SHORT TERM CONTRACT FOR THE BIOLOGICAL AND GENETIC SAMPLING AND ANALYSIS (ICCAT-GBYP 06/2011) WITHIN THE GBYP (Phase 2)

Final Report

for:

ICCAT

Scientific coordinator:

Dr. Haritz Arrizabalaga (AZTI-Tecnalia)

Pasaia, July 9th, 2012

PARTNERS:

Fundación AZTI-AZTI Fundazioa

Universidad de Cádiz

Instituto Español de Oceanografía

IFREMER

Universitá di Genova

University of Bologna

Necton

University of Cagliari

Euskal Herriko Unibertsitatea / Universidad del País Vasco

National Research Institute of Far Seas Fisheries

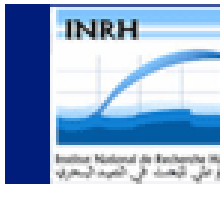
Hellenic Center for Marine Research

FEDERATION OF MALTESE AQUACULTURE PRODUCERS

Federation of Maltese Aquaculture Producers

Texas A&M University

SUBCONTRACTORS:



INRH

IPIMAR

BIOGENOMICS

ÍNDEX:

EXECUTIVE SUMMARY:

The main objective of this project is to enhance knowledge about Atlantic bluefin tuna population structure and mixing, but also focusses on age and reproductive dynamics. The sampling protocols and structure of the data bank were revised and agreed with ICCAT Secretariat.

At the end of the project, the consortium has sampled a total of 1916 bluefin tuna (10 larvae, 239 YOY, 446 juveniles, 552 medium size fish, and 669 large fish) from different regions (188 from the East Mediterranean, 270 from the Central Mediterranean, 732 from the Western Mediterranean, 597 from the Northeast Atlantic and 129 from the Central North Atlantic). From these individuals, 4309 biological samples were taken (1632 genetic samples, 1324 otoliths, 1078 spines and 275 gonads).

Following up on an earlier project, a panel of 384 SNPs was selected and 919 bluefin tuna were genotyped with Next Generation Sequencing – Transcriptome Sequencing (NGS-TS). In parallel, 192 bluefin tuna were genotyped with Next Generation Sequencing – Reduced Representation Sequencing and Genotyping (NGS-RRSG). Only the use of restricted panels of outlier and highly divergent SNP loci permitted to discriminate spawning population samples and assign individuals of feeding aggregates to originating populations. The NGS-TS approach did not clearly resolve genetic relationships among spawning populations and/or feeding aggregates. On the contrary, the NGS-RRSG was more promising than NGS-TS because it permitted to unequivocally discriminate reference samples from the Gulf of Mexico and Western Mediterranean. However, since in this approach the feeding aggregations clustered as a third different group, the dataset and analyses will need to be improved by adding more reference spawning samples and by selecting the most performing SNP loci, respectively. These improvements will enhance the NGS-RRSG power for resolving population structure of the spawning ABFT populations and assignment of individuals composing the feeding aggregates.

Regarding microchemistry, 600 otoliths from the Central North Atlantic, Bay of Biscay, Strait of Gibraltar, Balearics, Malta, Sardinia and the Adriatic Sea have been analyzed, results showing >99% of Eastern origin fish except in the Central North Atlantic were 84% of the fish were of Eastern origin.

Regarding age determination analyses, 374 otoliths and 375 spines have been interpreted already. Inter-reader agreement was high and age-length-keys were generated for both spines and otoliths. The comparison between ages estimated from different structures of the same specimen showed a good age agreement. This indicates that both structures may be used indistinctly for age determination of Atlantic bluefin tuna for the age ranges analyzed in this project.

A histological analysis was conducted on 188 individuals from the Strait of Gibraltar, Balearics, Malta, Sardinia and Ligurian Sea. However, the sampling was not adapted to the histological analysis (mostly occurring before and after the reproductive season), thus they contributed no essential information to the understanding of the Atlantic bluefin tuna reproductive biology.

In general, most of the objectives of the project were achieved and the analyses already started to provide some results on population structure, catch composition, age structure and reproductive ecology.

1. CONTEXT

On May 13th 2011, the consortium formed by Fundación AZTI-AZTI Fundazioa, Universidad de Cádiz, Instituto Español de Oceanografía, IFREMER, Universitá di Genova, University of Bologna, Necton, University of Cagliari, Euskal Herriko Unibertsitatea / Universidad del País Vasco, National Research Institute of Far Seas Fisheries, Hellenic Center for Marine Research, Federation of Maltese Aquaculture Producers and Texas A&M University, coordinated by Fundación AZTI-AZTI Fundazioa, presented a proposal to the call for tenders on biological and genetic sampling and analysis (ICCAT-GBYP 06/2011). This proposal was awarded by the Secretariat on May 27 2011. The final contract between ICCAT and the consortium represented by Fundación AZTI-AZTI Fundazioa was signed on July 18th 2011.

According to the terms of the contract, a short preliminary interim report was submitted to ICCAT by July 22nd 2011. The second deliverable was submitted by 21st of September, and included an update of the short preliminary interim report, which constitutes the Interim Report, as well as a powerpoint presentation, containing a summary description of the activities carried out till date.

A preliminary final report was submitted by 21st of November 2011. Following the terms of the contract, this report included a detailed report of all the activities, the complete list of the samples collected by area, the list of samples already analyzed, the list of samples to be stored and analyzed in the future, and a detailed scientific report of the results obtained. The difficulties encountered, as well as any other relevant information were also reported. A powerpoint presentation, the datasets and some photos were also provided. Subsequently, a final report was submitted by 30th of November, incorporating all comments by ICCAT, and reflecting the activities carried out since the notification of the proposal being awarded (27th of May 2011). ICCAT made some comments the 20th of December 2011 and these were incorporated in a revised final report submitted the 23th of December 2011.

In addition, the consortium requested a prorogation and an extension of the activities to be carried out. This request was accepted by ICCAT and, according to

the amended contract, a complete final report needs to be prepared at the end of the prorogation period, with details of all the activities, including those agreed for the extended contract. The present report was prepared in response to such requirement, and represents the final report.

2. SAMPLING

The sampling conducted under this project follows a specific design, aimed primarily at contributing to knowledge on population structure and mixing. 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).

2.1 Sampling protocols

The consortium thoroughly revised the preliminary sampling protocols that were attached to the proposal presented to ICCAT. Besides small modifications on the technical side, special efforts were made in order to standardize common parts of the 4 protocols (e.g. format, as well as sections "filling the data form" and "shipping procedures"). The final adopted set of protocols was distributed to all members of the consortium and is included as Appendix 1.

2.2 Structure of the data bank

The structure of the data bank was revised by the consortium. Moreover, a short meeting in Madrid with staff from ICCAT Secretariat was held to discuss the structure of the data bank. In this meeting, it was concluded that the structure of the data bank was correct for the purposes of the project and both the nature and dimensions of the database. On top of this, ICCAT Secretariat suggested that if sampling continued in subsequent years, the consortium might want to consider recording additional information on the number of fish caught, harvested or available for sampling each time a sample is taken. This would allow to know precisely how many fish are represented in a sample, while the actual scheme assumes that all fish sampled in a given strata would be representative of the catch in that strata. In the meeting, it was also identified that in order to include this extra information, the structure of the database would need to change substantially.

The consortium discussed the suggestion made by the Secretariat and agreed that this could be considered in subsequent years if it was proven that the extra bit of information was really going to be useful when interpreting the results of the project, which was unclear. In the meantime, the consortium decided to continue with the current structure during 2011 because several partners had already started sampling.

2.3 Sampling acomplished

A total of 4309 samples from 1916 bluefin tuna have been taken so far. Table 2.1 shows the number of bluefin tuna sampled in each strata (area/fishery/size class combination) and Table 2.2 and Figure 2.1 provide summaries by main region and size class.

The original plan was to sample 1950 individuals, thus the current sampling status represents 98% of the target in terms of number of individuals. By size class, the objectives for age 0, juveniles and large fish were accomplished (96%, 112% and 122% respectively) and the sampling for medium fish and larvae was below expectations (79% and 20%, respectively).

Table 2.1. Number of bluefin tuna sampled by area/fishery and size class. Empty cells indicate that no sampling was planned in that stratum. Green cells indicate strata were no sampling was planned but some sampling was finally accomplished.

		Larvae	Age 0	Juvenile	Medium	Large			
			<3 kg	3-25 Kg	25-100 Kg	>100 kg	Responsible	Target	%
Eastern Mediterranean	Northern Levantine Sea (juvenile-medium- large): Turkish PS			0	61	87	AZTI	150	99%
	Levantine Sea (larvae)	10	29				AZTI	100	39%
	Crete (medium-large fish): Greek LL			1	0	0	HCMR	100	1%
Central Mediterranean	Malta (medium-large): Maltese LL				22	79	FMAP	100	101%
	South of Sicily and Ionian Sea (medium-large): Italian PS and LL		21	48	50	0	NECTON	100	119%
	Adriatic Sea (small): Croatian and Italian PS		0	50			UNIBO	100	50%
Western Mediterranean	Balearic (medium-large): French and Spanish PS		121	43	34	0	IEO	100	198%
	South Tyrrhenian (medium-large): Italian PS		18		45	0	NECTON	100	63%
	Sardinia (medium-large): Italian Trap			14	142	66	UNICA	150	148%
	Gulf of Lyon, French artisanal/sport fisheries			51	53		IFREMER	100	104%
	Ligurian Sea, Italian artisanal/sport Fisheries			72	23		UNIGE	50	190%
	Tyrrhenian (small): Italian handline		50				NECTON	50	100%
	Southern Spain (juveniles & medium size): Spanish LL		0	0	0		IEO	150	0%
Northeast Atlantic	Gibraltar (small, medium-large): Spanish HL, traps, BB		0	0	18	94	IEO	200	56%
	Gibraltar: Portuguese traps				8	104	IPIMAR	100	112%
	Bay of Biscay (small): Spanish BB & French TW			167	79	41	AZTI	100	287%
	Western coast of Africa (medium-large): Morrocan Trap				0	86	INRH	100	86%
Central North Atlantic	Central and North (medium-large): Japanese & Taiwanese LL				17	112	NRIFSF	100	129%

Table 2.2: Number of bluefin tuna sampled by main region and size class. Empty cells indicate that no sampling was planned in that strata:

	Larvae	Age 0	Juvenile	Medium	Large	TOTAL	Target	%wrt target
East Med	10	29	1	61	87	188	350	54%
Central Med		21	98	72	79	270	300	90%
West Med		189	180	297	66	732	700	105%
NE Atl		0	167	105	325	597	500	119%
Central N Atl				17	112	129	100	129%
TOTAL	10	239	446	552	669	**1916**	1950	98%
Target	50	250	400	700	550	1950		
% wrt target	20%	96%	112%	79%	122%	98%		

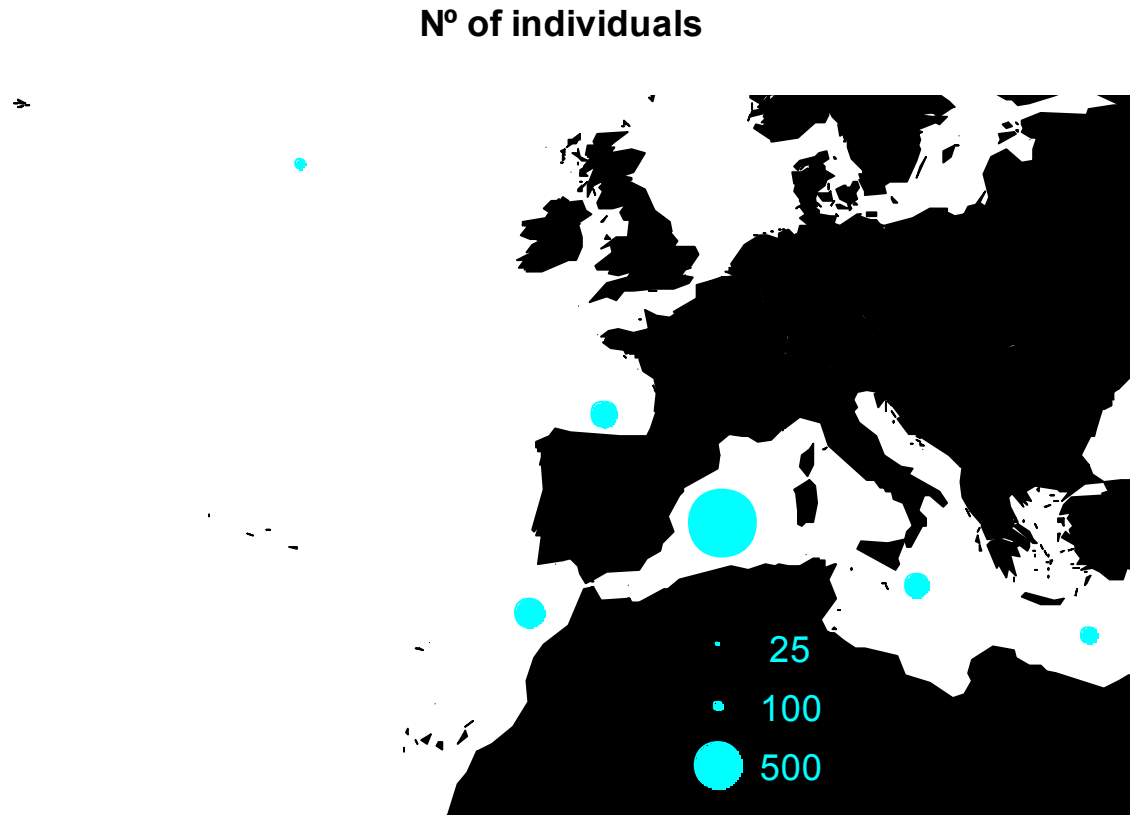


Figure 2.1: Number of individuals sampled, aggregated by main region. Positions of the dots are approximate averages across all samples. In the case of the North East Atlantic region, two dots are presented, one in the Atlantic side of the Strait of Gibraltar and the other in the Bay of Biscay.

In the Eastern Mediterranean, 54% of the target number of individuals has been sampled. Some reference samples (larvae and young of the year, YOY) as well as medium and large fish have been collected, but only one juvenile fish was collected. Although a significant number of larvae were caught, only 10 of them were confirmed to be bluefin tuna after conducting genetic tests. Some juveniles in the Levantine sea were expected to be caught in the albacore fishery, but this did not happen (partly due to low number of boats targeting albacore). In Greece the fishery was closed before the contract was signed, with little possibilities to accomplish the sampling.

In the Central Mediterranean, 90% of the target number of individuals was sampled. The main difficulty arouse from the fact that Italian fisheries were closed early (with respect to the date of contract signature) and samples in the South of Sicily and Ionian Sea could not be obtained. This also affected the sampling in Malta, since 80% of the quota was taken before the contract was signed, and most of the remaining fish were sampled. In addition, dedicated surveys targeting YOY bluefin in the Adriatic were not successful either. The consortium tried to accommodate this by sampling YOY in other areas of the Central Mediterranean. In some cases we had difficulties in getting permits to catch YOY for scientific purposes (e.g. in Malta). However, dedicated efforts by University of Cagliari allowed to sample 21 YOY in Sicily, partially compensating for the low sampling in this strata.

In the Western Mediterranean, 105% of the target number of individuals was sampled. The early closure (with respect to the date when the contract was signed) of the Italian, Spanish and French purse seine fisheries affected the sampling in this area importantly. Also, few longliners operated in southern Spain during the spawning season since they mostly sold their quota to purse seiners or traps, but some individuals were finally sampled from the bycatch in other longline fisheries around the Balearics. But overall, the sampling of the different size classes in the western Mediterranean is still significant thanks to the sampling accomplished in cages and other fisheries (traps, artisanal/sport fisheries) where special efforts allowed to sample beyond the initial target (i.e. in Sardinian and the Ligurian Sea by the University of Cagliari and Universitá di Genova, respectively). Also, special efforts were made to cover strata not initially planned (medium size fish in the Ligurian Sea by Universitá di Genova, YOY in the Balearics by IEO and AZTI and juveniles from the Balearics made available by the GBYP tagging program, also sampled by AZTI).

In the North East Atlantic, 119% of the target number of individuals was sampled. Some few strata could not be properly sampled, i.e. YOY and juveniles in Gibraltar and medium size fish

in Moroccan and Portuguese traps (due to the large size of the fish caught by traps in this area), but overall, juveniles and adults are well sampled, due to additional efforts conducted in baitboats in the Bay of Biscay and Portuguese traps (by IPIMAR and AZTI, respectively). Also, some fish in strata not initially planned were sampled (i.e. large fish sampled by AZTI in the Bay of Biscay baitboat fishery).

In the Central North Atlantic, the fishing season for Japanese longline vessels started in September. Scientific observers are expected to get otoliths and muscle from 80 fish in the area East of 45ºW during the 2011 fishing season. These samples will be provided in 2012, because it takes more than a year to obtain the samples from the observers when the vessels come back to Japan. However, Japanese scientists have provided otoliths collected from 129 bluefin caught in earlier fishing seasons (4 in 2009, 119 in 2010 and 6 in 2011), which were unscheduled samples for this project. These samples correspond to both the medium (n=17) and the large (n=102) size categories, and are caught mostly in the eastern Atlantic (east of 45ºW, n=121) but some are also caught in the western Atlantic (west of 45ºW, n=8).

Table 2.3 shows the number of tissues sampled in each strata (area/fishery combination) and Table 2.4 and Figures 2.2, 2.3, 2.4 and 2.5 provide summaries by main region and tissue type. In many cases, not all tissues (otoliths, muscle or fin for genetics, spine, and/or gonad, according to the sampling scheme) were collected from each single fish. However, both the total amount of samples (4309) as well as the number of samples by tissue type (1632 genetic samples, 1324 otoliths, 1078 spines and 275 gonads) is high and relatively well distributed over the different main regions (considering the circumstances explained in earlier paragraphs). It is expected that once 2011 samples from the Central North Atlantic arrive (during 2012), the distribution of samples by main region will improve considerably.

Table 2.3: Number of samples collected by area/fishery and tissue type:

		Otolith	Spine	Gonad	Muscle/Fin	Sampler
Eastern Mediterranean	Northern Levantine Sea (juvenile-medium- large): Turkish PS	100	105		105	AZTI
	Levantine Sea (larvae)	29	29		39	AZTI
	Crete (medium-large fish): Greek LL	1	1		1	AZTI
Central Mediterranean	Malta (medium-large): Maltese LL	85		12	100	FMAP
	South of Sicily and Ionian Sea (medium-large): Italian PS and LL	69	98		119	NECTON/UNIBO/UNICA
	Adriatic Sea (small): Croatian and Italian PS	50	50		50	UNIBO
Western Mediterranean	Balearic (medium-large): French and Spanish PS	195	152	43	197	IEO/AZTI
	South Tyrrhenian (medium-large): Italian PS		63		61	NECTON
	Sardinia (medium-large): Italian Trap	47	109	51	213	UNICA
	Gulf of Lyon, French artisanal/sport fisheries	100	98		104	IFREMER
	Ligurian Sea, Italian artisanal/sport fisheries	90	94	90	95	UNIGE
	Tyrrhenian (small): Italian handline	50	50		50	UNIB O
	Southern Spain (juveniles & medium size): Spanish LL	0	0	0	0	IEO
Northeast Atlantic	Gibraltar (small, medium-large): Spanish HL, traps, BB	98	78	81	56	IEO
	Gibraltar: Portuguese traps	104	34		106	IPIMAR
	Bay of Biscay (small): Spanish BB & French TW	141	117		286	AZTI
	Western coast of Africa (medium-large): Morrocan Trap	36			50	INRH
Central North Atlantic	Central and North (medium-large): Japanese & Taiwanese LL	129				NRIFSF

Table 2.4: Number of samples by main region and tissue type:

	Otolith	Spine	Gonad	Muscle/Fin	TOTAL
East Med	130	135		145	410
Central Med	204	148	10	269	631
West Med	482	566	184	720	1952
NE Atl	379	229	81	498	1187
Central N Atl	129				129
TOTAL	1324	1078	275	1632	4309

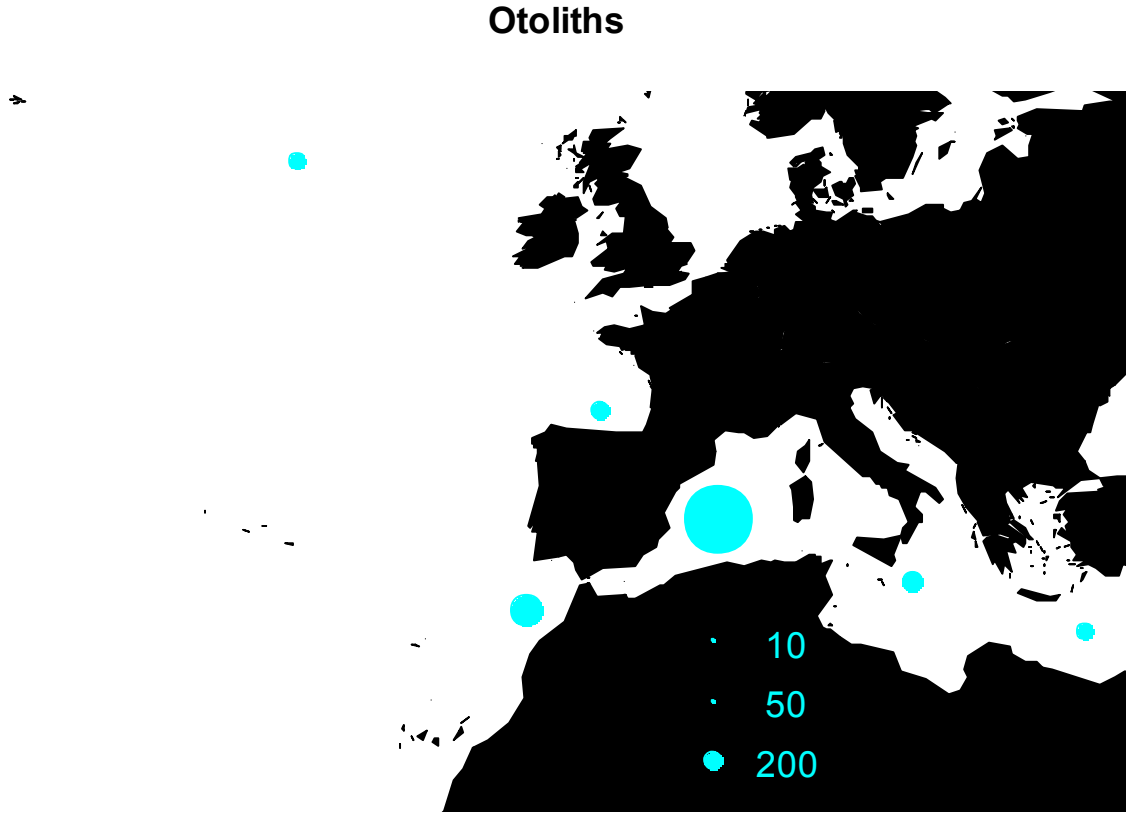


Figure 2.2: Number of individuals with otolith sampling, aggregated by main region. Positions of the dots are approximate averages across all samples. In the case of the North East Atlantic region, two dots are presented, one in the Atlantic side of the Strait of Gibraltar and the other in the Bay of Biscay.

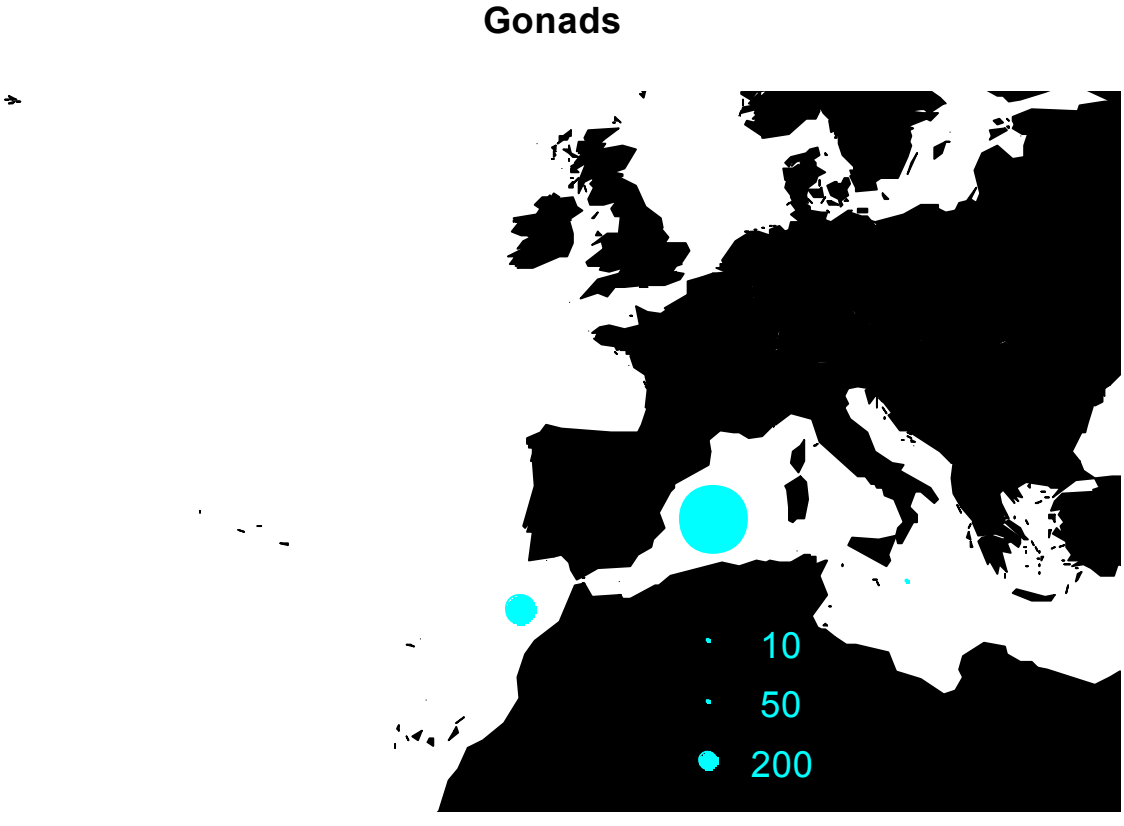


Figure 2.3: Number of gonads collected, aggregated by main region. Positions of the dots are approximate averages across all samples. In the case of the North East Atlantic region, two dots are presented, one in the Atlantic side of the Strait of Gibraltar and the other in the Bay of Biscay.

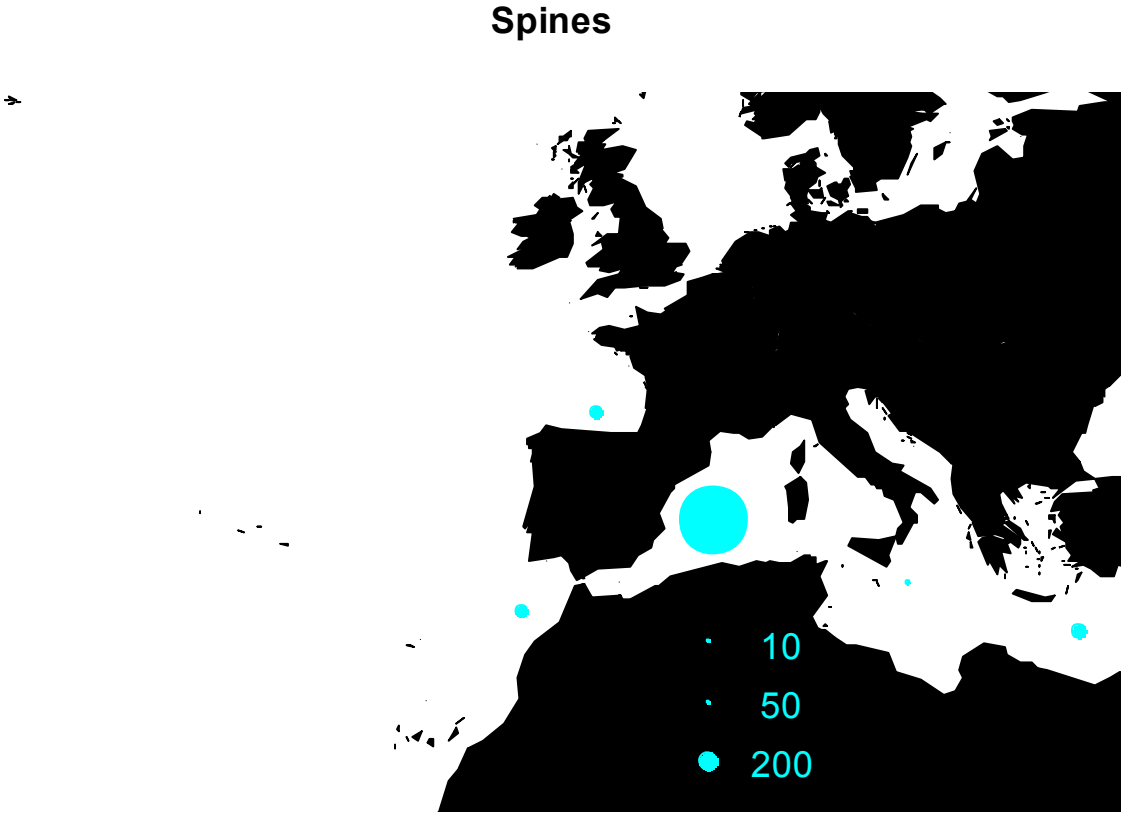


Figure 2.4: Number of spines collected, aggregated by main region. Positions of the dots are approximate averages across all samples. In the case of the North East Atlantic region, two dots are presented, one in the Atlantic side of the Strait of Gibraltar and the other in the Bay of Biscay.

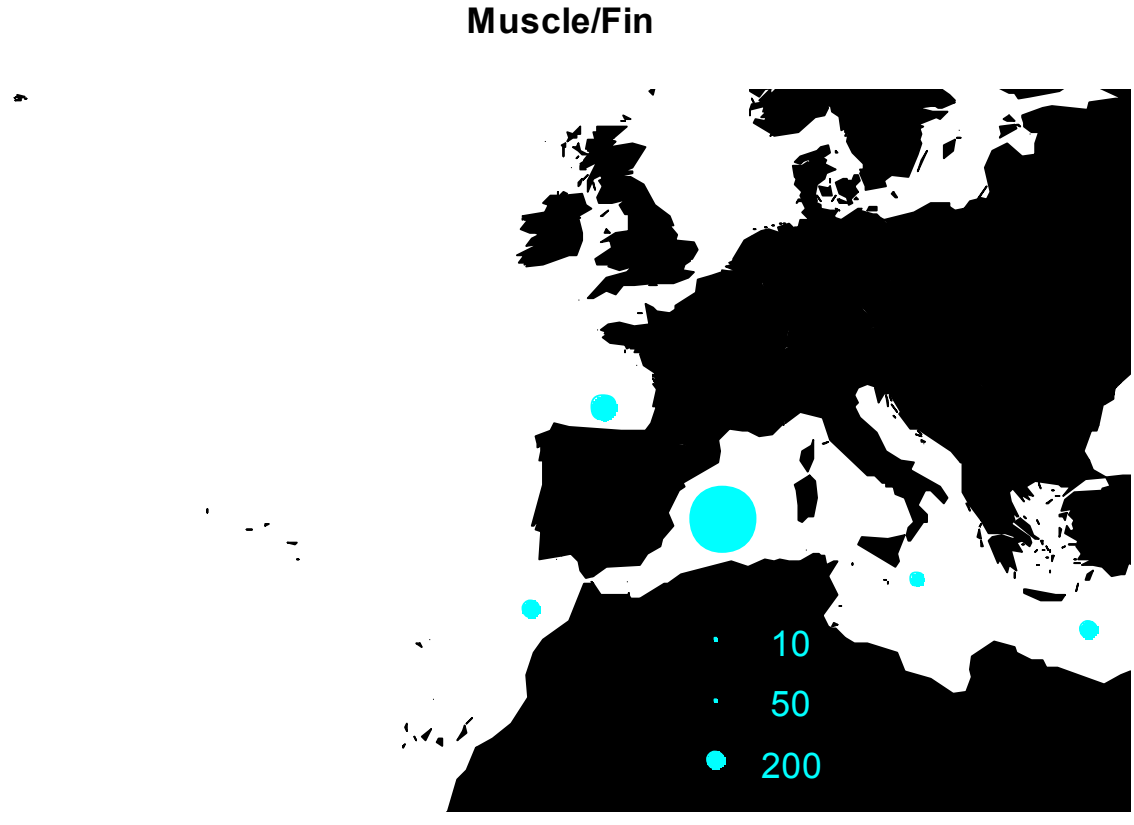


Figure 2.5: Number of muscle or fin tissue samples collected, aggregated by main region. Positions of the dots are approximate averages across all samples. In the case of the North East Atlantic region, two dots are presented, one in the Atlantic side of the Strait of Gibraltar and the other in the Bay of Biscay.

Most of these samples have been sent to AZTI, following the protocols (although some samples were directly sent to the analyst due to time constraints). This step allows for quality control of the samples and the coding, as well as fulfilling the requirement of having a centralized collection of samples for future use. The samples are conserved following the protocols and stored in the central facilities of AZTI-Tecnalia in Pasaia (contact persons: Igaratza Fraile and

Nicolas Goñi). The samples already distributed to other labs (for analyses under different tasks) are tagged in the database.

In the offer, the consortium agreed to spread the target of 50 individuals sampled in each strata in at least 5 different days, with no more than 10 individuals sampled per fishing day. Table 2.5 reveals some strata where less than 5 fishing days were sampled. Generally, these strata correspond to those where the target of n=50 individuals was not reached. However, sometimes this is not the case, and more than 50 individuals have been sampled in less than 5 days (e.g. Medium and Large Fish in the Levantine Sea, or Medium fish in the South of Sicily and Ionian Sea). For several strata, the percent of fishing days with more than 10 fish sampled exceeds 25%. In some cases, this is due to the fact that only one or two fishing days could be sampled (e.g. in South of Sicily and Ionian Sea, the Levantine Sea or the Thyrrenian). In other cases, however, in spite of having some days with more than 10 individuals sampled, there are more than 5 days sampled and the average number of individuals per day is less than 10, which allows to select and get a more representative sample for the strata (e.g. Medium fish in the Bay of Biscay). Finally, it should be stressed that some of the strata in Table 2.5 were not included in the original sampling design (e.g. large fish in the Bay of Biscay). As such, although they might not fulfill all the requirements from the point of view of an ideal sampling design, they could be considered as added value to the project. The sampling database just allows the analyst to select the samples to be analyzed based on these and other criteria.

Table 2.5: Summary of sampling effort. For each area and size class, the table indicates the number of individuals sampled (cells bellow n=50 in red), how many fishing days are represented in the sample (if less than 5, in red), number of days with more than 10 fish sampled, the percent of days with more than 10 fish per day (in red if >25%), and the mean (minimum-maximum) number of samples taken per fishing day (in red if >10).

Grand Area	Area	Size	N inds	N days	N days n>10	%	Mean (min-max) N per day
Cmed	Malta	Large	79	24	0	0%	3 (1-10)
		Medium	22	11	0	0%	2 (1-4)
	Sicily-Ionian	YOY	21	1	1	100%	21 (21-21)
		Juvenile	48	1	1	100%	48 (48-48)
		Medium	50	2	2	100%	25 (24-26)
	Adriatic	Juvenile	50	5	1	20%	10 (7-13)
CNAtl	Cental Atl	Large	112	33	0	0%	3 (1-8)
		Medium	17	15	0	0%	1 (1-2)
Emed	Aegean	Juvenile	1	1	0	0%	1 (1-1)
	Levantine	YOY	29	11	0	0%	3 (1-7)
		Large	87	1	1	100%	87 (87-87)
		Medium	61	1	1	100%	61 (61-61)
		Larvae	10	3	0	0%	3 (1-7)
NEAtl	Gibraltar	Large	94	6	3	50%	16 (2-47)
		Medium	18	5	0	0%	4 (1-8)
	Portugal	Large	104	33	2	6%	3 (1-15)
		Medium	8	8	0	0%	1 (1-1)
	Bay of Biscay	Juvenile	167	17	4	24%	10 (1-34)
		Large	41	3	2	67%	14 (10-19)
		Medium	79	10	3	30%	8 (1-26)
	Morocco	Large	86	8	3	38%	11 (6-15)
Wmed	Balearics	YOY	121	9	3	33%	13 (1-44)
		Juvenile	43	7	1	14%	6 (1-34)
		Medium	34	12	0	0%	3 (1-7)
	Thyrrenian	YOY	68	2	2	100%	34 (18-50)
		Medium	45	3	2	67%	15 (8-22)
	Sardinia	Juvenile	14	4	0	0%	4 (1-6)
		Large	66	4	2	50%	17 (1-36)
		Medium	142	8	6	75%	18 (4-35)
	Gulf of Lion	Juvenile	51	25	0	0%	2 (1-9)
		Medium	53	24	0	0%	2 (1-7)
	Ligurian	Juvenile	72	16	0	0%	5 (1-10)
		Medium	23	11	0	0%	2 (1-4)

3. ANALYSES

In the proposal, considering the extension, the consortium proposed to analyze a subset of 600 otoliths (for microchemistry), 405 spines and 405 otoliths for aging, 80 gonads, and 950 muscle/fin samples for genetic analyses (plus 10 positive controls). The number of samples obtained is above those numbers. However, the number of samples described in the previous section reflects all the samples that are collected at the time of writing this report. As reflected in the Interim Report, the late start of the project implied that samples collected, checked and made available for analyses were generally low. On top of this, the tight deadlines for conducting the analyses and the time needed to accomplish them urged to start analyses as soon as possible. This, in some cases, limited the samples that were analyzed to those that were first available. The prorogation of the contract partially alleviated this problem.

The following sections reflect the status of analyses conducted by the consortium under this circumstance. The samples that were not analyzed in 2011 remain stored in AZTI for future analyses, where a more optimized design of the different analyses can be approached.

4. GENETIC ANALYSIS OF ATLANTIC BLUEFIN TUNA USING NOVEL GENOMICS TOOLS

Task Leader: Fausto Tinti (UNIBO)

Participants

UNIBO: Alessia Cariani, Corrado Piccinetti, Eleonora Pintus, Marco Stagioni, Fausto Tinti.

UPV-EHU: Andone Estonba, Aitor Albaina.

UNICA: Piero Addis, Rita Cannas.

Biogenomics-KULeuven: Gregory Maes, Jeroen van Houdt.

IFREMER: Jean-Marc Fromentin.

AZTI: Haritz Arrizabalaga, Urtzi Laconcha, Igaratza Fraile, Nicolas Goñi

4.1 Introduction

The genetic analyses we carried out within the GBYP-Phase 2 aimed to improve knowledge on the bluefin tuna population spatial dynamics by means of population structure and individual assignment. This strategy potentially gives opportunity to trace back the origin of individuals and populations in marine fish species with high potential for dispersal as bony fish and bluefin tuna.

The bluefin tuna exhibited a very subtle genetic structure among populations, with very low genetic differentiation at neutral loci (Hauser and Ward 1998; Carlsson et al. 2004; 2007; Riccioni et al. 2010). The lack of a clear signal of reproductive isolation among population samples at neutral loci can be maintained by low rates of gene flow among populations and might compromise the identification of clear patterns of population structure as well as the assignment of individuals mixing in the foraging grounds to the originating population (Nielsen et al. 2009). Therefore, to resolve population structure and individual origin in bluefin tuna, the basic concept of genetic analyses should shift from a neutral variation-based approach to a new concept for population genetics with high resolution power. The conceptual basis predicts that environmental adaptation and selection can rapidly drive allele frequency of populations to diverge at the expressed genes level (i.e. under selection loci). Experimental evidence showed that marine fish populations inhabiting closed areas, can exhibit very high differentiation at the loci under selection (outlier loci; Odgen 2008; Nielsen et al. 2009). The main obstacle to implement this approach in non-model fish species is the lack of genomic data on expressed genes and loci. Such difficulty has been overcome by using Next Generation Sequencing (NGS) approaches, offering the opportunity to obtain several hundreds/thousands of Single Nucleotide Polymorphisms (SNPs) in expressed gene sequences in non-model fish species (Hauser and Seeb 2008; Freamo et al. 2011; Helyar et al 2011; Milano et al. 2011).

4.1.1 Genetic strategies and objectives

In the Genetic tasks of the GBYP 06/2011-Phase 2 program we used in parallel **two different genomic strategies,** both based on new Next Generation Sequencing (NGS) approaches, to increase the power and likelihood to detect bluefin tuna spatial population dynamics. In order to have greater chance to definitely resolve the population structure of Atlantic bluefin tuna and therefore to assess the ecological/reproductive interactions among subpopulations in the Mediterranean and adjacent Eastern Atlantic foraging grounds, it is essential to couple 1) the

use of a large number (from several hundreds to thousands) of species-specific high performance genetic loci (SNPs) developed by novel transcriptomic and genomic NGS technologies, and most of these loci can be potentially under positive selection; 2) a broad scale spatial, temporally replicated and multi-strata genetic sampling of the reproductive populations and ecological aggregates of the Atlantic bluefin tuna; 3) the production of individual bluefin tuna genomic and genetic data that can be integrated with biological, otolith microchemistry and fishery data.

The two following genetic strategies have been addressed in the Genetic tasks:

- **NGS-Transcriptome Sequencing** (henceforth NGS-TS). This strategy aimed to develop **384 highly-confident SNPs** based on joint transcriptome-genome sequencing and *in silico* SNP detection and subsequently, standard SNP genotyping of a set of 23 population samples (in total 919 individuals) selected among those collected in the Sampling tasks. The background genomic information that allowed the performing of Genetic tasks in this strategy were represented by hundreds of bluefin tuna candidate SNPs available within the GBYP-Phase 2 consortium and provided by the research project BFTbySNP carried out by the consortium formed by UNIBO, UPV/EHU, UNICA and AZTI, together with the technologies and expertise available in the GBYP partnership. This strategy has been proven to be successful to detect population structure and to improve population traceability and individual assignment in several marine fish as cod, herring, hake and sole (FishPopTrace 2012).
- **NGS-Reduced Representation Sequencing and Genotyping** (Davey et al. 2011; henceforth NGS-RRSG) to jointly **identify and genotype thousands of SNP loci** at the genome-wide level. The NGS-RRSG was carried out to a set of 8 population samples (in total 192 individuals) selected among those collected in the Sampling tasks. The main requirement for successful implementation of genetic traceability is the availability of many molecular markers and the development of novel validated genetic tests for the needs of fisheries enforcement. However, the current time to market of such tools is long. The process of marker discovery, validation and translation in population/stock diagnostic application is still labour-intensive and expensive. This limits the number of **diagnostic markers** identified to date in exploited fish species. The technological focus of NGS-RRSG was the rapid and cost-efficient **genome wide genetic marker discovery and genotyping** by NGS in bluefin tuna sampled in the GBYP-Phase 2 Sampling tasks. The novel sequencing technologies are capable of discovering, sequencing and genotyping not hundreds but thousands of markers across almost any genome of

interest in a single step, even in populations in which little or no genomic information is available (such as bluefin tuna). Several new methods are aimed at reducing the sequencing effort to screen thousands of SNPs at a much smaller cost than using whole genome sequencing or SNP-chip analyses, making rapid uptake and implementation of this technology realistic. Examples of Reduced Representation Sequencing and Genotyping technology (RRSG) are reduced-representation libraries (RRLs) or Complexity Reduction of Polymorphic Sequences (CRoPS), Restriction-site-Associated DNA sequencing (RAD-seq) and low coverage genotyping (Genotyping-By-Sequencing or GBS). **This task focuses on the technological and bioinformatic development of a RRSG technology for traceability purposes in the bluefin tuna**. RRSG approaches are quick, extremely specific, highly reproducible, and may reach important regions of the genome that are inaccessible to sequence capture approaches. The advantage of RRSG in species lacking a complete genome sequence is that a reference map can be generated in the process of sample genotyping, while genome enabled species can highly benefit from a cheap genotyping by sequencing approach, allowing the discovery of novel (regulatory) polymorphisms outside exons. RRSG is particularly useful, as it enables decrease the genomic regions queried to a scalable number of loci, typically from a few thousands to 100,000 and more, depending on the application envisaged (and restriction enzyme used). After optimization by Biogenomics (see www.biogenomics.eu for more information), the applied RRSG technology is specifically tailored to perform a genome-wide screening of the bluefin tuna genome and to find diagnostic markers (highly-confident SNP loci or outlier loci).

4.1.2 Genetic tasks

The Genetic tasks of the GBYP 06/2011 (ICCAT/GBYP PHASE 2) program we carried out are:

Task	NGS-TS	NGS-RRSG
1	**Population sample analysis design, DNA extraction and quality control** of bluefin tuna samples	
2	**Selection of a panel of 384 SNPs** among those available from previous transcriptomic and genomic data	Testing and using a novel high throughput technology (**Reduced Representation Sequencing and Genotyping** or RRSG) for large scale joint SNP development and genotyping, by genome sequencing of 8 population samples (in total 192 individuals)
3	**384-SNP genotyping of 23 population samples** (in total 919 individuals) followed by a thorough quality analysis	
4	**Bioinformatic and population genetic/genomic analysis** of data obtained in the tasks 2 and 3 of NGS-TS and in the task 2-3 of NGS-RGSS, through classical and advanced tests and methods for inferring population genetic structure and population origin assignment	

4.2 Results

Results obtained in the GBYP 06/2011 (ICCAT/GBYP PHASE 2) program are reported according to Genetic tasks and NGS strategies.

4.2.1 Task 1 (NGS-TS and NGS-RRSG) - Population sample analysis design, DNA extraction and quality control

Population sample analysis design

The samples analysed in the Genetic tasks (Table 4.1) were selected among those available in early October 2011 according to the following criteria and priorities: age and ecology (in descending order of priority: Larvae and Age0 individuals > Juveniles > Medium and Large individuals), sampling area (in descending order of priority: Eastern Mediterranean > Central Mediterranean > NE Atlantic > Western Mediterranean).

Table 4.1 – Samples used in the Genetic tasks (includes samples collected under GBYP 06/2011 and samples available from other programs). Samples in bold are those that have been genotyped in the second round accomplished during May 2012. Samples marked by light blue are samples analysed by both NGS-TS and RRSG.

POPULATION SAMPLE ANALYSIS DESIGN FOR GENETIC TASKS															
SAMPLING				SIZE CLASS						NGS-TS				NGS-RRSG	
SOURCE	REGION	AREA	PARTNER	Larvae	Age 0	Juv (<25 kg)	Med (25-100 kg)	Large (>100 kg)	Total	CODE SAMPLE	#SNPtyped 1st round	#SNPtyped 2nd round	#SNPtyped TOTAL	CODE SAMPLE	#SNPtyped
GBYP 2011	EMED	LS	CYPR	**10**(1)	**29**(1)				39	**EMED-LS-LA+0**		39	39	CYPR	24
GBYP 2011	EMED	LS	AZTI/ISTA				**40**	40	80	**EMED-LS-M** EMED-LS-L	25	55	80		
2007	EMED	LS	UNIBO				40		40	EMED-LS-M-2007	40		40		
GBYP 2011	CMED	MA	FMAP					40	40	CMED-MA-L	40		40		
GBYP 2011	CMED	AS	UNIBO			40			40	CMED-AS-J	40		40	UNIBO	24
GBYP 2011	CMED	SI	UNIBO			40			40	CMED-SI-J	40		40		
GBYP 2011	CMED	SI	NECTON				40		40	CMED-SI-M	40		40		
GBYP 2011	WMED	LI	UNIGE			**40**			40	**WMED-LI-J**		40	40	UNIGO	24
GBYP 2011	WMED	SA	UNICA				**40**	**40**	80	**WMED-SA-M** **WMED-SA-L**		80	80	UNICA	24
GBYP 2011	WMED	GL	IFREMER			**40**			40	**WMED-GL-J**		40	40		
GBYP 2011	WMED	TY	NECTON				40		40	WMED-TY-M	20	20	40		
GBYP 2011	WMED	TY	UNIBO		40				40	WMED-TY-0	30	10	40		
2009	WMED	BA	AZTI		40				40	WMED-BA-0-2009	40		40		
2010	WMED	BA	AZTI		40				40	**WMED-BA-0-2010**		40	40		
GBYP 2011	WMED	BA	IEO/AZTI		40	40			80	WMED-BA-0 WMED-BA-J	80		80	WMED	24
GBYP 2011	NEAtl	GI	IEO					40	40	NEAtl-GI-L	40		40		
GBYP 2011	NEAtl	PO	IPIMAR					40	40	NEAtl-PO-L	24	16	40	IPIMO	24

GBYP 2011	NEAtl	BB	AZTI			40	**40**		80	NEAtl-BB-J **NEAtl-BB-M**	24	56	80	BAYBIO	24
GBYP 2011	GOM		AZTI	29	16				45	GOM-LA+0	40		40	GMYOY	24
Total general				**39**	**205**	**240**	**240**	**200**	**924**	**23**	**523**[2]	**396**	**919**[3]	**8**	**192**

[1] *Thunnus thynnus* individuals for which the species ID has been checked before the second round of genotyping.

[2] individuals genotyped in the first round.

[3] 30 *Thunnus alalunga* individuals were added to the population sample analysis design and 11 positive controls for a grand total of 960 individuals to be SNP genotyped.

In the NGS-TS tasks we have initially selected 23 bluefin tuna population samples (919 individuals, plus 11 positive controls and 30 albacore individuals), of which 6 were reference samples of spawning populations (EMED-LS-LA+0, WMED-TY-0, WMED-BA-0-2009, WMED-BA-0-2010, WMED-BA-0 and GOM-LA+0) and 17 represented feeding aggregates in the regions/areas targeted by Sampling tasks (Table 4.1). Because delays caused by low DNA quality/quantity yields and species misidentification (see the paragraph *Problems of DNA quantity/quality yields and Species misidentification*), the SNP genotyping was conducted in two rounds (see Table 4.1 for indivuals genotyped in round).

In the NGS-RRSG we have selected and SNP-typed 8 population samples of which 3 were reference samples of spawning populations (CYPR, WMED, and GMYOY) and 5 were from feeding aggregates in the regions/areas targeted by Sampling tasks (Table 4.1).

DNA extraction and quality control

From individual tissues (muscle, finclip) and larvae collected according to the protocols defined by the Consortium, the total genomic DNA was extracted using the Nucleospin Tissue DNA extraction kit according to the manufacturer's conditions (Machery&Nagel GmBH, Düren, Germany) and assessed for quantity/quality yields.
In the NGS-TS, the genomic DNA has been extracted by the company subcontracted for the services of DNA extraction and SNP genotyping (i.e. Parco Tecnologico Padano, http://www.tecnoparco.org).
In the NGS-RRSG, the genomic DNA has been purified and analysed by the company subcontracted for the Reduced Representation Sequencing and Genotyping analysis (i.e. Biogenomics, www.biogenomics.eu, a K.U. Leuven contract research division under the supervision of Dr. Gregory Maes and Dr. Jeroen Van Houdt).

Problems of DNA quantity/quality yields and Species misidentification

Tissue samples were collected following the sampling protocol and stored in ethanol 96%. These were shipped to AZTI for storing and data basing. From AZTI, tissue aliquots were shipped to Biogenomics and to UNIBO for NGS-RRSG and NGS-TS tasks, respectively. While DNA extraction for NGS-RRSG was carried out in

November 2011, the DNA extraction for NGS-TS was started in February 2012 at Parco Tecnologico Padano (hereafter PTP) since the Illumina SNP assay was delivered at this date.

A first experiment of DNA extraction on 192 individual tissues was performed at PTP on February 13-18, 2012. This experiment yielded genomic DNA with low quantity and quality. Further DNA extraction experiments were carried out at the PTP, UNIBO and AZTI from February 23 to March 6, 2012, by introducing slight modifications in the DNA extraction protocols to improve the yields in quantity and quality.

The comparison between the DNA yields obtained by Biogenomics in the NGS-RRSG tasks carried out in November 2011 and those obtained by PTP, UNIBO and AZTI in the NGS-TS tasks in February-March 2011 revealed that 1) some samples had the same low yields independently from the date in which the DNA extraction was performed, 2) some samples yielded DNA of low quantity and quality in the NGS-TS extraction experiments while they gave high yields in the NGS-RRSG extraction experiments, suggesting a deterioration of tissue samples in the storage and shipping period (October 2011 – February 2012); 3) some samples yielded DNA of high quantity and quality in both NGS-TS and NGS-RRSG extraction experiments suggesting that no deterioration of tissue samples has occurred in the storage and shipping period (October 2011 – February 2012).

In addition, the population sample EMED-LS-LA, potentially formed by 40 bluefin tuna larvae collected in the Eastern Mediterranean (Levantine Sea), was instead prevalently composed by non-bluefin tuna larvae (see the Task 4 NGS-RRSG). The molecular species identification of the 83 individual larvae collected in the strata EMED-LS-LA revealed that only 10 were bluefin tuna. These larvae, together with the 29 Age0 bluefin tuna individuals (EMED-LS-0) collected in this area represent a reference sample for the spawning population of the Eastern Mediterranean and were therefore included in the second round of SNP typing.

4.2.2 Task 2 (NGS-TS) - The selection of a panel of 384 SNPs

Availability and exploitation of genomic and transcriptomic data

To address the specific objectives of the GBYP06/2011 genetic program and to retrieve the most reliable SNPs from the available datasets produced during an earlier project (BFTbySNP, by UNIBO, UNICA, UPV/EHU and AZTI), an additional in depth analysis of SNPs was first performed on the available (more than 5000) candidate SNPs, applying the following stringency/quality criteria for SNP selection: i) presence in several target populations, ii) showing sufficient polymorphism, iii) high quality and reliable (coverage depth), iv) present in both transcriptome and genome data, and v) flanking regions should be lowly polymorphic for the genotyping assay design.

Selection of 384 high performance SNP loci of bluefin tuna

The selection of the 384 high performance SNPs was carried out in multiple rounds. Since for the Genetic task 3 we selected a company providing genotyping service based on the IlluminaGoldenGate assay on the VeraCode BeadXpress format, in the SNP selection procedure the requirement needed for that assay were properly taken into account.

The selection of the 384 SNP loci was focused on picking reliable SNPs (so real polymorphisms and not sequencing errors) and SNPs suitable for IlluminaGoldenGate genotyping requirement, to have the best conversion rate (number of working and polymorphic SNPs) as possible. Currently, conversion rates for non validated SNPs developed from EST-sequencing of non model organisms are approximately 30% (Milano et al. 2011), but our sequencing approach is expected to increase significantly the conversion rate to 70-80 % validated polymorphic SNPs to be used for population assignment and population structure analyses.

The selected panel of 384 high performance SNPs was submitted to Illumina Inc. (http://www.illumina.com) for final evaluation and after positive assessment of the list, 10 SNP assay plates (usable for the genotyping of 960 individuals) were ordered on the 7th of November 2011, with Illumina Order Confirmation Number – 245827. However, due to the move of the factory, the Illumina Company communicated that the expected delivery was postponed to 1st of February 2012. Additionally they experienced some backlog issues for this product, so we received a negative expedite request from the planning department when we asked for an urgent delivery of the order. Consumables were shipped directly to the subcontracted company PTP on the 1st of February 2012.

4.2.3 Task 3 (NGS-TS) - The 384-SNP genotyping of bluefin tuna individuals

To carry out the core activity of this task (i.e. the individual genotyping at 384 high performance SNP loci) we have selected and subcontracted the company "Parco Tecnologico Padano" (http://www.tecnoparco.org) after launching a best bid procedure to seven different companies and evaluating their offer for cost/benefit tradeoffs. The Task 3 has experienced a significant delay and was started in February 2012 because several problems of different kind. First of all, the late signature of the contract caused a delay in the delivery of several samples. In addition, a further delay in the delivery of results for this task was caused by the technical problem of low DNA quality/quantity yields experienced during Task 1 (see the paragraph "Problems of DNA quantity/quality yields and species misidentification"), and because the lengthy in depth bioinformatic analysis required for the reliable selection of the 384 high performance SNPs (Genetic Task 2).

Additionally, the ordering schedule of Illumina has been prolonged due to the moving of the whole company's production unit to other offices. This added 6 additional weeks to the SNP array production schedule before the contracted company could start genotyping. Despite our careful and streamlined analysis protocol, the first set of genotyping data was obtained in March 2012 instead of November 2011. However, on the positive side, the new timing of the analyses allowed making some modifications to the final set of samples that were genotyped, based on the final outcome of the sampling process. Moreover, the approach allowed for bluefin tuna samples collected from other strata to be genotyped in the second round using the assay plates ordered at the Illumina Inc.

4.2.4 Task 2-3 (NGS-RRSG) - Reduced Representation Sequencing and Genotyping of bluefin tuna individuals

Set up of the RRSG strategy and pipeline for bluefin tuna

Given the original development of RRSG technologies in model organisms and to assess the results on fishes, we started the task by performing a preliminary test case using the model fish species *Gasterosteus aculeatus* (three-spined stickleback).

The RRSG test in the three-spined stickleback permitted to set up a feasible pipeline of conducting RRSG in bluefin tuna with success. After an evaluation of these initial results and the test of various restriction enzymes in tuna, it was decided to genotype 192 individuals using a RRSG approach on an Illumina Hi-seq 2000 to discover highly-divergent SNP markers which would provide high power for resolving population structure and population assignment.

Reduced Representation Sequencing and Genotyping of 192 bluefin tuna individuals

To apply the RRSG technique on bluefin tuna, several steps need to be followed: testing of DNA quality and quantity, test of restriction enzymes for optimal fragment number and finally the full analyses using the 192 barcodes. The late arrival of all samples needed for this analysis made that this task experienced a significant delay and could only be started in mid-October 2011. The full dataset was generated over January and February 2012 and is now presented hereunder.

4.2.5 Task 4 (NGS-TS) - Bioinformatic and population genetic data analysis

Quality check genotypes

Table 4.2: Average of percentage of missing values for each of the 23 population samples.

Population sample	#_call rate <80%_ind	Average of % MV_ind
CMED-AS-J		*0.75*
CMED-MA-L	19	*7.92*
CMED-SI-J	4	*4.49*
CMED-SI-M	1	*5.68*
EMED-LS-L		*0.32*
EMED-LS-LA+0		*1.01*
EMED-LS-M		*0.26*
EMED-LS-M-2007	1	*1.69*
GOM-LA+0	2	*0.62*
NEAtl-BB-J		*1.05*
NEAtl-BB-M	1	*0.57*
NEAtl-GI-L	11	*8.75*
NEAtl-PO-L	1	*1.99*
WMED-BA-0		*1.55*
WMED-BA-0-2009		*1.28*
WMED-BA-0-2010	1	*1.67*
WMED-BA-J	12	*6.45*
WMED-GL-J	1	*2.35*
WMED-LI-J		*0.37*
WMED-SA-L	1	*1.40*
WMED-SA-M	2	*4.52*
WMED-TY-0		*2.70*
WMED-TY-M	2	*5.04*
Total/Average	**59**	***2.71***

DNA extraction and SNP genotyping were performed at the subcontracted company Parco Tecnologico Padano with the IlluminaGoldenGate assay on the VeraCode BeadXpress format and results were visualized and analysed with the GenomeStudio Data Analysis Software package.

268 SNPs (out of the 384 selected loci) with good quality of clustering were chosen. We excluded SNPs that were monomorphic (54), did not work, did not cluster or had

a percentage of missing value >10% (62) in the 919 bluefin tuna individuals. This yielded a conversion rate of ≈70%.

With the purpose of having a more informative, reduced dataset, we excluded 59 individuals prevalently of the CMED-MA-L, NEAtl-GI-L and WMED-BA-J population samples with a call rate < 80% (Table 4.2). At the end, we obtained a dataset with **860 individuals** genotyped at **268 SNPs.** The average percentage of missing value for each population ranged from 0.26 in the EMED-LS-M to 8.75 in the NEAtl-GI-L (Table 4.2).

In order to assess the accuracy and the quality of the SNP genotyping procedure, we included 11 replicates of the same individual in each locus. In 17 SNP loci out of 268, the biological replicates did not produce the same genotype. However in 15 loci this was caused by the failure of the amplification of one or more replicates. Therefore, only two out of 268 SNPs showed non consistency genotypes, for a total of 3 erroneous genotypes over a total of 268x11= 2948 data points, giving an estimate of the genotyping error rate of $1.02 * 10^{-3}$.

Genetic diversity and Hardy-Weinberg Equilibrium

Basic descriptive statistics of genetic diversity in the 23 population samples were calculated over the complete dataset of 268 loci (Table 4.3; Figure 4.1). Values are almost similar across all population samples, with CMED-MA-L and NEAtl-GI-L having the lowest Hobs while WMED-BA-0-2009 the highest (Figure 4.1). Most population samples showed a negative F_{IS} value and this is caused by an excess of heterozygote genotypes with respect to the expectation under the Hardy-Weinberg Equilibrium (HWE) (Table 4.3).

Table 4.3: Descriptive statistics of genetic diversity in the 23 population samples at the 268 SNP loci. Reference samples (Larvae and Age 0) are boxed.

Population sample	N	Hexp	Hobs	AR	Fis		%P
CMED-AS-J	40	*0.356*	*0.358*	*1.965*	*0.005*		*100.00%*
CMED-MA-L	21	*0.348*	*0.294*	*1.964*	*0.180*	*	*99.63%*
CMED-SI-J	36	*0.361*	*0.369*	*1.972*	*-0.007*		*100.00%*
CMED-SI-M	39	*0.352*	*0.361*	*1.960*	*-0.012*		*99.63%*
EMED-LS-LA+0	39	*0.354*	*0.358*	*1.963*	*0.002*		*100.00%*
EMED-LS-L	40	*0.356*	*0.358*	*1.965*	*0.008*		*100.00%*
EMED-LS-M-2007	39	*0.357*	*0.364*	*1.969*	*-0.006*		*100.00%*
EMED-LS-M	40	*0.348*	*0.364*	*1.960*	*-0.032*		*99.63%*
GOM-LA+0	38	*0.360*	*0.368*	*1.972*	*-0.009*		*100.00%*
NEAtl-BB-J	40	*0.353*	*0.353*	*1.966*	*0.012*		*100.00%*
NEAtl-BB-M	39	*0.359*	*0.364*	*1.973*	*-0.001*		*100.00%*
NEAtl-GI-L	29	*0.347*	*0.292*	*1.950*	*0.179*	*	*98.88%*
NEAtl-PO-L	39	*0.359*	*0.369*	*1.968*	*-0.015*		*100.00%*
WMED-BA-0	40	*0.356*	*0.367*	*1.963*	*-0.017*		*100.00%*
WMED-BA-0-2009	40	*0.364*	*0.375*	*1.973*	*-0.015*		*99.63%*
WMED-BA0-2010	39	*0.357*	*0.363*	*1.970*	*-0.003*		*100.00%*
WMED-BA-J	28	*0.355*	*0.314*	*1.961*	*0.135*	*	*99.63%*
WMED-GL-J	39	*0.356*	*0.368*	*1.967*	*-0.020*		*99.63%*
WMED-LI-J	40	*0.358*	*0.360*	*1.970*	*0.008*		*100.00%*
WMED-SA-L	39	*0.358*	*0.364*	*1.972*	*-0.003*		*100.00%*
WMED-SA-M	38	*0.349*	*0.339*	*1.960*	*0.040*	*	*100.00%*
WMED-TY-0	40	*0.356*	*0.353*	*1.972*	*0.019*	*	*100.00%*
WMED-TY-M	38	*0.348*	*0.361*	*1.962*	*-0.023*		*99.63%*

N: sample size; Hexp: expected heterozygosity; Hobs: observed heterozygosity; AR: allelic richness; Fis: fixation index (* indicates $p<0.05$); %P: % of polymorphic loci.

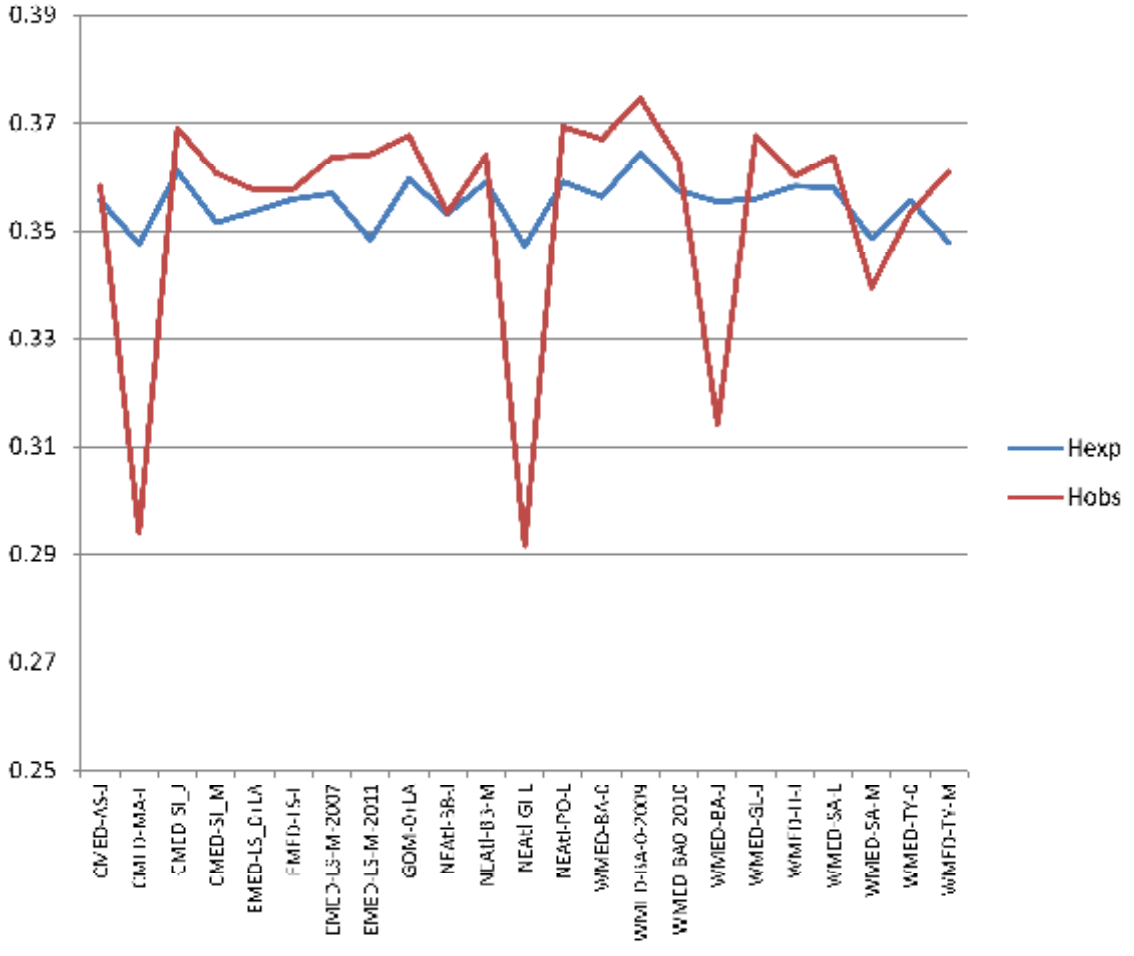


Figure 4.1: Observed (Hobs) and expected (Hexp) heterozygosity in the 23 population samples.

We tested the departure from the HWE per population and per locus. Among population samples (Figure 4.2), the CMED-MA-L, WMED-BA-J and NEAtl-GI-L samples displayed the higher rate of loci significantly deviating from the HWE.

None of the loci deviated from the HWE in all the 23 population samples (Figure 4.3). The percentage of HW disequilibria is low and can be therefore neglected at this stage of analysis.

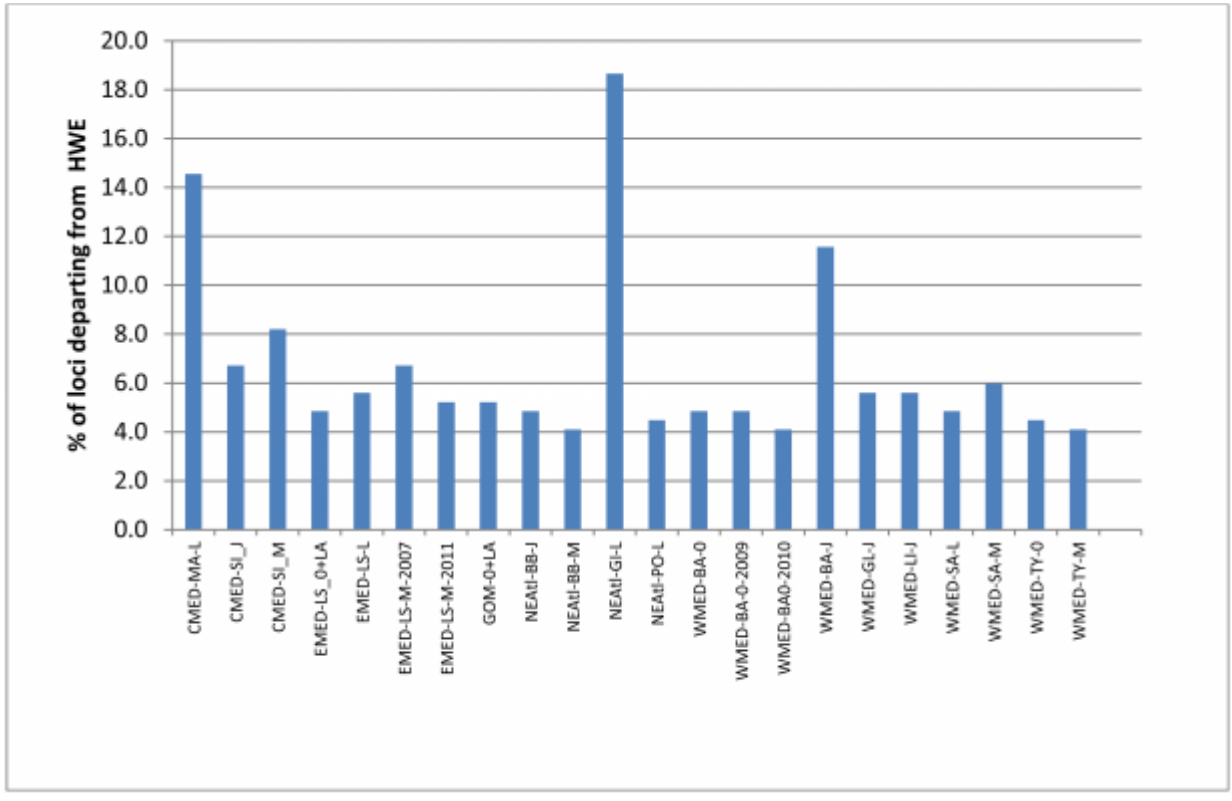


Figure 4.2: Percentage of SNP loci departing from HWE ($p<0.05$) in the 23 population samples.

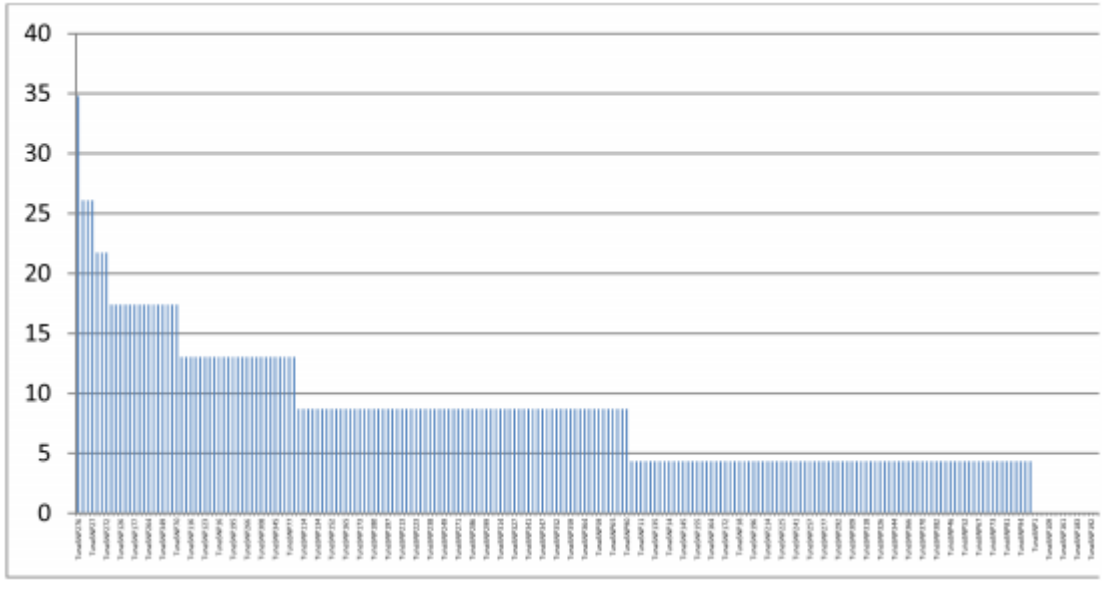


Figure 4.3: Percentage of population samples significantly deviating from HWE ($p<0.05$) at each locus.

Genetic differentiation

The **genetic differentiation among population samples at the 268 loci was extremely low**, with an overall F_{ST} value of -0.00010 (p = 1.00000+-0.00000). The **pairwise F_{ST} values among samples calculated over the 268 SNP loci were very low** (from -0.00491 to -0.0051) and a few were significant, mainly for sample CMED-SI-J. According to these values, **the distribution of single-locus F_{ST} estimates per locus** (Figure 4.4) **showed that more than half of the 268 loci displayed negative values** and **only 52 loci were above a value of 0.5%** (F_{ST}> 0.005).

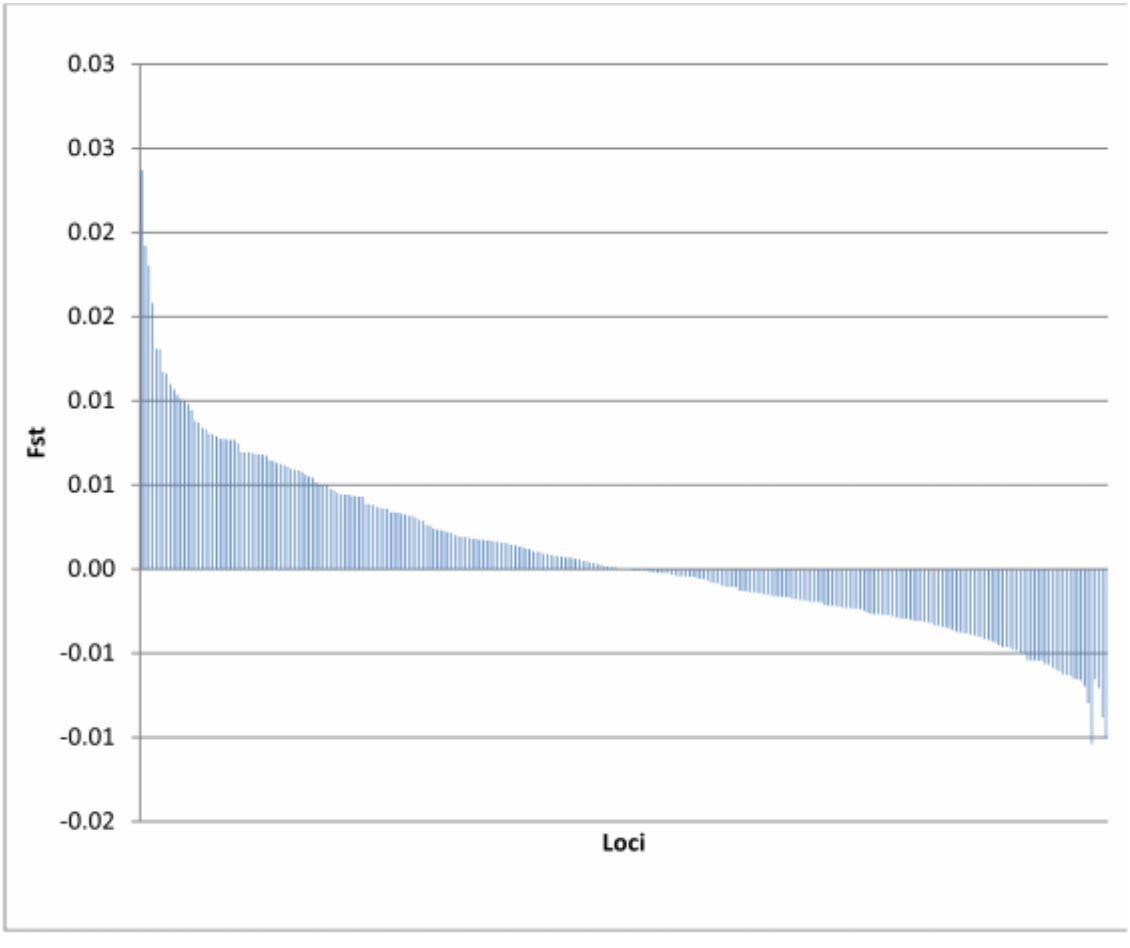


Figure 4.4: Distribution of single-locus F_{ST} values among the 23 population samples at the 268 loci.

In order to assess signals of genetic differentiation among the bluefin tuna population samples, we decided to use a restricted SNP panel formed by the 52 loci with F_{ST}>0.005 in the downstream population genetic data analyses. **Based on this**

panel, few of the pairwise F_{ST} values were significant (p<0.05), but without a clear pattern of population differentiation (e.g. the **expected differentiation between WMED and GOM reference samples**). On average the CMED-MA-L, CMED-SI-J, CMED-SI-M and EMED-LS-M showed a somehow higher level of differentiation (Figure 4.5a). The MDS plot based on the F_{ST} matrix showed a great level of genetic differentiation among samples but without significant grouping. The CMED-MA-L, CMED-SI-J, CMED-SI-M and EMED-LS-M samples resulted distant from the other population samples that grouped in the plot centre (Figure 4.5b).

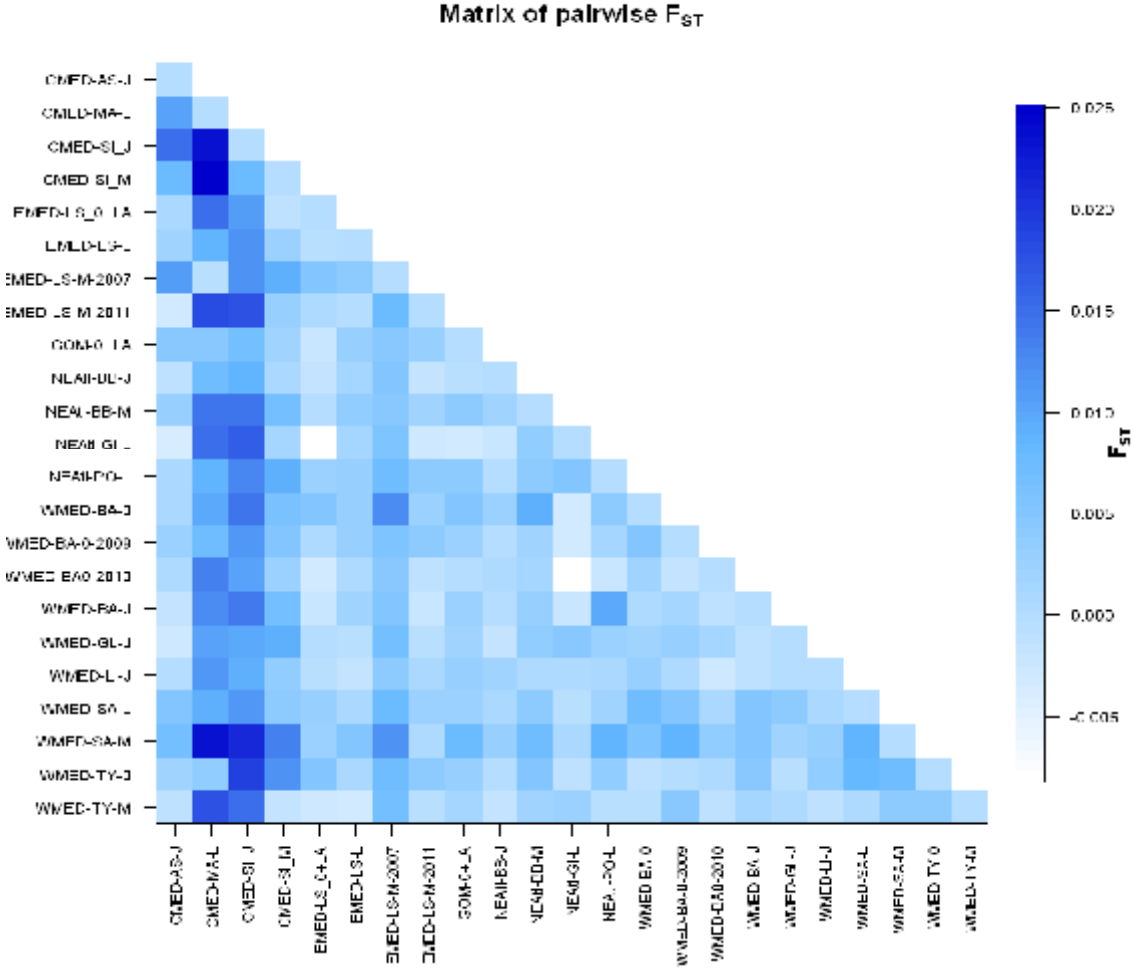


Figure 4.5a: Genetic differentiation among all 23 population samples at the 52 loci with F_{ST}>0.005. Pairwise F_{ST} matrix.

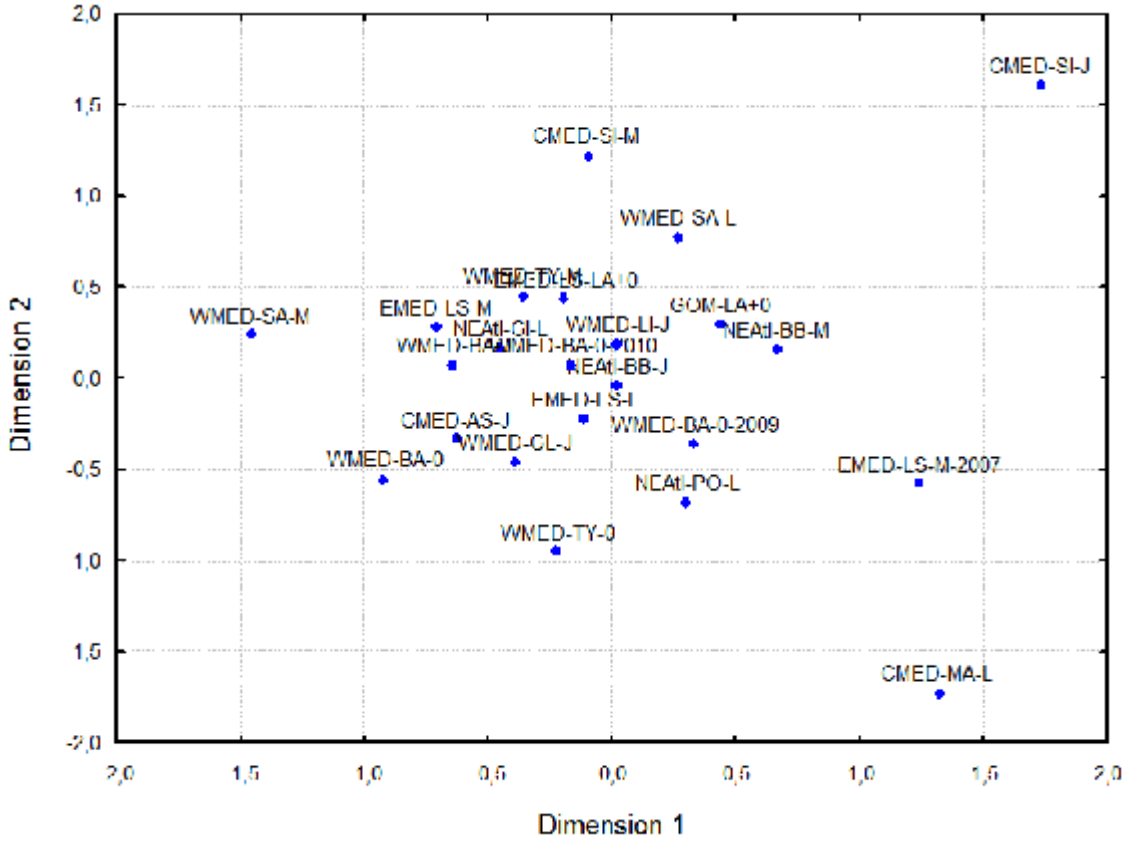


Figure 4.5b: MDS plot on genetic differentiation among all 23 population samples at the 52 loci with F_{ST}>0.005.

Similarly, the distribution of single-locus F_{ST} estimates per locus in the six reference samples of the spawning populations (i.e. EMED-LS-LA+0, GOM-LA+0, WMED-BA-0, WMED-BA-0-2009, WMED-BA-0-2010, WMED-TY-0) showed that more than half of the 268 loci displayed negative values and only 63 loci were with F_{ST}>0.005 (Figure 4.6).

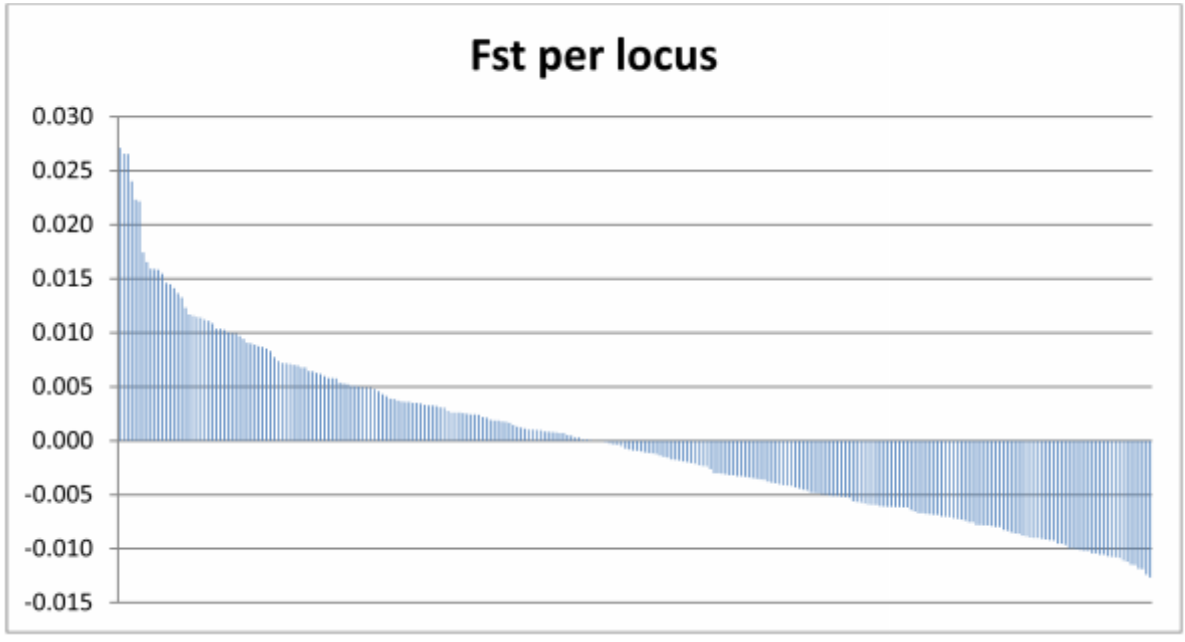


Figure 4.6: Distribution of single-locus F_{ST} values among the six reference samples of the bluefin tuna spawning populations at the 268 loci.

We have then selected a restricted SNP panel of 63 loci with F_{ST}>0.005 among the six reference samples that was used in the downstream genetic analyses. Based on this restricted panel, the F_{ST} values among the six reference samples pairwise raised up and ranged from 0.1% to 1.2%, being most significant (p<0.05; Figure 4.7b). The main differences were found between WMED-BA-0-2009 and GOM-LA+0, followed by WMED-BA-0 and EMED-LS-LA+0. The same pattern is reflected in the MDS plot based on the F_{ST} values (Figure 4.7b).

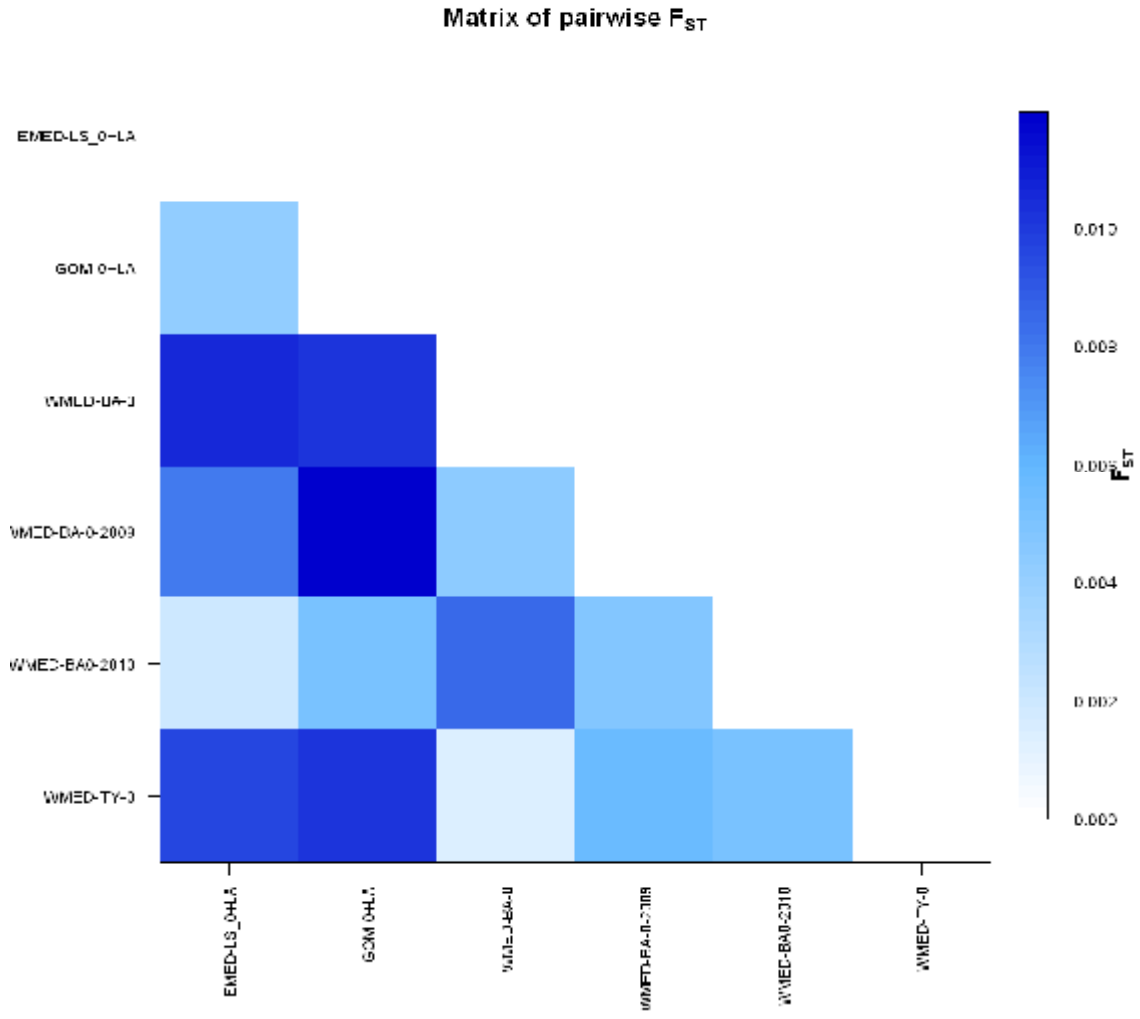


Figure 4.7a: Pairwise F_{ST} matrix among the six reference samples of the spawning populations at the 63 loci with F_{ST}>0.005.

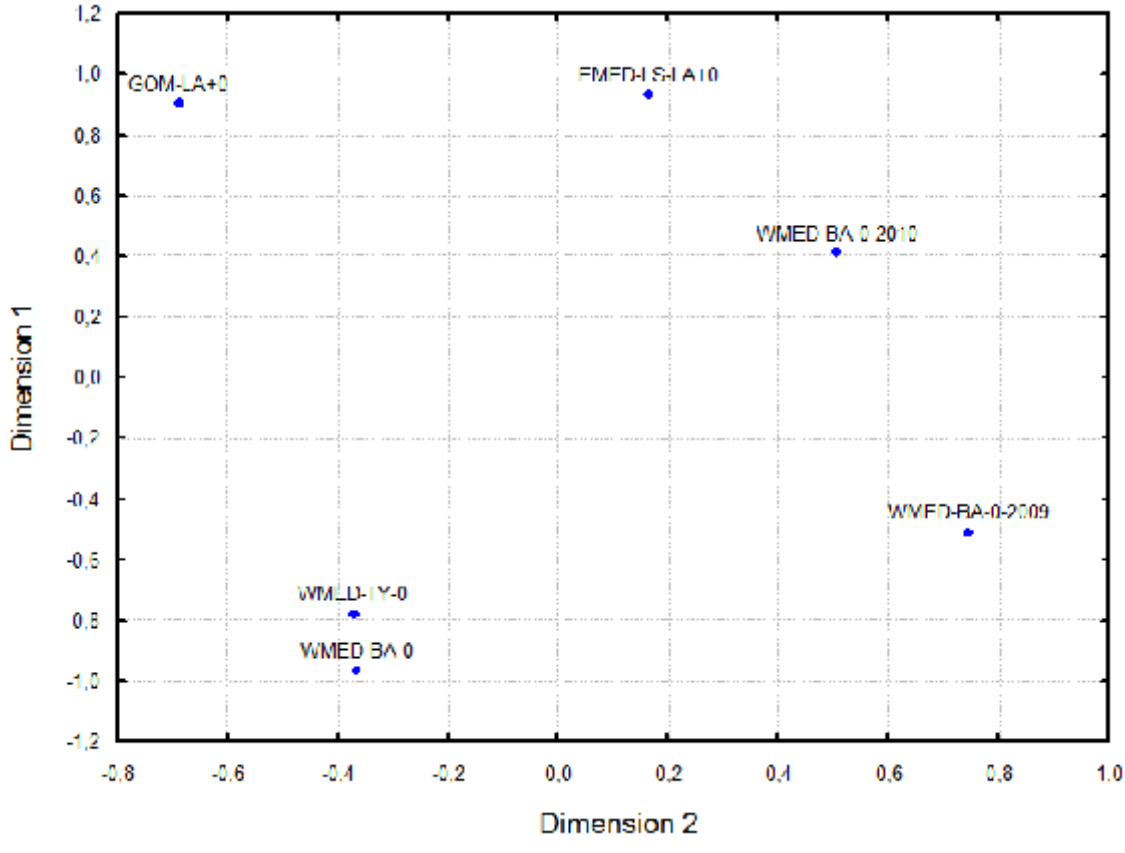


Figure 4.7b: MSD plot on genetic differentiation among the six reference samples of the spawning populations at the 63 loci with F_{ST}>0.005.

Outlier loci analysis

Outlier loci analysis was performed using the LOSITAN software to detect loci potentially under selection, likely located or linked to expressed genes, as it is for the 268 SNPs derived from NGS-TS.

Over the 23 population samples, **13 SNP loci were identified as potentially under positive selection, all of them being included in the panel of 52 SNP loci with F_{ST}>0.005** (Figure 4.8). Gene annotation analysis of these selected loci is ongoing and the biological functions linked to these markers will be further evaluated.

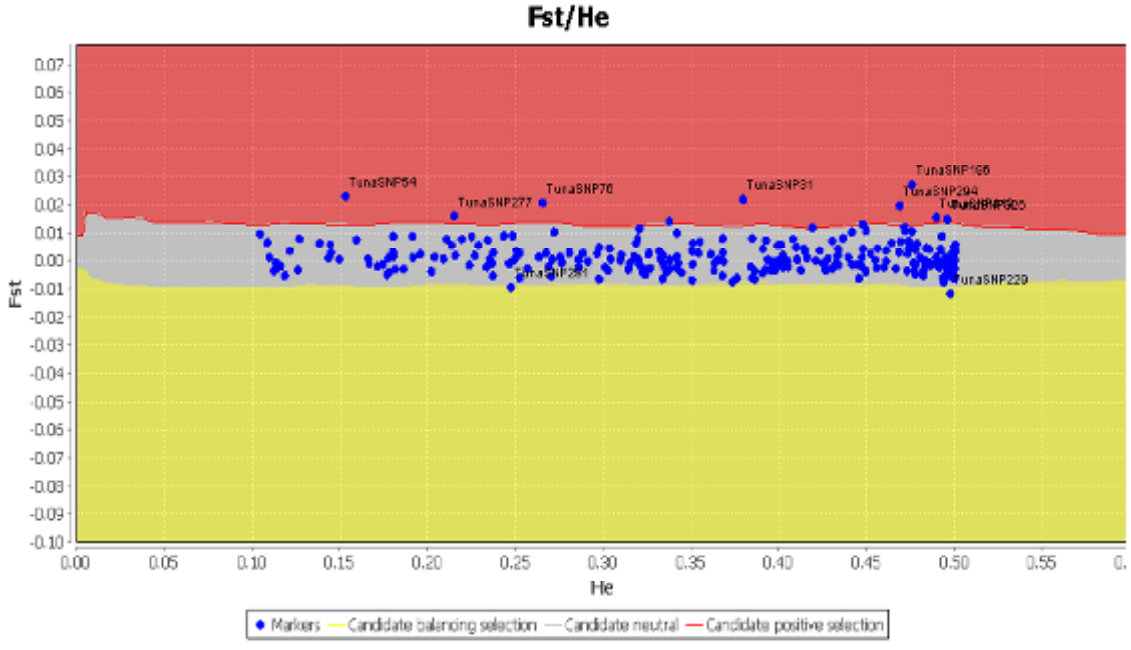


Figure 4.8: Outlier loci analysis among all 23 populations.

When performing the same analysis on the six reference samples of the spawning populations (i.e. EMED-LS_0+LA, GOM-0+LA, WMED-BA-0, WMED-BA-0-2009, WMED-BA0-2010, WMED-TY-0), **none of the loci were detected as being under selection** (Figure 4.9). This lack of resolution of the analysis is probably due to the very low F_{ST} value distribution, affecting the power of outlier loci detection.

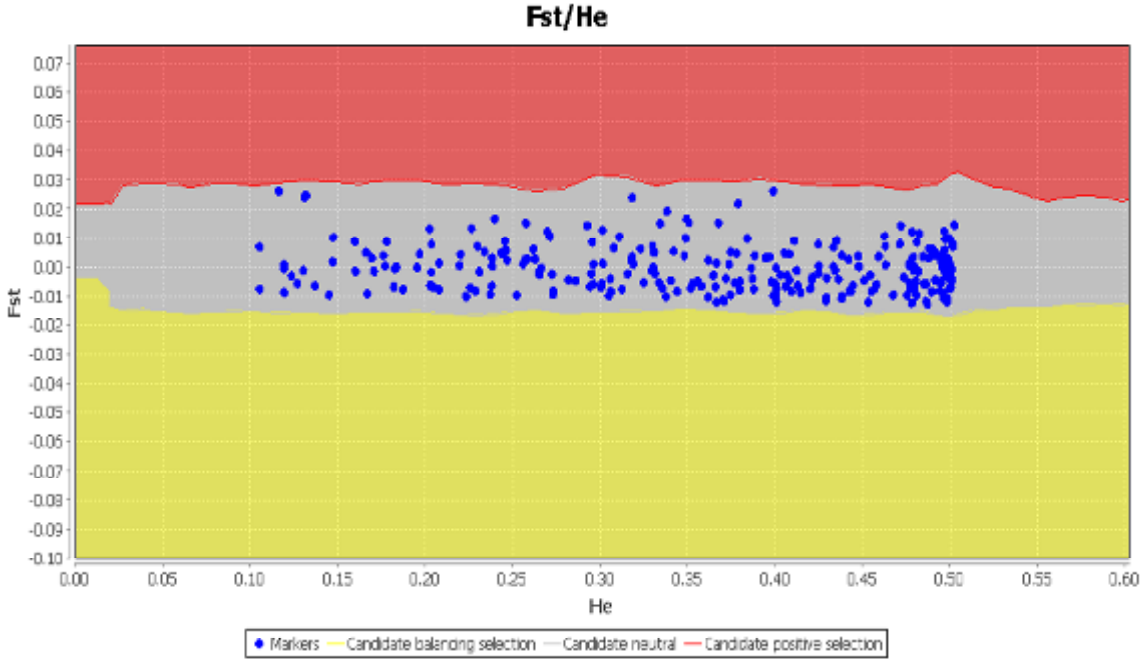


Figure 4.9: Outlier analysis among the six reference samples of the spawning populations.

Multivariate analyses

A Factorial Correspondence Analysis (AFC), implemented in the software Genetix, was performed on the 860 individuals of the 23 population samples with the entire panel of 268 SNP loci (Figure 4.10). This analysis revealed **no evidence of genetic structure**. Therefore, we performed **an AFC using the reduced panel of 52 SNPs with F_{ST}>0.005 but, similarly to the previous analysis, it did not give a clustering pattern of population samples** (Figure 4.11).

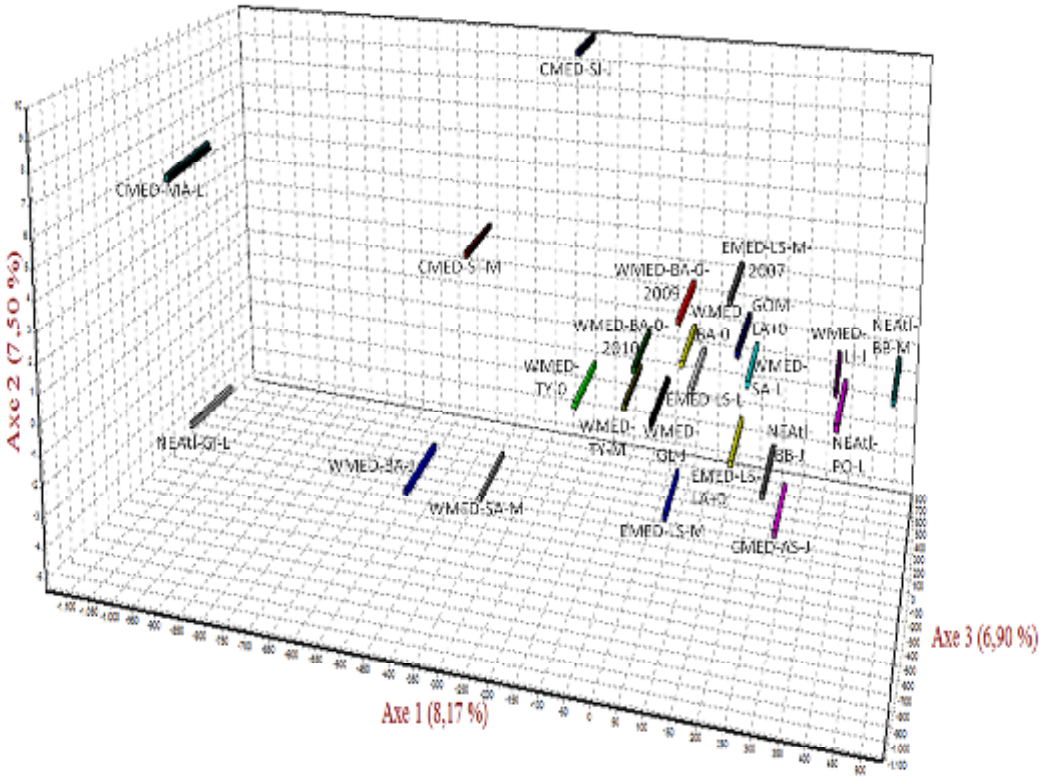


Figure 4.10: AFC analysis of all 23 population samples at all 268 loci.

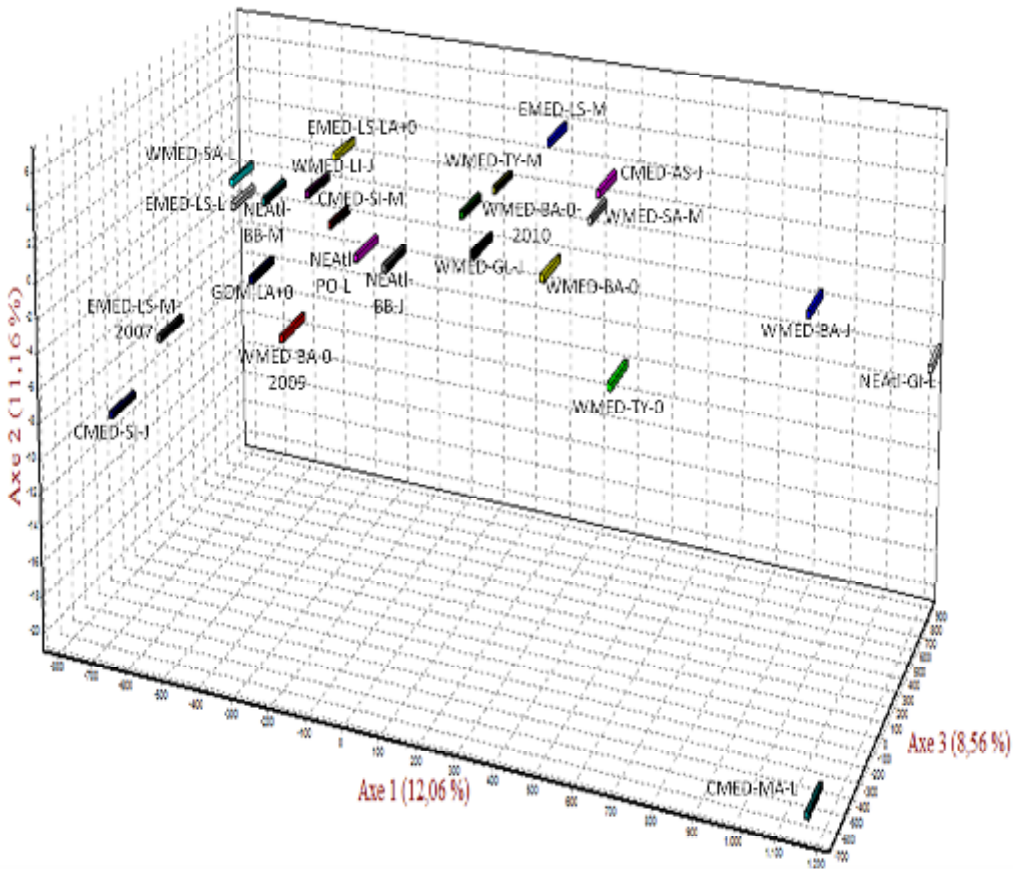


Figure 4.11: AFC analysis of all 23 population samples with the restricted panel of 52 loci with F_{ST}>0.005.

Due to the lack of any structuring signal over all 23 population samples, we assessed the genetic differentiation among the six reference samples representing the spawning populations. An **AFC was performed on these six reference samples with the entire panel of 268 SNP loci and then with a reduced panel of 63 SNP with F_{ST}>0.005, without detecting signal of structuring** (Figures 4.12 and 4.13).

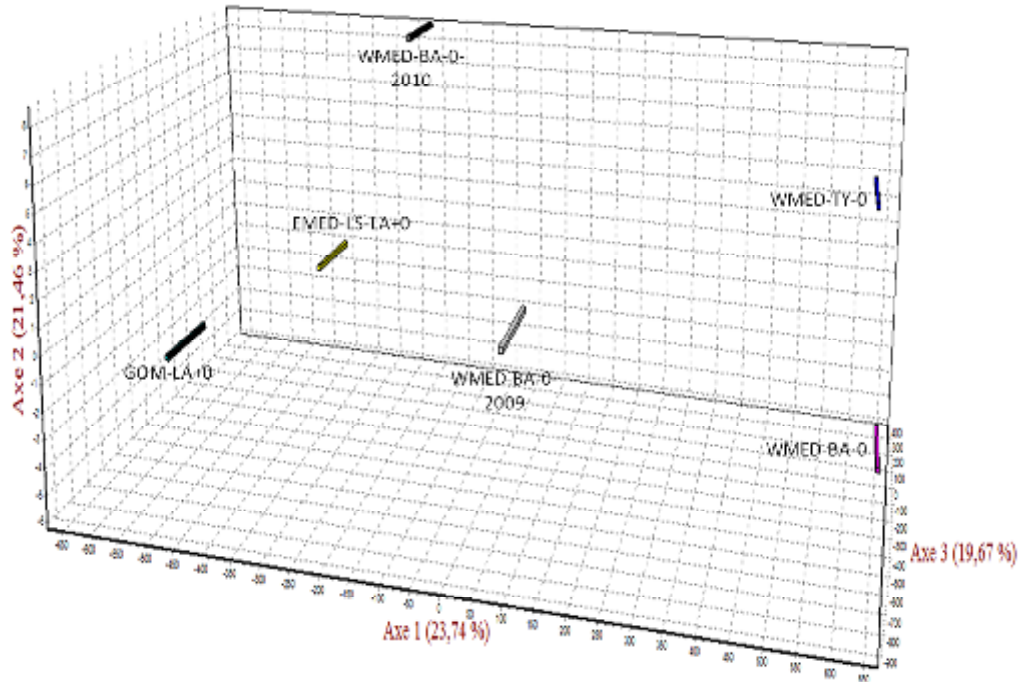


Figure 4.12: AFC analyses of six reference samples using all 268 loci.

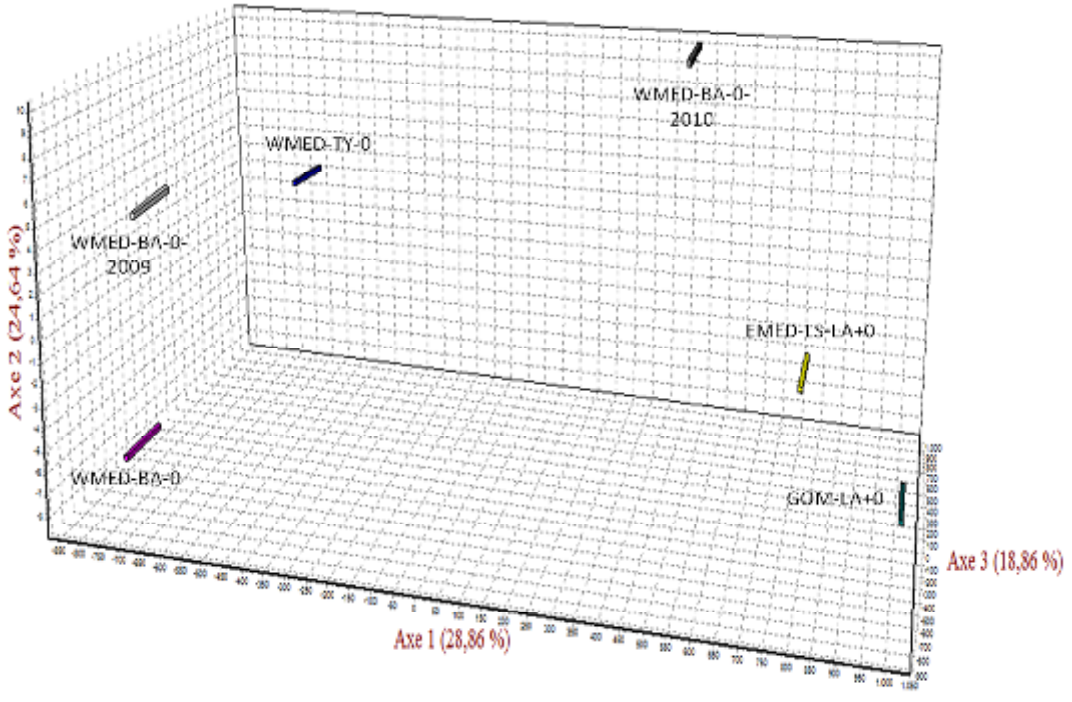


Figure 4.13: AFC analyses of six reference samples using the restricted panel of 63 SNP loci with F_{ST}>0.005.

A **Discriminant Analysis of Principal Components (DAPC) was performed on all 23 populations with the entire panel of 268 SNP loci and with the restricted panel of 52 loci with F_{ST}>0.005**. This method transforms data using principle component analysis (PCA) to create uncorrelated variables for input into Discriminant Analysis (DA). **DA maximizes between-group variation and minimizes within-group variation for assessment of between-group variation**. DAPC is free of assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium and provides graphical representation of divergence among populations. DAPC was implemented using the *adegenet* package (Jombart 2008) in R version 2.14.1 (R Development Core Team 2010). DAPC allowed the search for the most likely number of clusters/groups in the dataset. This can be achieved using k-means, a clustering algorithm which finds a given number (say, k) of groups maximizing the variation between groups, B(X). To identify the optimal number of clusters, k-means is run sequentially with increasing values of k, and di□erent clustering solutions are compared using Bayesian Information Criterion (BIC). Ideally, the optimal clustering solution should correspond to the lowest BIC. In practice, the 'best' BIC is often indicated by an elbow in the curve of BIC values as a function of k. Moreover, being based on the Discriminant Analysis, DAPC also provides membership probabilities of each individual for the di□erent groups based on the retained discriminant functions. While these are di□erent from the admixture coe□cients of software like STRUCTURE (see later), they can still be interpreted as proximities of individuals to the di□erent clusters. Membership probabilities also provide indications of how clear-cut genetic clusters are. Loose clusters will result in fairly flat distributions of membership probabilities of individuals across clusters, pointing to possible admixture.

Again, **the DAPC analysis did not discriminate the bluefin tuna population samples** (Figure 4.14a,b). We then performed **the DAPC on the six reference samples of the spawning populations** (i.e. EMED-LS-LA+0, GOM-LA+0, WMED-BA-0, WMED-BA-0-2009, WMED-BA-0-2010, WMED-TY-0) **with the entire panel of 268 loci and with the restricted panel of 63 SNPs with F_{ST}>0.005**. In this case, both DAPCs **showed a certain but very weak degree of structuring, with GOM-LA+0 separated from the five Mediterranean population samples, which partially overlapped** (Figure 4.15a,b).

Supposing an efficiency of the DAPC with the restricted panel of 63 SNPs with F_{ST}>0.005 in resolving population structuring, we performed a DAPC on all 860 individuals of 23 populations with this panel detecting no signal of structure (Figure 4.16).

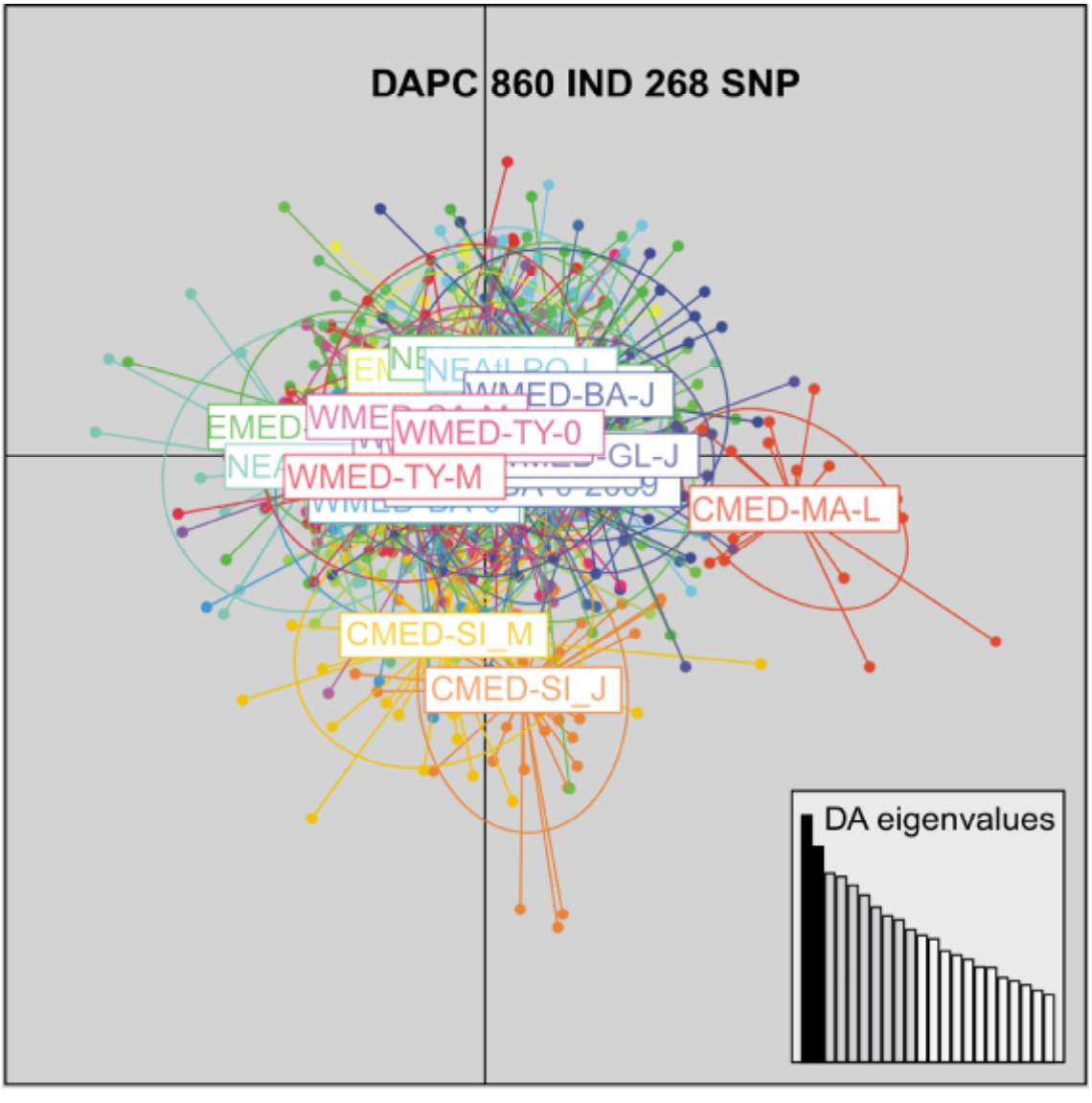


Figure 4.14a: DAPC of the individuals of 23 population samples with all 268 loci.

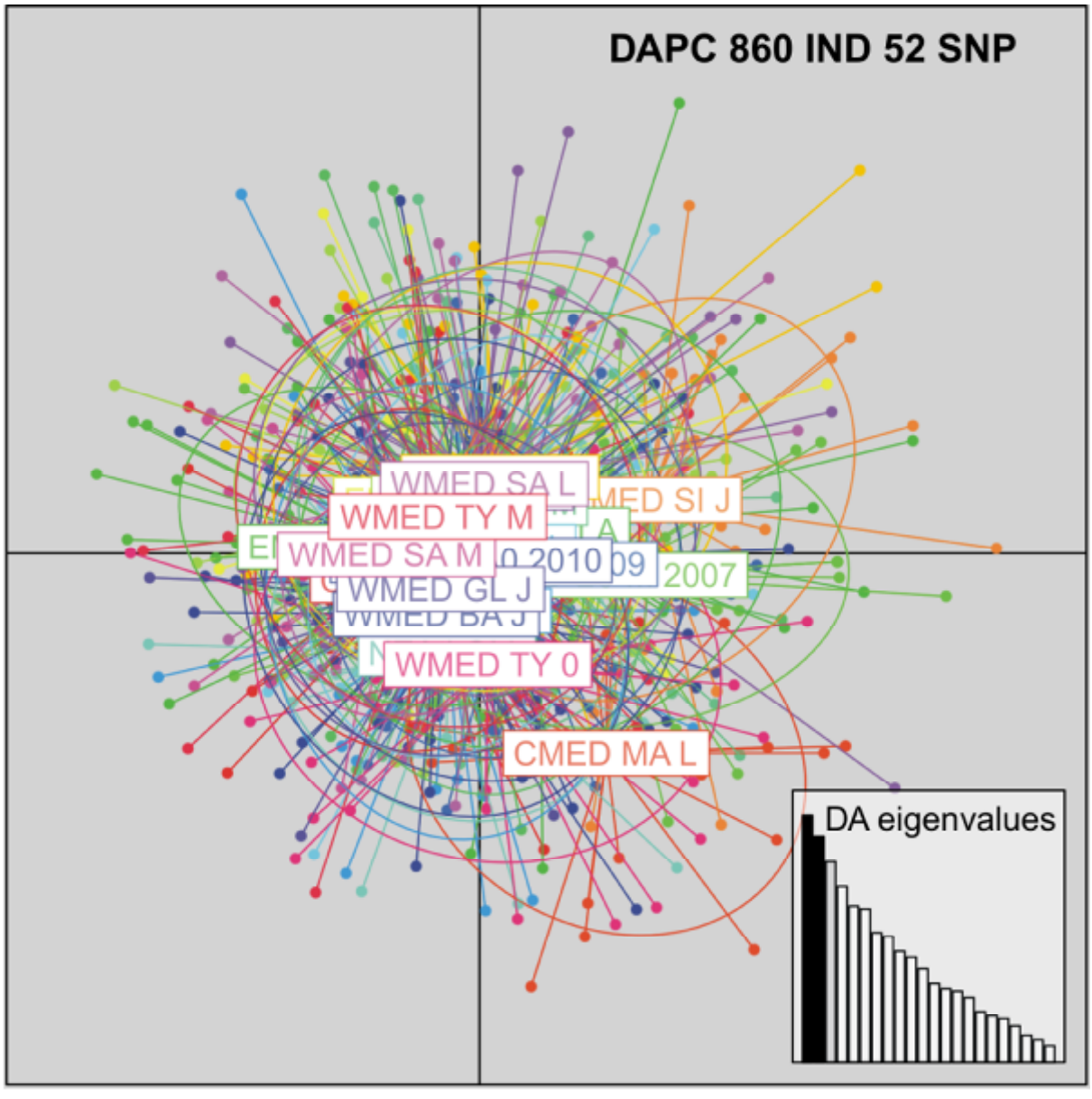


Figure 4.14b: DAPC of the individuals of 23 population samples with the restricted panel of 52 loci with F_{ST}>0.005.

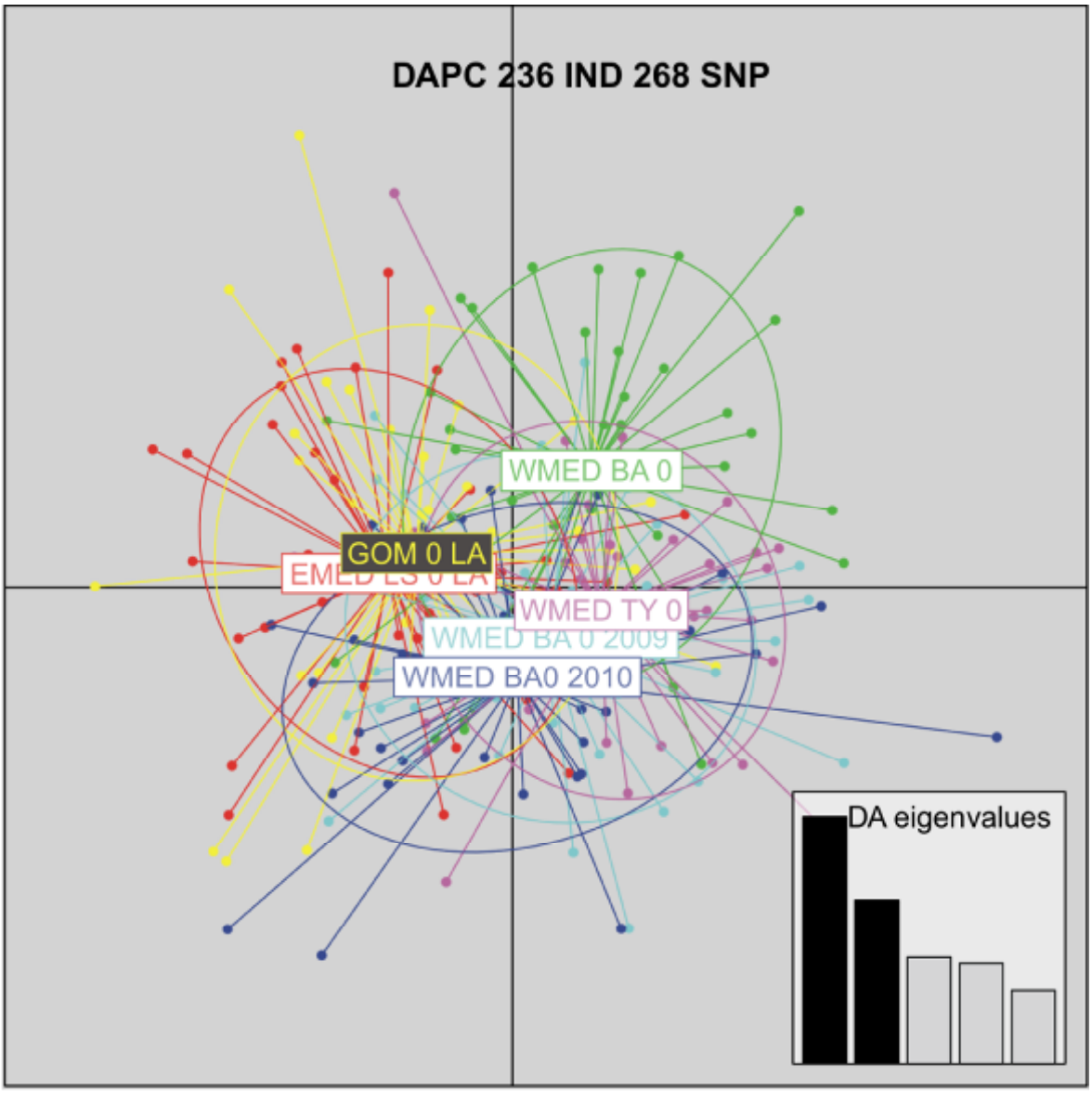


Figure 4.15a: DAPC on the six reference samples with all 268 loci.

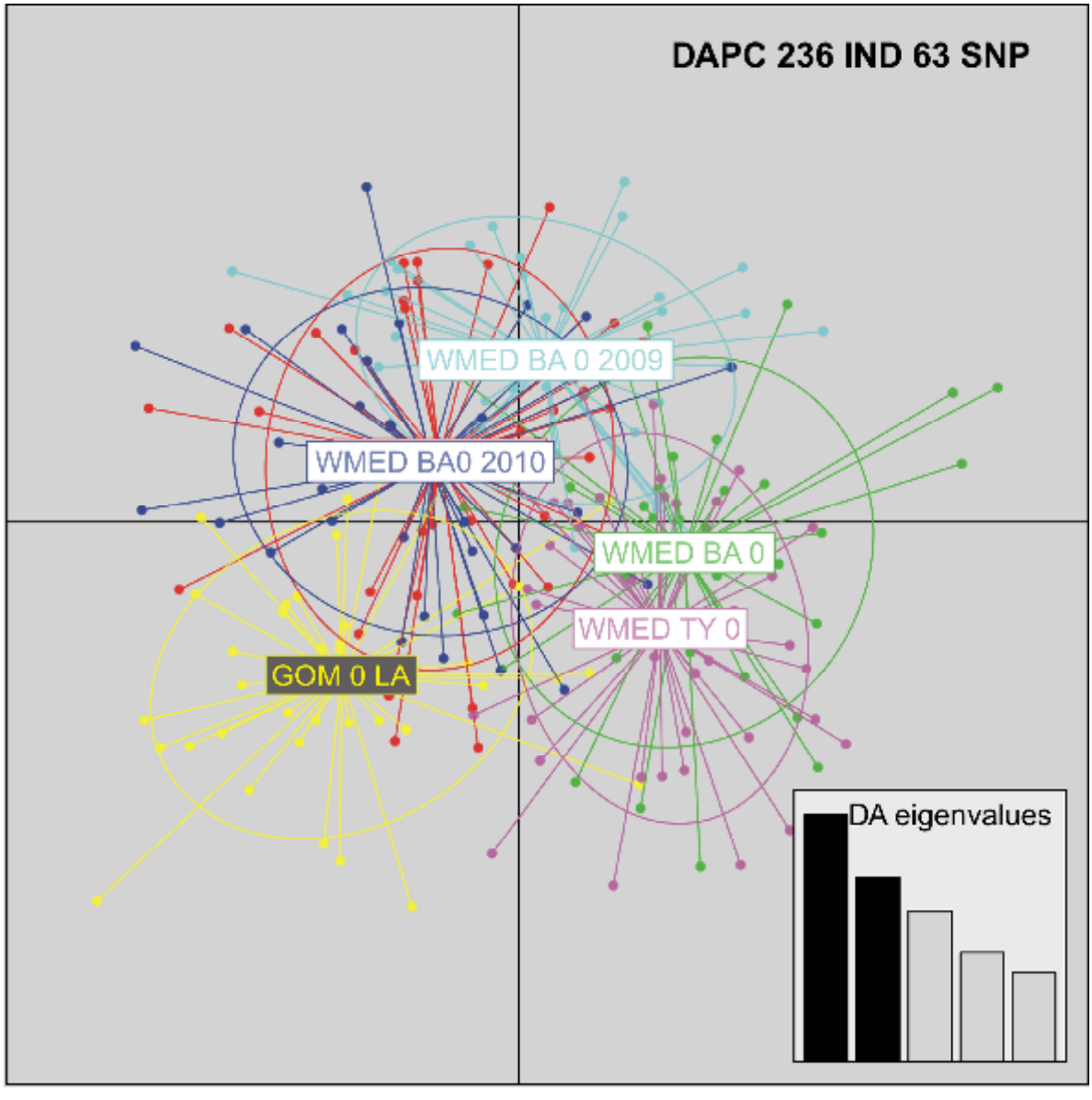


Figure 4.15b: DAPC on the six reference samples with the restricted panel of 63 loci with F_{ST}*>0.005.*

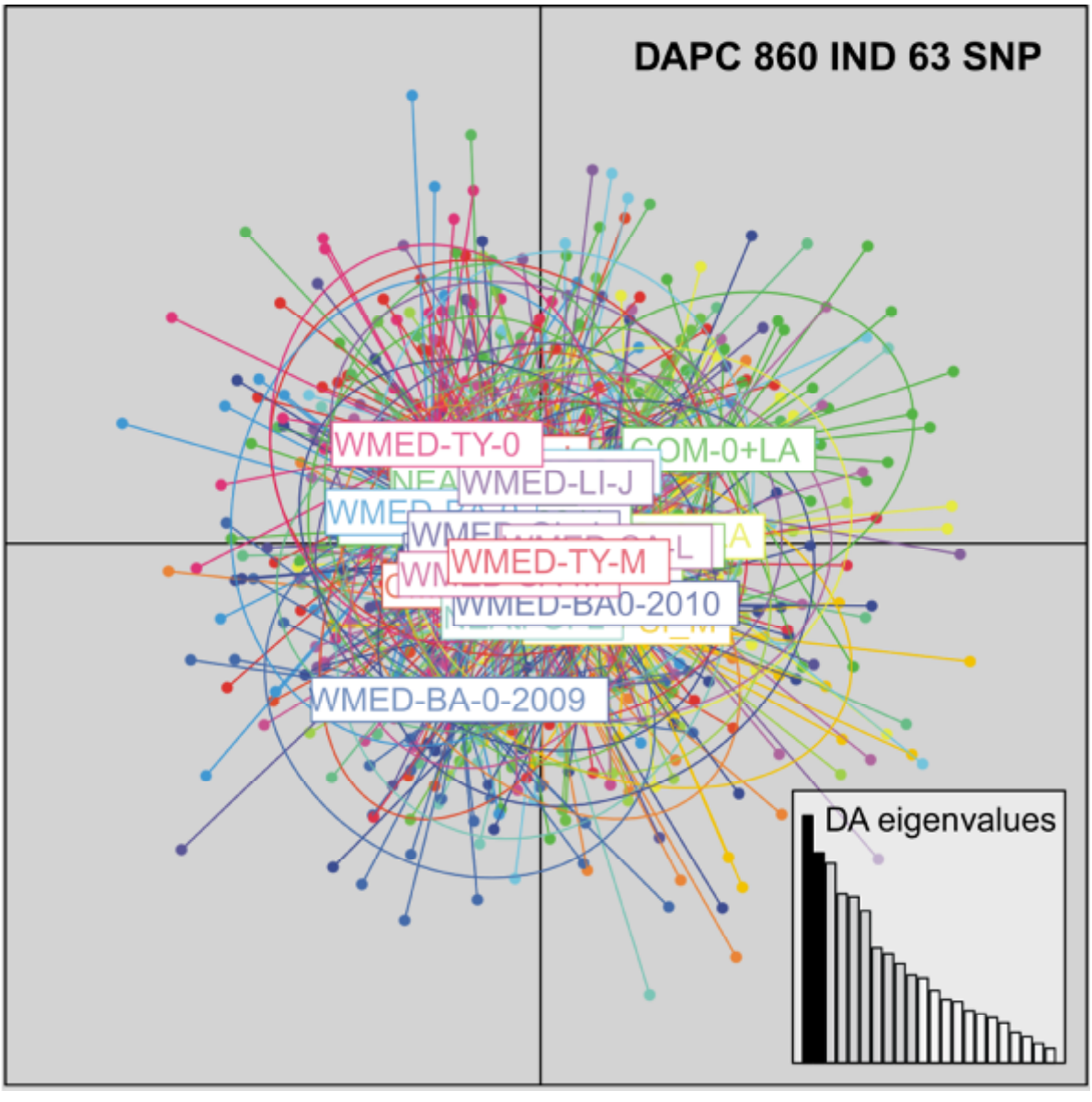


Figure 4.16: DAPC on all individuals of 23 populations with the restricted panel of 63 SNPs with F_{ST}>0.005.

Clustering and assignment analysis

Based on the results of DAPC analysis on the six reference samples (i.e. EMED-LS-LA+0, GOM-LA+0, WMED-BA-0, WMED-BA-0-2009, WMED-BA-0-2010, WMED-TY-0), **we performed a clusters analysis on these samples to assess the number of different groups maximizing the variation between populations**.

In the six reference samples, **the restricted panel of 63 SNP loci with F_{ST}>0.005 provided the best number of clusters with K= 2 and K=3**, based on results of the

Bayesian Information Criterion (BIC). However, when we assigned individual membership probabilities to the different groups based on the retained discriminant functions **no clear subdivision of the reference samples of putative spawning populations was observed, with a mixed contribution of each cluster to the six populations samples** (Figure 4.17).

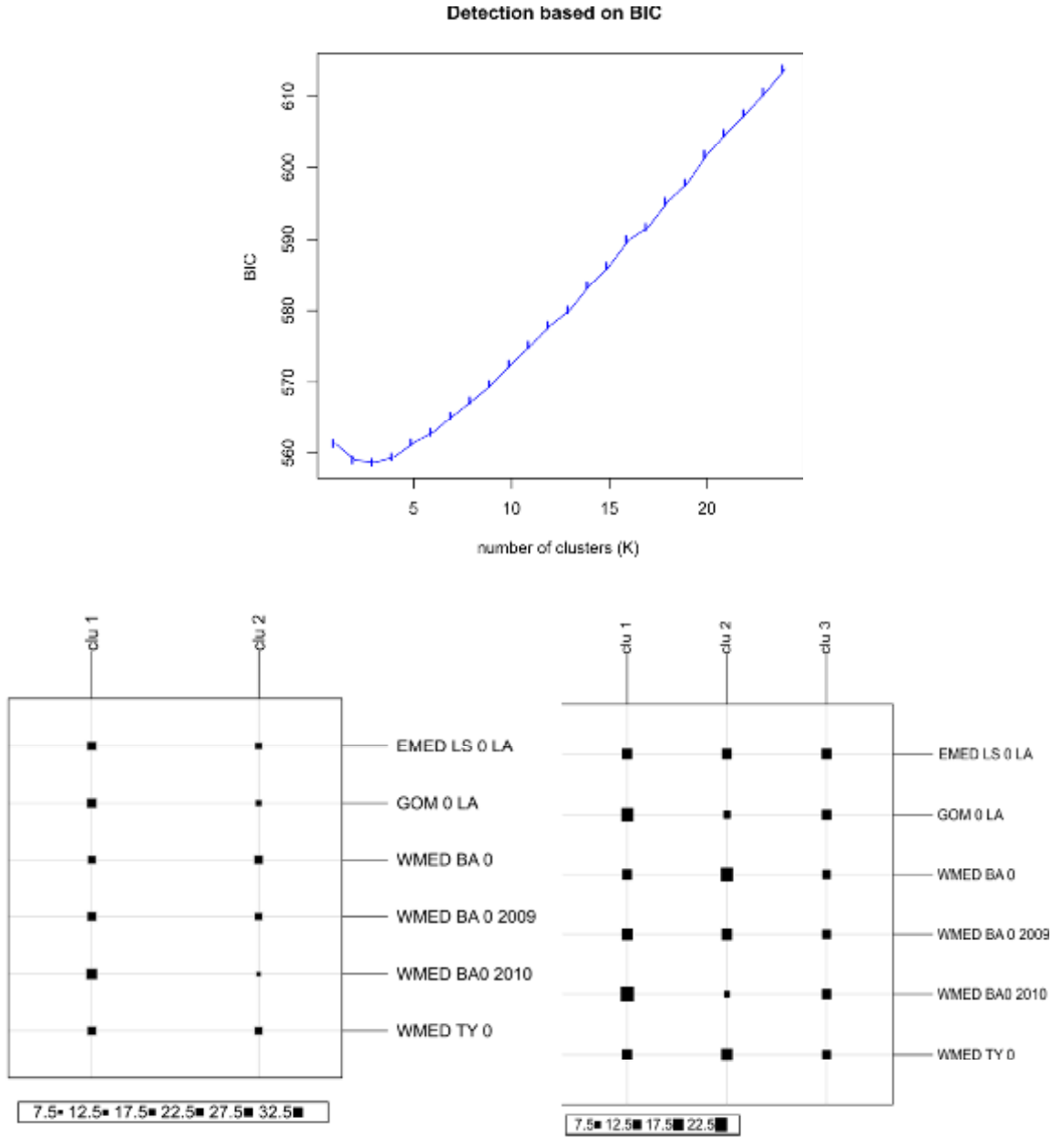
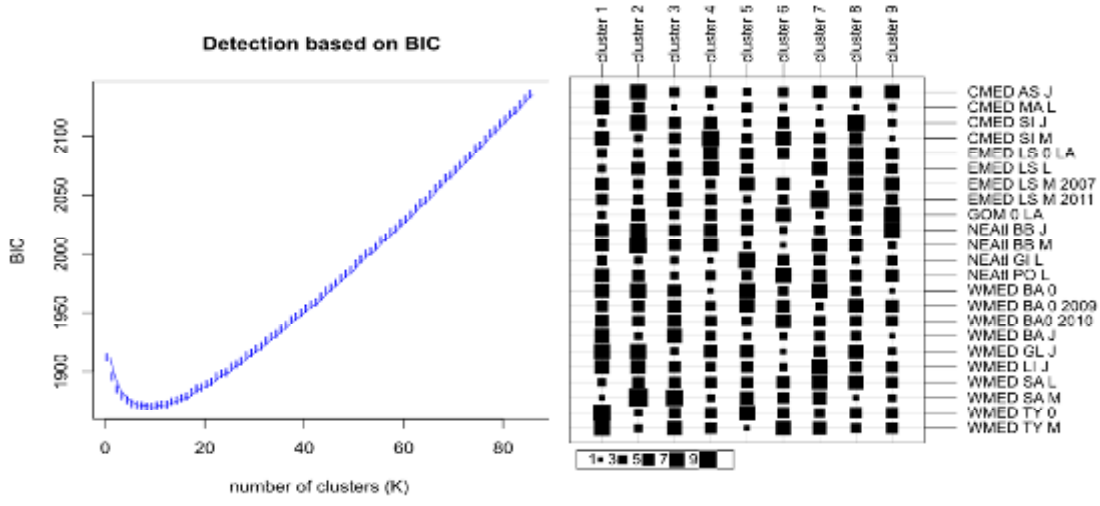


Figure 4.17: DAPC cluster analysis of the six reference samples with the restricted panel of 63 SNP with FST>0.005. Above panel: BIC values; below-left panel: individual membership for each cluster with K=2; below-right panel: individual membership for each cluster with K=3.

Similarly, **the lack of clustering was also reflected in the analysis of DAPC clusters with the complete dataset of the 23 population samples, with both restricted panels of loci with F_{ST}>0.005** (Figure 4.18). A fuzziness of clusters was obtained where no indication of potentially distinct biological populations can be retrieved, even if the 63 SNP dataset identified a lower numbers of groups maximizing the variation between populations.

A1) A2)



B1) B2)

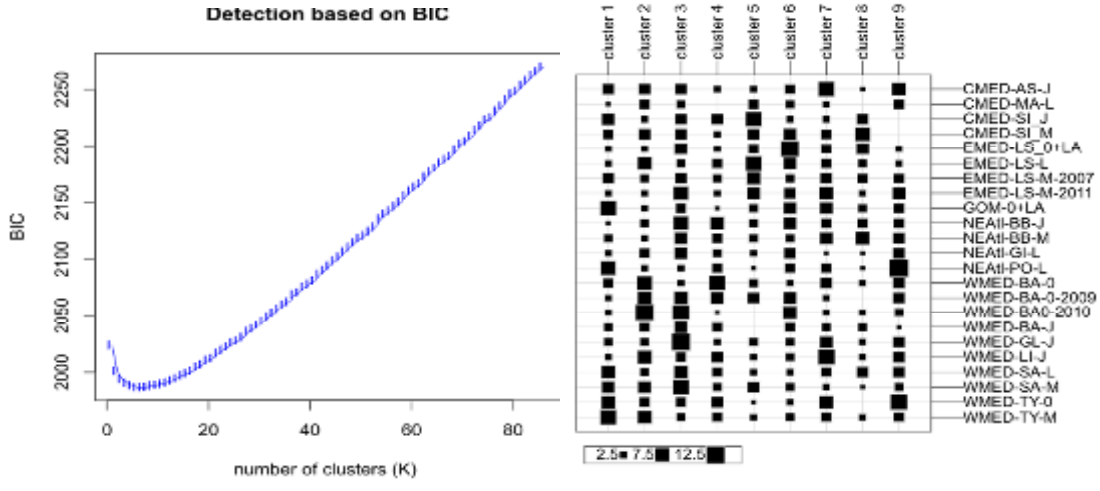
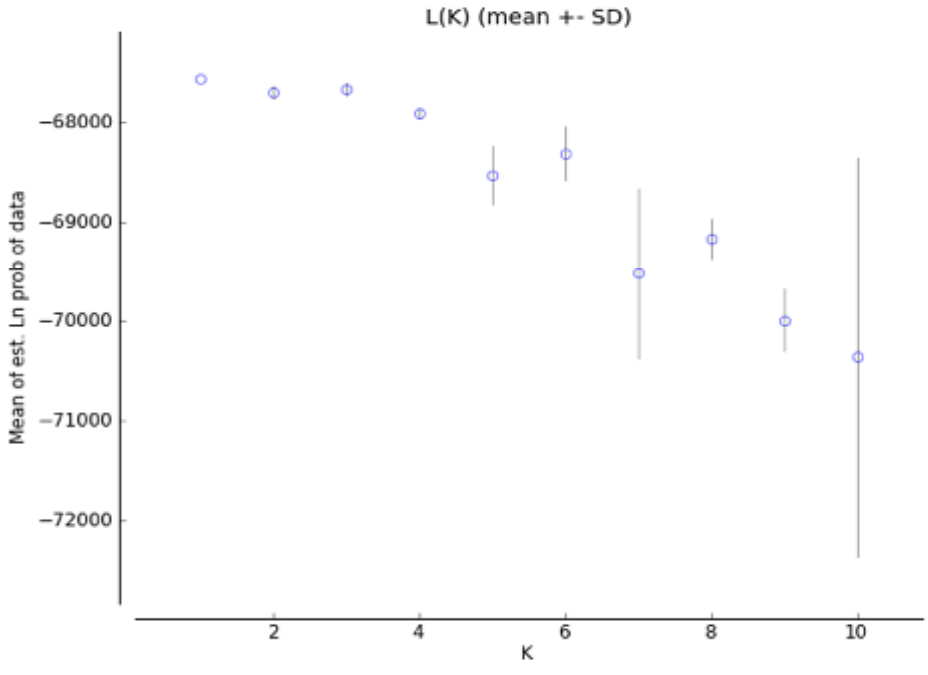


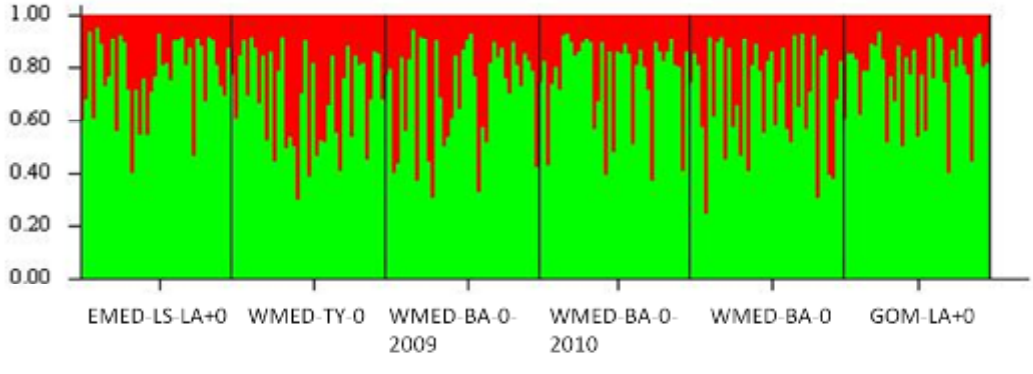
Figure 4.18: DAPC cluster analysis of the complete dataset of 23 population samples with A) the restricted panel of 52 SNP loci with F_{ST}>0.005; B) the restricted panel of 63 SNP loci with F_{ST}>0.005. 1) BIC values; 2) individual membership for each cluster.

The cluster analyses were also performed with the Bayesian approach implemented in the STRUCTURE software. We assessed the clustering of the six reference samples with two separate analyses performed, using the entire panel of 268 loci and the restricted panel of 63 SNP loci with F_{ST}>0.005. **No clustering of individuals of the six reference samples was detected with the entire panel of 268 loci** (Figure

4.19a,b). **Even if the clustering likelihood increases with the reduced dataset** of 63 SNP loci with F_{ST}>0.005**, no significant subdivision of individuals in the identified 3 clusters was observed** (Figure 4.19a,b).

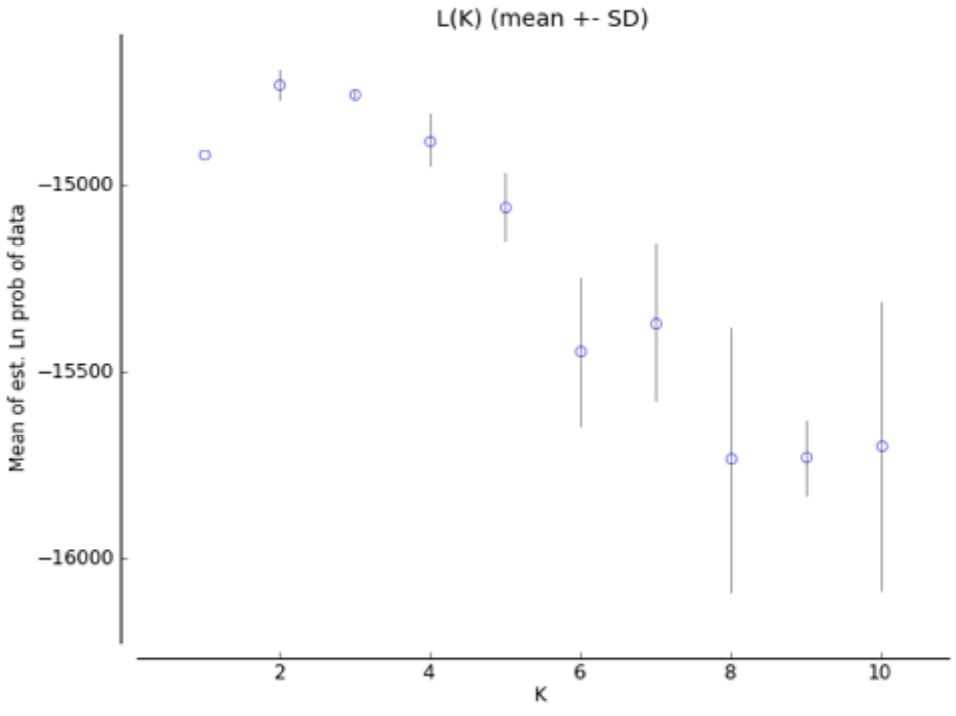


A)

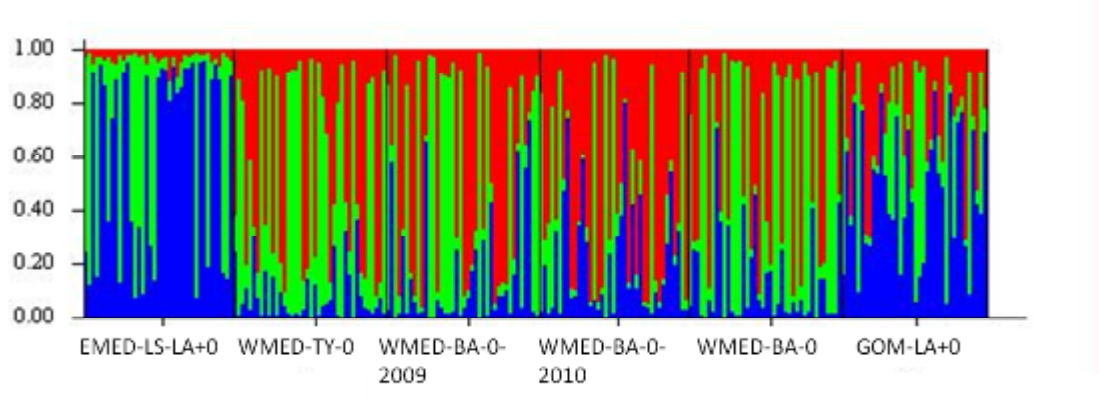


B)

Figure 4.19a: Individual based clustering of the six reference samples based on the entire panel of 268 loci. The mean likelihood of the data for each K (panel A) and the individual membership to each cluster (K=2, panel B) are shown.



A)



B)

Figure 4.19b: Individual based clustering of the six reference samples based on the restricted panel of 63 SNP loci with FST>0.005. The mean likelihood of the data for each K (panel A) and the individual membership to each cluster (panel D) are shown.

We tried to assign individuals to the reference samples with the software GeneClass2 by pooling the four population samples from WMED (even if they were significantly differentiated). **The assignment performed with the restricted panel of 63 SNP with F_{ST}>0.005 showed a prevalent contribution of the WMED reference**

samples to the feeding aggregates, except as expected in the reference populations from the Levantine Sea and Gulf of Mexico (Table 4.4 and Figure 4.20).

Table 4.4: Individual assignment analysis of the 23 population samples towards the geographical reference populations. In this analysis, the four reference samples from the WMED were pooled.

Population sample	N	EMED	WMED	GOM	% EMED	% WMED	% GOM
EMED-LS-LA+0	39	30	6	3	76.92	15.38	7.69
EMED-LS-M	40	13	21	6	32.50	52.50	15.00
EMED-LS-L	40	14	15	11	35.00	37.50	27.50
EMED-LS-M-2007	39	8	21	10	20.51	53.85	25.64
CMED-AS-J	40	14	17	9	35.00	42.50	22.50
CMED-MA-L	21	3	13	5	14.29	61.90	23.81
CMED-SI-J	36	4	23	9	11.11	63.89	25.00
CMED-SI-M	39	13	14	12	33.33	35.90	30.77
WMED-TY-0	40	4	32	4	10.00	80.00	10.00
WMED-TY-M	38	11	19	8	28.95	50.00	21.05
WMED-LI-J	40	11	22	7	27.50	55.00	17.50
WMED-SA-M	38	7	26	5	18.42	68.42	13.16
WMED-SA-L	39	12	16	11	30.77	41.03	28.21
WMED-GL-J	39	12	19	8	30.77	48.72	20.51
WMED-BA-0-2009	40	6	32	2	15.00	80.00	5.00
WMED-BA-0-2010	39	12	26	1	30.77	66.67	2.56
WMED-BA-0	40	7	28	5	17.50	70.00	12.50
WMED-BA-J	28	9	16	3	32.14	57.14	10.71
NEAtl-GI-L	29	10	13	6	34.48	44.83	20.69
NEAtl-BB-J	40	7	23	10	17.50	57.50	25.00
NEAtl-PO-L	39	11	19	9	28.21	48.72	23.08
NEAtl-BB-M	39	14	21	4	35.90	53.85	10.26
GOM-LA+0	38	3	4	31	7.89	10.53	81.58
Total	**860**	**235**	**446**	**179**	**27.33**	**51.86**	**20.81**

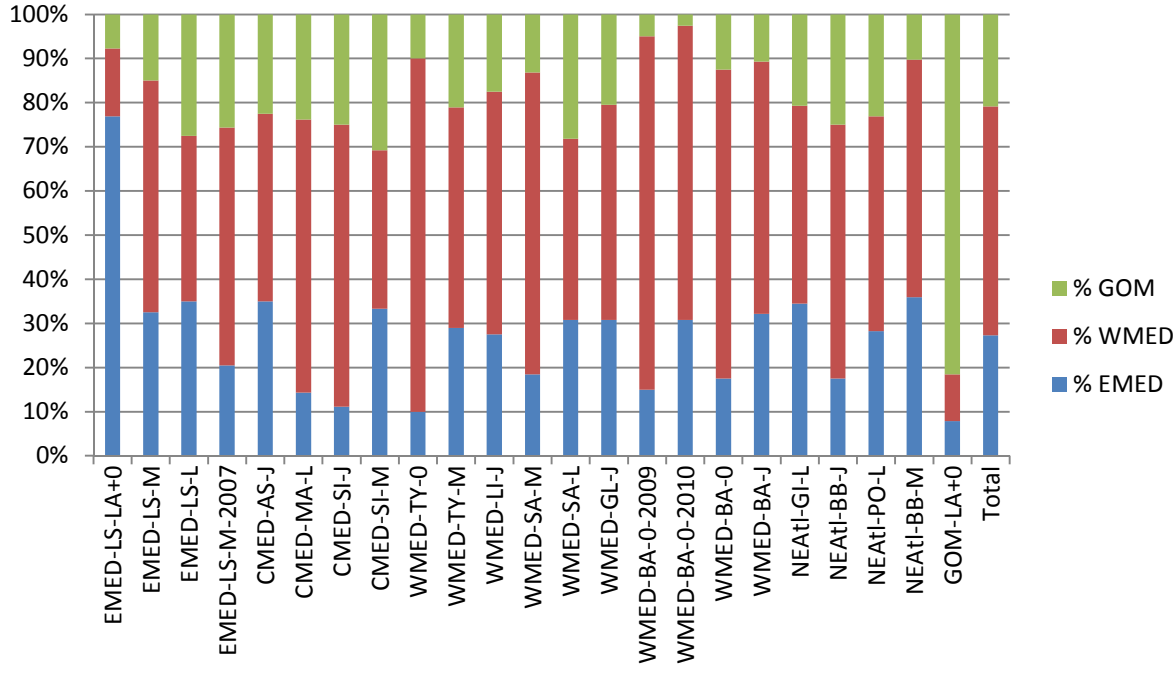


Figure 4.20: Barplot of percentage of individual assignment of the 23 population samples towards the geographical reference populations. In this analysis the four reference samples from the WMED were pooled.

4.2.6 Task 5 (NGS-RRSG) - Bioinformatic and population genomic data analysis

We firstly carried out the bioinformatic analysis of the RRSG data, allowing the simultaneous discovering and genotyping of a large number of genetic markers (we expected > 10000 SNPs) for connectivity/traceability purposes. To do so, we analysed raw sequencing data, summarized this into tags, whereafter a set of SNPs per individual could be genotyped. These SNP genotypes can then be analysed in specific downstream population genomic software.

For population genomic analyses, commonly-used descriptive statistics to estimate genetic diversity and differentiation among samples were performed, using population genetic statistical tests implemented in the up to date versions of population genetic software, improved in order to deal with large datasets and to increase the resolution of analysis (Genepop, Genetix, FSTAT v 2.9.3.2, Arlequin v

3.5). Outlier analyses were performed using various software (Lositan, Bayescan) to define the most discriminative markers (those with the highest Fst values). Multiple approaches were applied for population structure analysis, as this analysis represents a crucial step for genetic stock identification and for further analyses concerning individual assignment to population of origin. Bluefin tuna spatial population structure was explored using Bayesian MCMC clustering approaches implemented in the software Structure v2.3.3 and BAPS, and PCA and MDS methods using specific packages (*Adegenet*) of the R software, in order to discriminate population units.

Sequencing and bioinformatic analyses

A total of 192 GBS libraries were multiplexed and sequenced in a 100-cycle Paired End run on an IlluminaHiseq 2000 at KU Leuven-Biogenomics. Basic statistics were reported in Table 4.5.

Table 4.5: Basic statistics of the 192 GBS libraries of bluefin tuna.

Measure	**Value /1 library**	**Value /2 library**
Filename	GBS_tuna_R1.pf.fastq	GBS_tuna_R2.pf.fastq
File type	Conventional base calls	Conventional base calls
Encoding	Sanger / Illumina 1.9	Sanger / Illumina 1.9
Total Sequences	180222515	180222515
Filtered Sequences	0	0
Sequence length	101	101
%GC	48	47

Although a high number of SNPs was found initially with the first run, the DNA quality/quantity of all individuals was not always 100 % optimal, leading to a lower representation of 50 % of the analyzed individuals. As such, **not all populations were represented equally to allow a population genetic analysis of > 10000 SNPs** (only 6,000 SNPs were available for 70% of all individuals per population). Hence, we

decided to perform **a second sequencing run to allow a higher number of SNPs to be analyzed**. The total number of sequencing reads per individual for the two runs are given in Figure 4.21.

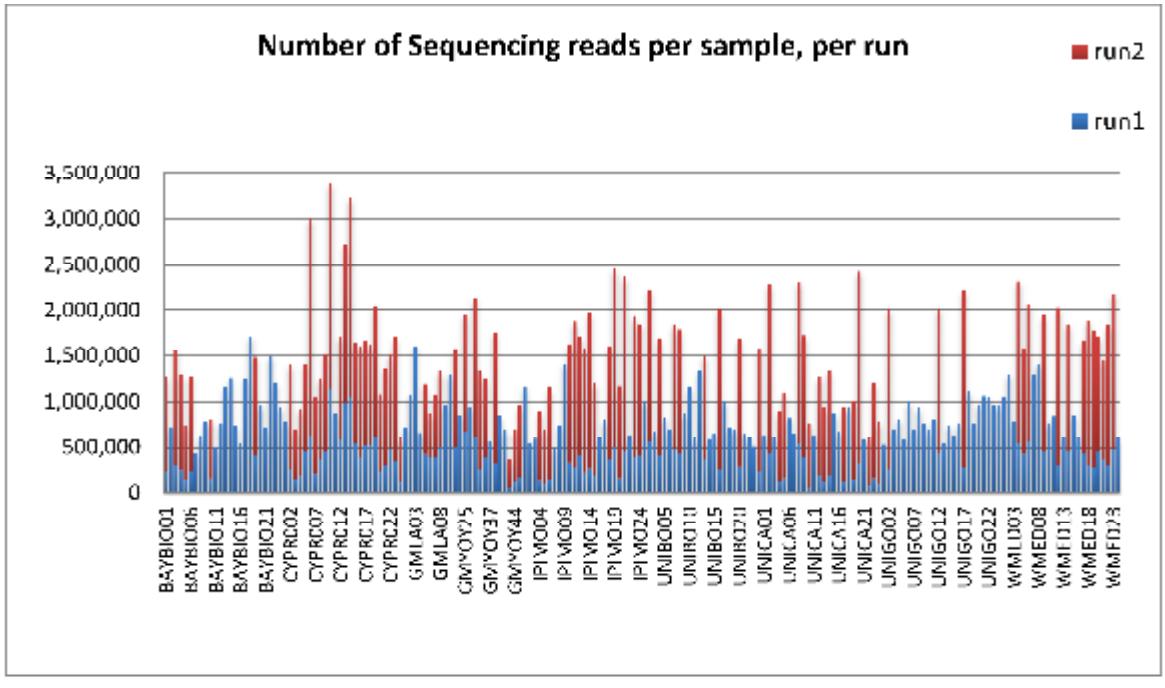


Figure 4.21: Sequencing reads per sample in the two RRSG runs carried out in the 192 bluefin tuna samples.

The "de novo" assembled "reference genome" of tuna consisted of **508757 contigs** with a total length of 1.5 GB. Compared to the Pacific bluefin tuna genome (~800MB[1]) there is about 50% redundancy in this assembled genome.

SNP discovery

SNPs were discovered by comparing tags that mapped to the same contig. Only dinucleotide SNPs (e.g. A/G but not A/G/T) were retained (default of Tassel pipeline). In total, **324433 of these dinucleotide SNPs** are discovered **in 88817 contigs**. **On average ~50% of these SNPs are called per individual in all groups, except for the CYPR population sample** (Figure 4.22). Several possible explanations for this phenomenon were investigated.

[1] http://www.intl-pag.org/19/abstracts/P05p_PAGXIX_647.html

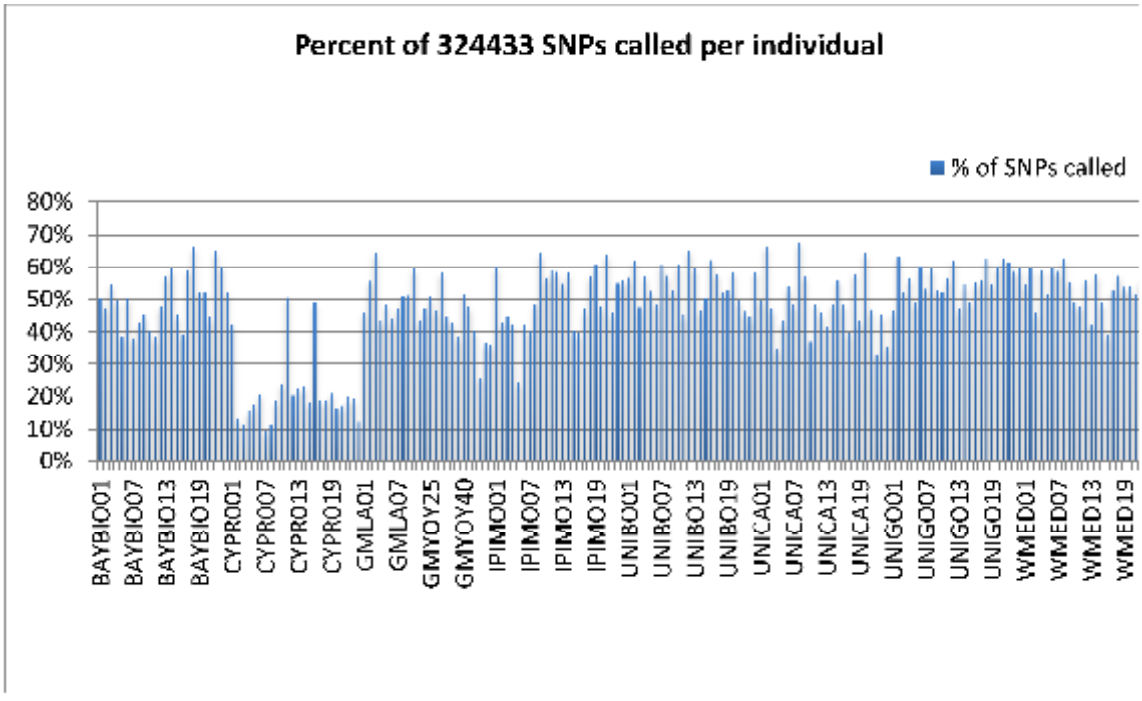


Figure 4.22: Percentage of SNPs called per individual. Note the low percentage exhibited by almost all CYPR individuals.

The CYPR sample

In the first run, the poor performance of the CYPR population sample (in terms of the number of SNPs found) was already noted. All CYPR individuals were therefore included in the second run. This, however, did not alleviate the problem, as can be seen in Figure 4.23. In general, there was an almost linear increase in the number of SNPs called with the number of individuals sequenced. This linear increase can also be observed for the CYPR population sample, although at a much slower rate. It can therefore be concluded that the number of sequencing reads for CYPR was not the problem. Only three individuals of the CYPR population sample fell into the normal distribution along with the other populations. It therefore could be expected that the CYPR individuals were genetically very dissimilar to those of the other populations (except for these 3 individuals). Sequence quality of the CYPR individuals were checked and this did not seem to be the main issue as QC metrics were comparable

to individuals of other populations. Ultimately, after a barcoding exercise, it was found that **21 of the 24 CYPR larvae were not bluefin tuna**.

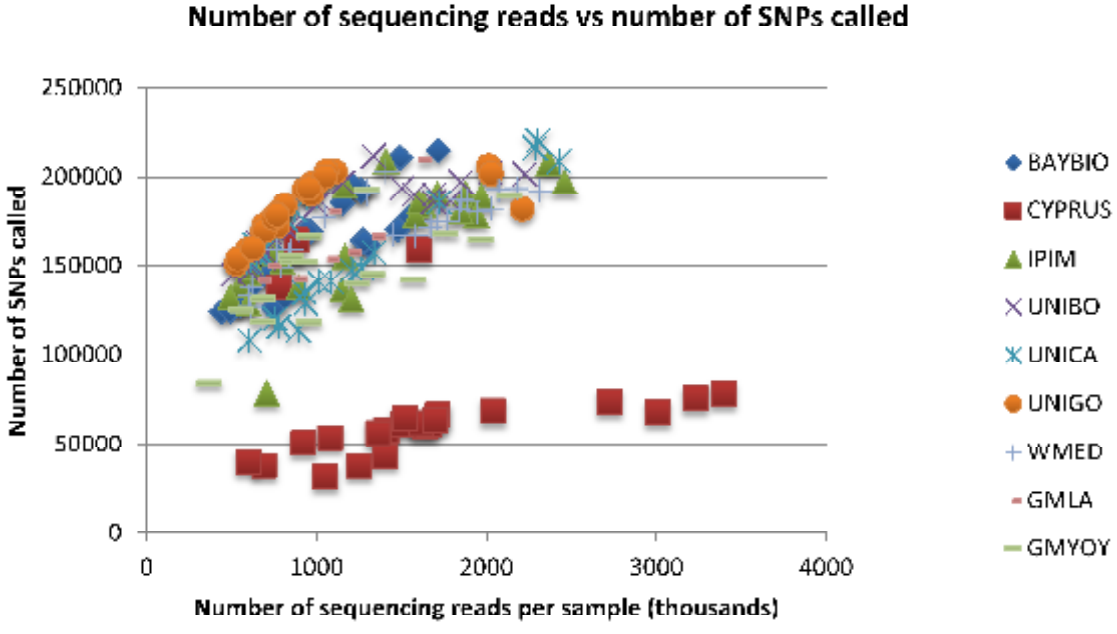


Figure 4.23 Plot illustrating the number of sequencing reads vs the number of SNPs in the 8 population samples. In this plot, the population sample was subdivided according to size classes (Larvae: GMLA; Age0: GMYOY).

SNP filtering and format conversion

Based on this issue, **SNPs were filtered according to the representation of a minimum number of individuals per population**. Only if a minimum of **70 % of the individuals per population was called, the SNP was selected**. This dataset resulted in **6857 usable SNPs** for intra- and interspecies comparisons. In addition, due to the low number of SNPs found in the CYPR population, another selection was carried out in which no presence of any CYPR individuals was required. This dataset resulted in **27316 usable SNPs** for intraspecies/interpopulation comparisons. The results of these two selections are reported in the Table 4.6.

Table 4.6: SNPs filtered from the initial RRSG SNP panel based on the two selection criteria.

Selection 1 (Criterion: >70% per population)		**Selection 2 (Criterion: >70% per population without CYPR)**	
Number of SNPs	6857	Number of SNPs	27316
Number of contigs	1725	Number of contigs	6844
SNPs/contig	3.98	SNPs/contig	3.99

Both datasets for selection 1 and selection 2 were formatted in Genepop files to perform the downstream population genomic analyses.

The population genomic analyses of the **27316 SNPs** can be subdivided into the following: Genotype Quality Check (QC), Genetic diversity and Hardy-Weinberg equilibrium, Genetic differentiation, Outlier analysis, Multivariate and Cluster analysis, and finally Assignment and traceability analyses.

Genotype Quality Check

We first explored the **number of alleles** per population and the **differences between expected (Hexp) and observed (Hobs) heterozygosities** (Figure 4.24). Heterozygosities were estimated over all populations and over all loci. These preliminary tests permitted to identify problematic populations and/or loci before the final analysis.

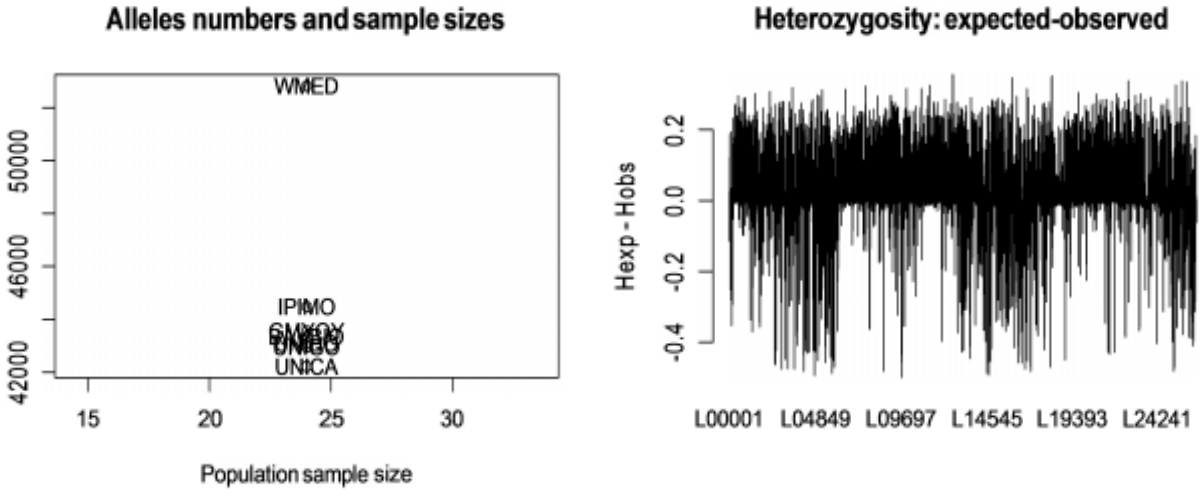


Figure 4.24: Descriptive genetic diversity per population (left plot) and per locus (right plot).

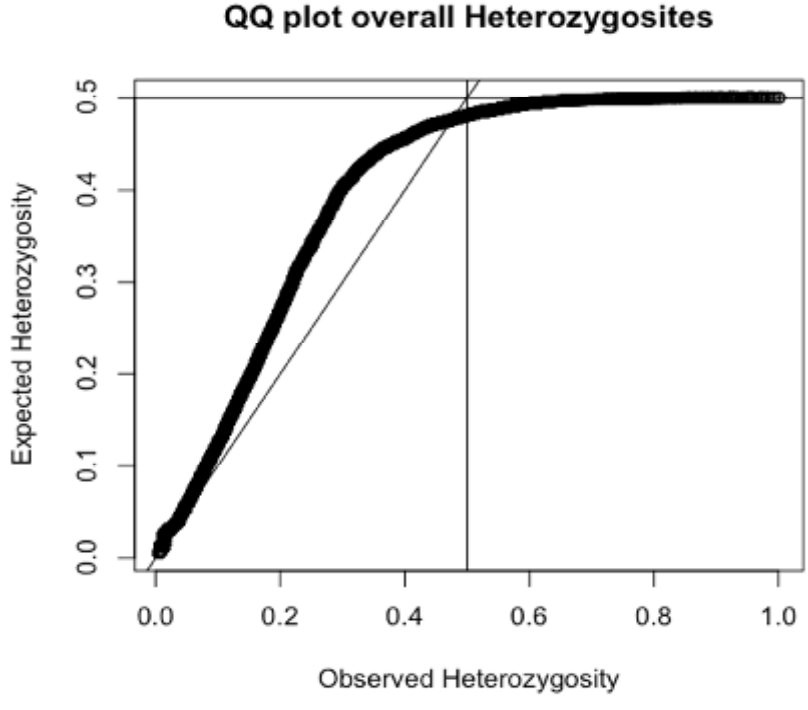


Figure 4.25: QQ-plot of observed vs expected heterozygosities.

From the Figure 4.24 (left plot) it appeared that **WMED showed a much larger number of alleles than the other populations**. Although SNPs are biallelic, this means that the WMED population (YOY) has a much higher proportion of polymorphic loci and, thus, genomic diversity, than the other population samples. Looking at the differences between observed and expected heterozygosities, we clearly see a significant number of loci showing as well lower as higher observed heterozygosities than expected (Figure 4.24, right plot). To further explore the levels of observed and expected heterozygosities, we plotted the quantiles of Hobs versus the quantiles of a theoretical distribution (here, Hexp) in a QQplot (Figure 4.25).

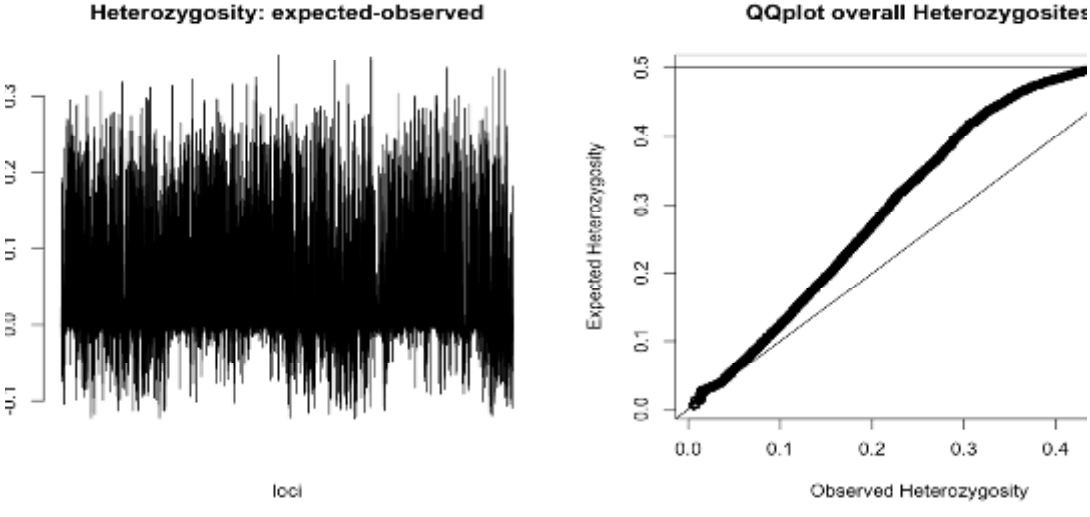


Figure 4.26: Descriptive statistics (left panel) and QQ-plot (right panel) of filtered loci (Hobs > 0.5) with observed vs expected heterozygosities.

Firstly, we observed that **many loci have Hobs larger than 0.5** (i.e. a theoretical maximum limit for bi-allelic loci as SNPs), which is possible under a scenario of selection, but **most likely due to paralogous genes**. The GBS method is based on short-read mapping on a reference "genome" or using a *de novo* assembly. Paralogous and duplicated regions can easily be mapped together when a good reference is lacking, yielding full heterozygous loci. This was easily picked up and **we therefore removed all loci with Hobs > 0.5** and plotted again Hexp-Hobs and QQ plot

(Figure 4.26). The new plots look better, more like the expected values, despite some lower Hobs levels than HWE. In total, **around 1000 loci were removed**.

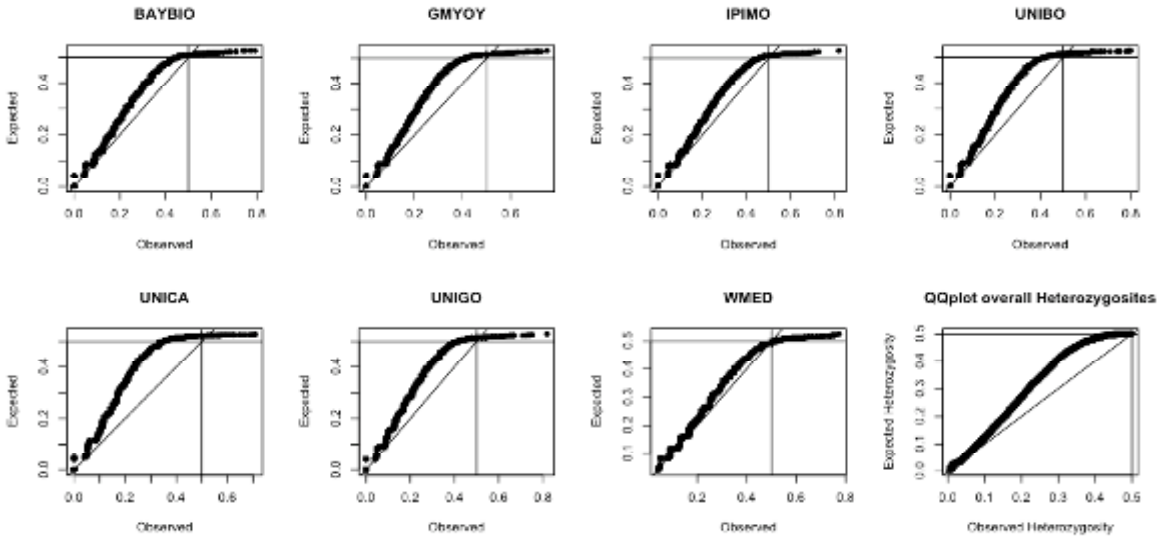


Figure 4.27: QQ-plot of observed vs expected heterozygosities for each of the seven population samples, and all combined.

To test the difference between observed and expected heterozygosities, we performed a simple t-test between Hexp and Hobs with highly significant results over all populations (t = 66.7403, df = 25355, p<2.2e-16; mean difference: 0.02). To overview the population specific estimates and per locus, a QQplot was performed per population sample (Figure 4.27). A slight bias towards Hexp (heterozygote deficiency) was detected, with the UNICA population sample having the strongest deviation. All the differences were highly significant. In addition, some loci still showed Hobs higher than 0.5. Indeed, we only removed those loci that showed overall Hobs > 0.5 (over all populations; see bottom right panel of Figure 4.27), to discard clear problematic loci. However, there were still loci with Hobs > 0.5 and this issue will be investigated more in depth at a later stage.

Genetic diversity and Hardy-Weinberg equilibrium

To further quantify the level of genetic homogeneity of the dataset, we calculated several basic statistics (Table 4.7). Here also, it appeared that WMED is much more diverse and in addition, it displayed low overall F_{IS} values.

Table 4.7: Descriptive statistics for the seven bluefin tuna population samples estimated over the 27316 SNP loci selected (excluding CYPR).

Sample	Number of alleles	Hobs	Hexp	F_{IS}	AR
BAYBIO	41233	0.0877	0.1071	0.1196	1.5752
GMYOY	41442	0.0822	0.1092	0.1585	1.5893
IPIMO	42364	0.0961	0.1158	0.1120	1.6237
UNIBO	40983	0.0787	0.1046	0.1653	1.5715
UNICA	40108	0.0668	0.1041	0.2420	1.5538
UNIGO	40839	0.0777	0.1041	0.1672	1.5625
WMED	50712	0.1422	0.1523	0.0278	1.9095
Overall	42526	0.0902	0.1139	0.2081	1.6265

Hobs: observed heterozygosity; Hexp: expected heterozygosity; F_{IS}: degree of deviation from Hardy-Weinberg disequilibrium; AR: allelic richness.

We analyzed the distribution of F_{IS} per population (Figure 4.28) to assess the proportion of SNP loci that deviated from the Hardy Weinberg Equilibrium (HWE). It appeared that **40% of all loci showed significant deviations from what is expected under HWE** (F_{IS}> 0). This has to be further investigated and would be due to a mixing of populations (i.e. the Walhund effect), some technical artifacts, or simply a violation of theoretical assumptions of HWE. This will be taken into account once the discriminatory loci will be selected in the following analyses.

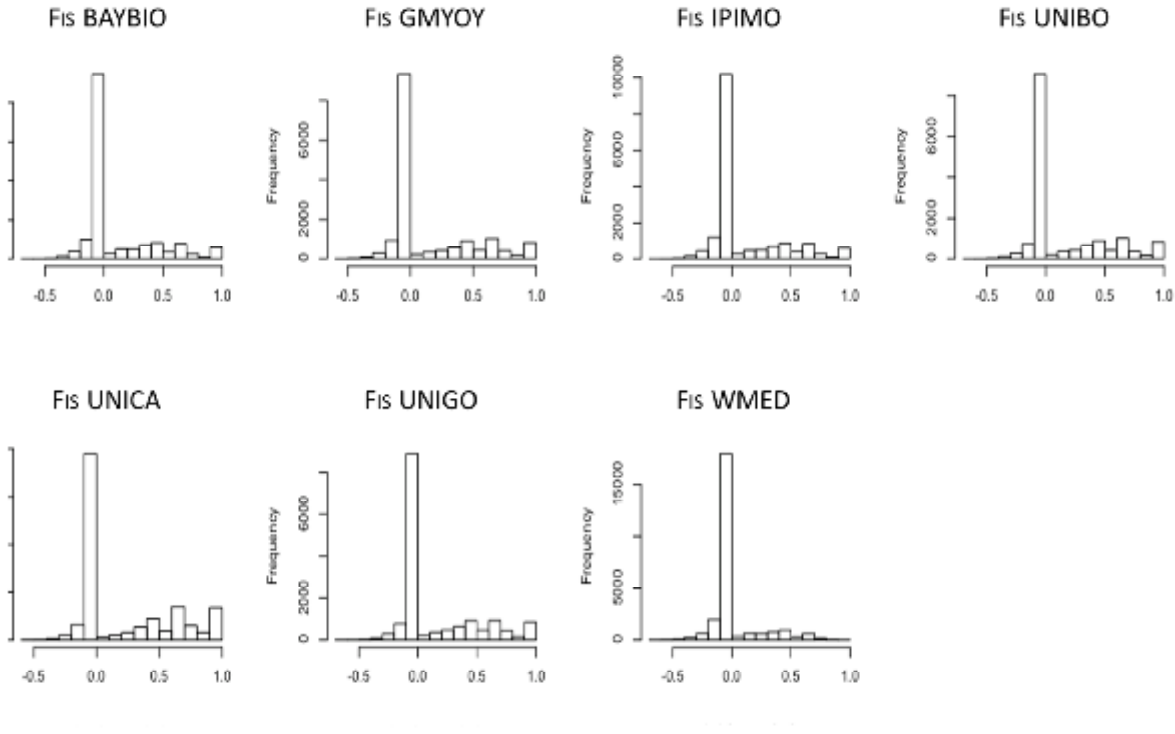


Figure 4.28: F_{IS} distribution for all loci per population in the 7 bluefin tuna population samples.

Genetic differentiation and multivariate analysis with 27316 SNP loci

Overall and pairwise **genetic differentiation** among population samples was estimated using Weir and Cockerham's (1984) estimator using the whole set of 27316 SNP loci. The genetic differentiation analysis **showed a very low overall value (F_{ST} = 0.0047, not significant).** However, this value is expected at the overall level, given the fact that feeding population samples have a potentially mixed composition and only WMED and GMYOY are potentially real spawning populations. It was expected that pairwise differentiation estimates should be more informative to detect genetic differentiation between populations; however, they resulted very low and not significant (F_{ST}< 1%; Table 4.8).

Table 4.8: Pairwise genetic differentiation (F_{ST}) among the seven populations over 27316 SNP loci.

Populations	BAYBIO	GMYOY	IPIMO	UNIBO	UNICA	UNIGO
GMYOY	0.0057					
IPIMO	0.0019	0.0045				
UNIBO	0.0025	0.0074	0.0038			
UNICA	0.0048	0.0059	0.0039	0.0064		
UNIGO	0.0018	0.0042	0.0022	0.0031	0.0034	
WMED	0.0051	0.0044	0.0027	0.0070	0.0076	0.0065

We performed a multivariate analysis (Multidimensional Scaling, MDS) based on the pairwise differentiation values to assess the genetic differentiation among samples. The MDS pattern was not entirely clear, with three feeding populations clustering in the middle and four samples (including the reference samples WMED and GMYOY) at the outer borders of the plot (Figure 4.29).

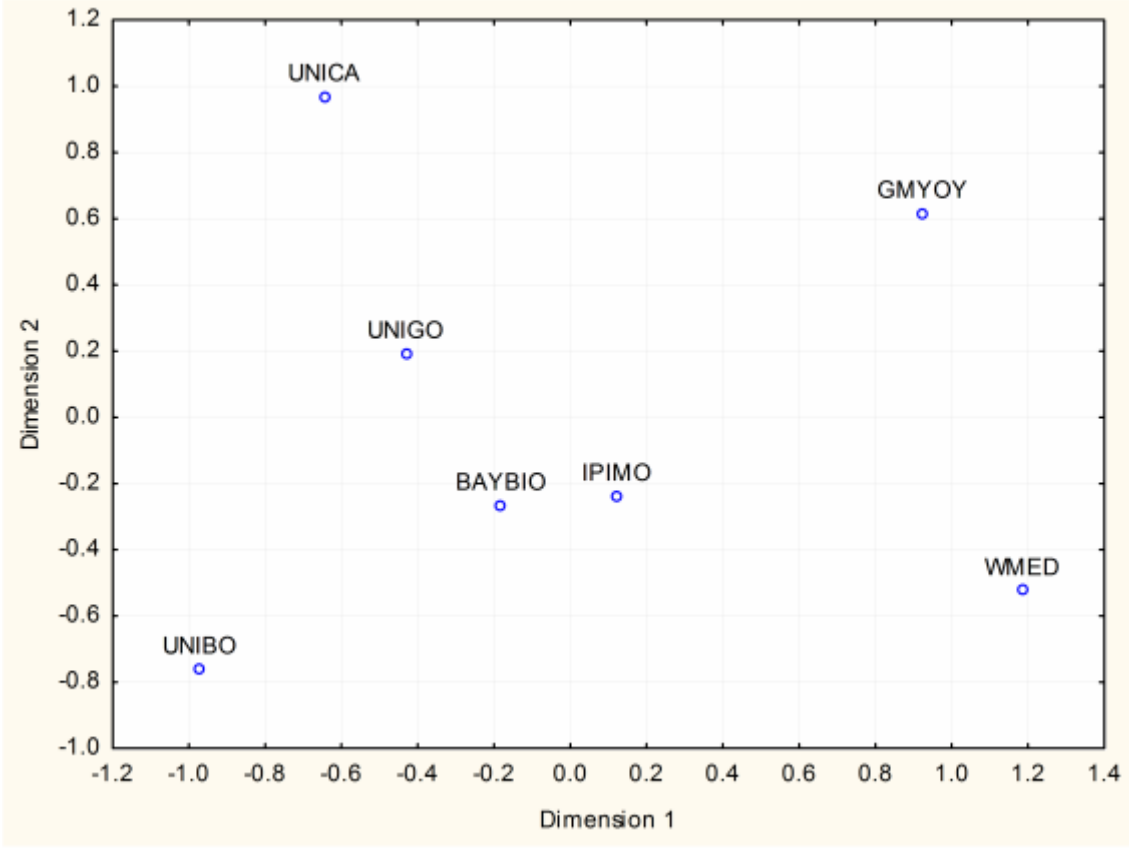


Figure 4.29: MDS plot of pairwise FST values among seven bluefin tuna population samples.

Because the low resolution obtained with all 27316 SNP loci for the genetic differentiation among samples, **we have further selected a sub set of highly-differentiating loci (outlier loci) that may yield better discrimination among bluefin tuna samples**.

Outlier loci analysis and F_{ST} ranking

To detect potential loci under positive selection, likely providing an increased power for traceability in a high gene flow system such as bluefin tuna, we performed an outlier analysis. This was first done on the full dataset and subsequently fine tuned for the selected dataset with Hobs < 0.5. Given that the only two potential spawning population samples are WMED and GMYOY, we **firstly calculated F_{ST} values and performed an outlier analysis between those two populations**.

We therefore analysed the WMED and GMYOYusing the LOSITAN software and we generated a list of neutral and outlier loci (directional and balancing selection). In total, **4198 loci showed p >0.95**, but using the FDR correction, **1683 loci remained informative as outliers** (Figure 4.30).

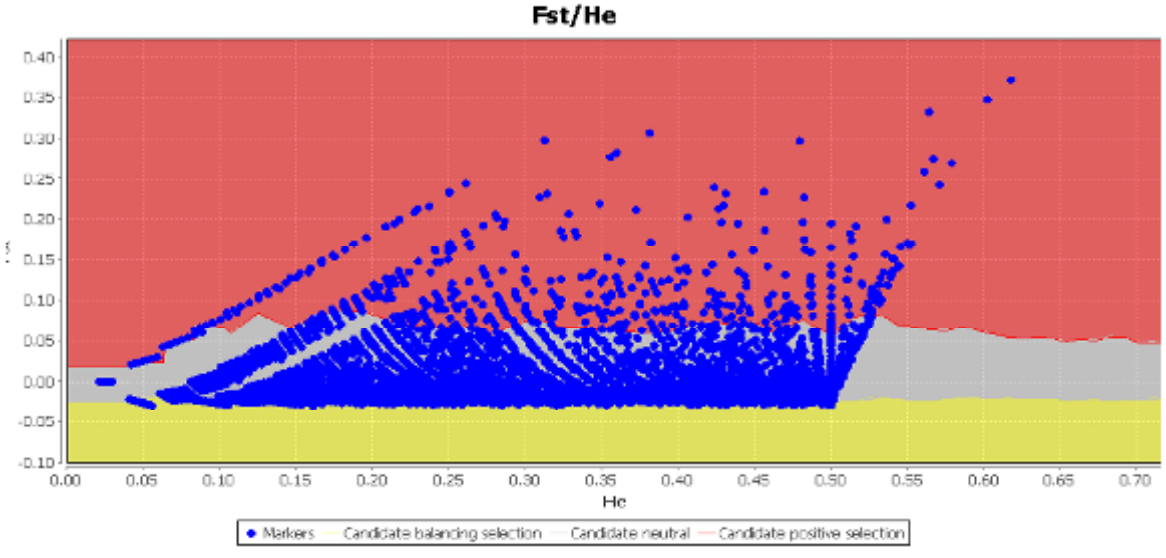


Figure 4.30: Results of the outlier analysis between WMED and GMYOY spawning populations.

Additionally, to reduce the number of loci for assignment analyses, we analyzed the distribution of single-locus F_{ST} values (mean F_{ST} = 0.4%) to overview the number of loci potentially under selection. Ranking all loci, we saw a significant number of loci with high F_{ST} values: **1934 loci with F_{ST}>5%** (which corresponded to 7% of the whole number of SNP loci) and **396 loci with F_{ST}>10%** (Figure 4.31). The highest observed genetic differentiation value between the two populations was **F_{ST} = 0.36**. Beside the full SNP panel, we also used these subpanels of highly-divergent SNP loci to perform multivariate, clustering and assignment analyses.

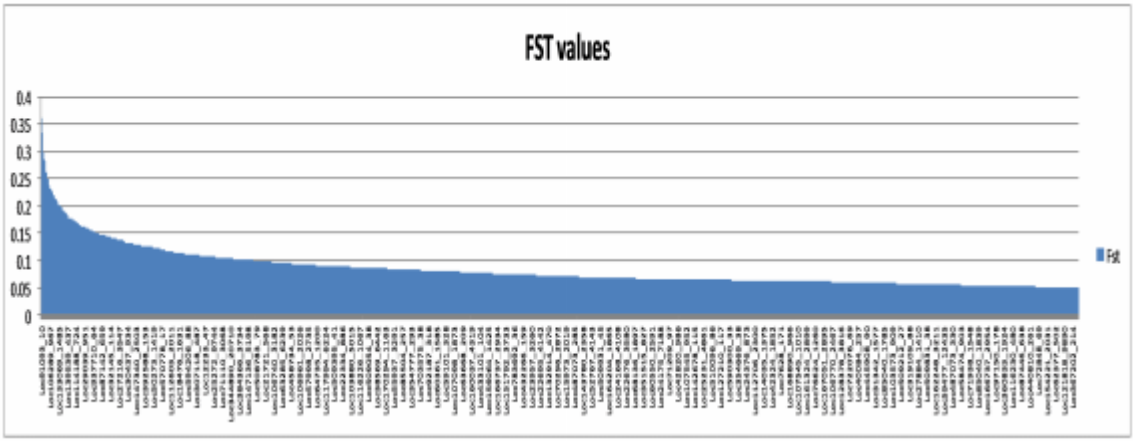


Figure 4.31: Distribution of ranked pairwise genetic differentiation values (F_{ST} > 5%) between WMED and GMYOY populations.

Multivariate and Cluster analysis using the full and outlier SNP panels

A Discriminant Analysis of Principal Components (DAPC, Jombart et al. 2010) was performed on multi-locus genotype data for all sample populations. This method transforms data using principal component analysis (PCA) to create uncorrelated variables for input into Discriminant Analysis (DA). **DA maximizes between-group variation and minimizes within-group variation for assessment of between-group variation.** DAPC is free of assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium and provides graphical representation of divergence among populations. DAPC was implemented using the *adegenet* package (Jombart 2008) in R version 2.14.1 (R Development Core Team 2010). **DAPC allowed the search for the most likely number of clusters/groups in the dataset.** This can be achieved using k-means, a clustering algorithm which finds a given number (say, k) of groups maximizing the variation between groups, B(X). To identify the optimal number of clusters, k-means is run sequentially with increasing values of k, and di□erent clustering solutions are compared using the Bayesian Information Criterion (BIC). Ideally, the optimal clustering solution should correspond to the lowest BIC. In practice, the 'best' BIC is often indicated by an elbow in the curve of BIC values as a function of k. Moreover, being based on the Discriminant Analysis, DAPC also provides membership probabilities of each individual for the di□erent groups based on the retained discriminant functions. While these are di□erent from the admixture coe□cients of softwares like STRUCTURE (see later), they can still be interpreted as

proximities of individuals to the di□erent clusters. Membership probabilities also provide indications of how clear-cut genetic clusters are. Loose clusters will result in fairly flat distributions of membership probabilities of individuals across clusters, pointing to possible admixture.

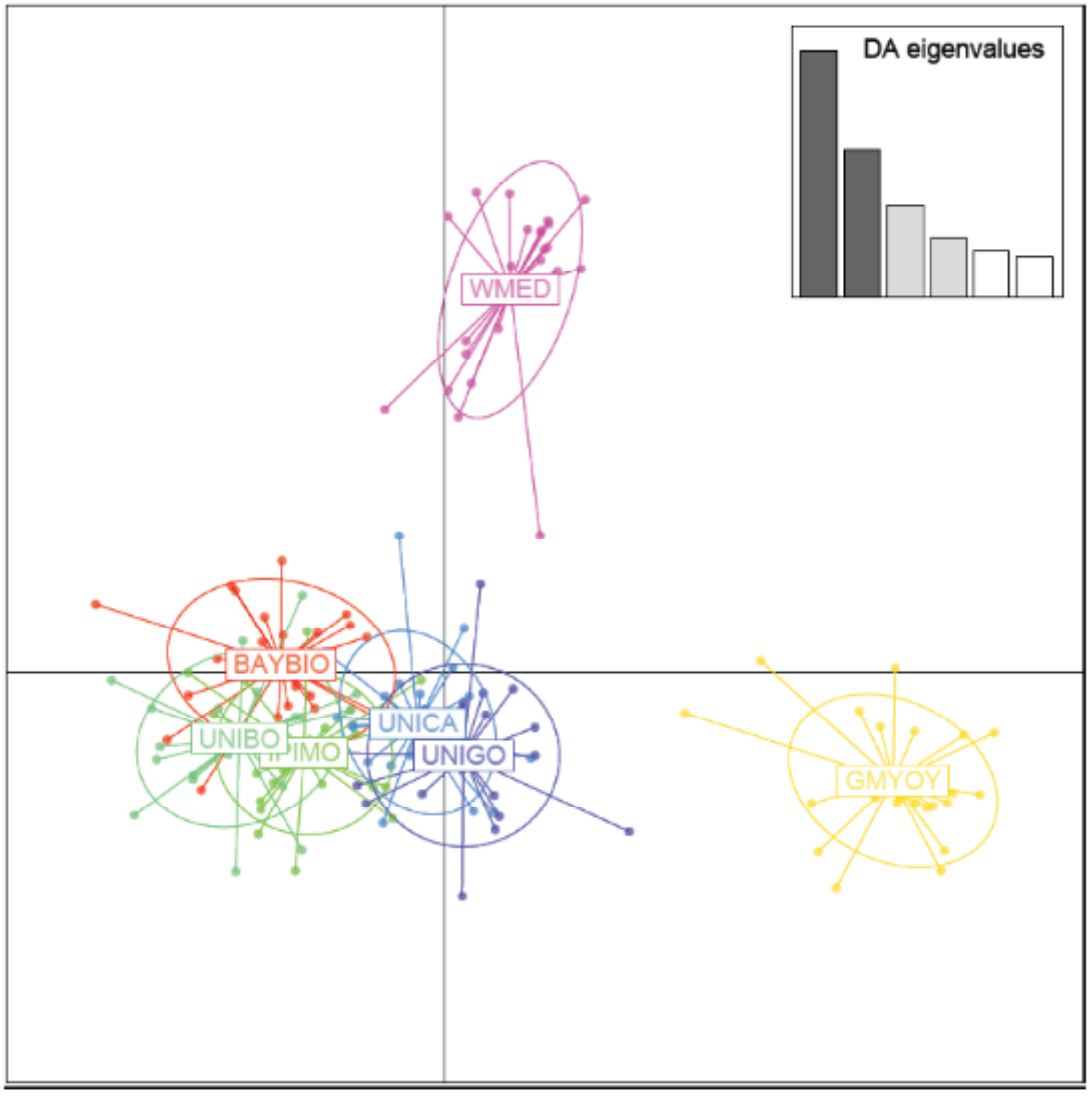


Figure 4.32: DAPC analysis of the seven bluefin tuna population samples with all 27316 loci.

Discriminant Analysis of Principal Components (DAPC) showed that a very high percentage of the variability in the data could be explained by the first two axes (Figure 4.32). In this figure, and on the contrary of basic F_{ST} values, **multilocus genotypes based on all loci showed a clear degree of structuring**. We clearly observed

three clusters of populations, namely the **WMED, the GMYOY and the remaining populations**. Given WMED and GMYOY are expected to represent pure populations, without the possibility to have migrated towards/being mixed with other populations, this results can be cautiously interpreted as showing two distinct spawning aggregations. **The feeding population samples formed one similar cluster**.

When we performed **DAPC analyses with the highly-divergent SNP panels** (e.g. F_{ST}>5 %, 1,936 SNPs; F_{ST}>10%, 391 SNPs), **we obtained similar patterns with an increased differentiation between WMED and GMYOY** (Figure 4.33). However, the feeding population samples even more overlapped in the middle of the graphs.

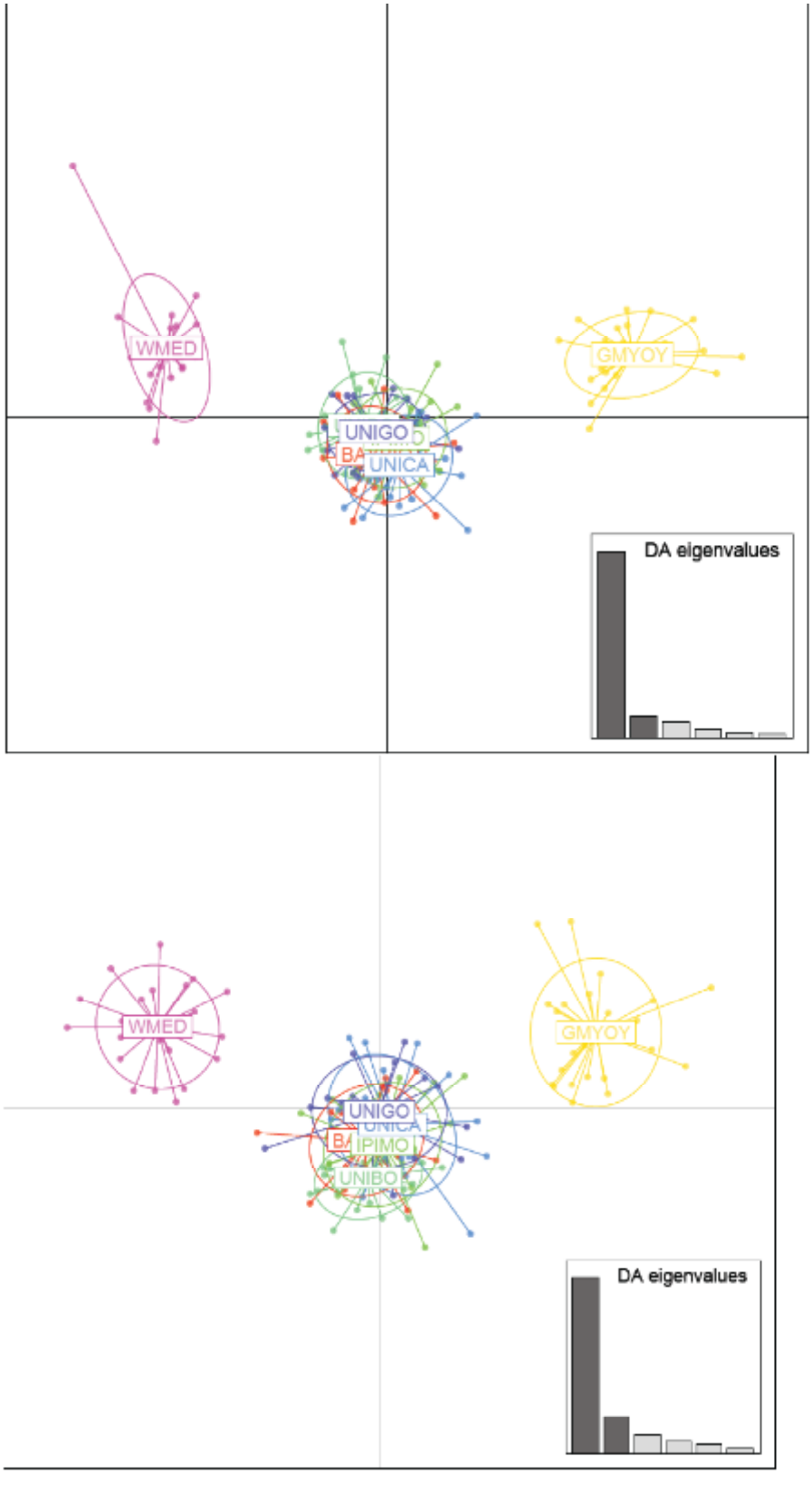
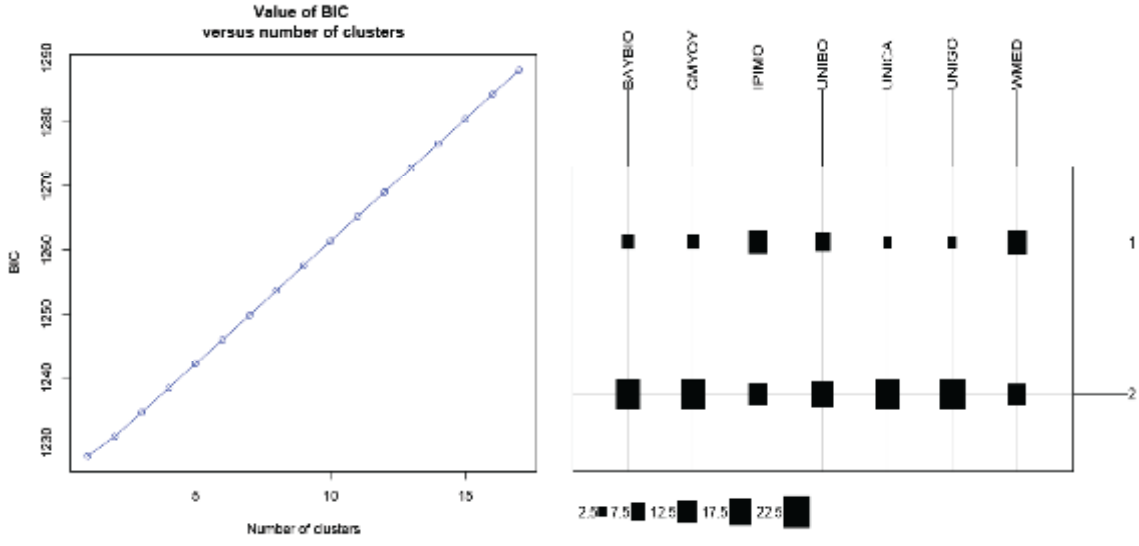


Figure 4.33: DAPC analysis of the seven analyzed populations with SNP panels with F_{ST}>5% (upper panel) and F_{ST}>10% (lower panel).

The explained variation was very high for both selected datasets. For the F_{ST}>10% and F_{ST}>5% SNP panels, the first two axes explained 85 % and 88% of the genomic variation, respectively.

Looking for the number of clusters detected by the full and selected SNP panels (we tested only the F_{ST}>10% panel), we found that the full panel showed no clear discrimination (low F_{ST}), while the selected F_{ST}>10% panel identified two clusters since the BIC drop at K = 2 (Figure 4.34). **Individual membership proportions, showed WMED and GMYOY as separate clusters, with most feeding populations showing a somewhat higher membership to the GMYOY population**. Further individual analyses (cluster analysis using Structure) are required to allow the definition of membership to the defined clusters.

A



B

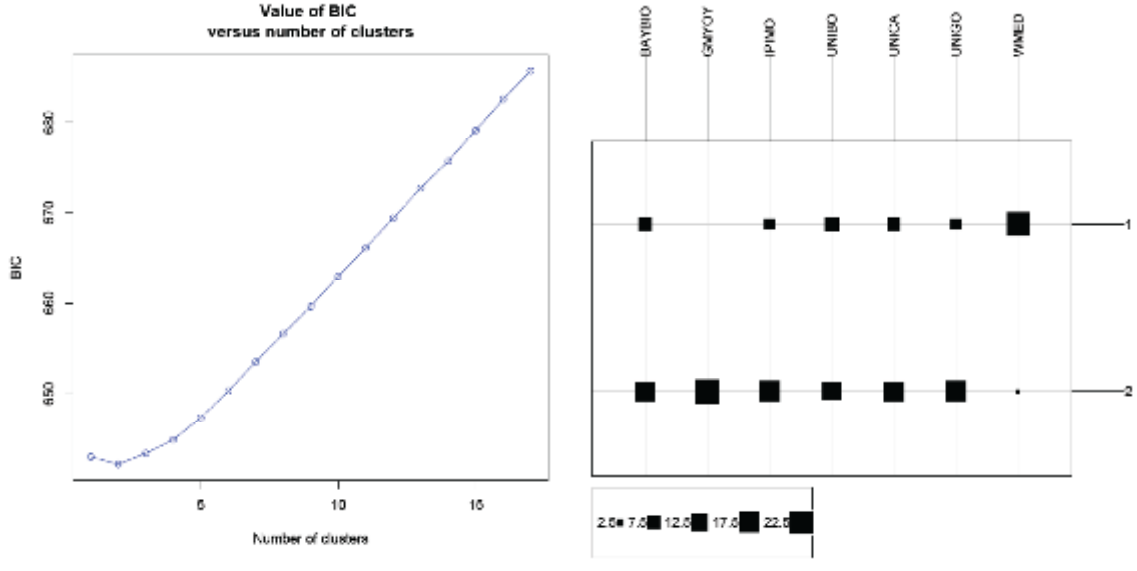


Figure 4.34: DAPC analysis of the seven bluefin tuna population samples with panels A) all-loci and B) F_{ST}>10%. Cluster 1 = WMED; Cluster 2 = GMYOY.

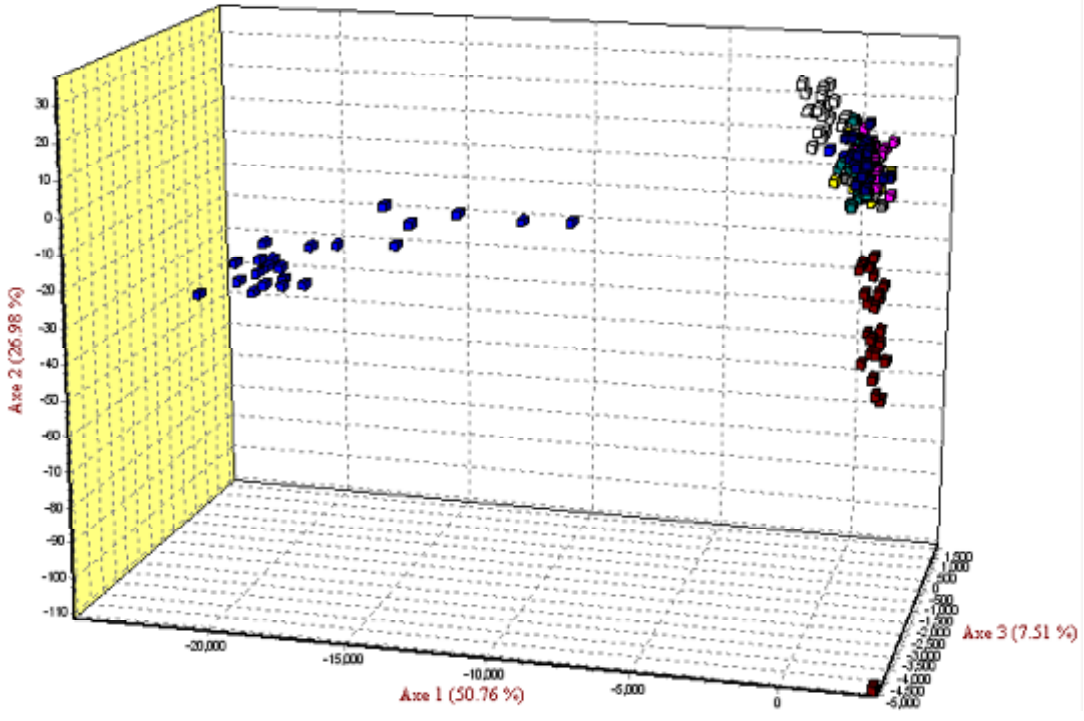


Figure 4.35: AFC analysis of the eight analyzed population samples. Blue dots (left) = CYPR; Brown dots (right-down) = WMED; white dots (right-top) = GMYOY. The remaining genotypic clusters correspond to the rest of feeding bluefin tuna population samples.

We performed a Factorial Correspondence Analysis (AFC) with the selected SNP panel with F_{ST}>10%, firstly on all populations (including CYPR to overview species/population genotypic distances; Figure 4.35), and then with only the validated bluefin tuna population samples (Figure 4.36). In the first analysis, a high degree of discrimination was observed with the CYPR fully separated by the other population samples. Among them, **the WMED population sample was more separated from the feeding population samples than the GMYOY** (Figure 4.35).

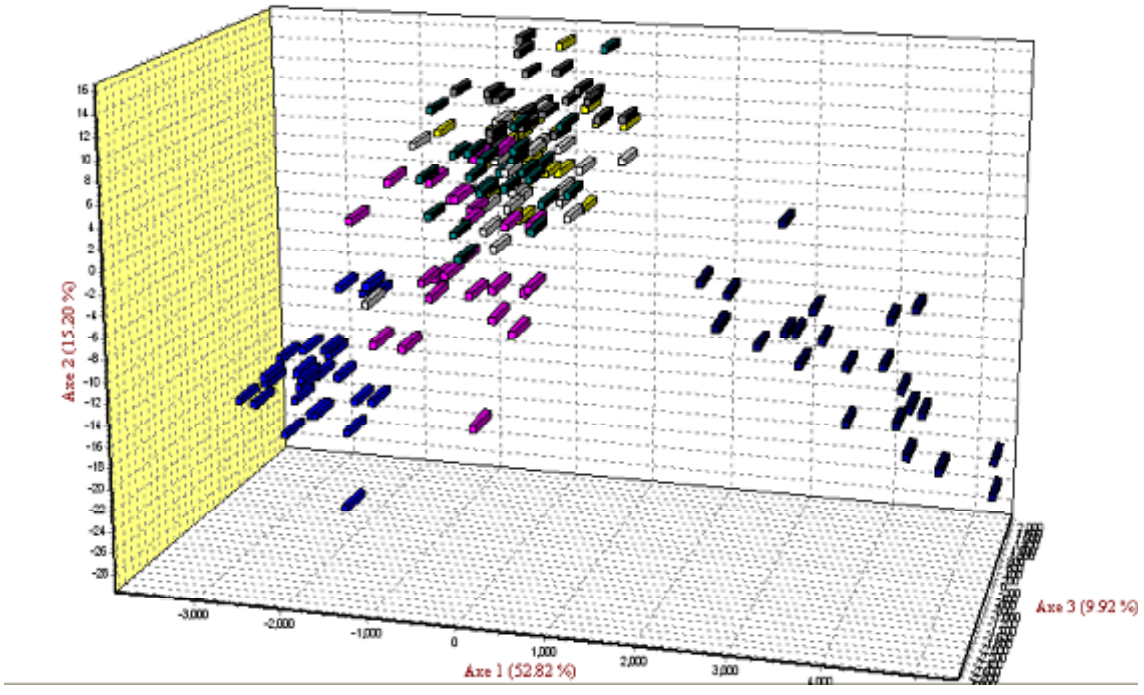


Figure 4.36: AFC analysis of 7 bluefin tuna population samples. Dark blue dots (right) = WMED; light blue (left-down) = GMYOY; the other dots represent the feeding population samples (the purple UNICA appears mostly related to GMYOY).

After removing CYPR, **the WMED sample was clearly separated from all other population samples and GMYOY was also separated, but it appeared more similar to the feeding population samples than WMED**. As other analyses (i.e. DAPC, individual proportion assigned to clusters), this **AFC analysis showed the relatedness between GMYOY and the feeding populations** (similar genetic variability and more similar genotypes).

Although in the AFC analyses **the WMED population sample seems more discrete than the GMYOY, this was not the case when looking at pairwise F_{ST} values based on the outlier SNP panel with F_{ST}>10%** (Table 4.9).

Table 4.9: Pairwise F_{ST} values among seven bluefin tuna populations estimated using the F_{ST}>10% panel (in bold when p<0.05).

Populations	BAYBIO	GMYOY	IPIMO	UNIBO	UNICA	UNIGO
GMYOY	**0.07198**					
IPIMO	-0.00003	**0.06097**				
UNIBO	-0.00237	**0.07255**	0.00356			
UNICA	0.00215	**0.06034**	0.00205	0.00847		
UNIGO	-0.00271	**0.06142**	0.00255	0.00131	0.00232	
WMED	**0.03056**	**0.14702**	**0.0377**	**0.03628**	**0.03772**	**0.03692**

With the F_{ST}>10% SNP panel (Table 4.9) the genetic differentiation between the two spawning populations was F_{ST} = 0.14702 (p<0.05). However, the **GMYOY population sample showed a higher differentiation than WMED with respect to the feeding population samples** (GMYOY vs feeding population samples: F_{ST}> 6%, p<0.05 for all values; WMED vs feeding population samples: F_{ST}> 3%, p<0.05 for all values). **None of the feeding aggregations showed a significant genetic differentiation among each other.**

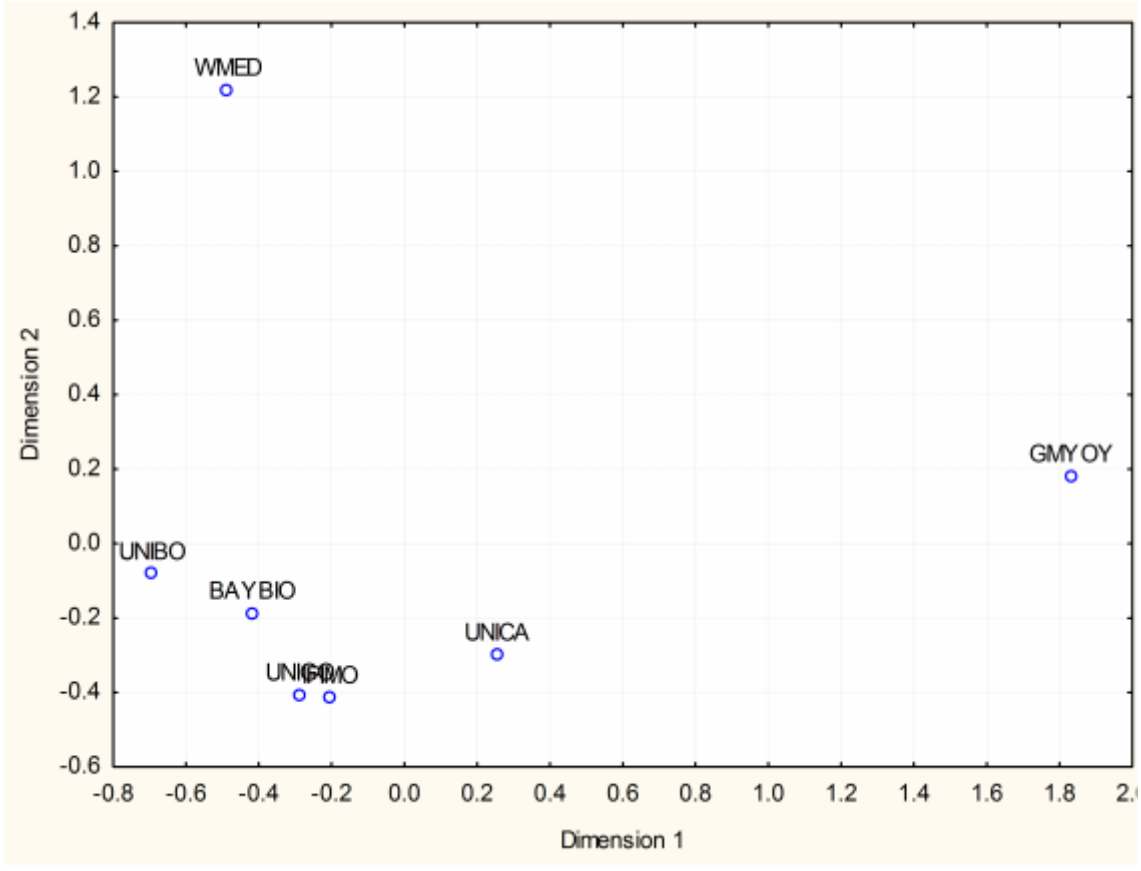


Figure 4.37: MDS plot of pairwise F_{ST} values estimated using the F_{ST}>10% SNP panel.

The MDS and CMDS plots of pairwise F_{ST} values estimated using the F_{ST}>10% and F_{ST}>5% SNP panels (Figures 4.37 and 4.38) revealed the intermediate position of the feeding population samples with respect to the spawning population samples (WMED and GMYOY).

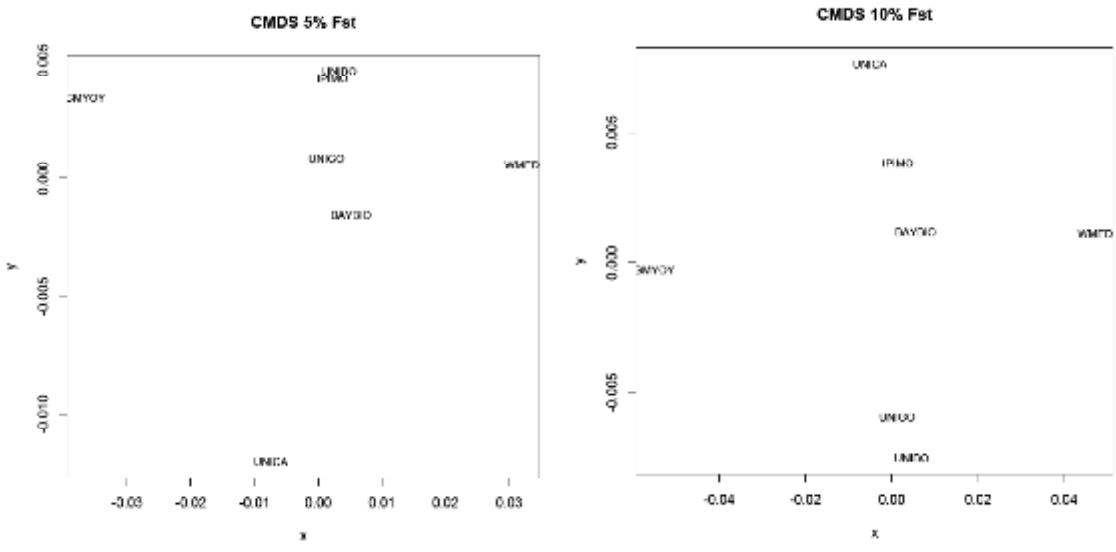


Figure 4.38: CMDS plot of pairwise FST values estimated with the FST>5% (left plot) and FST>10% (right plot) SNP panels.

Assignment analysis and traceability panel definition

Given the need for a smaller panel for population traceability, **we based our SNP marker selection on the ranked F_{ST} values of F_{ST}>5% and F_{ST}>10% between WMED and GMYOY**. These panels were used for performing an assignment analysis using GENECLASS software and a model based clustering (without a priori population information) implemented in STRUCTURE software.

- GENECLASS analysis: Using *a priori* population information (GMYOY and WMED spawning populations) to perform a self-assignment and mixed-population analysis, **GENECLASS was able to assign with 100% of accuracy all individuals to spawning populations** with both F_{ST}>5% and F_{ST}>10% panels. Reducing the **dataset to the 30 loci with the highest F_{ST}** (but also with high F_{IS}, so to be tested), **also a 100% assignment to spawning population samples** was achieved. With the **30 highest F_{ST} loci (but with low F_{IS}, 10% max), a 96 % assignment success** for spawning populations was achieved.

When we tried to assign individuals of the feeding population samples to the two spawning population samples using the full 391 loci panel, we obtained mixed proportions with a predominance of individuals assigned to the WMED spawning population sample (Table 4.10).

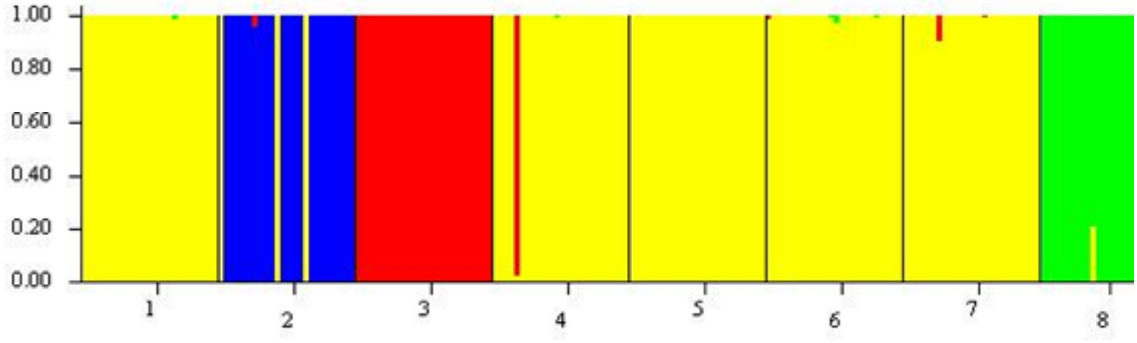
Table 4.10: Number and percentage of individuals of the feeding population samples assigned to the spawning population samples.

	Spawning populations			
Feeding population	GMYOY	WMED	% GMYOY	% WMED
BAYBIO	7	17	29 %	71 %
IPIMO	8	16	33 %	67 %
UNIBO	8	16	33 %	67 %
UNICA	11	13	46 %	54 %
UNIGO	11	13	46 %	54 %
Total individuals	**45**	**75**	38 %	63 %

These results were contradictory to those obtained by the DAPC and individual proportion analyses, namely a higher proportion of individuals assigned to the WMED than to the GMYOY. However, when performing a probability analysis of assignment to these two populations, results showed that the probability of feeding populations to be assigned to both populations is equal (similar to a confidence interval around the assignment score). This was tested using a simulation approach, where 1000 genotypes from the allele frequency distribution are generated and test genotypes arecompared to this novel dataset. Then, a likelihood of membership is calculated besides the basic assignment score (see Table 4.10). Hence, if the likelihood is equal for both clusters, more stringent and in depth analyses need to be performed to select the most informative loci. This prompts for further analyses without *a priori* population definition in the feeding samples, to test whether other spawning populations might be present, differing from both WMED and GMYOY.

- STRUCTURE analysis: When we performed a model-based Bayesian clustering analysis without *a priori* population definition (no admixture model, allele frequency correlated) and using all samples (including CYPR), we detected a number of clusters that lied between 4 and 5 (Figure 4.39). The number of clusters after removing CYPR (only validated bluefin tuna population samples) resulted K = 3 or K = 4.

A



B

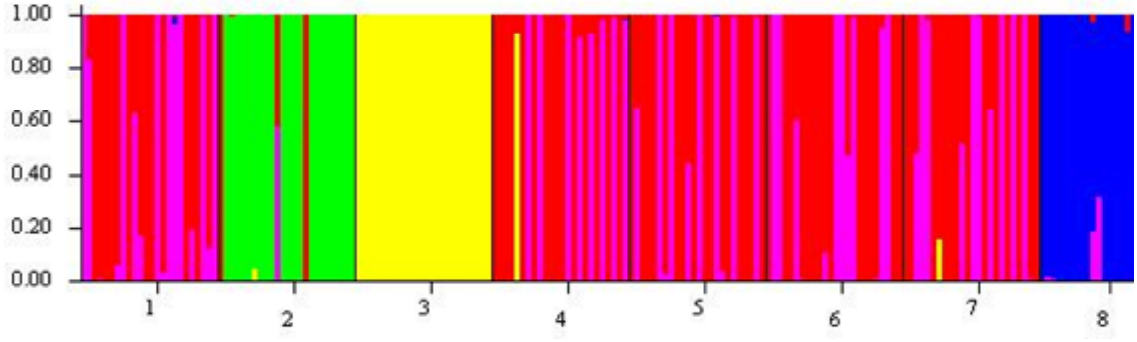


Figure 4.39: Individual based Bayesian clustering of all 192 individuals (including CYPR) based on the F_{ST}>10% panel. A) K = 4 ; B) K=5. 1 = BAYBIO; 2 = CYPR; 3 = GMYOY; 4 = IPIMO; 5 = UNIBO; 6 = UNICA; 7 = UNIGO; 8 = WMED.

The STRUCTURE results showed that:

1) In the CYPR sample, an incorrect species was easily identified using the current SNP panel. This finding is important when larvae samples are being used. Three individuals do show the correct genotype and are confirmed Atlantic bluefin tuna.
2) The GMYOY and WMED populations are clearly distinct populations.
3) The mixed juvenile/adult populations look as a separate group, potentially representing a mix of two additional genetically distinct populations (to be validated). However, we believe that further in depth testing of this dataset,

together with additional data from other sources and the NGS-TS dataset will help improving the feeding population assignment scores. One individual of the mixed feeding population 4 (large fish from the Northeast Atlantic) seems originating from the Gulf of Mexico.

All in all, the NGS-RRSG strategy and tasks yielded a wealth of novel genomic data for bluefin tuna. Additionally, this happened in a very cost-efficient way, although the downstream data-analyses increased in complexity. Over 25000 novel SNPs were jointly developed and genotyped in all target populations. Overall the following main results were obtained:

- In total 26316 selected SNP loci genotyped in 70% of all individuals were analyzed for 7 populations, including two supposed discrete spawning populations (excluding CYPR).
- Overall, approximately 2000 outliers with F_{ST}>3% were detected, a valuable number of loci for connectivity/traceability and adaptation studies.
- The WMED population exhibited a high genetic diversity level, nearly double than the other spawning and feeding populations.
- Multivariate results (DAPC, AFC, CMDS) showed three main clusters: WMED, GMYOY and other populations using all loci and the powerful restricted panels of outlier and highly divergent loci (with high F_{ST} values).
- Using the F_{ST}>10% panel (391 loci), a 100% assignment power for WMED and GMYOY populations was achieved. Using the 30 highest F_{ST} loci (but with high F_{IS}, so to be checked) also 100 % assignment to both populations was achieved. With the 30 highest F_{ST} loci with low F_{IS} (10% max), a 96 % assignment success for both populations was achieved.
- Feeding/mixed populations were more challenging to assign at this point and seem to form a heterogeneous group. Some analyses had slightly different results in membership proportions, to be assessed more in depth using additional analyses and data sources for validation purposes. A selected SNP validation step is required to confirm SNP power.

The model based analysis without a priori population assignment implemented in STRUCTURE, unveiled 3-4 clusters within the confirmed bluefin tuna populations (removing CYPR as misidentified species larvae): For K=3, WMED, GMYOY and other feeding aggregates. For K=4, WMED, GMYOY, and 2 other potential mixed/co-occurring clusters in feeding aggregates (with full genotypes, not 50/50). Additional analyses are needed to confirm the membership of feeding populations combining various analyses and datasets.

4.3 Discussion of the genetic results

The Genetic strategies and tasks we carried out in the GBYP-Phase2 have significantly increased the genomic and genetic resources for bluefin tuna and achieved some important biological and ecological issues for bluefin tuna population structure and population assignment.

Although the genetic strategies and tasks were proven highly confident from a technological side (i.e. providing high quality genomic data and high rate of individual genotyping with significantly high conversion and call rates), the two Next Generation Sequencing approaches and strategies we performed (i.e. the Transcriptome Sequencing, TS, and the Reduced Representation Sequencing and Genotyping, RRSG) were comparatively quite different in resolving bluefin tuna genetic and genomic variation over the broad spatial scale and among the population samples representing the putative spawning populations and feeding aggregates. Both strategies led to discover and genotype high fractions of genomic SNP loci that did not differentiate at all bluefin tuna individuals. This issue allowed the detection of extremely low and not significant levels of genetic differentiation among bluefin tuna population samples using the entire, large panels of SNPs we developed. Only the use of restricted panels of outlier and highly divergent SNP loci (from thousands to hundreds in the NGS-RRSG; only few tens in the NGS-TS) permitted to discriminate spawning population samples and assign individuals of feeding aggregates to originating populations. In addition, the species misidentification of a larvae reference sample of the bluefin tuna population spawning in the Eastern Mediterranean and the low DNA quality/quantity yields affected and prevented the completion of the population sample analysis plans at this stage of analysis.

Bluefin tuna population structure: genetic and genomic differentiation of spawning populations

The NGS-TS approach, based on the discovery, selection and genotyping of 384 species-specific SNPs located in expressed sequences did not resolve genetic variation of the bluefin tuna population samples consistently with a biological hypothesis of at least two spawning populations inhabiting the Mediterranean and the Gulf of Mexico. An extremely low and insignificant level of genetic differentiation detected among the 6 reference samples of the two well known putative spawning populations (Gulf of Mexico and Mediterranean) did not allow the complete discrimination of the analyzed larvae and young-of-the-year individuals collected from the Gulf of Mexico, Balearic, South Tyrrhenian and Cyprus area in several genetic analyses. Only the DAPC analysis (i.e. a spatial analysis that maximizes the differences between groups and minimizes those within groups), based on a restricted panel of 63 SNPs with F_{ST}>0.005 allowed the separation of the Mediterranean bluefin tuna Age 0 and larvae from those collected in the Gulf of Mexico. The differentiation observed between the three temporal replicates of the Balearic area (higher than the differences observed between these samples and that from the South Tyrrhenian) undermined the full validity of this result.

On the contrary, the restricted panels of outlier and highly divergent SNP loci obtained in the NGS-RRSG tasks (i.e. 1934 loci with F_{ST}>5%, which corresponded to 7% of the whole number of SNP loci, and 396 loci with F_{ST}>10%) permitted to unequivocally discriminate Age 0 and larvae collected from the two putative spawning populations (Gulf of Mexico and Western Mediterranean) with several statistical tests and analyses. Using only the 30 loci with highest F_{ST} values, we achieved 96% - 100% assignment of the individuals to the two spawning populations. No temporal replicates were included in the population sample analysis design of the NGS-RRSG and therefore the checking of the temporal stability of this unequivocal separation is of priority, together with a validation test of these outlier SNPs, to set up a promising and powerful genomic tool for resolving population structure and individual assignment in bluefin tuna. As well, the lack of a valid reference sample from the Eastern and Central Mediterranean in the NGS-RRSG analysis should be overcome to assess the population structure within the Mediterranean.

Bluefin tuna population structure: genetic and genomic differentiation of feeding populations

With both genetic strategies and tasks, the strata samples representing the feeding populations (or aggregates) resulted genetically undifferentiated, either using the powerful restricted panel of outlier and highly divergent loci, or the most powerful statistical tests and analysis. Likely, in the NGS-TS, the lack of significant genetic differences among feeding populations was due to the low power of resolution of the panels of SNPs with F_{ST}>0.005 (both the 52 and the 63-SNP loci panels), since they resulted significantly undifferentiated from the reference samples of the spawning populations. On the contrary, in the NGS-RRSG, feeding populations formed a third cluster separated and equidistant from the two well-separated spawning populations (Gulf of Mexico and Western Mediterranean) in several spatial clustering analyses and with the powerful restricted panels of outlier loci.

The assignment analyses carried out in both strategies provided contradictory results, which suggested further deep investigations on the composition of these feeding populations. The NGS-TS assignment of the individuals of feeding populations to spawning populations, based on the restricted panel of 63 loci with F_{ST}>0.005, showed a prevalent contribution of the spawning population of the Western Mediterranean compared to those of the Gulf of Mexico or Eastern Mediterranean, but with low accuracy. In the NGS-RRSG parallel task, and using the more powerful restricted panels of outlier loci, we obtained equal chance of being assigned to either the Gulf of Mexico or the Western Mediterranean.

In agreement with this issue, the genetic differentiation between the reference sample of the Western Mediterranean and the feeding populations was on average significantly lower than that detected between the Gulf of Mexico reference sample and the feeding populations (3% vs 6%, respectively). On the contrary, AFC and DAPC – BIC analyses displayed a closer genetic relationship between Gulf of Mexico and feeding populations and higher individual membership proportions to the Gulf of Mexico spawning population in the feeding populations, respectively.

Lastly, as the DAPC, the model-based Bayesian clustering analysis without *a priori* population definition implemented in STRUCTURE identified the feeding

populations as a third separate cluster, potentially representing a mix of two additional genetically distinct populations.

4.4 Recommendations and further activities

The promising results obtained with the genomic strategies and tasks carried out in the GBYP-Phase 2 recommend important key-actions and guidelines to reach the objectives to assess population structure and population assignment in the Atlantic bluefin tuna.

The problem of low DNA quality/quantity yields detected across several samples faced in the GBYP Phase 2 will be remediated in the Phase 3 by assuring most performing tissue storage and shipping conditions. The replacing of initial storage buffer after some days and the reduction of shipping time by using Express Courier Service will reduce the risk of degradation of DNA and increase the probability to retrieve high quality extracted DNA from tissues. Performing the DNA extraction as close as the tissue samples are collected, the storage of DNAs at -80°C, and, whenever possible, the shipping of dried genomic DNAs rather than tissue in ethanol 96% will also improve the quality/quantity yields of DNA. To avoid species misidentification at stages were morphological species assignment could be problematic, DNA barcoding of larvae will be addressed before SNP-typing to avoid the inclusion of non-bluefin tuna samples in the batches of individuals that will be SNP-typed.

So far, the NGS-TS strategy appeared to be not powerful in bluefin tuna because the lack of large and highly divergent (outlier) panels of SNP loci. On the contrary, more advances were achieved with the NGS-RRSG strategy, even if the current large dataset generated in GBYP-Phase 2 requires additional improvement, both on the dataset and on the analyses. The improvement of the dataset for the NGS-RRSG strategy will be achieved by the replacement of the mostly misidentified Larvae sample EMED-LS-LA from the Levantine Sea (the easternmost ABFT spawning area in the Mediterranean) with the sample formed by correctly identified Age0 individuals (EMED-LS-0), already available. Moreover, additional spawning population samples already collected (i.e. temporal replicates from the Balearic Islands WMED-BA-0-2009 and WMED-BA-0-2010) and those that will be collected

in 2012 will be included to increase the robustness of the dataset regarding the reference populations spawning in the Mediterranean. Given the improvement of the dataset, more performing data analyses will be conducted selecting outlier loci that will allow the discrimination between a more representative set of reference populations. The selection of most performing outlier SNP loci will be conducted taking into consideration both interannual variability and the whole range of ABFT spawning populations. The genotyping of population samples representing feeding aggregates with the validated and powerful panels of outlier SNP will be also a priority, to reach the final goals of assessing population structure and assignment in the Atlantic bluefin tuna.

References

- Carlsson, J; McDowell, JR; Diaz-Jaimes, P; Carlsson, JEL; Boles, SB; Gold, JR; Graves, JE (2004) Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnusthynnusthynnus*) population structure in the Mediterranean Sea 13 (11): 3345-3356
- Carlsson, Jens; McDowell, Jan R.; Carlsson, Jeanette E. L.; Graves, John E. (2007) Genetic identity of YOY bluefin tuna from the eastern and Western Atlantic spawning areas. JOURNAL OF HEREDITY 98 (1): 23-28
- Davey J., Paul Hohenlohe, Paul Etter, Jason Boone, Julian Catchen and Mark Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. NATURE REVIEWS GENETICS, 12:499-510
- FishPopTrace -Fish Populations and Traceability: The Structure of Fish Populations and Traceability of Fish and Fish Products (European Community's Seventh Framework Programme (FP7/20072013) under grant agreement n° KBBE212399 , Final report submitted July 2011
- Freamo, Heather; O'Reilly, Patrick; Berg, Paul R.; Lien, Sigbjorn; Boulding, Elizabeth G. (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. MOLECULAR ECOLOGY RESOURCES 11:254-267
- Hauser, L; Ward, RD (1998) Population identification in pelagic fish: the limits of molecular markers. ADVANCES IN MOLECULAR ECOLOGY 306: 191-224
- Hauser, Lorenz; Seeb, James E. (2008) Advances in molecular technology and their impact on fisheries genetics. FISH AND FISHERIES 9 (4): 473-486
- Helyar, S. J.; Hemmer-Hansen, J.; Bekkevold, D.; Taylor, M. I.; Ogden, R.; Limborg, M. T.; Cariani, A.; Maes, G. E.; Diopere, E.; Carvalho, G. R.; Nielsen, E. E. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges MOLECULAR ECOLOGY RESOURCES 11: : 123-136

- Milano Ilaria, MassimilianoBabbucci, Frank Panitz, Rob Ogden, Rasmus O. Nielsen, Martin I. Taylor, Sarah J. Helyar, Gary R. Carvalho, Montserrat Espiñeira, MiroslavaAtanassova, FaustoTinti, Gregory E. Maes, TomasoPatarnello, FishPopTrace Consortium and Luca Bargelloni. (2011) Novel Tools for Conservation Genomics: Comparing Two High-throughput Approaches for SNP Discovery in the Transcriptome of the European Hake. PLOS One 6(11): e28008. doi:10.1371/journal.pone.0028008
- Nielsen, Einar E.; Hemmer-Hansen, Jakob; Larsen, Peter Foged; Bekkevold, Dorte (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. MOLECULAR ECOLOGY 18 (15): 3128-3150
- Ogden, Rob. (2008) Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. FISH AND FISHERIES 9 (4): : 462-472
- Riccioni, Giulia; Landi, Monica; Ferrara, Giorgia; Milano, Ilaria; Cariani, Alessia; Zane, Lorenzo; Sella, Massimo; Barbujani, Guido; Tinti, Fausto (2010). Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 107 (5): 2102-2107

5. OTOLITH MICROCHEMISTRY ANALYSES: ORIGIN OF ATLANTIC BLUEFIN TUNA IN THE EASTERN ATLANTIC AND MEDITERRANEAN USING $\delta^{13}C$ AND $\delta^{18}O$

Task Leader: Jay R. Rooker (TAMU)

Participants

AZTI: Igaratza Fraile, Haritz Arrizabalaga

5.1 Introduction

Several novel tools are currently being used to investigate the natal origin and stock structure of Atlantic bluefin tuna, including electronic tags, molecular genetics, and otolith chemistry. Of the three, chemical markers in otoliths (ear stones) have significant potential for determining natal origin and population connectivity of bluefin tuna (Rooker et al. 2007). This is due to the fact that otoliths precipitate material (primarily calcium carbonate) as a fish grows, and the chemical composition of each newly accreted layer is often associated with physicochemical conditions of the water mass they inhabit. As a result, material deposited in the otolith during the first year of life serves as a natural marker of the individual's nursery or place of origin. Previous studies have demonstrated that trace elements and stable isotopes in otoliths can be used to determine the origin of bluefin tuna from different regions in the Atlantic Ocean and its marginal seas (Mediterranean Sea and Gulf of Mexico; see Rooker et al. 2008a,b, Schloesser et al. 2010). Results from these studies indicate that trans-Atlantic movement is more significant than previously assumed, with a considerable fraction of adolescents in US water originating from spawning/nursery areas in the east (Mediterranean Sea).

Here, we investigated the origin of bluefin tuna collected in the eastern Atlantic (Bay of Biscay) and a variety of locations in the Mediterranean Sea using stable $\delta^{13}C$ and $\delta^{18}O$ isotopes in otoliths.

5.2 Material and Methods

As otoliths become available from sampling activities, different sets of otoliths were prepared for analyses. A first set of otoliths (juveniles from the Bay of Biscay and medium and adult fish from Gibraltar and Malta) was sent to TAMU for milling and analyses in October 2011. In parallel, a second set of otoliths (juveniles from the Adriatic, juvenile and medium size fish from the Bay of Biscay, medium size fish from Malta, Medium and Large fish from Sardinia and large fish from the Central Atlantic) was milled in AZTI and the power was sent to TAMU in November 2011 for chemical analyses. A third set of otoliths from the Central North Atlantic was

milled in AZTI and shipped to TAMU in December 2011to conduct the chemical analyses.

Otolith handling followed the protocols previously described in Rooker et al. (2008b). Briefly, following extraction by GBYP participants, sagittal otoliths of bluefin tuna were cleaned of excess tissue with hydrogen peroxide and deionized water. One sagittal otolith from each bluefin tuna specimen was embedded in Struers epoxy resin (EpoFix) and sectioned using a low speed ISOMET saw to obtain 1.5 mm transverse sections that included the core. Following attachment to a sample plate, the portion of the otolith core corresponding to approximately the yearling periods of bluefin tuna was milled from the otolith section using a New Wave Research MicroMill system. A two-vector drill path based upon otolith measurements of several yearling bluefin tuna was created and used as the standard template to isolate core material following Rooker et al. (2008b). The pre-programmed drill path was made using a 500 μm diameter drill bit and 15 passes each at a depth of 50 μm was used to obtain core material from the otolith. Powdered core material was transferred to silver capsules and later analyzed for $\delta^{13}C$ and $\delta^{18}O$ on an automated carbonate preparation device (KIEL-III) coupled to a gas-ratio mass spectrometer (Finnigan MAT 252). Stable $\delta^{13}C$ and $\delta^{18}O$ isotopes were reported relative to the PeeDee belemnite (PDB) scale after comparison to an in-house laboratory standard calibrated to PDB.

Region-specific estimates of nursery origin of bluefin tuna were based on comparing otolith 'cores' (corresponds otolith material deposited during the first year of life or yearling period) of juvenile and adult bluefin tuna to the baseline or reference samples of yearling bluefin tuna. Estimate of origin for juvenile (3-25 kg) and adult (medium and large (>25-100 kg and > 100 kg, respectively) bluefin tuna were obtained using the maximum likelihood mixed-stock analysis program HISEA developed by Millar (1990). The baseline data set used for mixed-stock analysis was otolith $\delta^{13}C$ and $\delta^{18}O$ of yearling samples collected in the east and west from 1998-2009, with recent samples (e.g. 2008 and 2009) supplied by AZTI scientists. Otolith cores of juvenile and adult bluefin tuna collected in the Mediterranean and eastern Atlantic were then used to estimate the origin of these recruits in the bootstrap

mode of HISEA with 1000 simulations, which provided non-parametric estimates of the reliability of predicted contributions from eastern and western nurseries.

5.3 Results and Discussion

Otolith $\delta^{13}C$ and $\delta^{18}O$ in the cores of yearling bluefin tuna from eastern and western nurseries were distinct (MANOVA, $p<0.001$) and used here as baseline data for mixed-stock runs to predict the origin of unknown individuals. Mean otolith $\delta^{13}C$ of yearlings in the updated baseline sample were relatively similar for individuals collected in the east (-8.54 ppt) and west (-8.73 ppt). Conversely, otolith $\delta^{18}O$ of yearlings was markedly different between the east and west, with bluefin tuna of eastern origin having more enriched values (mean -0.81 ppt) relative to yearlings collected in the west (-1.35 ppt).

Otolith $\delta^{13}C$ and $\delta^{18}O$ were measured in the otolith cores of 600 bluefin tuna (both juveniles and adults) from a variety of locations in the Atlantic Ocean (Central North Atlantic, Bay of Biscay) and in the Mediterranean Sea (Strait of Gibraltar, Balearic Sea, Adriatic Sea, Malta, Sardinia). Outside the Mediterranean, estimates of nursery origin were determined for medium and large category bluefin tuna collected in the Central North Atlantic, and mixed-stock analysis indicated that mixing of eastern and western stocks occurred in this region. Maximum-likelihood estimates (MLE) of bluefin tuna collected in the Central North Atlantic were largely comprised of individuals from the 'eastern' or Mediterranean nursery (84.1%) (Table 5.1 and Figure 5.1). Still, a significant number of 'western' bluefin tuna were present in our 'unknown' sample from the Central North Atlantic, indicating migrants from both eastern and western populations mix in this region. Alternatively, the presence of western migrants in the Bay of Biscay was rare (< 1%) for both juvenile and adult category bluefin tuna, suggesting that the Bay of Biscay fishery is supported by local production of bluefin tuna in the Mediterranean Sea. Rooker et al. (2008a) and Dickhut et al. (2009) reported significant east to west movement of juvenile bluefin tuna to the US Atlantic; however, results from our analysis of juveniles from the GBYP indicate that movement of juvenile bluefin tuna produced in the Gulf of Mexico (western origin) to the Bay of Biscay is limited or insignificant.

Table 5.1. Maximum-likelihood predictions of the origin of bluefin tuna collected from Central and Eastern North Atlantic Ocean. Estimates are given as percentages and the mixed-stock analysis (HISEA program) was run under bootstrap mode with 1000 runs to obtain standard deviations around estimated percentages (□ %). Bay of Biscay samples are shown as two size class: juvenile (10-25 kg) and adult (includes both medium and large categories or all individuals > 25 kg).

		Predicted Origin		
Region	**N**	**% East**	**% West**	**% Error**
Central North Atlantic	117	84.1	15.9	7.9
Bay of Biscay (Juvenile)	135	99.1	0.9	0.9
Bay of Biscay (Adult)	122	99.0	1.0	1.2

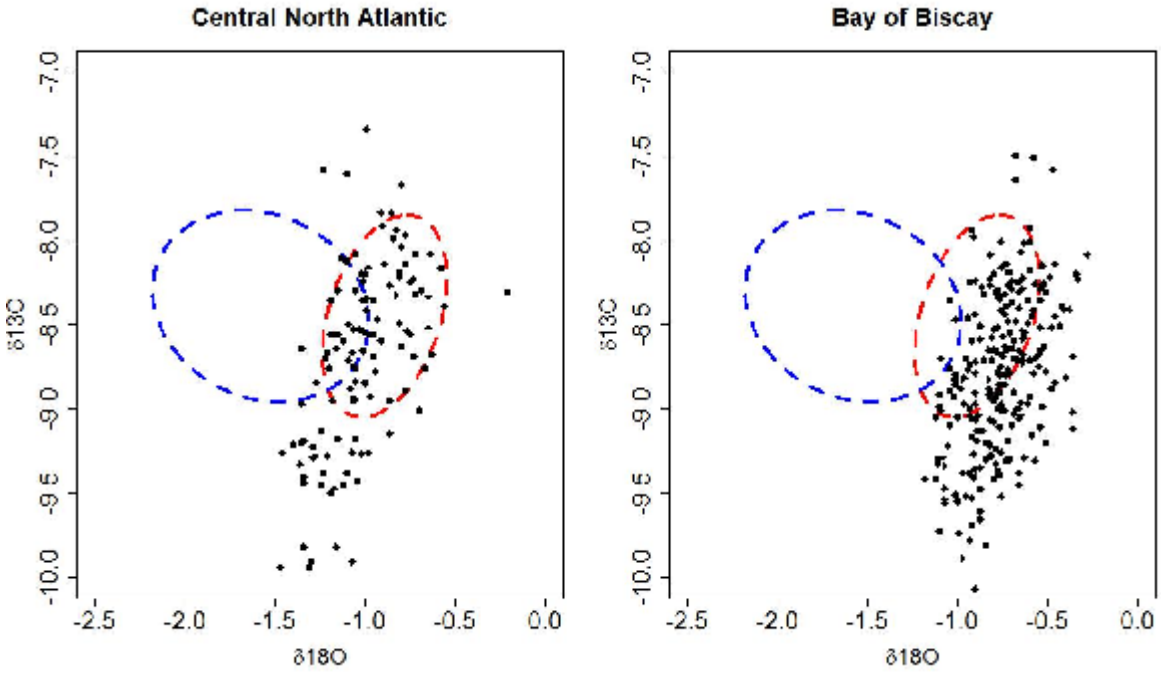


Figure 5.1: Otolith $\delta^{13}C$ and $\delta^{18}O$ of bluefin tuna from the Central North Atlantic (n = 117) and Bay of Biscay (n = 257). Points shown in relation to confidence ellipses (with p=0.6827) based on otolith $\delta^{13}C$ and $\delta^{18}O$ of yearling bluefin tuna from each region (baseline developed with yearling samples collected during 1998-2009; Blue = West (n =103), Red = East (n = 176)).

At the point of entry into the Mediterranean (Strait of Gibraltar), the occurrence of western migrants was essentially nil (0.1%), with 99.9% of the bluefin tuna predicted to be of eastern origin (Table 5.2). Within the Mediterranean, the trend continued and the predicted origin was 100% 'eastern' fish for all collection regions examined within the Mediterranean Sea. Mixed-stock analysis of samples collected in several locations clearly indicated local production support bluefin tuna fisheries in these regions (Table 5.2). Maximum-likelihood estimates (MLE) were 100% 'eastern' origin for juvenile bluefin tuna from the Adriatic Sea and adult bluefin tuna collected in Malta, Sardinia, and the Balearic Sea. Standard deviations around estimated means for bluefin tuna in all regions within the Mediterranean Sea were < 0.1%, indicating the degree of confidence in the 100% prediction is high.

Table 5.2. Maximum-likelihood predictions of the origin of bluefin bluefin tuna collected in the Strait of Gibraltar and within the Mediterranean Sea. Estimates are given as percentages and the mixed-stock analysis (HISEA program) was run under bootstrap mode with 1000 runs to obtain standard deviations around estimated percentages (± %). Nearly all bluefin collected in the Adriatic Sea were juveniles while all other regions are comprised primarily of adults in the medium and large categories (> 25 kg).

		Predicted Origin		
Region	**N**	**% East**	**% West**	**% Error**
Strait of Gibraltar	38	99.8	0.2	0.1
Balearic Sea	39	100	0.0	0.0
Malta	82	100	0.0	0.0
Sardinia	20	100	0.0	0.0
Adriatic Sea	47	100	0.0	0.0

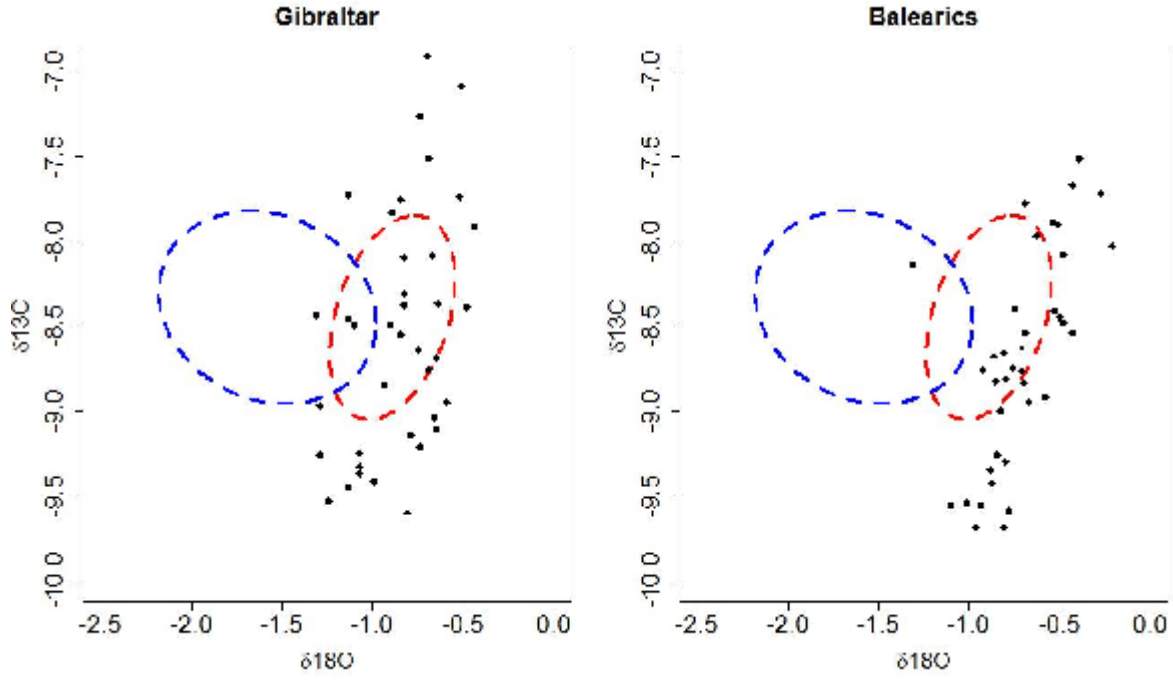


Figure 5.2: Otolith $\delta^{13}C$ and $\delta^{18}O$ of bluefin tuna from the Strait of Gibraltar (n = 38) and the Balearic Sea (n = 39). Points shown in relation to confidence ellipses (with p=0.6827) based on otolith $\delta^{13}C$ and $\delta^{18}O$ of yearling bluefin tuna from each region (baseline developed with yearling samples collected during 1998-2009; Blue = West (n =103), Red = East (n = 176)).

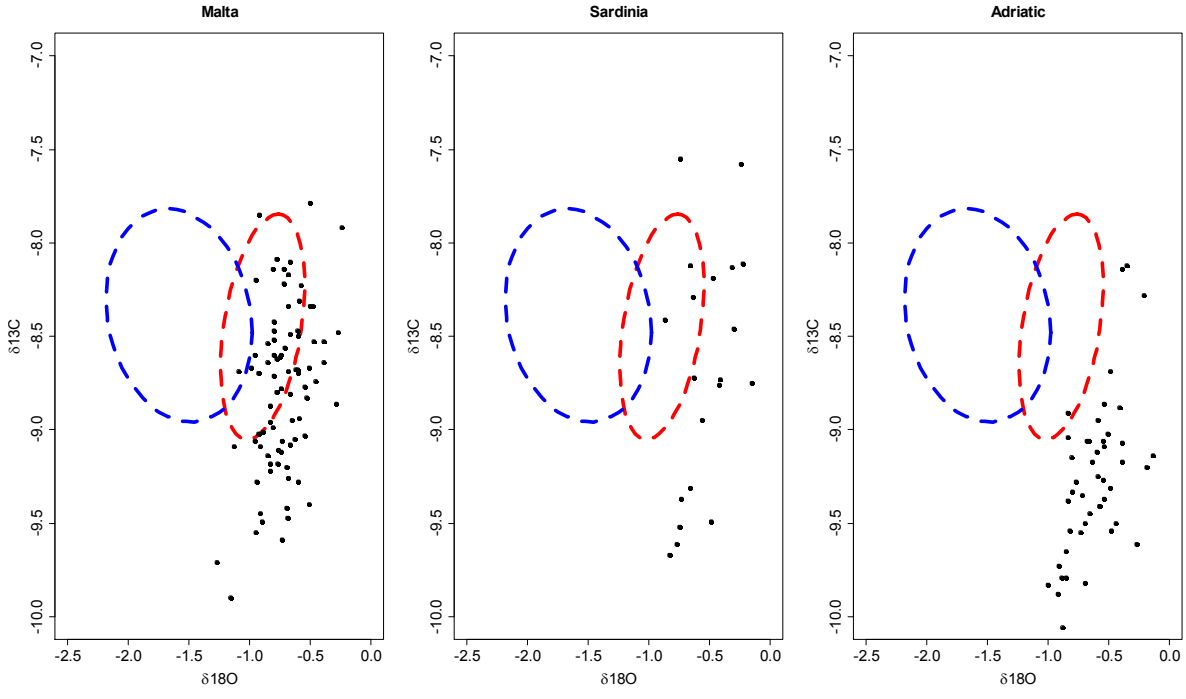


Figure 5.3: Otolith $\delta^{13}C$ and $\delta^{18}O$ of bluefin tuna from Malta (n = 82), Sardinia (n = 20) and the Adriatic Sea (n = 47). Points shown in relation to confidence ellipses (with p=0.6827) based on otolith $\delta^{13}C$ and $\delta^{18}O$ of yearling bluefin tuna from each region (baseline developed with yearling samples collected during 1998-2009; Blue = West (n =103), Red = East (n = 176)).

References

- Dickhut RM, Deshpande AD, Cincinelli A, Cochran MA, Corsolini S, Brill RW, Secor DH, Graves JE. 2009. Atlantic Bluefin Tuna (*Thunnus thynnus*) Population Dynamics Delineated by Organochlorine Tracers RID A-8468-2011. *Environmental Science and Technology* 43:8522-8527.
- Rooker JR, Alvarado Bremer JR, Block BA, Dewar H, De Metrio G, Kraus RT, Prince ED, Rodriquez-Marin E, Secor DH (2007) Life history and stock structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Reviews in Fisheries Science* 15: 265-310
- Rooker JR, Secor DH, DeMetrio G, Kaufman JA, Belmonte Rios A, Ticina A (2008a) Evidence of trans-Atlantic mixing and natal homing of bluefin tuna. *Marine Ecology Progress Series* 368: 231-239

- Rooker JR, Secor DH, DeMetrio G, Schloesser R, Block BA, Neilson JD (2008b) Natal homing and connectivity in Atlantic bluefin tuna populations. *Science* 322: 742-744
- Schloesser RW, Rooker JR, Louchuoarn P, Neilson JD, Secor DH (2009) Interdecadal variation in seawater $\delta^{13}C$ and $\delta^{18}O$ recorded in fish otoliths. *Limnology and Oceanograph*y 54: 1665–1668
- Schloesser RW, Neilson JD, Secor DH, Rooker JR (In press) Natal origin of Atlantic bluefin tuna (*Thunnus thynnus*) from the Gulf of St. Lawrence based on $\delta^{13}C$ and $\delta^{18}O$ in otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*

6. AGE DETERMINATION ANALYSES

Task Leader: Enrique Rodriguez-Marín (IEO)

Participants

IEO: Luque, P.L., Quelle, P. and Ruiz, M.

6.1 Introduction

Biological studies on age and growth of fish are crucial components for describing their life cycle (age at maturity, age at recruitment, longevity, etc.). Age determination is an essential feature in fish stock assessment to estimate the rates of mortalities and growth. Assessment of Atlantic bluefin tuna (*Thunnus thynnus*) using age structured models has proved useful in establishing a diagnosis of stock status.

The biological sampling of this project includes information on age composition of the samples carried out for population structure purpose, and at the same time, for obtaining information about the age composition of the bluefin tuna catches. To estimate the age of the catch the following approach was selected: to measure the length using a representative sample of the catch (i.e. through random sampling) and applying an age-length-key (i.e. through length-stratified sampling) to convert size into age distributions. This approach has been also applied for estimating southern bluefin tuna (*Thunnus maccoyii*) age composition (Anon., 2002).

We used two calcified structures for the age interpretation of bluefin tuna: sagittal otolith and first dorsal fin spine (hereby spine). This procedure was employed because sampling of either structure depends on dockside practices. Otoliths represent an advantage in relation to other calcified structures because all ages can be interpreted since there is no nucleus resorption. Otoliths can be used to age giant bluefin tuna, while other hard parts are difficult to interpret from 10 years upwards (Rodriguez-Marin et al., 2007). Conversely, spines are easier to collect and prepare than otoliths. In this Project we attempted to provide a preliminary age-length key based on otoliths and another one based on spines.

6.2 Material and Methods

A total of 381 otoliths and 437 spines were prepared for biometry analysis and age determination of bluefin tuna (Table 6.1). Samples were collected from May to November 2011 and from specimens caught in the central and western Mediterranean Sea (Adriatic Sea and waters around Sardinia, Malta and Balearic

Islands) and north-eastern Atlantic Ocean (Atlantic waters near the Strait of Gibraltar and Bay of Biscay). Bluefin tuna juveniles were caught by bait boats and adults by longliners and traps. Fin spine and sagittal otoliths extraction and conservation were carried out following the present project sampling protocols. There is a difference of 61 samples between the contract and amendment specifications for the age determination analyses (810) and the final number of calcified structures age estimations presented in the report (749). This difference in number is due to three main reasons: broken samples (nearly 5 % in spines), unreadable samples and size ranges already covered with enough number of samples. Furthermore, a special effort was done to face the comparison between age interpretation from otoliths and spines coming from the same fish, in order to obtain a representative number of samples by age.

Table 6.1. Summary of bluefin tuna calcified structures used for age determination by length range. Paired structures refer to hard parts coming from the same specimen. Length was measured as straight fork length (SFL) in cm. "No length" means that length was not recorded or that there was a clear error in the measurement.

Length (SFL, cm)	Spines	Otoliths	Paired structures
20-29	10	10	10
30-39	10	10	10
40-49	6	6	6
50-59	12	10	7
60-69	10	11	5
70-79	12	13	9
80-89	33	23	10
90-99	10	12	10
100-109	23	19	13
110-119	71	30	25
120-129	37	17	9
130-139	29	24	9
140-149	29	17	9
150-159	20	15	8
160-169	11	12	9
170-179	4	8	4
180-189	21	21	11
190-299	17	21	10
200-209	15	20	10
210-219	16	22	8
220-229	15	20	10
230-239	17	20	8
240-249	3	10	1
250-259	6	3	3
260-269			
270-279		1	
No length		6	
Total number	437	381	214

Spine preparation and age interpretation criteria were performed according to Rodriguez-Marin et al. (2012). Spine section location was established at 1.5 times the condyle base width. Sections were obtained using a precision rotating diamond saw and mounted on glass slides. It is easy to identify the translucent and opaque bands formed on the spine of young individuals. However, in fish over two years old, the central area of the spine begins to reabsorb and the bands consequently disappear. To overcome the problem of nucleus reabsorption with age, the

translucent band diameters measured from spines without reabsorption (i.e. spines from young specimens) had to be used to assign an age to the first inner visible translucent band in reabsorbed spines (Figure 6.1). Age was estimated by counting the translucent bands which are deposited annually between November and April (Luque et al., under review). For the interpretation of the border of the spine section we followed Rodriguez-Marin et al. (2007) criterion, in which a bluefin tuna with a translucent band formed at the edge of the spine section and caught at the beginning of the year was interpreted as having one year more, although there were still five or six months before its true date of birth (Rooker et al., 2007), whereas when the fish was caught in autumn, this band was not considered as one year more.

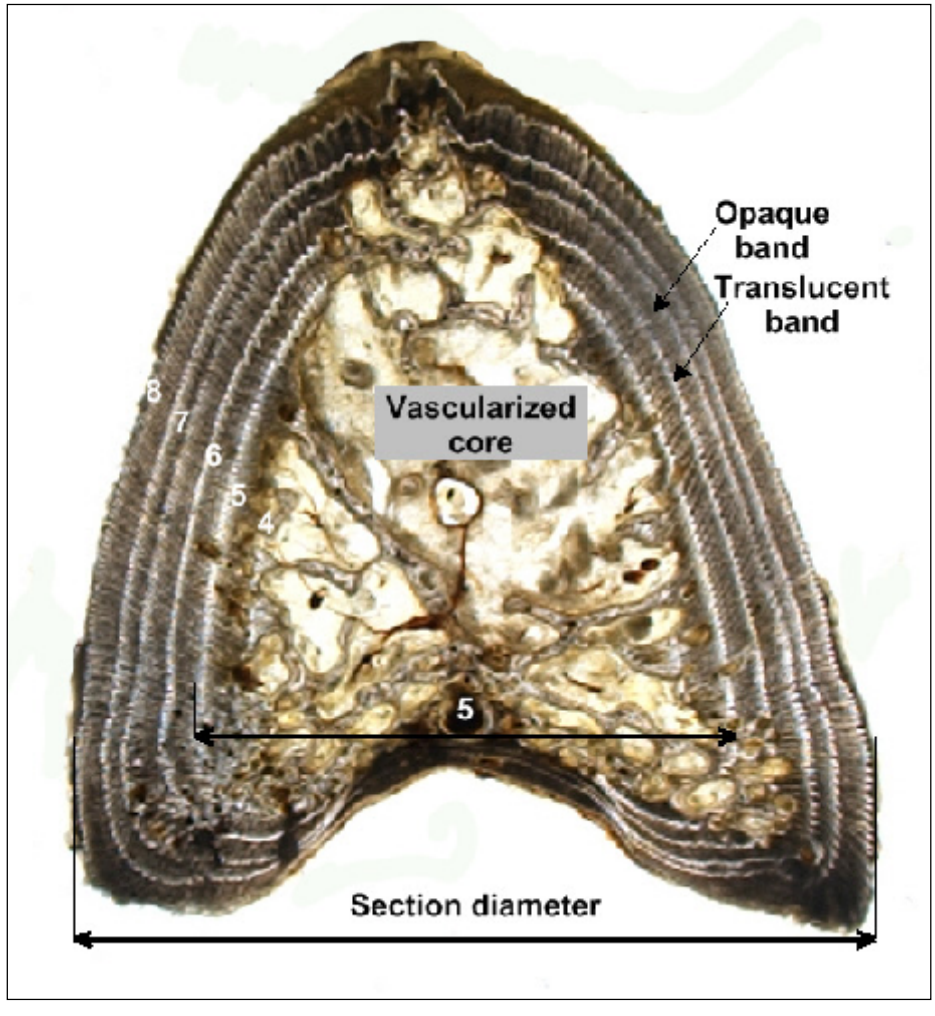


Figure 6.1. Spine section image of an Atlantic bluefin tuna aged 8 years old. Numbers reflect translucent bands considered annually formed. A false translucent band appears after 6th band.

Spines direct ageing was carried upon digital images that were captured using a binocular lens magnifier connected by digital camera NIKON. An image analyzer (Nis-elements D 3.0 Nikon software) was used to measure the maximum spine diameter as well as diameter for successive growth bands. Spines sections were read by two independent readers. For those spines that there was a disagreement between readers, an additional reading was achieved and the final estimated age assigned was the consensus among readers.

Otoliths were also sectioned by embedding them in a matrix resin within a mould. Three sections of 300-400 µm were obtained in the core area of each otolith (Figure 6.2). Encased otolith sections were mounted on glass slides and polished before imaging. The same procedure described for spines was used to obtain digital images of otoliths. Age interpretation was performed on digitally enhanced images using Adobe Photoshop and annulus counts were made along the longest (ventral) arm of the sectioned sagittae otolith. Otolith sections were examined using reflected light and age was estimated by counting the translucent bands. We applied the same border interpretation criterion described previously for spines.

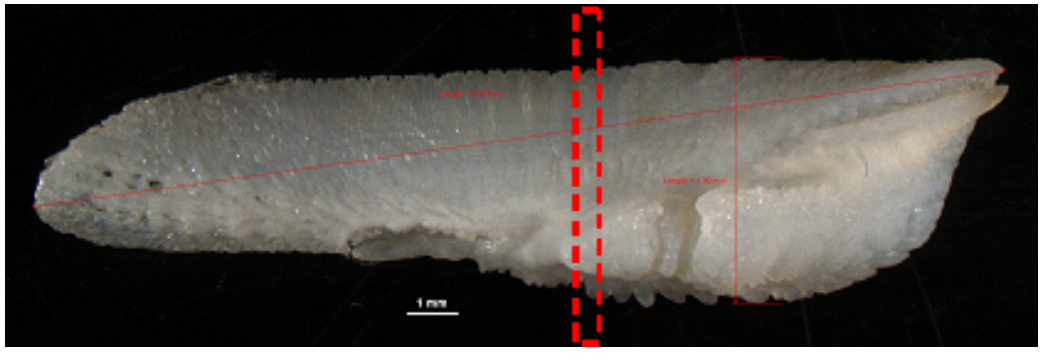


Figure 6.2. Right sagittal otolith of a bluefin tuna showing location of optimal section (discontinous red line) and whole otolith measurements. The anterior part of the otolith is narrower and is called rostrum (left) and the rear part is wider and is called postrostrum (right). The protuberance in this side of the otolith is called antirostrum, and is used as guide for the sectioning location.

Quality and edge identification for both calcified structures was annotated. Samples from different months and geographic areas were combined for constructing the age-length keys.

Calcified structures biometry

Several biometric measures were recorded for each structure in order to analyze the relationship between the growth of the hard part and the specimen sampled. Spine diameter and total spine length were measured. For otoliths the longest and widest axes of the sagittal otolith were measured by placing the whole sagittal otolith sulcus side down and using reflected light on a black background (Figure 6.2). Weight was also recorded. Incomplete otoliths were not used for this biometric analysis. Length of otolith ventral arm was measured as the distance from the outer edge of the first inflection point (elbow) to the apical end (Figure 6.3). This measurement was obtained in order to identify the first annulus (for this we use specimens with presumably one year old based on its length) and to analyze the allometry of this otolith ventral arm in relation to the length of the fish. Linear and power regression functions were tested for the relationships mentioned above, using the coefficient of determination (r^2) as a goodness index.

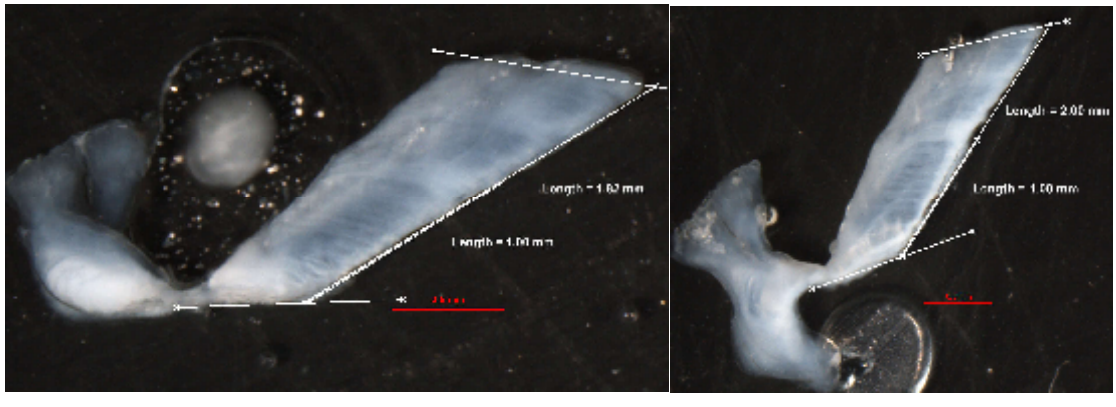


Figure 6.3. Otoliths transverse sections of an Atlantic bluefin tuna aged 3 years old. Left photo shows the section before the antirostrum and the right one shows the section obtained at the tip of atirostrum.

Comparing age estimates among readers and among calcified structures: agreement, precision and relative accuracy

Comparisons of age estimates between readers for spine and otolith and between calcified structures were carried out. Readers were scored into two categories

according to their reading experience as high and low experienced readers. Age readings were analysed using the method developed by Eltink et al. (2000). This analysis compares estimated ages from each reader with the modal age, i.e. the best approach available to the true age. In the present study, the modal age considered was the consensus between readers for each structure (inter-reader comparisons) and estimates from spines sections (inter-structures comparison). Two indices were used to estimate spine and otolith ageing precision among readers and structures, the Average Percent Error (APE) and the Coefficient of Variation (CV). APE was estimated by using the Beamish (1981) recommended formula as follow:

$$APEj = 100 \times \frac{1}{R} \sum_{i=1}^{R} \frac{[X_{ij} - \overline{X}_j]}{\overline{X}_j}$$

where Xij is the ith age determination of the jth fish, $\overline{X}j$ is the mean age estimate of the jth fish and R is the number of times each fish was aged.

The mean CV was estimated by using the European Fish Ageing Network (EFAN) software (Eltink et al., 2000) recommended formula:

$$CV = \frac{100}{n} \frac{\sqrt{\sum_{i=1}^{R} \frac{(Xij - \overline{Xi})^2}{R-1}}}{\overline{Xi}}$$

where n is the number of spine or otolith, R is the number of readers, Xij is the j value of age estimation for spine or otolith i and $\overline{Xi}$ is the average age calculated for each structure.

Moreover, in the absence of known age specimens, the relative accuracy was estimated by the relative bias. This bias is a systematically over or underestimation of age compared to the modal age. Another statistic estimated by the Eltink's workbook and related to the precision, is the weighted mean percentage agreement (PA), which indicates agreement with respect to the modal age. For testing differences in estimates among readers, an inter-reader bias test was also applied.

6.3 Results and Discussion

Relationship between otolith and spine size and fish length

Biometric relationships for otoliths are described in Figure 6.4. The goodness of fit was high, despite increasing variation in data in all size-length relationships for fish over 180 cm SFL. Regression functions showed high determination coefficients (r^2), with better potential than linear relationship, between otolith size (length, height and weight) and fish length. Also a curvilinear relation was detected between the length of the ventral otolith arm and fish length. Comparing the otolith ventral arm length of the three sections in the core area, a decreasing trend in length was found as the section was closer to the front of the otolith or rostrum.

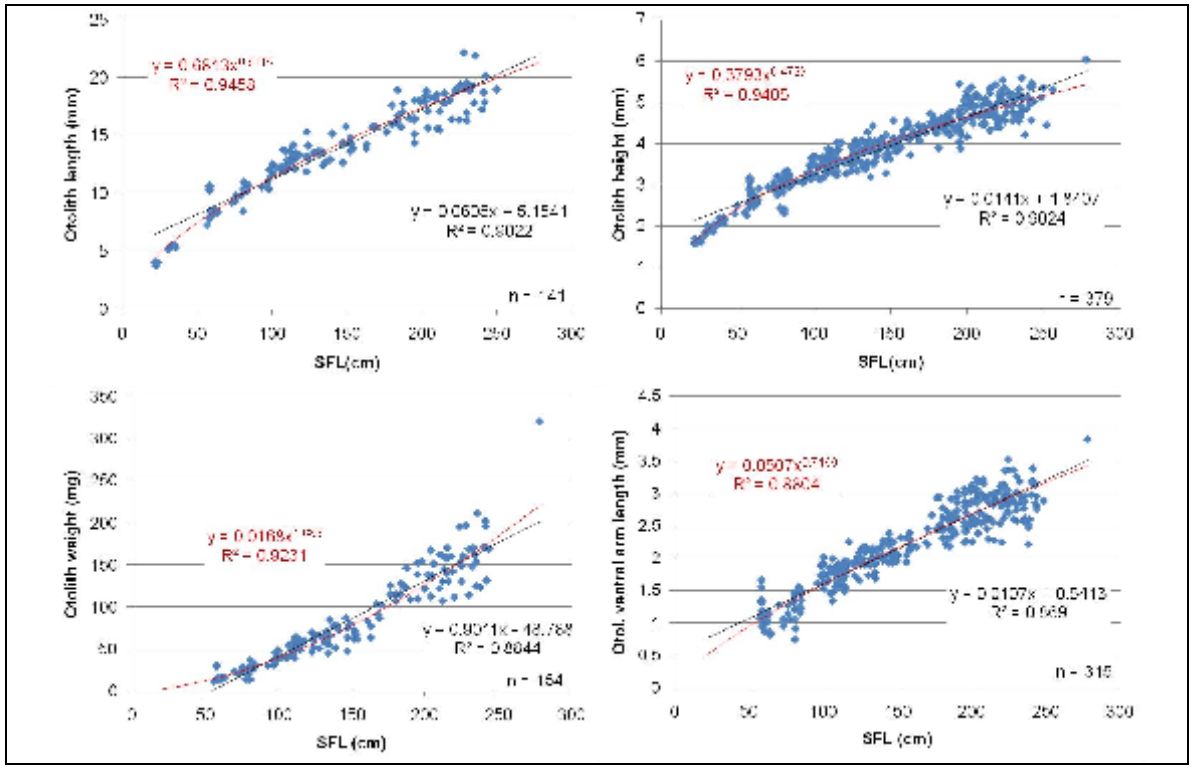


Figure 6.4. Biometric relationships between otolith measurements and Atlantic bluefin tuna straight fork length (SFL).

Both linear and power equations fit adequately the spine length and diameter versus fish length relationship (Figure 6.5). The goodness of fit between the spine

diameter and SFL showed that the fish body length and the size of the calcified structure were closely related.

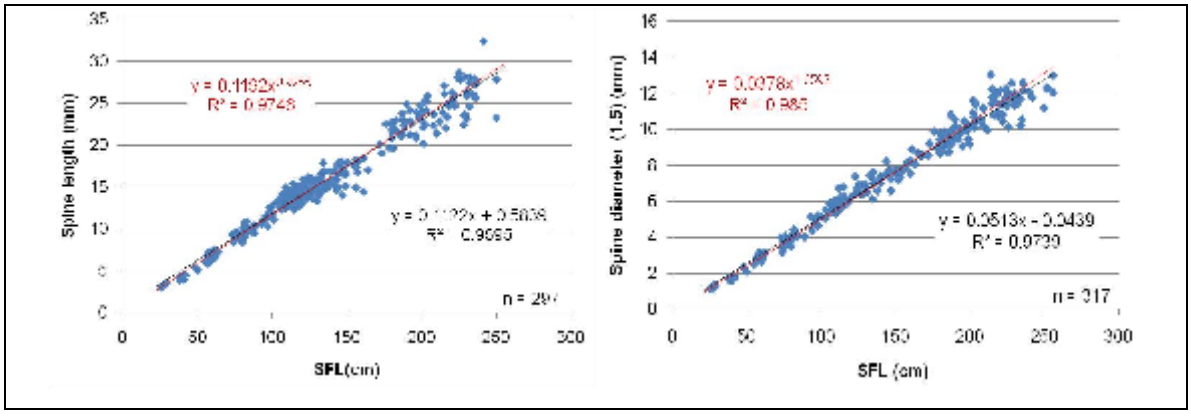


Figure 6.5. Biometric relationships between fin spine length and diameter versus Atlantic bluefin tuna fork length (SFL).

Precision of age estimates

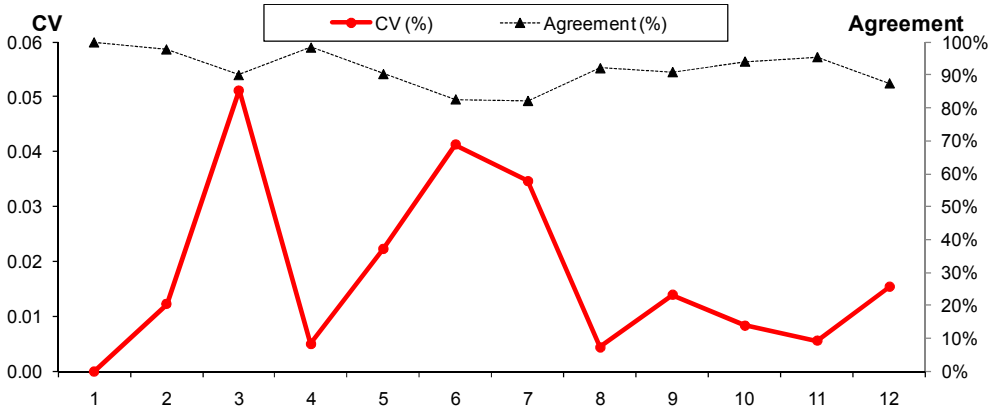
To estimate the reproducibility of age estimates for each calcified structure (i.e. ageing precision), the APE and CV were estimated for all comparisons analyzed (Table 6.2). Overall, for each calcified structure, both indices were low for the inter-reader comparisons analyzed, with CV values of 1.9% for spines and 2.2% for otoliths, corresponding to an APE of 1.55% and 1.52%, respectively. Whereas, when age estimates between spine and otolith coming from the same specimens were compared, CV and APE were slightly higher, with values of 7.1% and 4.34%, respectively for Reader 1 (PL); 7.1% and 5.28% for Reader 2 (ER) and 7.8% and 5.53% for otolith readings combined.

Table 6.2. Summary of parameters obtained from the five inter-reader comparisons analysed. The table shows Relative bias, Coefficient of variation (CV), the Average percent error (APE), Percent agreement (PA) and p significance level of the inter-reader bias test (n. s.: not significant).

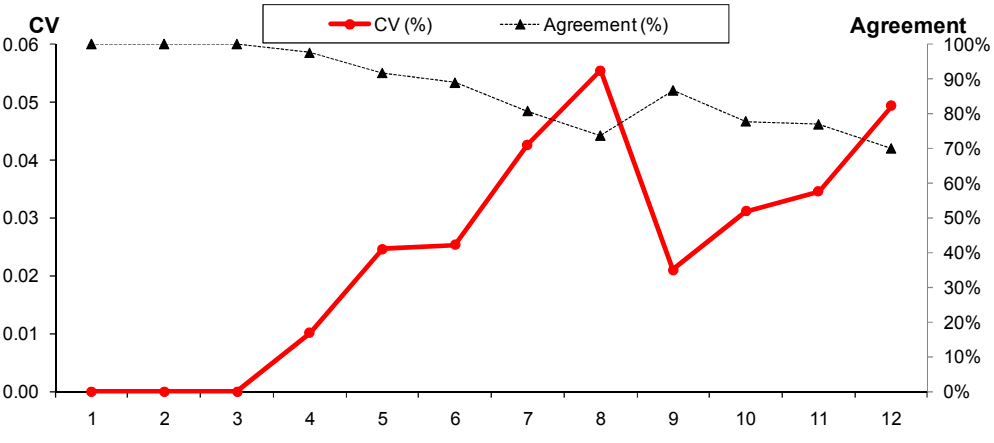
						Relative accuracy		Precision			
Readers	comparison	calcified structures	Reader experience	n	Age range	Relative bias		CV (%)	APE (%)	PA (%)	Inter-reader bias test
MR_PL	between readers	spine	high	243	1-12	0.02		1.9	1.55	91.4	n.s.
PL_ ER	between readers	otolith	low	194	1-12	0.04		2.2	1.52	88.4	n.s.
PL	between spine consensus and otoliths	both	high/low	181	1-12	-0.03		7.1	4.34	73.5	n.s.
ER				170	1-12	0.02		7.1	5.28	71.8	n.s.
PL-ER combined				187	1-12	-0.02		7.8	5.53	78.6	n.s.

Furthermore, the CV estimated in each comparison did not show an increasing pattern with age throughout the age range for all well sampled ages (i.e. from 1 to 11years old), except for the comparison between otolith readings (Figure 6.6).

a. Between readers (PL vs MR) for spines



b. Between readers (PL vs ER) for otoliths



c. Between spines (consensus) and otoliths (combined)

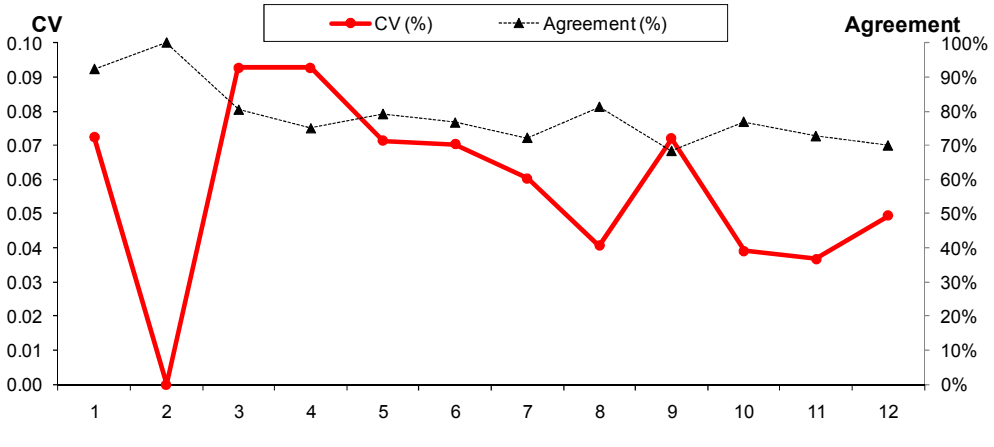


Figure 6.6. The coefficient of variation (CV%) and percent agreement plotted against MODAL age.

The inter-reader bias test was non-significant in all the comparisons. An overall high PA was achieved in all comparison analyzed, ranging from 71.8% to 91.4% with a low overall relative bias from -0.03 to 0.04, indicating a good proxy in terms of relative accuracy in age estimation using either both structures, particularly important in the absence of calcified structures of known age.

Age estimates

Otoliths age length key (ALK) is displayed in Table 6.3. This ALK was built with the age interpretation from two readers. In samples where both readers did not coincide, the final estimated age was selected from the reader who gave better quality to the sample, obtaining a combined reading. The target objective for sampling 10 specimens by 10 cm length range was nearly achieved. The ages covered in this ALK are remarkable, since young ages, mainly 1 to 6 years old samples, are difficult to interpret in bluefin tuna (Rodriguez-Marin et al., 2007). This difficulty in interpreting the age of young specimens also occurs in southern bluefin tuna (*Thunnus maccoyii*) age estimations from this calcified structure (Anon, 2002).

Table 6.3. Age-length key based in age interpretation from Atlantic bluefin tuna otolith sections. Numbers represent percent by number by length class (SFL, cm).

	Age class																			
Length class	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	*n*
20-30	100																				10
30-40	100																				10
40-50	100																				6
50-60	20	80																			15
60-70		100																			11
70-80		54	46																		13
80-90		38	52	10																	21
90-100			17	75	8																12
100-110			21	53	21	5															19
110-120				20	53	27															30
120-130				6	56	38															16
130-140					42	46	13														24
140-150					24	24	47	6													17
150-160						23	69		8												13
160-170						8	8	58	25												12
170-180								25	50	25											8
180-190								10	52	38											21
190-200								5	38	33	19	5									21
200-210									15	45	30	5	5								20
210-220										29	48	24									21
220-230									15	15	30	20	20								20
230-240											20	45	20	15							20
240-250											20	50		20			10				10
250-260												100									3
260-270																					
270-280																				100	1
Total *n*	29	38	23	28	44	34	21	13	33	35	32	28	9	5			1			1	374

0-20%
20-50%
50-100%

For spines, the ALK is presented in Table 6.4. Each spine was read by two readers and a reading consensus was used for age interpretation discrepancies. The target objective for sampling 10 specimens by 10 cm length range was achieved in almost all length classes.

Table 6.4. Age-length key based in age interpretation from Atlantic bluefin tuna spine sections. Numbers represent percent by number by length class (SFL, cm).

	Age class																			
Length class	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	*n*
20-30	100																				10
30-40	100																				10
40-50	100																				6
50-60	10	90																			10
60-70		83																			6
70-80		25	75																		12
80-90			85	15																	34
90-100				100																	10
100-110				38	56	6															16
110-120				11	64	25															36
120-130					31	66	3														29
130-140					4	82	14														28
140-150					7	43	43	7													28
150-160						42	47	11													19
160-170						8	42	50													12
170-180							25		75												4
180-190								25	60	15											20
190-200								12	35	47	6										17
200-210								7	7	40	27	20									15
210-220										29	57	7	7								14
220-230										27	13	47	7	7							15
230-240											53	33	13								15
240-250											33	33		33							3
250-260											33	50	17								6
260-270																					
270-280																					
Total *n*	14	17	39	25	44	73	32	18	22	25	26	20	5	2			0			0	375

0-20%
20-50%
50-100%

Comparison between age estimates from different calcifies structures coming from the same specimen.

The mean age estimates between spines and otoliths from samples collected from the same specimen, plotted as a function of the otoliths-based age are shown in Figure 6.7. The comparison of otoliths and spines age interpretation showed a good fit to a linear relationship between both age estimations, indicating good age agreement.

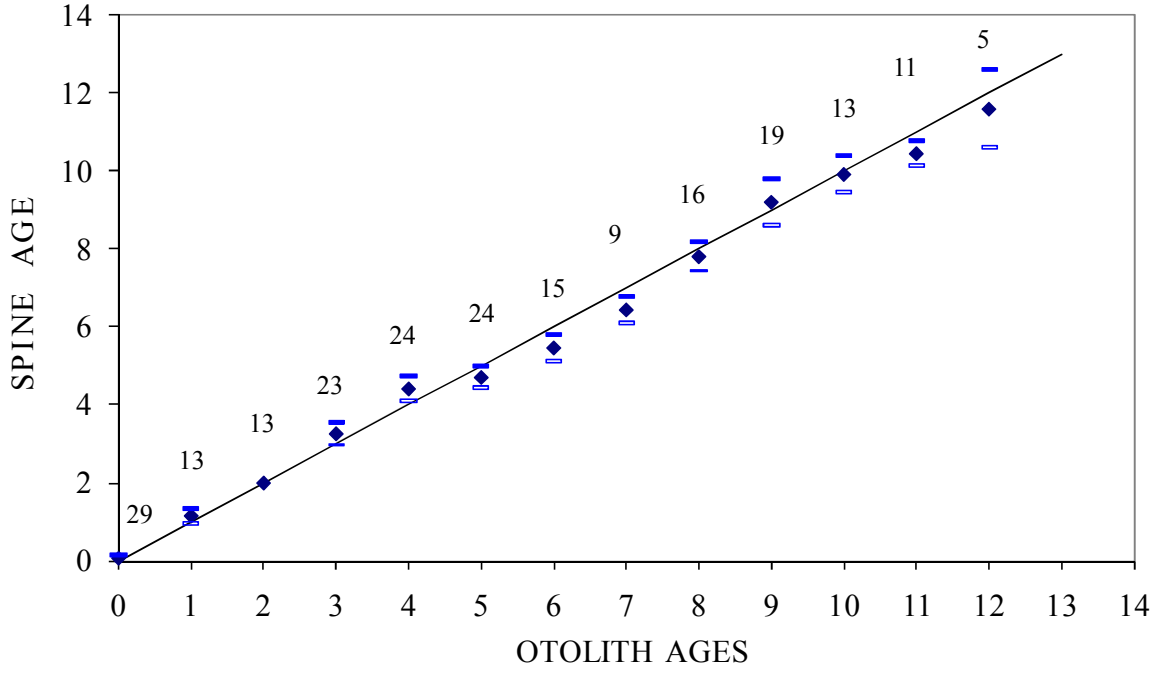


Figure 6.7. Bias comparison between spine and otolith age interpretations. Spines age readings are presented as the mean age and 95% confidence interval corresponding to otolith age readings (numbers above values represent number of calcified structures used, total number: 214).

Conclusions

Inter-reader agreement was high in either both spines and otoliths in the present study. Precision between both structures age readings, described by Coefficient of Variation (CV) and Average Percent Error (APE), was also high with low values of both indices.

The good age agreement between age estimations from spines and otoliths from the same fish, indicates that both structures could be used for age determination of Atlantic bluefin tuna for the age ranges analyzed in this project.

None of these two calcified structures can be excluded for routine direct ageing because in certain fisheries, fish processing or fish market practices would hinder the sampling of either structure. This means that more ageing comparison studies

are needed on the calcified structures of the same specimen, increasing the sampling of specimens larger than 250 cm SFL.

Recommendations for future sampling for age interpretation from calcified structures

Nearly 5% of the whole first dorsal fin spines were considered unusable because they were broken or fractured at the base of the structure. This damage to the structure could have been due to a non proper method of extraction of the whole spine from the fish or to a crushing when samples were stored in the envelopes. For future sampling, special care should be taken in order not to twist the spine structure in its base during the extraction (as it was recommended in the project sampling protocols). If required, it is recommended to use a sharp knife or scalpel to cut the strong ligaments that support the spine base deep in the fin insertion.

It is also important that the date of specimen collection refers to the date of the capture of the fish, not to the date when the biological sampling was conducted. For age interpretation we need to apply a type of border criterion at the edge of the structure and this implies to take into account the real date of capture for final age estimation.

References

- Anon. 2002. A manual for age determination of southern bluefin tuna *Thunnus maccoyii*. Otolith sampling, preparation and interpretation. In: Report of the Direct Age Estimation Workshop. 11-14 June 2002. Victoria, Australia. Meeting Report of the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). 36p.
- Beamish, R.J. 1981. Use of fin-rays sections to age walleye Pollock, Pacific Cod, and Albacore, and the importance of this method. Transactions of the American Fisheries Society 110, 287–299.
- Eltink, A.T.G.W., Newton, A.W., Morgado, C., Santamaria, M.T.G. and Modin, J. 2000. Guidelines and tools for age reading (PDF document version 1.0). Internet: http://www.efan.no.Rodriguez-Marin, E., N. Clear., J.L. Cort., P. Megafonou., J.D. Neilson., M. Neves dos Santos., D. Olafsdottir., C. Rodriguez-Cabello., M. Ruiz. and J. Valeiras. 2007. Report of the 2006 ICCAT Workshop for bluefin tuna direct ageing. ICCAT Collective Volume of Scientific Papers 60(4): 1349-1392.

- Luque, P.L., Rodríguez-Marin, E., Ruiz, M., Quelle, P., Landa, J., Macías, D. 2012. A review of direct ageing methodology using dorsal fin spine from Atlantic bluefin tuna (*Thunnus thynnus*). (Fish and Fisheries, under review).
- Rodríguez-Marín, E., N. Clear., J.L. Cort., P. Megafonou., J.D. Neilson., M. Neves dos Santos., D. Olafsdottir., C. Rodriguez-Cabello., M. Ruiz. and J. Valeiras. 2007. Report of the 2006 ICCAT Workshop for bluefin tuna direct ageing. ICCAT Collective Volume of Scientific Papers 60(4): 1349-1392.
- Rodriguez-Marin, E., Luque, P. L., Ruiz, M., Quelle, P. and Landa, J. 2012. Protocol for sampling, preparing and age interpreting criteria of Atlantic bluefin tuna (*Thunnus thynnus*) first dorsal fin spine sections. ICCAT Collective Volume of Scientific Papers (in press). SCRS/2011/175.
- Rooker, J. R., Alvarado Bremer, J.R., Block, B.A., Dewar, H., De Metrio, G., Corriero, A., Kraus, R.T., Prince, E.D., Rodriguez-Marin, E. and Secor, D.H.. 2007. Life history and stock structure of Atlantic bluefin tuna (*Thunnus thyunnus*). Reviews in Fisheries Science 15:265–310.

7. HISTOLOGICAL ANALYSES OF GONADS OF BLUEFIN TUNA

Task Leader: Antonio Medina (UCA)

Participants

AZTI: Nicolas Goñi, Igaratza Fraile, Hilario Murua

IEO: David Macías

7.1 Introduction

The study of ovary maturation has become increasingly important in the study of fish population dynamics, especially in the case of a species spawning in constrained areas and periods, such as bluefin tuna. Microscopic analyses of bluefin tuna gonads can be used for a variety of purposes, e.g. focused on identification of spawning grounds along the Mediterranean Sea, determination of maturation stages – and of their variability within the Mediterranean –, and fecundity estimates.

The present report summarizes the histological analyses conducted under this project both for females and males gonads.

7.2 Material and Methods

A total of 275 gonads have been sampled (Table 7.1). Initially, for 2011 it was proposed to analyze a subset of 20 samples collected in longline fisheries in Crete, Malta, South of Sicily and Ionian Sea, and southern Spain. However, the final selection of samples to analyze histologically was constrained by sample availability (analyses were conducted on all strata where gonads were sampled).

As a result, a histological analysis was conducted on 189 gonads to determine their reproductive status (Table 7.1).

Table 7.1: Number of gonads sampled in the Central and Western Mediterranean and in the Eastern Atlantic, and number of gonads analyzed.

Partner code	Area	Fishing gear	Nº of gonads sampled			Nº of gonads histologically analyzed	
			Females	Males	Indeterminate	Females	Males
FMAP	Malta	Longline & purse seine	6	4		6	
IEO	Baleares	Longline	23	20		23	20
UNICA	Sardinia	Trap	26	25		26	
UNIGE	Ligurian Sea	Longline	33	20	37	33	
IEO	Gibraltar Strait	Trap	46	35		46	35
TOTAL			134	104	37	134	55

7.3 Results

The classification schemes used in the present report (modified from Schaefer, 1996) are shown in Tables 7.1 and 7.2 for females and males, respectively. Depending on the most advanced cohorts of oocytes encountered in the ovary (MACOs), bluefin females can be classed into one of five maturation stages. The ovaries of resting (quiescent) or pre-maturing (early vitellogenic) females, which contain previtellogenic or small vitellogenic oocytes as the MACOs, are reproductively inactive. The ovaries of active non-spawning females contain advanced vitellogenic oocytes and minor, if any, α atresia. Active females are classified as active spawning if the ovaries show additional evidence of either recent spawning (postovulatory follicles are present) or imminent spawning (migratory-nucleus or hydrated oocytes can be identified in the ovarian parenchyma). Females are considered inactive mature when they have entered into regression following a phase of reproductive activity, in which case the ovary encloses either previtellogenic or early yolked oocytes plus α and/or β atresia, or advanced yolked oocytes plus major atresia (Table 7.2).

Male tuna are considered mature when a significant amount of spermatozoa are present in the sperm duct (Table 7.3).

Table 7.2. Classification of ovaries and reproductive state of female bluefin tuna.

Stage	Histological description (females)	Physiological state
1 - Quiescence	Perinucleolar oocytes Lipid stage oocytes	Inactive
2 – Early vitellogenesis	Early vitellogenesis	Inactive
3 - Late vitellogenesis	Late vitellogenesis	Active, non-spawning
4 - Spawning	Migratory nucleus, GV breakdown and/or Hydrated oocytes and/or Post-ovulatory follicles	Active, spawning
5 - Spent	α atresia ($\geq$50%) of late vitellogenic oocytes	Inactive

Table 7.3. Classification of testes and reproductive state of male bluefin tuna.

Stage	Histological description (males)	Physiological state
1 - Quiescence	Germinal cysts containing spermatogonia, few spermatocytes, and rare spermatids and spermatozoa	Inactive
2 - Early spermatogenesis	All stages: spermatogonia; increase of spermatocytes and spermatids; few spermatozoa	Inactive
3 - Late spermatogenesis	Many spermatid cysts; abundance of spermatozoa; sperm in lumina	Active
4 – Mature/Spawning	Lumen of seminiferous tubules and main sperm duct filled with spermatozoa	Active
5 - Spent	Residual spermatozoa in lumina	Inactive

Specimens from the Strait of Gibraltar.

Samples from migrating bluefin tuna in the area of the Strait of Gibraltar (eastward run: tuna swimming from the Atlantic to Mediterranean spawning grounds) were collected from trap catches in the Strait of Gibraltar between the 1st and the 15th of June, 2011. A total number of 81 bluefin tuna (35 males and 46 females) were sampled and histologically analyzed.

Females

46 ovaries were examined. Nine of these specimens (**19.6**%) showed late vitellogenic oocytes and lacked maturation oocytes or postovulatory follicles, whereas the amount of atretic follicles was low or moderate (from 7% to 44%, average of 29.7%) (Fig 7.1A). Therefore, these individuals were classified as reproductively **active but non-spawning**, a reproductive state that reflects the **natural condition** of migrating bluefin tuna spawners as they pass through the Strait of Gibraltar (Medina et al., 2002). The thirty seven other individuals examined (**80.4**%) were found to be inactive **(spent)**, as their ovaries contained abundant α-atresia (≥50%) of large yolked oocytes, indicating that they were entering regression (Figure 7.1B). It is impossible from a mere histological evaluation to ascertain whether these specimens would be capable to resume ovarian maturation and eventually spawn later on through the season or they would skip the spawning function altogether.

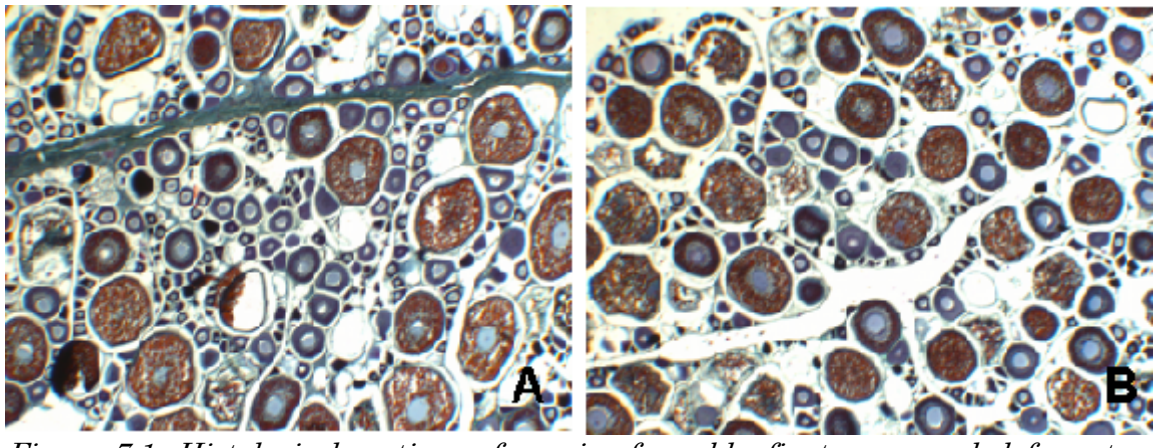


Figure 7.1. Histological sections of ovaries from bluefin tuna sampled from trap catches, showing previtellogenic and vitellogenic oocytes; the frequency of α-atresia of vitellogenic follicles can be from low to moderate (<50%) in active, non-spawning fish (A) or abundant (>50%) in inactive, mature individuals (B).

Males

All the 35 testes examined were histologically **active**, showing all developing stages of the male germ cell line (from spermatogonia to late spermatids and spermatozoa) at the cortical region (proliferative region). Some of these fish appeared to be in the middle of the spermatogenetic process, showing abundant spermatocyte cysts as well as cysts containing all spermatid stages (Figure 7.2A), whereas others were apparently in advanced spermatogenesis as the flagellate spermatid cysts were clearly predominant over earlier spermatogenic stages. In all cases the lumina of the testicular lobules, the secondary duct system and the main sperm duct were fully packed with sperm masses (Fig. 7.2B). The histological structure of these testes was similar to that described previously in male bluefin tuna spawners captured in traps as they enter the Mediterranean Sea to spawn.

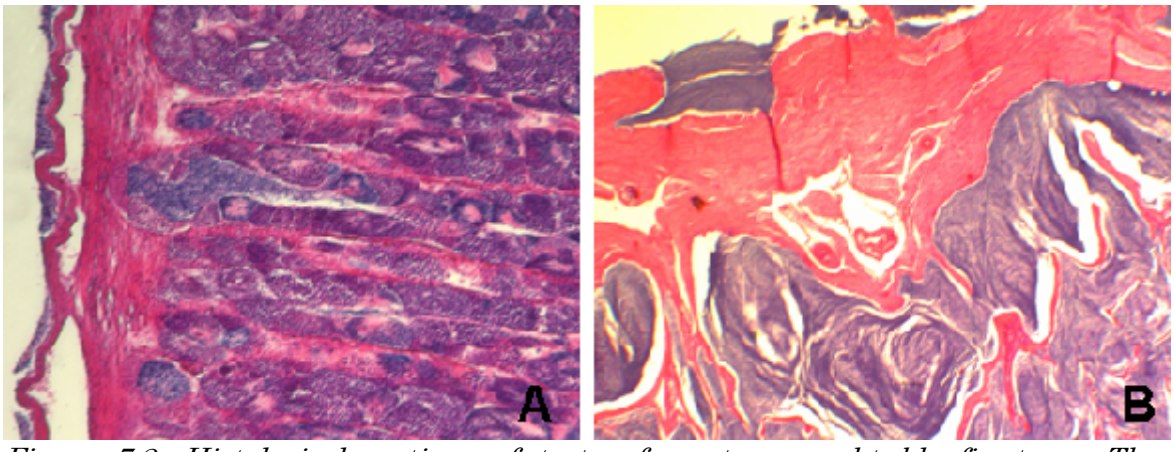


Figure 7.2. Histological sections of testes from trap-caught bluefin tuna. The periferal region contains male cell cysts at all developmental stages (A), and the network of sperm ducts appear full of densely packed sperm masses (B).

Specimens from the Balearic Islands. Long-line fishery.

A total number of 43 bluefin tuna (20 males and 23 females) were sampled and histologically analyzed from the longline fishery in the Balearic Sea between the 15th of August and the 27th of September, 2011.

Females

All the ovaries analyzed (n = 23) were quiescent (inactive), containing only previtelogenic (perinucleolar stage) oocytes (Figure 7.3). Histological analyses did not allow us to determine whether the reproductively inactive state observed in all females was due to immaturity (young age) or just reflected a physiological state of quiescence. The latter possibility appears most likely, as the size range of the sampled specimens exceeds the first-maturity size established for eastern bluefin tuna (Corriero et al., 2005).

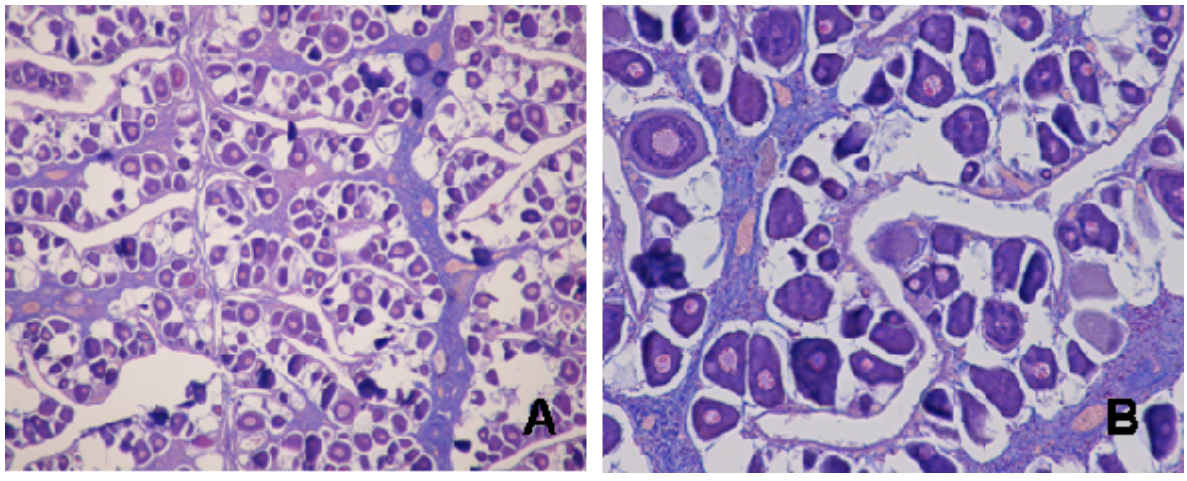


Figure 7.3. Histological sections of ovaries from bluefin tuna captured by longline. The most advanced oocytes found in the whole sample are lipid-stage oocytes (primary oocyte growth). No vitellogenic oocytes were present in the whole sample.

Males

All the males examined but one (n = 19) were inactive (spent), displaying lumina of testicular lobules and ducts either completely empty or containing only residual spermatozoa (Figure 4B). No active germ cell cysts were found in the peripheral region of the testes (Figure 4A). The presence of residual sperm in the testicular ducts along with the large size of the sampled specimens lead us think that these fish had started maturation but further they went spent. One of the males sampled posed some doubts as conspicuous masses of sperm were enclosed in the central testicular ducts, though neither developing germinal cysts nor spermatozoa were observed in the distal end of the testicular lobules.

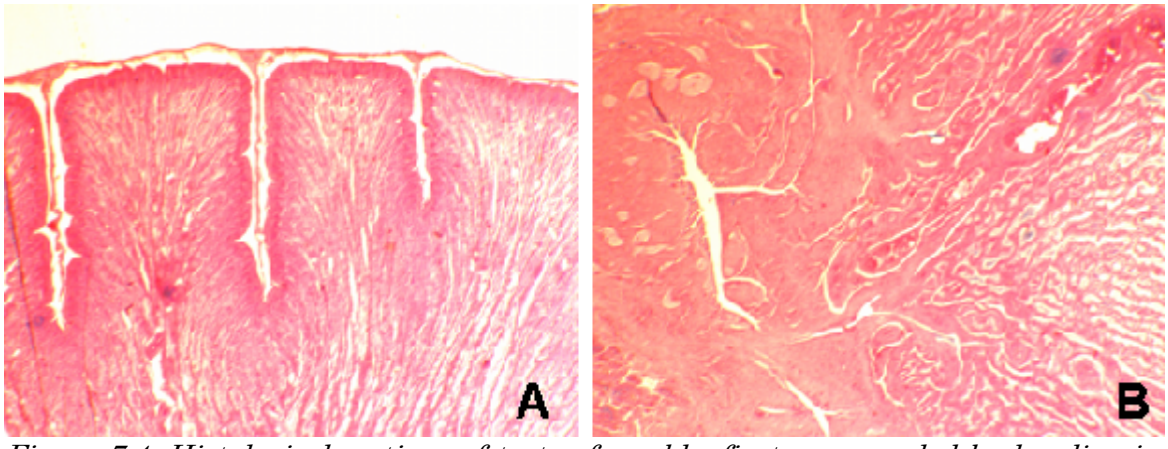


Figure 7.4. Histological sections of testes from bluefin tuna sampled by longline in the Balearic Archipelago. No proliferative activity is observed at the peripheral region underneath the tunica (A): cysts of male germ cells are absent and the lumen of testicular lobules appear empty at their blind ends (A). The central main sperm duct and the network of secondary ducts are empty or contain only residual spermatozoa (B).

Specimens from Malta. Purse Seine and Longline fishery.

A total number of 10 bluefin tuna (6 females and 4 males) were sampled but only the 6 females were histologically analyzed. The fish were caught at sea with purse seine gear the 10th of June, and sampling occurred at cages between the 26th of June and the 19th of July 2011. The size range of the samples individuals ranged between 146 and 240 cm which can be considered adults individuals. On these samples, only females were analysed.

Females

All except one ovaries analyzed were in an inactive physiological stage, containing only previtelogenic (perinucleolar stage and cortical alveoli stage) oocytes. One of those females showing cortical alveoli oocytes as the most advanced oocytes had high levels of atresia (> 50 %) which is interpreted as female entering the resting stage after the spawning season and, hence, is classified as resting mature inactive female. One female showed early/late vitellogenesis in conjunction with high levels of atresia which is also a signal of a female in a resting inactive mature phase.

Although histological analysis of females in previtellogenic stages did not allow us to determine whether the reproductively inactive state observed in all females was due to quiescence or immaturity, or just reflects a physiological state of resting, the latter possibility appears most likely because 2 females showed high levels of atresia and the size range of the sampled specimens exceeded the first-maturity size established for eastern bluefin tuna (Corriero et al., 2005).

Specimens from Sardinia. Trap fishery.

A total number of 51 bluefin tuna (26 females and 25 males) were sampled but only the female ovaries were histologically analyzed from the Italian Trap fishery in the Sardinian Sea between the 14th of May and the 11th of June 2011. The size range of the sampled individuals ranged between 118 and 231 cm-s. On these samples, only females were analysed.

Females

16 out of 26 ovaries analysed (61.5%) showed late vitellogenic oocytes as the most advanced oocytes without presence of matured oocytes or postovulatory follicles, whereas the amount of atretic follicles was low or moderate (less than <50 % of atresia in 11 individuals and > 50 % in 5) (Fig 5a, b). Therefore, the individuals classified with low level of atresia can be considered reproductively active but non-spawning females, a reproductive state that may reflects the natural condition of migrating bluefin tuna spawners (Medina et al., 2002) before spawning or the progress to a postspawning condition (i.e. the atresia levels will increase as time progressed). However, the individuals in late vitellogenesis with high levels of α-atresia (≥50%) can be considered to be inactive mature females (spent) in a recovery stage, as their ovaries contained abundant of large yolked atretic oocytes, indicating that they were entering regression (Figure 5b). In both cases, it is impossible from a mere histological evaluation to ascertain whether these specimens would be capable to resume ovarian maturation and eventually spawn later on through the season or they would skip the spawning function altogether.

The rest of the ovaries analyzed (n = 10) contained only previtelogenic (perinucleolar stage and cortical alveoli oocytes with low levels of atresia. Although it is difficult to ascertain histologically between immature and resting individuals, as those samples had very low levels of atresia, the gonad was compact and the gonad wall was thin it can be considered that those samples were in immature stage.

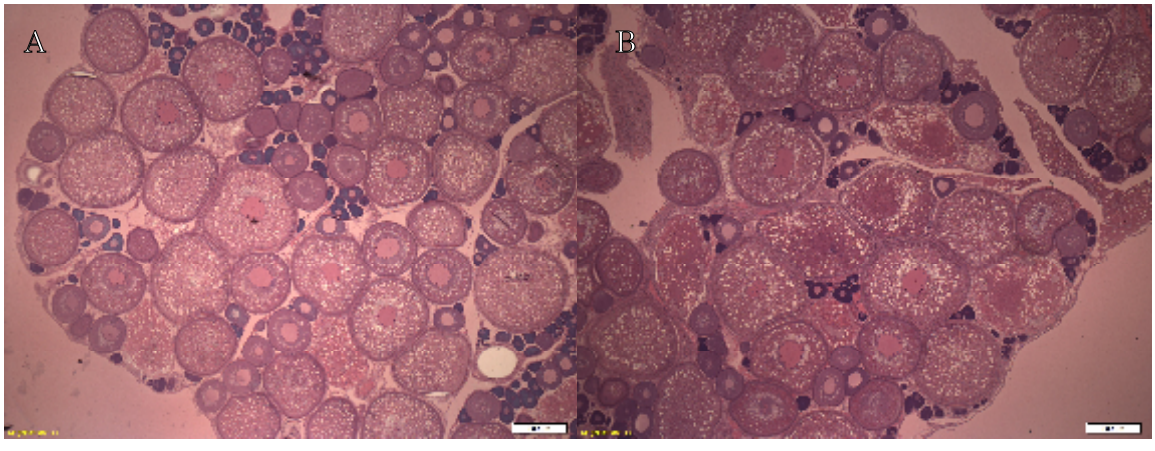


Figure 7.5.- (A) bluefin tuna gonad presenting late vitellogenic oocytes as the most advanced oocytes with atretic levels less than <50 % and (B) bluefin tuna gonad presenting late vitellogenic oocytes as the most advanced oocytes with atretic levels less than > 50 %.

Specimens from the Ligurian Sea. Longline artisanal fishery.

A total number of 90 bluefin tuna (33 females, 20 males, and 37 indeterminate) were sampled from the Italian artisanal Longline fishery in the Ligurian Sea between the 17th of May and the 1st of September 2011. Only females were histologically analyzed because the ovaries of males and indeterminate fish were too small to get any valuable information. The size range of the samples individuals ranged between 75 and 154 cm, however, the samples size range varied between sampling periods being in the range of 75-115 cm in May and June whereas varied from 100 and 154 cm between July and September. On these samples, only females were analysed.

Females

32 out of 33 ovaries analysed (94 %) contained only previtelogenic (perinucleolar stage and cortical alveoli oocytes) with low levels of atresia and, therefore, were considered to be quiescence or immature females. Although it is difficult to ascertain histologically between immature and resting individuals, as those samples had very low levels of atresia, the gonad was compact, the gonad wall was thin and the size was lower than the length at first maturity (Corriero et al., 2005), thus they can be considered as immature females. One individual examined in July 2011 with a length of 154 cm showed late vitellogenesis with high levels of α-atresia (≥50%) and was found to be inactive mature in a spent condition.

7.4 Discussion

The present results from the histological analysis of bluefin tuna caught by trap as they enter the Mediterranean Sea to spawn (eastward run) are quite unexpected. Albeit the male reproductive organs were **apparently active** and similar in structure to what has been reported earlier (Abascal et al., 2004), apparent signs of reproductive **impairment** were present in most of the sampled ovaries. An extremely high (80.4%) percentage of females were sexually mature but inactive, showing ovaries that had entered into regression, reabsorbing numerous vitellogenic oocytes through atretic processes. These observations are inconsistent with others made on eastward migrant bluefin tuna caught in the Strait of Gibraltar by trap (Medina et al., 2002). The most likely reason for the present results is that a great number of the tuna captured in 2011 (well in excess of the quota assigned to Spanish traps) were retained for **days to weeks in the trap facilities**. Stressful conditions of captivity along with the interception of the natural migratory route could probably result in the interruption of gametogenesis (mostly oogenesis) and further resorption of oocytes. Under these circumstances, the fish sampled in 2011 **should not be considered as representative of the normal reproductive population**. And even in the case that trap-caught bluefin tuna are sacrificed immediately after capture, we do not foresee significant new information from trap fisheries in terms of reproductive assessments.

The bluefin tuna captured by longline in the **Balearic** Sea were **quiescent**. These fish are assumed to have spawned in the reproductive season (June-July), entered into regression in late July-August, and finally reached the resting stage by August-September. Similarly, the bluefin tuna captured by longline and purse seine in the **Maltese** Sea were **quiescent**. These fish are assumed to have spawned in the reproductive season (June-July), entering the regression in late July. In future studies aiming at characterizing and assessing reproductive traits of bluefin tuna eastern populations, spawners **should be sampled throughout the reproductive season**. Although samplings of bluefin tuna in the Mediterranean Sea after July are useful to comply with other objectives of the ICCAT-GBYP 06/2011 project, they contribute no essential information to the understanding of the bluefin tuna reproductive biology. The bluefin tuna captured by artisanal longline in the **Ligurian** Sea were **mostly immature** and small individuals which are not contributing with essential information to understand the reproductive biology of this species.

The bluefin tuna captured by the trap fishery in the **Sardinian Sea in May and June** were mostly in advanced vitellogenic stages. Although some of them showed high level of atresia (postspawning and resting condition), most of them (42 % of the total sampled females) showed low levels of atresia which could be interpreted as fishes in **mature active condition** and capable of spawning during the current season (July). Those samples could be used to give some information about the reproductive biology of bluefin tuna.

The transfer of a big deal of the longline quota to purse seine and trap fisheries has reduced the availability and suitability of these samples for reproductive studies.

Histological samples of gonads from several areas have not been obtained because some of the fisheries were already closed at the onset of the project, or quotas were transferred in some key fisheries, making the sampling impossible or yielding gonad samples **out of the reproductive season**. Due to the spatially and temporally constrained reproduction of the bluefin tuna, there are specific areas of its broad distribution range that prove to be essential for the understanding of the species' reproductive biology and the determination of the reproductive potential of the

stocks and the age of first maturity. Of particular interest are the spawning grounds (Balearic Archipelago, South Tyrrhenian Sea-Malta, and Levantine Sea), where the preferred sampling gear is the purse seine, which specifically targets schools of breeders at their reproductive peak. Specimens captured by longline shortly before, during and shortly after the reproductive season can be very useful as well. Further efforts should be made to allow scientifically meaningful sampling onboard purse-seine and long-line boats.

References

- Abascal FJ, Megina C, Medina A (2004) Testicular development in migrant and spawning bluefin tuna (*Thunnus thynnus* (L.) from the eastern Atlantic and Mediterranean. Fish Bull 102:407–417
- Corriero A, Karakulak S, Santamaria N, Deflorio M and 11 others (2005) Size and age at sexual maturity of female bluefin tuna (*Thunnus thynnus* L. 1758) from the Mediterranean Sea. J Appl Ichthyol 21:483–486
- Medina A, Abascal FJ, Megina C, García A (2002) Stereological assessment of the reproductive status of female Atlantic northern bluefin tuna, *Thunnus thynnus* (L.), during migration to Mediterranean spawning grounds through the Strait of Gibraltar. J Fish Biol 60:203–217
- Schaefer KM (1996) Spawning time, frequency, and batch fecundity of yellowfin tuna, *Thunnus albacares*, near Clipperton Atoll in the eastern Pacific Ocean. Fish Bull 94:98–112

8. SUMMARY AND RECOMMENDATIONS

This section summarizes the main achievements of the project, as well as the main difficulties encountered and some future prospects and recommendations.

The consortium aimed to sample 1950 individuals and get 6350 biological samples (1950 genetic samples, 1900 otoliths, 1900 spines and 600 gonads). At the end of the project, the consortium had sampled a total of 1916 bluefin tuna (10 larvae, 239 YOY, 446 juveniles, 552 medium size fish, and 669 large fish) from different regions (188 from the East Mediterranean, 270 from the Central Mediterranean, 732 from the Western Mediterranean, 597 from the Northeast Atlantic and 129 from the Central North Atlantic). From these individuals, 4309 biological samples were taken (1632 genetic samples, 1324 otoliths, 1078 spines and 275 gonads).

The consortium aimed to perform 950 NGS-TS analyses, 160 NGS-RRSG analyses, 600 microchemical analyses, 810 aging analyses and 80 histological analyses. At the end of the project, the consortium has conducted 919 NGS-TS analyses, 192 NGS-RRSG analyses, 600 microchemical analyses, 749 aging analyses and 188 histological analyses.

Following up on an earlier project, a panel of 384 SNPs was selected and assay genotyping plates ordered to ILLUMINA for the NGS-TS. DNA extraction was conducted and both NGS-TS and NGS-RRSG genotypes are available. Only the use of restricted panels of outlier and highly divergent SNP loci (from thousands to hundreds in the NGS-RRSG; only few tens in the NGS-TS) permitted to discriminate spawning population samples and assign individuals of feeding aggregates to originating populations. The NGS-TS approach did not clearly resolve genetic relationships among spawning populations and/or feeding aggregates. On the contrary, the NGS-RRSG permitted to unequivocally discriminate Age0 and larvae from the Gulf of Mexico and Western Mediterranean. However, since in this approach the feeding aggregations clustered as a third different group, the dataset and analyses need to be improved by adding more reference spawning samples and by selecting the most performing SNP loci, respectively. These improvements will

enhance the NGS-RRSG power for resolving population structure of the spawning ABFT populations and assignment of individuals composing the feeding aggregates.

Regarding microchemistry, 600 otoliths from the Central North Atlantic, Bay of Biscay, Strait of Gibraltar, Balearics, Malta, Sardinia and the Adriatic Sea have been already analyzed, results showing >99% of Eastern origin fish except in the Central North Atlantic were 84% of the fish were of Eastern origin.

Regarding age determination analyses, 374 otoliths and 375 spines have been interpreted already. Inter-reader agreement was high and age-length-keys were generated for both spines and otoliths. The comparison between ages estimated from different structures of the same specimen showed a good age agreement. This indicates that both structures may be used indistinctly for age determination of Atlantic bluefin tuna for the age ranges analyzed in this project.

A histological analysis was conducted on 188 individuals from the Strait of Gibraltar, Balearics, Malta, Sardinia and Ligurian Sea. However, the sampling was not adapted to the histological analysis (mostly occurring before and after the reproductive season), thus they contributed no essential information to the understanding of the Atlantic bluefin tuna reproductive biology.

When accomplishing the tasks of the project, the main difficulty came from the late announcement, resolution, and signature of the contract (which is linked to the first payment). By the time the contract was signed, several fisheries where closed due to having reached their quotas or due to time area closures (see Section 2.3). In the meantime, some partners were able to accomplish an important fraction of the proposed sampling, but in some cases it was difficult to allocate all necessary resources to accomplish this. Special sampling efforts were made to try to overcome this to the extent possible, and finally a large part of the proposed sampling was accomplished.

The late start of the project also affected most downstream analyses. Since these take considerable time (especially the optimization of genetic markers and assay ordering to third party companies), they needed to start (in general) before the

sampling was finished, thus the samples to be analyzed necessarily were selected among those that were available. As explained in Section 3, this can be overcome in subsequent years, given that important 2011 samples remain available for the future.

Overall, and in spite of the difficulties faced during this first year, the project already started to provide some results on population structure, catch composition, age structure and reproductive ecology.

For subsequent years, if this program continues, it is recommended to start several months prior to the start of the fishing season (e.g. February) so as to be able to appropriately organize the time for a general meeting of the whole consortium, improve coordination within the consortium, and avoid problems derived from late signature of the contract. In case of difficulty to start earlier, it would be desirable to be able to extend the deadline for the final report so as to be able to properly analyze the samples obtained.

On top of this, and because the sampling needs to go beyond the fraction of the population that is allowed to catch under the current management regime, it would be desirable if the administrations could facilitate the process of getting scientific permits for the project team to access the non-catchable fraction of the population (e.g. reference samples such as larvae and YOY individuals, as well as juveniles).

9. APPENDICES

9.1 Appendix 1: Sampling Protocols