediterranean Sea, one of the primary spawning areas of bluefin tuna. The project successfully achieved its objectives, ensuring the reception and quality checking of 3,998 larval samples, and preparing 3,822 for analyses.

1.1 Larvae identification and sorting

The collection of Atlantic bluefin tuna larvae in the main spawning area of the western Mediterranean provides an opportunity to provide samples of the early life stages of this species to the biological sample bank. Since 2019, one of the collectors used in the sampling of larvae using Bongo net is preserved in ethanol. This preservation method ensures that larvae can be used for other purposes rather than species identification. Sample sorting, initial ID and curation are critical to the success of obtaining high quality samples. This activity is focused on providing larval samples preserved in ethanol to the GBYP, ensuring an adequate number of samples are available for future analyses.

The stations and larvae were selected following discussions with the team responsible for statistical analysis and modeling within the ABFT eastern stock CKMR feasibility study

and considering previous discussions with the expert group participating in the GBYP workshops focused on close-kin issues.

Fish species were identified after sampling using a 90 cm Bongo net equipped with a 500-micron mesh size and preserved in ethanol. Sampling was conducted at different stations during the 2022 oceanographic survey in the primary tuna spawning ground in the western Mediterranean, the Balearic Sea.

The results from the analyses of samples preserved in formalin during Tunibal survey, which provide a spawning stock biomass index based on larval abundances, were considered to identify the ethanol-preserved samples containing bluefin tuna larvae suitable for genetic analyses. Based on recommendations from statistical and modeling experts, larvae from six stations sampled during the 2022 survey were selected. These stations represented different orders of magnitude in larval abundance, as determined by the number of larvae observed in the formalin samples.

Tuna larvae were separated from the rest of zooplankters in the selected samples, and then Atlantic bluefin tuna (ABFT) larvae were identified. The larvae were then placed in vials according to the abundance in the samples: individually if the abundance was up to 10 larvae, in groups of 10 if the abundance was up to 100 larvae, and in groups of 100 if the abundance was up to 1,000 larvae. The final count of larvae, their developmental stage, and the corresponding vial information are provided in Table 1.1. These vials were finally sent to AZTI for further analyses.

Table 1.1: Details on the larvae sorted and sent to AZTI´s laboratories for genotyping during Phase-14

ID	Survey	Station	Order	Bongo	Net	Lat (ºN)	Lon (ºE)	Date	Nº	Pre-flexion	Flexion	Post-flexion	Yolk-sac
1	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	300	300			
2	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	300	300			
3	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
4	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
5	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
6	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
7	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
8	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
9	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
10	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
11	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
12	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
13	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
14	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
15	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
16	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
17	TU0622	1402	14	B90	500	38.83250	0.99867	17/06/2022	100	100			
18	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
19	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
20	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
21	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
22	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
23	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
24	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
25	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
26	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
27	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
28	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
29	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	100	100			
30	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	9		9		
31	TU0622	1587	18	B90	500	38.49983	0.99983	18/06/2022	1			1	
32	TU0622	808	80	B90	500	39.99883	4.47333	24/06/2022	25	25			
33	TU0622	790	116	B90	500	40.01650	2.08617	28/06/2022	5	5			
34	TU0622	1127	5	B90	500	39.33300	1.43517	17/06/2022	16	16			
35	TU0622	1127	5	B90	500	39.33300	1.43517	17/06/2022	2		2		
36	TU0622	1400	15	B90	500	38.83167	0.78350	17/06/2022	72				72
37	TU0622	1593	29	B90	500	38.49900	1.65183	19/06/2022	364	364			
38	TU0622	893	89	B90	500	39.83217	4.03750	25/06/2022	125	125			
39	TU0622	1222	7	B90	500	39.17067	1.22033	17/06/2022	26	26			
40	TU0622	719	78	B90	500	40.16517	4.26450	24/06/2022	46	46			
41	TU0622	701	110	B90	500	40.16650	2.30200	27/06/2022	7	7			

1.2 Sample preparation and maintenance of the GBYP tissue bank and associated catalogue

During this Phase 14 AZTI team have received approximately 3998 larval samples from the Balearic Sea, provided by the IEO-CSIC team, along with finlet, caudal, and keel tissue samples from one medium and five large ABFT individuals captured in Portuguese traps. (Table 1.2.1 and Figure 1.2.1), provided by IPMA. All samples underwent rigorous quality checks to ensure their suitability for subsequent analyses. No significant issues were encountered during the reception process, and all samples met the required quality standards.

Table 1.2.1: Fish and larvae samples received during Phase 14.

ICCAT MSE Region	Area	Sampling institute	Size class sampled					TOTAL
			Larvae	Age 0	J	M	L	
			-	<3 kg	3-25 kg	25-100kg	>100 kg	
MED	Balearic Sea	IEO	3998					3998
SATL	Str. Gibraltar	IPMA				1	5	6
TOTAL			3998			1	5	4004

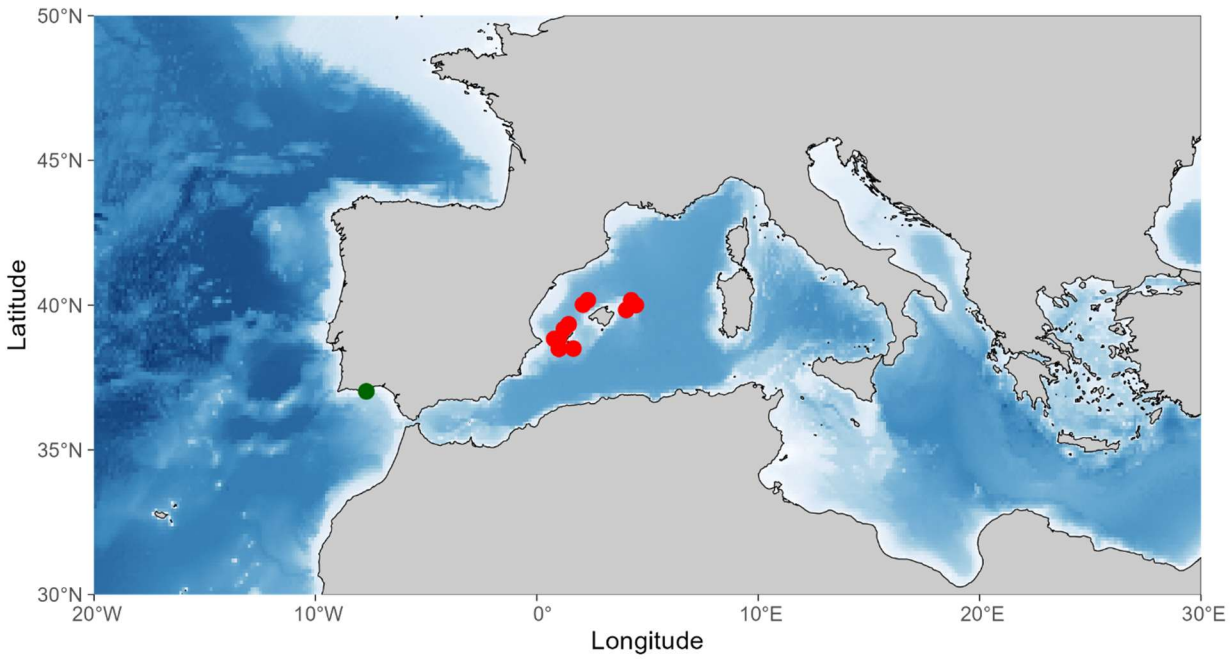

Figure 1.2.1: Map showing larvae (red dots) and adult (green dot) ABFT samples capture positions received during Phase 14.

Of the samples received in Phase 14, 3,686 were selected for genotyping. Additionally, 136 larval samples from Phase 13 were included, bringing the total number of larvae samples analyzed to 3,822. Selection of samples for genotyping followed the criteria established by ICCAT and CKMR modelers group. The remaining larvae were stored appropriately according to the specified protocols in the AZTI Biobank (Table 1.2.2).

Table 1.2.2: Larvae stored at AZTI Biobank that have not been prepared nor analysed yet.

	Survey			
Station	**TB0620**	**TU0622**	**TU0623**	**Total**
701		5		**5**
719		15		**15**
808		15		**15**
893		29		**29**
1127		12		**12**
1222		7		**7**
1331			5	**5**
1402	82		46	**128**
1587		9		**9**
1593		173		**173**
1139	81			**81**
1239	8			**8**
703	12			**12**
876	10			**10**
878	124			**124**
962	5			**5**
964	35			**35**
966	103			**103**
Total	**460**	**265**	**51**	**776**

For sample preparation, tubes containing multiple larvae were poured into petri dishes embedded in ethanol, and larvae were individualized using tweezers and placed into separated wells in plates of 96 wells (Figure 1.2.2). Larvae were manipulated with the help of a Petri disc and tweezers washed in ethanol after each use. 70-100µl of ethanol 96% were pipetted into each well for sample conservation. Plates were labelled and sealed using silicon lids and film and transported in dry ice to the genotyping facilities (Xenetica Fontao) in four different batches to speed up the process (Figure 1.2.2). DNA of each sample was extracted using the Biosprint 96 DNA Blood Qiagen Kit and genotyped using a GeneTitan™ MC machine in batches of 384 samples to produce raw CEL files for each sample. Genotyping of the samples required between 14 to 25 days from shipping to data obtention.

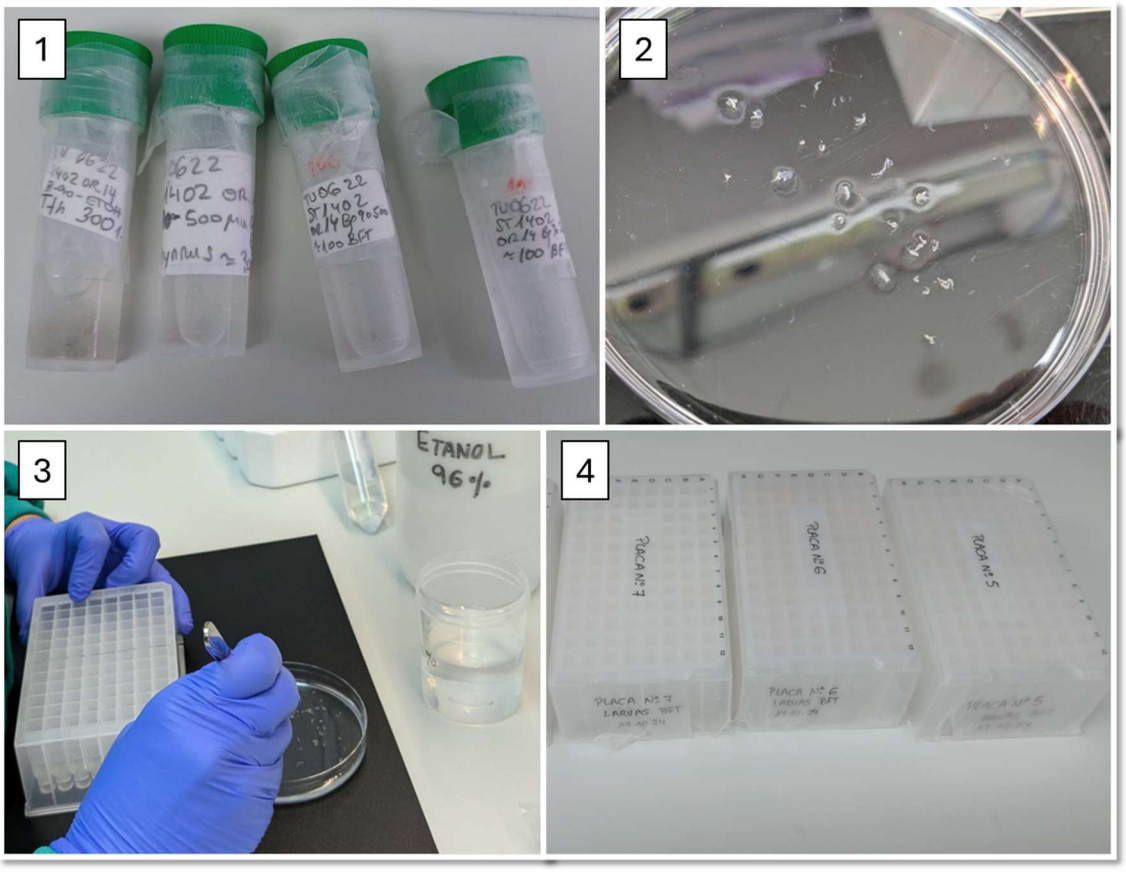

Figure 1.2.2. Preparation process of larvae samples for posterior genotyping. Tubes containing samples from selected stations (1) are poured into Petri dishes and rinsed with 96% ethanol using a Pasteur pipette to recover all samples (2). Larvae are individually collected from the Petri dish using tweezers, which are cleaned after each collection, and placed into individual wells of 96-well plates (3). The plates are then labelled and covered with specific silicon lids and film, making them ready for shipment to the genotyping facilities (4).

The list of larval samples from each station, prepared and sent for genotyping, is detailed in Table 1.2.3. Additionally, 18 tissue samples from six bluefin tuna individuals captured in Portuguese traps were sent for genotyping to assess their suitability for CKMR kinship analysis. The data derived from these samples will be analyzed, if so decided, within the next GBYP Phase. In total, 3,840 samples were genotyped within this call for tenders.

Table 1.2.3. Number samples sent for genotyping during Phase 14.

Size class	TUNIBAL22 - Station	Number of samples
Larvae	701	2
Larvae	719	21
Larvae	790	5
Larvae	808	10
Larvae	893	96
Larvae	1127	6
Larvae	1222	19
Larvae	1400	57
Larvae	1402	2183
Larvae	1587	1232
Larvae	1593	191
Large	-	18
TOTAL		**3840**

The metadata associated with samples received during Phase 14 were cross-checked for accuracy and completeness. After verification, the metadata were integrated into the ICCAT-GBYP tissue bank's information system, ensuring seamless management and accessibility (https://aztidata.es/BioTuna/). Additionally, the catalogue of samples stored in the GBYP Tissue has been updated and refined, following ICCAT-GBYP guidelines (See Anex II). Currently, the catalogue contains information from 33,769 individuals. The consortium has continued to provide appropriate storage for all biological samples collected during previous phases, including otoliths and spines stored in the general store, gonads preserved in Bouin's or Hollande solution, muscle and fin tissues stored in freezers, with duplicates in separate buildings for added security and larvae that had not yet being processed. The tissue bank facilitates microchemical, genetic, histological, and morphological analyses, aligning with recommendations from the SCRS and the GBYP Steering Committee, with a key focus on stock-piling samples suitable for future studies.

Currently, the consortium provides storage of 17,349 otolith pairs, 7,981 spines, 2,046 gonad samples and 23,811 muscle or fin tissue samples (Table 1.2.4).

Table 1.2.4. Number of samples stored in AZTI laboratories as part of the ICCAT-GBYP tissue bank

Region	Area	Otoliths	Spines	Gonads	Muscle/fin	Total
MED	Western Mediterranean	8114	4140	1987	9190	**23431**
	Central Mediterranean	5600	1377	50	4765	**11792**
	Eastern Mediterranean	1342	860	7	1691	**3900**
GOM	Gulf of Mexico	0	0	0	207	**207**
SEATL	Gibraltar Strait	1010	482	0	1822	**3314**
	Madeira and Canary islads	129	0	0	380	**509**
NEATL	Bay of Biscay	393	516	2	1139	**2050**
NoS	Norwegian Sea and Skagerrak	207	604	0	1265	**2076**
CNATL	Central Atlantic (east of 45ºW)	485	2	0	2305	**2792**
	Central Atlantic (west of 45ºW)	69	0	0	687	**756**
NWATL	Gulf Saint Lawrence, Gulf of Maine, New Foundland, Western Atlantic	0	0	0	360	**360**
Total		**17349**	**7981**	**2046**	**23811**	**51187**

2. TOWARDS DETERMINING THE FREQUENCY OF THE DIFFERENT ATLANTIC BLUEFIN TUNA MITOCHONDRIAL HAPLOTYPES

Task Leader: Natalia Diaz-Arce (AZTI)

Participants:

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

The implementation of Close-Kin-Mark-Recapture (CKMR) models for stock abundance estimation requires the detection of kin pairs in the population. Three types of kin-pairs are included: full-siblings, half-siblings and parent-offspring pairs. For half-siblings and parent-offspring pairs, in addition to the number of pairs found, information about the sex of the shared parent is relevant for the model. Fish cells contain genomic information into the nuclear and the mitochondrial genomes. While the nuclear genome is inherited at equivalent proportions from both parents, the entire mitochondrial genome is transferred from the mother. Therefore, while nuclear genomic markers enable the identification of different kin pairs, the mitochondrial genome, if sufficiently variable within the population, can provide information about the sex of the shared parent in half-sibling pairs or the sex of the parent in parent-offspring pairs. Thus, if a pair of half siblings share the same mitochondrial haplotype, they likely share the same mother. Conversely, if their haplotypes differ, they must share the same father. However, if the mitochondrial genome variability is too low in the population and different variants are found at high proportions, two individuals could share the same mitochondrial genome without necessarily sharing the same mother, complicating these inferences. Here, we prepared three different datasets based on genomic sequences obtained from three different sources. These datasets will enable future analysis of mitochondrial variability in the ABFT Mediterranean population, which is the first step in determining the feasibility of cost-effective assays targeting suitable mitochondrial genomic regions to determine mitochondrial haplotypes.

2.1 Material and Methods

2.1.1 Pool-sequencing data

Pool whole genomic sequencing data consist of genomic sequences covering the whole genome of a pool of individuals. In these pools, groups of individuals captured in the same location and of the same age class were included in variable numbers (Table 2.1). This dataset allows for the identification of genomic variables and to estimate their frequencies within the pools.

To ensure high-quality sequencing data, the following steps were undertaken: *fasta* files (text-based formats for representing nucleotide sequences), including raw pool-sequencing data from sample pools indicated in Table 2.1, were filtered using the software trimmomatic (Bolger, Lohse et al. 2014). Reads with a length < 50 bp were discarded after trimming regions with an average sequencing quality below 25, screened in windows of 5 bp.

The resulting paired reads were mapped against the most updated version of the reference genome of the ABFT (Accession number GCA_963924715), which contains the mitochondrial genome, using bwa (Li 2013). Separated *bam* files containing only properly paired reads mapped against the mitochondrial genome were produced using *samtools* (Li, Handsaker et al. 2009).

Coverage depth, or the number of sequence reads supporting each specific base (nucleotide) of the mitochondrial genome was calculated at each position for each of the 13 pools of individuals using *bedtools* (Quinlan and Hall 2010).

2.1.2 Individual whole genome sequencing data

The *bam* files including whole genome sequencing data for 25 ABFT individuals, mapped to the reference genome of ABFT used in GBYP-13 for Task 3.1, were filtered. Separate files including only reads mapped to the mitochondrial genome were created. The average coverage depth at the mitochondrial genome per individual was calculated using *bedtool*s (Quinlan and Hall 2010).

2.1.3 Genbank downloaded sequences

The publicly available GeneBank database was set to filter to provide sequences of the species *Thunnus thynnus* and from the mitochondrion genetic compartment. The primary objective was to determine if the mitochondrial genome of the ABFT shows sufficient genetic diversity to identify maternal lineages and to design specific genetic assays for haplotype characterization. To explore genetic diversity across the entire mitochondrial genome, only long sequences with lengths between 2,000 and 17,000 base pairs were retained. The obtained sequences were mapped against the reference genome of the ABFT following the same procedure used for the pool-sequencing and whole genome sequencing data.

2.2 Results

2.2.1 Pool-sequencing data

Coverage depth (i.e. the number of sequences reads supporting each specific nucleotide) obtained from the pooled sequencing data was variable between pools. The average coverage depth at the mitochondrial genome per individual ranged from 2.8 to 66.5. These numbers divided by the number of individuals included in the pool indicate the average number of reads expected to be recovered from each individual. This means that each individual haplotype is expected to be sequenced at least 2.8 times (Table 2.1, Figure 2.1).

Table 2.1. Number of individuals included in each sequenced pool. Pools contained individuals from the same location and different size-classes. Average number of reads supporting each mitochondrial locus (Av. Depth Coverage) is indicated.

Region	Size-class	Number	Av. Depth Coverage
Western Mediterranean	V-Larvae	37	202.7
Western Mediterranean	0-YOY	50	476.6
Western Mediterranean	L-Large	50	683.4
Central Mediterranean	V-Larvae	50	142
Central Mediterranean	V-Larvae	50	352.2
Central Mediterranean	L-Large	50	601.2
Eastern Mediterranean	V-Larvae	10	303
Eastern Mediterranean	V-Larvae	50	467.2
Eastern Mediterranean	L-Large	44	710.9
Gulf of Mexico	V-Larvae	50	206.1
Gulf of Mexico	L-Large	50	1001.7
Slope Sea	V-Larvae	15	997.9
Slope Sea	V-Larvae	38	381.4

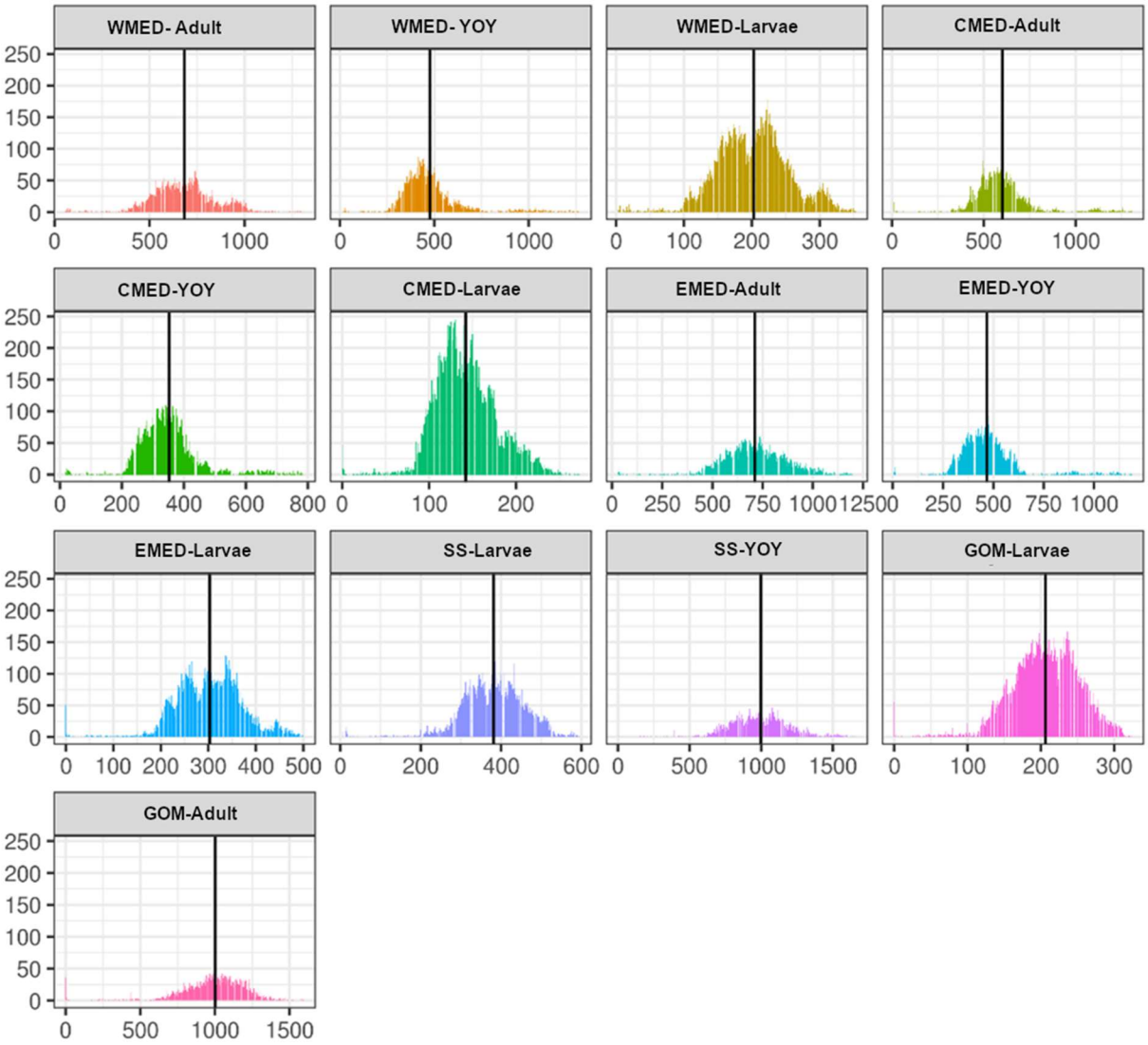

Figure 2.1. Histogram showing the number of positions (axis y) and coverage depth (axis x) for each of the 13 pools of individuals analyzed. Vertical lines show the average coverage depth of the mitochondrial genome for each pool.

2.2.2 Individual whole-genome sequencing data

Average coverage depth at the mitochondrial genome ranged from 233 to 1512, meaning that the recovery of whole individual mitochondrial haplotypes at high confidence is expected to be possible using this dataset (Table 2.2).

Table 2.2. Average coverage depth at each mitochondrial locus for each individual sequenced using whole genome sequencing, indicating its catch location and size-class (V=veliger/larvae, 0=Young-of-the-Year, L=Large).

Individual Nº	Size-class	Catch region	Average depth coverage
1	L-Large	CMED-Central Mediterranean	1220
2	L-Large	CMED-Central Mediterranean	733
3	0-YOY	CMED-Central Mediterranean	748
4	L-Large	EMED-East Mediterranean	1512
5	L-Large	GOM-Gulf of Mexico	752
6	L-Large	GOM-Gulf of Mexico	266
7	L-Large	GOM-Gulf of Mexico	606
8	L-Large	GOM-Gulf of Mexico	487
9	L-Large	GOM-Gulf of Mexico	474
10	L-Large	GOM-Gulf of Mexico	444
11	L-Large	GOM-Gulf of Mexico	799
12	L-Large	GOM-Gulf of Mexico	812
13	L-Large	GOM-Gulf of Mexico	779
14	0-YOY	SS-Slope Sea	1227
15	0-YOY	SS-Slope Sea	1032
16	0-YOY	SS-Slope Sea	529
17	0-YOY	SS-Slope Sea	784
18	0-YOY	SS-Slope Sea	1195
19	0-YOY	SS-Slope Sea	1069
20	V-Larvae	SS-Slope Sea	253
21	V-Larvae	SS-Slope Sea	564
22	V-Larvae	SS-Slope Sea	886
23	L-Large	WMED-West Mediterranean	434
24	V-Larvae	WMED-West Mediterranean	393
25	0-YOY	WMED-West Mediterranean	936

2.2.3 Data downloaded from GenBank

In total 10 complete *Thunnus thynnus* mitochondrial genomes sequences were obtained from the GenBank database (Table 2.3). All of them mapped against the mitochondrial genome included in the reference genome of the ABFT.

Table 2.3. Name, length and accession number list of publicly available downloaded sequences of the complete mitochondrial DNA of Atlantic bluefin tuna.

Name	Length	Accession Number
1. Thunnus thynnus mitochondrion, complete genome	16,529 bp	KF906720.1 GI:577027873
2. Thunnus thynnus mitochondrial DNA, complete genome	16,528 bp	AP006034.1 GI:1173213948
3. Thunnus thynnus mitochondrion, complete genome	16,527 bp	GU256522.1 GI:281428546
4. Thunnus thynnus mitochondrion, complete genome	16,527 bp	NC_014052.1 GI:295065594
5. Thunnus thynnus isolate DM353 mitochondrion, complete genome	16,527 bp	MT410869.1 GI:1898139670
6. Thunnus thynnus genome assembly, organelle: mitochondrion	16,527 bp	OZ004756.1 GI:2663408035
7. Thunnus thynnus thynnus mitochondrial DNA, complete genome	16,526 bp	AB097669.1 GI:32351881
8. Thunnus thynnus thynnus mitochondrion, complete genome	16,526 bp	AY302574.2 GI:33622385
9. Thunnus thynnus thynnus mitochondrion, complete genome	16,526 bp	NC_004901.2 GI:34577043
10. Thunnus thynnus voucher TT02-2312 mitochondrion, complete genome	16,526 bp	JN086149.1 GI:336327049

2.3 Conclusions

The individual and pooled whole-genome sequencing prepared datasets provide sufficient coverage which will serve to detect polymorphic positions in the bluefin tuna mitochondrial DNA, if existent, among hundreds of individuals. This is the first step in determining whether the Mediterranean ABFT mitochondrial variability allows for assessing if two half-siblings share the same mother. These data, together with longer publicly available sequences, should be sufficient to confidently recover entire haplotypes. This could enable the design of specific assays to detect these haplotypes in newly sampled individuals, provided there is enough variability.

3. SIBSHIP EVALUATION AMONG ATLANTIC BLUEFIN TUNA SAMPLED WITHIN THE BALEARIC SEA

Task Leader: Natalia Diaz-Arce

Participants:

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

The success in the implementation of CKMR models depends on appropriate sampling design. One of the main sample sources considered in the design of the CKMR model for the eastern Atlantic bluefin tuna are larval surveys yearly conducted in the Balearic Sea. One relevant feature of this source to be considered for the definition of the sampling strategy is the level of sibship within sampled stations. Here, we have genotyped 3,822 larvae samples obtained during Task 1.1 using the Atlantic bluefin tuna SNP array (developed in Phase 10), providing genotype information at > 6,000 neutral SNP markers that can be used to identify the presence of kin pairs among the analyzed samples. The stations and number of samples of each station were determined following the proposal from the team currently in charge of the development of the CKMR model for the eastern Atlantic bluefin tuna. This approach aims to determine the level of sibship within the selected stations and to evaluate the suitability of this sample source for the future implementation of the model. The genotyped individuals were analyzed together with previously genotyped larvae and young-of-the-year from the Mediterranean Sea captured between the years 2018 to 2023. Parent offspring, full sibling, and half sibling pairs within the analyzed samples were determined using genotype information.

3.1 Material and Methods

3.1.1 Obtention of genotype tables

The raw data contained in the CEL files produced by the genotyping company was downloaded and processed using the Axiom Analysis Suite software. Samples and SNPs with genotyping quality values below default threshold values and samples with genotyping rates <97% were excluded. Filtered genotype tables were exported into PLINK format. Replicate samples from adult individuals genotyped for tests on DNA quality from

different sampling sources were excluded from these analyses. The genotype table containing all individuals genotyped in this GBYP-Phase 14 was merged with the genotype table produced in GBYP-Phase13 (Task 3.2) which contained 490 and 134 additional larvae and YOY individuals from the Mediterranean Sea captured between 2018 and 2023. Only neutral markers genotyped in at least the 90% and showing a Minimum Allele Frequency (MAF) > 0.05 were kept. Individual heterozygosity was calculated to check for potentially contaminated samples.

3.1.2 Kinship analysis

The final dataset was first screened for the presence of replicates and kin-pairs using the R package CKMRsim (Anderson EC 2024). If replicates were found, one of the pairs was removed from the dataset before conducting pairwise comparisons for kin-pair detection.

CKMRsim allows for the comparison of all possible sample pairs and tests the hypothesis of sharing different types of relationships by estimating the log ratio between the probabilities of each type of relationship. Initially, potential Parent-Offspring (PO) pairs were identified by testing the probability of each pair being PO versus being unrelated (U). If potential PO pairs were found, they were then compared against the probability of being Full Siblings (FS) to confirm their PO status. After excluding identified PO pairs, the probability of each remaining pair being Half Siblings (HS) was compared against the probability of being Half Aunt Niece (HAN). At each step, the predicted distribution of log ratio values for each type of relationship was calculated. Threshold values to classify each type of relationship were chosen based on visual inspection of these distributions. The False Discovery Rate (FDR) for each selected threshold was estimated to ensure that FDR was much less than 0.1 times the number of paired comparisons, as recommended by the developers of CKMRsim.

3.2 Results

3.2.1 Genotype tables

The percentage of the newly genotyped samples failing the quality control filtering step ranged between 6.8-18.2 % for each of the 384 sample genotyped plates, meaning that failing samples were not agglutinated in the same plate. Samples that passed filtering criteria showed high genotyping or call rates (>99%) (Table 3.1).

Table 3.1. Number of samples included, those failing quality control (QC) filters and the number and percentage of samples kept in the final dataset from each 384-sample genotyped plate. Also, Average call rate for passing samples is indicated.

Plate Number	Samples included	Samples failing QC	Samples that passed	% of passing samples	Av. call rate for passing samples
1	384	69	315	82.0	99.088
2	384	45	339	88.3	99.705
3	384	62	322	83.9	99.664
4	384	36	348	90.6	99.713
5	384	60	324	84.4	99.641
6	384	70	314	81.8	98.907
7	384	43	341	88.8	99.467
8	384	26	358	93.2	99.629
9	384	28	356	92.7	99.644
10	384	34	350	91.1	99.663

The final dataset included larval and YOY Mediterranean samples genotyped in the current and previous phases contained 3,978 individuals and 6,278 SNPs. Three sample pairs were found to be replicates, showing genotype mismatches at 2, 31 and 85 SNPs respectively. All pairs of replicates involved larvae from the same original tube and processed in the same plate and were assumed to be fragments of the same larvae placed in different wells of the same plate. After removing one of each pair from the final dataset, a total of 3,975 samples were retained (Table 3.2). No sample showed increased heterozygosity values, indicating either the absence of cross-contamination or that cross-contaminated samples were filtered out during the quality control step.

*Table 3.2. Summary table showing the number of 3,978 samples of each age class and catch year included in the final genotype table. * Indicates category from which 3 samples were removed because they were identified as replicates.*

	2018	2020	2021	2022	2023
YOY	40	16	5	65	8
Larvae	0	368	0	3427*	49

3.2.2 Kinship among larvae and Young of the Year captured in the Mediterranean Sea from 2018 to 2023

No PO pairs were found, which could be clearly distinguished from FS (False Positive Rate=1.18e-223 and False Negative Rate=0). This indicates a very high level of accuracy in distinguishing PO pairs from FS pairs. In total, 229 FS and 1,019 HS pairs were identified, which could be clearly distinguished from HS (False Positive Rate =9.33e-223 and False Negative Rate=0, see Figure 3.1) and HAN (False Positive Rate =1.08e-42 and False Negative Rate =0, see Figure 3.2) respectively.

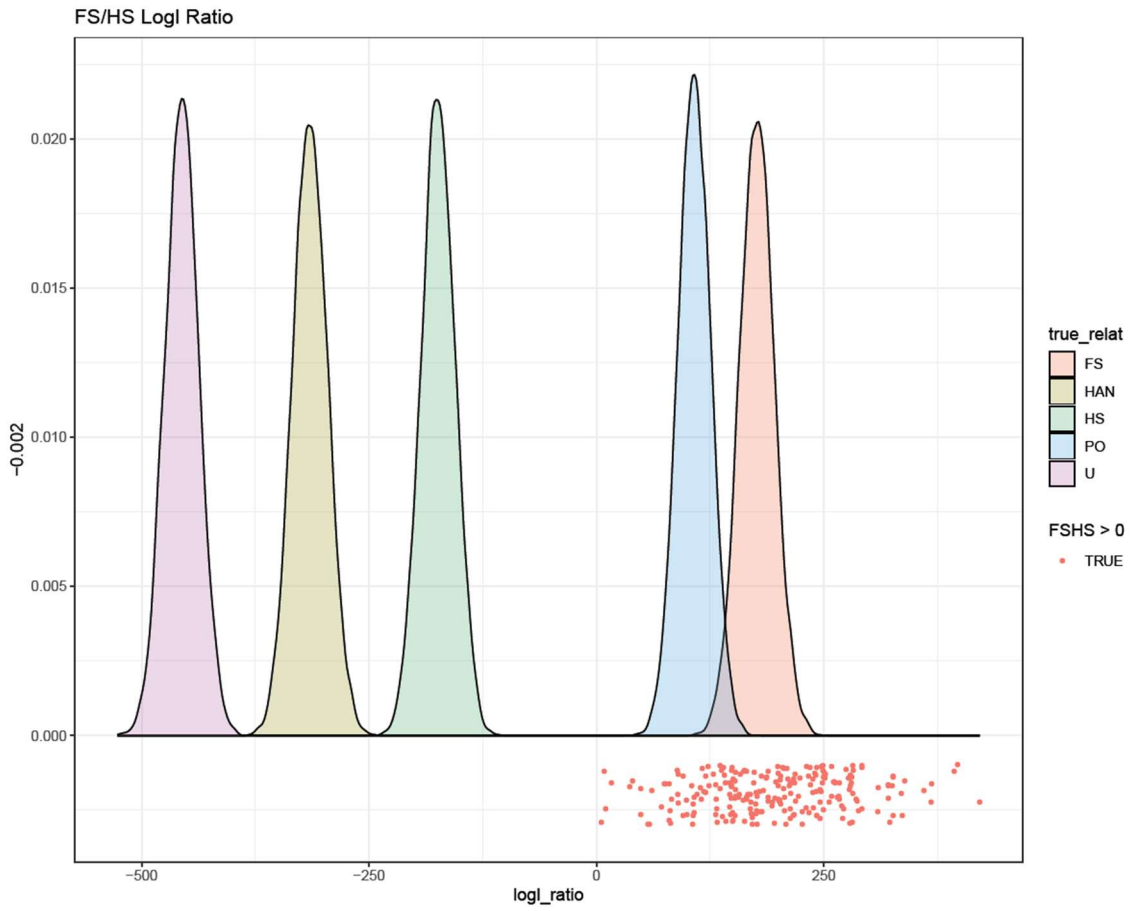

Figure 3.1. Distribution of the estimated logl ratios between FS and HS for each of the following pair types: parent-offspring (PO), half sibling (HS), full sibling (FS), half-aunt-niece (HAN) and unrelated (U). In the x axis the red dots show the logl ratio values for the 229 identified FS pairs from the analyzed dataset. The graph shown no overlap between the FS and HS distributions, meaning that these two types of kinship can be clearly distinguished using these values setting 0 as a threshold.

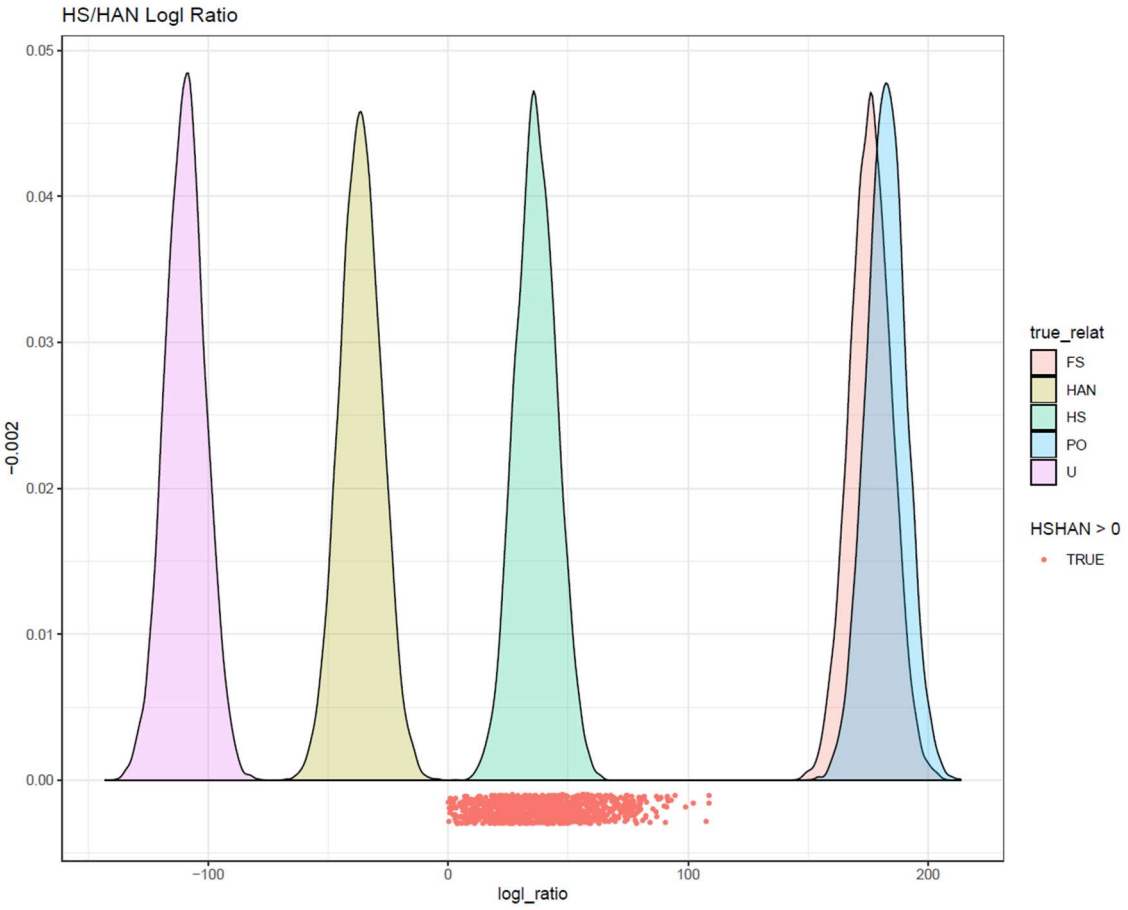

Figure 3.2. Distribution of the estimated logl ratios between HS and HAN for each of the following pair types: parent-offspring (PO), half sibling (HS), full sibling (FS), half-aunt-niece (HAN) and unrelated (U). In the x axis the red dots show the logl ratio values for the 1019 identified HS pairs from the analyzed dataset. The graph shown no overlap between the HS and HAN distributions, meaning that these two types of kinship can be clearly distinguished using these values setting 0 as a threshold.

Among the 229 FS pairs, 16 and 69 samples were involved in three and two FS pairs respectively. Similarly, among the 1019 HS relative pairs, 13 individuals were involved in 5 HS pairs, 51 were involved in 4 HS pairs, 130 were involved in 3 HS pairs and 299 were involved in 2 HS pairs. This means that the number of parental individuals shared among the analyzed samples is lower than the number of found sibling pairs.

All kin-pairs between samples analyzed in previous phases were consistent with those pairs detected in the current phase. All FS pairs involved larvae captured in the same year: 228 pairs were between larvae captured in 2022, and 1 pair was between larvae

captured in 2020. However, among the 1,019 detected HS pairs, 10 involved samples collected in both 2020 and 2022. Additionally, a few FS and HS pairs were detected involving samples from different stations collected in the same year (Table 3.3).

Table 3.3. Number of full sibling (FS) and half sibling (HS) identified between larvae stations captured in years 2020 and 2022. Lines in bold indicate pairs detected within the same station/year.

YearA	StationA	YearB	StationB	FS	HS
2020	**794**	**2020**	**794**	**1**	**4**
2020	**882**	**2020**	**1331**	**0**	**1**
2022	**792**	**2022**	**792**	**0**	**2**
2022	**1400**	**2022**	**1400**	**2**	**2**
2022	**1402**	**2022**	**1402**	**205**	**891**
2022	**1587**	**2022**	**1587**	**14**	**81**
2022	**1593**	**2022**	**1593**	**3**	**0**
2022	1402	2022	1400	4	16
2022	1402	2022	1127	0	1
2022	1402	2022	1222	0	1
2022	1402	2022	1587	0	8
2022	1587	2022	1400	0	1
2022	1587	2022	1593	0	1
2020	794	2022	1402	0	5
2020	794	2022	1587	0	2
2020	794	2022	1593	0	1
2020	882	2022	1593	0	1
2020	1331	2022	1402	0	1

In particular, the stations selected by the developers of the CKMR model for eastern ABFT showed high proportions of within-station sib-ship (Table 3.4). Expectedly, the number of detected kin-pairs increased with the number of analyzed individuals (Figure 3.3).

Table 3.4. Number of within-station FS and HS pairs found among larvae from selected stations from TUNIBAL survey from 2022. The number of samples proposed to be included in the analysis and the number of samples included in the analysis (after genotype quality filtering) is indicated for each station.

Station	Samples proposed	Samples analyzed	FS	HS
1402	2247	1958	205	891
1587	1168	1037	14	81
808	10	10	0	0
790	5	5	0	0
1127	6	6	0	0
1400	57	54	2	2
1331	6	17	0	0
792	0	56	0	2
1593	156	168	3	0
893	88	81	0	0
1137	0	0	0	0
1222	19	12	0	0
719	21	21	0	0
701	2	2	0	0

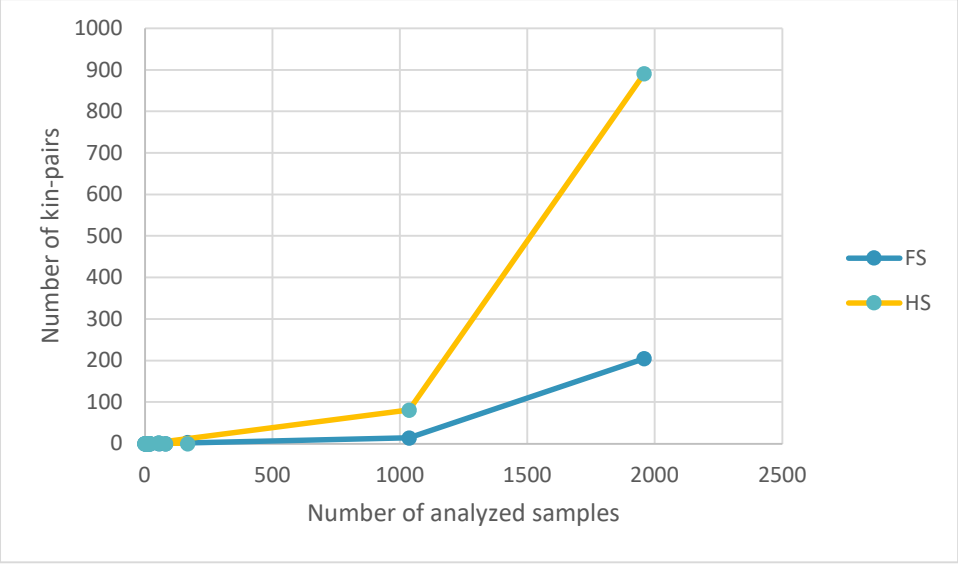

Figure 3.3. Number of within-station FS and HS pairs found among larvae from selected stations from TUNIBAL survey from 2022. The x axis indicates the number of samples analyzed from each station.

3.3 Conclusions

The ABFT array developed during GBYP phase 10 allows for accurate detection of full sibling and half sibling pairs among Mediterranean larvae.

Some sampled stations from the Balearic Sea show high sibship, which could be explained by biased sampling effort towards high larval density stations. Further analyses are needed to assess their suitability for being including within ABFT CKMR dedicated sampling program.

The detection of full and half-sibling pairs among larvae collected at different stations confirms that a single adult individual can spawn across multiple stations within the same spawning season. This information should be considered when designing the larval sampling strategy for the implementation of the eastern ABFT CKMR model.

The detection of half sibling pairs involving larvae captured in different years in the Balearic Sea confirms spawning fidelity of the same adult individuals to the same area.

Bibliography

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(16): 2078-2079.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics 26(6): 841-842.

4. ANNEX

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