FINAL REPORT FOR PHASE THREE OF THE ICCAT SHORT-TERM CONTRACT: SWORDFISH BIOLOGICAL SAMPLES COLLECTION FOR GROWTH, REPRODUCTION AND GENETICS STUDIES

Kyle Gillespie[1], Alex Hanke[1], Rui Coelho[2,3], Daniela Rosa[2,3], Oliana Carnevali[4], Giorgia Gioacchini[4], David Macias[5]

SUMMARY

This report details the third phase of biological sampling and associated analysis undertaken as part of an international swordfish biology program. The program was established in 2018 and sampling was conducted for swordfish in the North and South Atlantic and Mediterranean. Fish were sampled for size, sex, and maturity. Anal fin spines, otoliths, gonads, and tissues were obtained for ageing, growth, maturity and genetic analyses. These data will be used to inform ICCAT assessment and the ongoing management strategy evaluation process. In this report we examine sampling representativeness relative to spatial and temporal patterns in recent catch data. Samples were obtained from a broad temporal and spatial range, however, some improvements are required in spatial-temporal coverage.

KEYWORDS

swordfish, biological sampling, growth, reproduction, genetics, sampling representativeness

[1] Fisheries and Oceans Canada, Saint Andrews Biological Station, 125 Marine Science Drive, Saint Andrews, NB E5B 0E4 CANADA. Kyle.Gillespie@dfo-mpo.gc.ca
[2] IPMA – Instituto Português do Mar e da Atmosfera. Av. 5 de Outubro s/n, 8700-305 Olhão, Portugal
[3] CCMAR – Centro de Ciências do Mar da Universidade do Algarve, Campus de Gambelas 8005-139 Faro, Portugal
[4] Università Politecnica delle Marche, Ancona, Italy
[5] Instituto Español de Oceanografıa (IEO), Puerto Pesquero s/n. Fuengirola (Málaga)

Executive summary

In 2018, ICCAT's Swordfish Species Group initiated a biological sampling program in the North and South Atlantic and Mediterranean. The aim of the program was to collect biological data that would support research critical to the assessment and management of this highly migratory and internationally managed species. In the first phase of the program, an international group of institutions developed a sampling protocol, collected samples from 1810 fish, developed a relational database for sample data, and identified strategies for optimizing further sample collection and data analysis. In phase two of the program, 1433 samples were collected as well as initial analysis for the three project sub-areas: ageing and growth, reproduction and maturity, and genetics. In phase three of the program, 916 additional samples were collected and further analysis for the three project sub-areas continued. Fish were sampled for size, sex, and maturity; the location and date of capture (and/or landing) were recorded; calcified parts (otoliths and/or anal fin spines) and/or tissues were collected, processed, and archived for analysis. Samples were collected for each stock and in many cases these data were representative of the major swordfish catch locations and timing, however, in some locations and seasons additional sampling coverage is required. In this report we assess where and when additional sampling effort is needed, we provide basic analysis of data collected to date, and we suggest next steps for sample collection and analysis. These data will make contributions to our understanding of patterns of growth, maturity, movement and mixing among the three swordfish stocks under ICCAT management and will be critical for devising management plans that maximize swordfish yield and support stock productivity.

Introduction

Swordfish are an important fisheries resource in the Atlantic and Mediterranean but there remain significant unknowns regarding their basic biology, how the three stocks differ biologically, and the level of mixing between stocks. In 2018, ICCAT's Swordfish Species Group initiated a sampling program to collect biological data for swordfish in the Atlantic and Mediterranean. Sampling has continued in 2019, 2020, and 2021. The objective of this ongoing program is to improve knowledge of the stock distribution, age and sex, growth rate, age at maturation, maturation rate, spawning season and location and diet. It is expected that the program will contribute to the next major advance in the assessment of swordfish status by permitting the development of more spatially and biologically realistic population models used in both Atlantic and Mediterranean populations assessments and within the ICCAT Management Strategy Evaluation (MSE) for North Atlantic swordfish. This should translate into more reliable advice on stock status for an internationally and collectively managed resource.

Among all phases of this data collection program, 20 institutions from 14 ICCAT CPCs/Cooperating Non-Contracting Parties (**Appendix 1**) collected biological data via existing national fisheries at-sea monitoring programs and through targeted port sampling. This consisted largely of opportunistic collection of anal fin spines, tissues samples, otoliths, size, sex and maturity information. Samples were collected for all three Atlantic swordfish stocks in seven of the eight billfish sampling areas and across all four seasons.

This report provides an overview of the third phase of this program, basic analysis of data collected to date and recommendations for next steps for data collection and analysis of samples. We provide a review of sampling methodology and analysis of sample spatial and temporal representativeness relative to fishing effort for each stock. We also provide basic descriptions of length frequencies and sex ratios.

Program objectives

As indicated by the Swordfish Species Group, data collected in this program will support the following objectives:

- Resolve the spatial-temporal distribution of the three known swordfish stocks found within the Atlantic Ocean and Mediterranean Sea using a genetic analysis of tissue sampled from the catch of participating CPCs.
- Resolve the age and size at maturity of the three known swordfish stocks found within the Atlantic Ocean and Mediterranean Sea using samples/measurements provided by participating CPCs.
- Characterize the age composition of the catch and validate the growth curves for each swordfish stock.
- Determine the spawning period and areas of each stock.

- Identify the seasonal and spatial species composition of the swordfish diet using stomach content and/or tissues.
- Develop a protocol/template based on genetic analysis that will allow for the assignment of tissue samples to a particular stock.
- Develop a biological database that links the sample information to the age, stock origin, gender, size, diet and maturity data of each fish.
- Update the ICCAT Manual with new pertinent information.

Contract deliverables

Deliverable	**Status**
1. Ensure participation in a technical workshop on setting reference sets for spine and otoliths aging and calibration of reproduction, as well as to allow training of the Swordfish Species Group team members to be involved in the processing and data analysis of the samples collected. The workshop is tentatively scheduled to take place in Italy (October or November 2020).	Complete March 2021
2. SCRS documents and/or power point presentations at 2020 Swordfish Species Group meeting (September 2020) regarding the: a) Distribution of the collected samples by area, season, and sex will be made to the SCRS; b) Any updates on the protocols for sampling, aging and assignment of maturity stage; c) Report on the level of completion of sample collection and processing;	Complete
3. Labeled anal spines, otoliths and tissue samples that are to be shipped according to the updated protocols established during the technical workshop.	Final shipping in progress
4. A relational database containing the sample data that has undergone thorough QA/QC is to be provided. This database will reside at ICCAT Secretariat and will be made available for distribution upon request.	Complete
5. Shipping and processing of samples determined to be analyzed by the selected laboratories	Complete
6. Analysis of the samples and reporting of final findings.	Analysis of many samples is complete and reporting occurred at the March 2021 technical workshop, the SWO 2021 intersessional, and in this document
7. In case the 2021 Swordfish intersessional meeting occurs after May 2021, a draft final report shall be submitted to the Secretariat by 18 June 2021 at the latest. The Swordfish Species Group coordinator and rapporteurs, the SCRS Chair or Vice-chair, in consultation with the ICCAT Secretariat, will review will provide comments and communicate any necessary revisions (if applicable) to the Contractor and/or inform of approval within 5 days of the submission(s). The Contractor shall submit the revised final documents (if changes are request) together with the regular invoice, within 5 days after the aforementioned 5 days period. This report shall be formatted as a SCRS paper and include: a) Executive Summary; b) Full description of the work carried out; c) Description of final results; d) Proposals of further activities to be developed for achieving the objectives of the project; e) References and literature cited.	Complete
8. May the contract duration be of a maximum of 12-month, the final report shall be updated taking into account the comments provided by the	Complete

ICCAT Secretariat, the Swordfish Species Group coordinator and rapporteurs and the SCRS Chair and Vice-Chair, be submitted to the Secretariat by 25 June 2020 at the latest.	

Methods

Swordfish were sampled in a combination of existing national swordfish sampling programs and targeted port sampling programs. Detailed information on the sampling protocol, sampling definitions and maturity scale, can be found in Gillespie et al. (2020). To accommodate data available to port sampling programs (and some at-sea programs), the original sampling requirements were modified in Contract Amendment #1 and Contract Amendment #2 in phase one of this program.

In addition to data required for a full or partial sample, in some cases supplementary data were collected: in some fish, stomach contents were identified and quantified, while in other fish, otoliths were collected. A subset of these otoliths and limited number of anal fin spines were processed and aged by Fish Aging Services, a subcontractor of this project.

Results

Through all phases of this program, 4159 samples have been collected, covering all three stocks. A total of 1810 new and historical samples were sampled from all three Atlantic swordfish stocks in phase one of the program, and another 1433 new and historical samples in phase two of the program (table 1). In phase three, 916 new and historical samples were added to the sample collection. The majority of samples collected in this program are considered "Full" and are associated with an anal fin spine for aging, piece of tissue for genetic analysis, and contain data on fish size, sex, location and date. "Partial" samples lacked some combination of these data but always contained either an anal fin spine or a tissue sample (for sample definitions see Gillespie et al., 2020).

Sample spatial coverage

Samples were collected in several of the major fishing areas in the North and South Atlantic and Mediterranean (figure 1). Sampling in the North Atlantic was concentrated in three areas: the Scotian Shelf, in the Western Atlantic; along the 39°N parallel, in the Eastern Atlantic; and off the Western coast of Morocco in the Eastern Atlantic (figure 2). All three of these are major areas for swordfish catch. Samples obtained near the Strait of Gibraltar will be of particular relevance in future genetic analyses to understand mixing between Atlantic and Mediterranean stocks. In phase three of the program, a significant number of samples were obtained from the US east coast (billfish sampling area 92), however gaps remain in the Gulf of Mexico (BIL91) and the Caribbean (BIL93). Samples were also added from the coastal waters of Venezuela. In the cases of the Gulf of Mexico and Caribbean, there is relatively little swordfish catch, however, we anticipate that future sampling efforts will include data from these areas.

Sampling in the South Atlantic occurred between 5°N and 6°S, stretching from the coast of Brazil to the Gulf of Guinea (figure 3). More than half the samples were obtained in this zone which spans two billfish sampling areas (Bil96 and 97). This is an area of significant swordfish catch in distant water fishing fleets. This is also as an assumed mixing area for North Atlantic and South Atlantic stocks. In addition, samples were collected in the waters south of Brazil and off the coast of South Africa and Namibia. The south coast of Brazil and stretching east along the 30°S parallel is a major area for swordfish catch but was not sampled by this program.

Mediterranean sampling occurred in three regions: the Balearic Sea, in the western Mediterranean; the Tyrrhenian and Adriatic Seas, in the central Mediterranean; and the Greek Islands (figure 4). Sampling coverage of these sea appears somewhat representative of catch. More samples are required in the very western region of the Mediterranean, in the Alboran Sea and approaching the Strait of Gibraltar where there is suspected mixing between North Atlantic and Mediterranean stocks. Additional sampling is also required in the eastern Mediterranean in the Ionian and Aegean Seas.

Length frequencies

Length frequencies for fish sampled in this program are plotted by stock, aggregated and disaggregated by sex. The overall size distribution of sampled fish was similar to that of the size distribution estimated by task 2 size data for 2014-2018 (figure 5). Size frequencies for sampled fish in the North and South Atlantic and Mediterranean were roughly similar to estimated size frequencies observed in catch data (figure 6). As expected, females in all stocks were, on average, larger than males. Extrapolating the shape of these length frequencies by sex to overall catch may aid in estimating spawning stock. For a small proportion of sampled fish (~7%), only curved fork length (CFL) or standard fork length (SFL) was available. These lengths are excluded from this analysis, pending an appropriate length-length conversion.

Sex and maturity

The sex of fish was determined via macroscopic observation and through histological analysis. 86.5% of samples were assessed for sex, while in the remaining 13.5% of samples, gonads were not available for assessment or were in a state where sex was ambiguous (**table 2**). Sex data are not typically collected in national sampling programs, nor are these data required in ICCAT reporting, making it difficult to assess the representativeness of these data. In all regions, females outnumber males in the sample. The most extreme difference in sex ratio was observed in the Mediterranean, where only 30% of fish were assessed as male. This region also had the greatest level of uncertainty, where sex was unknown in approximately 30% of fish. Imbalance in sex ratios may be a result of inherent spatial zonation between sexes or it may be a result of males being classified as "unknown" at higher rates than females. For example, a large proportion of the sampled fish come from more northerly water where female swordfish are known to be at higher abundances.

Maturity was assessed on a six-point scale (see Gillespie et al., 2020). Nearly a third of fish sampled had maturity states that were labelled as "undetermined" and these data require further verification. In some cases, histological data are available for samples and in these cases, macroscopic assessments of gonads will be compared to histological data.

Progress on the age and growth component of the swordfish biology programme

Sample processing

Sectioning of spines and otoliths was performed at Fish Ageing Services (FAS). Preparation of spines followed Quelle et al. (2014). The second anal fin spine was embedded individually in resin for sectioning, two sections of approximately 0.5 mm were made at one distance of the condyle width (1D) and at half distance of the condyle width (0.5D) (**Figure 7**). Smaller spines were sectioned with a modified gem cutting machine high speed saw, using a single pro slicer diamond blade, while larger spines were sectioned using an Isomet with a diamond wafering blade. Spine sections were preserved in a polyplex clear ortho casting resin and photographed under a dissecting microscope with a digital camera.

Before processing, whole otoliths were measured for length and width and photographed using a Leica M80 with transmitted light and 5x magnification. Otoliths were prepared for annual and daily age readings in thin transverse sections by grinding down the otolith in a 3-step process (**Figure 8)**. Firstly, the otolith was fixed on the edge (end) of a slide using thermoplastic mounting media (Crystalbond 509) with the anterior side of the otolith hanging over the edge. Care was taken to ensure that the primordium was just on the inside of the glass edge. The otolith was then ground down to the edge using 400 and 800 grit wet and dry paper. The slide was then reheated and the otolith was removed and placed (ground side down) on another slide and Crystalbond was allowed to cool. Once cooled the otolith section was ground horizontally to the grinding surface using varying grades (400, 800 & 1500 grit) of wet and dry sandpaper and finally 5um lapping film. During this process, the otolith preparation was continuously checked for the appropriate thickness (220µm – 250µm for annual readings or 50-80µm for daily readings). Otolith sections were preserved in a polyplex clear ortho casting resin and photographed at a 40x magnification using a Leica M80 dissecting microscope illuminated with transmitted light.

Sectioned spine, whole and sectioned otoliths are presented in **Figure 9**.

According to the project database updated in May 2021, a total of 3497 spine samples (1414 males, 1832 females, 251 specimens with undetermined sex) were collected for this study from the North, South Atlantic and Mediterranean Sea. A total of 985 otolith samples (558 males, 414 females, 13 specimens with undetermined sex) were collected for this study from the North, South Atlantic and Mediterranean Sea. Sample size with summary statistics by sex and stock is presented in **Table 3** for spines and otoliths. The samples size frequency distribution is then shown in **Figure 10** for spines and **Figure 11** for otoliths.

From the collected spine and otolith samples, 698 spines, 177 otoliths for annual readings and 1 for daily readings from the Atlantic and 69 spines, 44 otoliths for annual readings and 6 for daily reading from the Mediterranean have been processed under previous phases of the project. In phase 3, 317 spines and 208 otoliths from the Atlantic and 30 spines from the Mediterranean have been sent for processing. In total, 1114 spines and 429 otoliths for annual readings and 7 for daily readings have been processed or are under processing from the North, South and Mediterranean stock (**Figure 12**).

Reference set

Under phase 3, a reference set for spines and otoliths has started to be developed. Initial agreement has been achieved on a set of 7 individuals during the 2021 ICCAT Swordfish Biology workshop (**Appendix 2**). Progress has been made on preliminary guidelines for age reading of both spines and otoliths and spine measurements (**Appendix 3**)

Reproduction and Maturity

See ICCAT document SCRS/2020/135 (Saber et al., 2020).

A preliminary analysis of L_{50} comparing Macroscopic and Microscopic data was conducted in 2020 (SCRS 2020/135). Altogether, 2434 data on sex and macroscopic maturity for swordfish from North, South Atlantic, and the Mediterranean Sea have been collected to date covering an ample size range (58 to 261 cm LJFL). About 498 gonad samples have been collected from the North Atlantic and the Mediterranean Sea. A total of 322 samples of gonads, 262 from the North Atlantic and 62 from the Mediterranean Sea have been processed for microscopic maturity. Further analysis will be conducted after increasing the sample size. This document also provides a preliminary analysis of the samples collected to date, and recommendations on next steps for data and sample collections. The descriptions of length frequencies by month/season and by stock of the swordfish sampled for maturity data are also provided.

The main topic to consider during the workshop was the establishment of reference sets for microscopic stages. A calibration exercise among experts was performed to reach this objective. For this, the group had the advice of invited experts Jessica Farley (CSIRO) and Freddy Arocha (UDO). See **Appendix 4** for additional details.

Genetics

The swordfish genome assembly was completed using a sequencing strategy that combined Oxford Nanopore ((MinION) and Illumina (NovaSeq 6000) technologies following standard analysis in a well-established bioinformatics workflow.

By comparing the swordfish genome with that of other 19 fish species, it was identified the percentage of swordfish-specific genes and the percentage genes shared with other 19 fish species. A Gene Ontology Enrichment Analysis (GOEA) was performed on several swordfish-specific orthologous groups to highlight their involvement on Biological Process, Molecular Function and Cellular component. Finally, the new assembled genome was used as a reference genome to guide the ddRAD analysis. Accordingly, the rationale behind this strategy was based on: 1) the better performances (i.e., precision) of the genotyping when guided with a reference genome, and 2) the finer scale of resolution and the expanded set of biological questions that can be addressed when a reference genome is available.

Double digest restriction-site associated DNA (ddRAD) sequencing technology was applied to obtain more than 40000 SNPs for the analysis of genetic differences among 672 samples collected from NA, SA and MED stocks. In particular, from NA were analyzed 322 samples of which 54 samples from BIL92, 12 samples from BIL93, 44 samples from BIL94A, 182 samples from BIL94B and 30 samples from BIL94C. From SA were analyzed a total

of 105 samples of which 11 from BIL96 and 94 from BIL97. Finally, from MED were analyzed 243 samples of which more than 100 from Balearic Islands. Samples were selected homogeneously not only on the basis of the catch area but also on the basis of gender, gonad maturity, length/weight, and period of catch.

To analyze genetic differentiation among samples, several statistical analyses including Principal Component Analysis (PCA), discriminant Analysis of principal Component (DAPC), pairwise genetic distances (heatmap matrix), NEIGHBOR-JOINING Cladogram were applied. Regarding genetic differentiation index such as, Fixation index (FST), Heterozygosity (both observed and expected), Observed heterozigosity related to single codifying genes, Inbreeding coefficient (FIS) and Allelic richness (both mean and total) were also calculated. Genetic structure was evaluated quantifying allelic frequencies clusters and their distribution among samples. Two populations were clearly identified among the whole samples analyzed and considerable evidence on the presence of subpopulations within the two populations emerged from the first 288 samples analyzed, to be confirmed once the 672 samples will be statistically analyzed.

Finally Whole Genome sequencing (WGS) analysis are still ongoing on 30 samples, 10 from NA, 10 from SA and 10 from MED in order to identify a set of SNPs that can be used to assign an unknown sample to one of the stocks and to identify sex-specific regions to assign sex to an unknown sample.

Complete details on genome assembly are found in **Appendix 5**.

Sampling database
Sample data are maintained in an Excel database that has undergone thorough QA/QC.

Sampling recommendations
In many regions, sample data was representative of catch data both spatially and temporally. Despite often good spatial and temporal coverage, we note some gaps that require additional sampling effort. In the North Atlantic, the Gulf of Mexico, Caribbean and mid-Atlantic represent a gap in sampling coverage. In the South Atlantic, additional sampling is required further from the equator, in more southerly regions. In the Mediterranean, sampling coverage is required near the Strait of Gibraltar and in the east, near Greece and Turkey. Amendments to sampling design have added flexibility for port sampling and we anticipate that this will help achieve greater sampling coverage for this species.

Program next steps
As ICCAT stock assessment and MSE processes become more analytically intensive, there is a need for greater accuracy in estimates of biological parameters such as size and age of maturity and stock mixing. These data are critical for devising management plans that maximize yield and support stock productivity. This sampling program is an initial step in reducing uncertainty in important biological parameters. The first and second year of biological sampling has produced data that is currently undergoing further analysis in a number of areas: tissue samples for genetic analysis for stock boundary definition and mixing; anal fin spines for aging correction analysis so as to estimate age structure in each stock; sex, size, age, maturity data are being analysed to refine maturity ogives; spatial and temporal abundance data combined with age and size data are helping define movement patterns by age class and are being used to update age-length-sex keys by area. Sample data with greater spatial and temporal coverage, particularly in regions suggested here, will further refine these parameter estimates. While there are some minor spatial-temporal sampling gaps, the primary focus of further work in this program should be analysis of samples collected to date. Significant effort has been invested in collection of samples and there is now a significant volume of samples that ready for analysis

Acknowledgements

This work was carried out under the provision of the ICCAT Science Envelope and the ICCAT – European Union Grant Agreement No. SI2.819116 - Strengthening the scientific basis for decision-making in ICCAT, and the ICCAT-US Data Fund.

References

Farley, J., Clear, N., Kolody, D., Krusic-Golub, K., Eveson, P., Young, J. 2016. Determination of swordfish growth and maturity relevant to the southwest Pacific stock. R 2014/0821: 118 pp.

Gillespie K., Hanke A., Coelho R., Rosa D., Carnevali O., Gioacchini G., Saber S. 2020. Final report for phase two of the ICCAT short-term contract: swordfish biological samples collection for growth, reproduction and genetics studies. *ICCAT Collect. Vol. Sci. Pap. ICCAT, 77(3)*: 136-161.

Quelle, P., González, F., Ruiz, M., Valeiras, X., Gutierrez, O., Rodriguez-Marin, E., Mejuto, J. 2014. An approach to age and growth of south Atlantic swordfish (*Xiphias gladius*) stock. *ICCAT Collect. Vol. Sci. Pap.* 70(4): 1927-1944.

Table 1. Total number of samples realized in this sampling program by stock.

Project Phase	Stock	Total Samples	Full samples	Partial samples	Fin spines	Tissue samples	Stomach contents	Gonads	Otoliths
1	N-Atl	365	184	166	192	348	27	51	17
	S-Atl	894	816	76	887	825	61	0	353
	Med	551	98	441	142	501	101	67	50
	Total	**1810**	**1098**	**683**	**1221**	**1674**	**189**	**118**	**420**
2	N-Atl	524	372	106	467	468	73	127	302
	S-Atl	521	295	173	491	488	145	0	223
	Med	388	0	388	60	388	0	121	24
	Total	**1433**	**667**	**667**	**1018**	**1344**	**218**	**248**	**549**
3	N-Atl	741	271	469	734	301	1	117	0
	S-Atl	66	0	59	66	66	46	46	29
	Med	92	0	90	90	21	71	0	0
	Total	**916**	**271**	**618**	**890**	**388**	**118**	**163**	**29**
	Total	**4159**	**2036**	**1968**	**3129**	**3406**	**525**	**529**	**998**

Table 2. Sex ratios by stock for samples collected in this program.

Sex	Total (n = 4159)	N-ATL (n = 1630)	S-ATL (n=1481)	MED (n = 1031)
Percent male	36.1	34.5	42.5	29.4
Percent female	46.3	51.2	47.9	36.3
Percent sex unknown	17.6	14.3	9.6	34.3

Table 3. Number of individuals, minimum, maximum and mean lower jaw fork length (LJFL, cm) and standard deviation (SD) by sex for each of the stocks, presented for the collected spines and otoliths of swordfish.

	Stock	Sex	N	Min	Max	Mean	SD
Spine	North	Male	728	45	265	134.6	25.9
		Female	1068	63	283	154.1	40.3
		NA	84	62	240	159.0	42.5
	South	Male	600	17	210	145.5	27.1
		Female	683	76	275	168.2	35.2
		NA	111	68	274	141.5	41.5
	Mediterranean	Male	86	78	180	126.0	27.6
		Female	81	72	200	125.3	30.1
		NA	56	93	189	130.9	20.3
Otolith	North	Male	104	92	204	125.0	17.6
		Female	213	94	261	140.3	26.3
		NA	2	124	156	140.0	22.6
	South	Male	276	78	207	144.3	28.1
		Female	308	89	246	175.9	33.6
		NA	8	74	205	161.6	44.6
	Mediterranean	Male	34	76	168	112.3	18.6
		Female	37	72	200	124.1	30.8
		NA	3	94	151	119.3	29.0

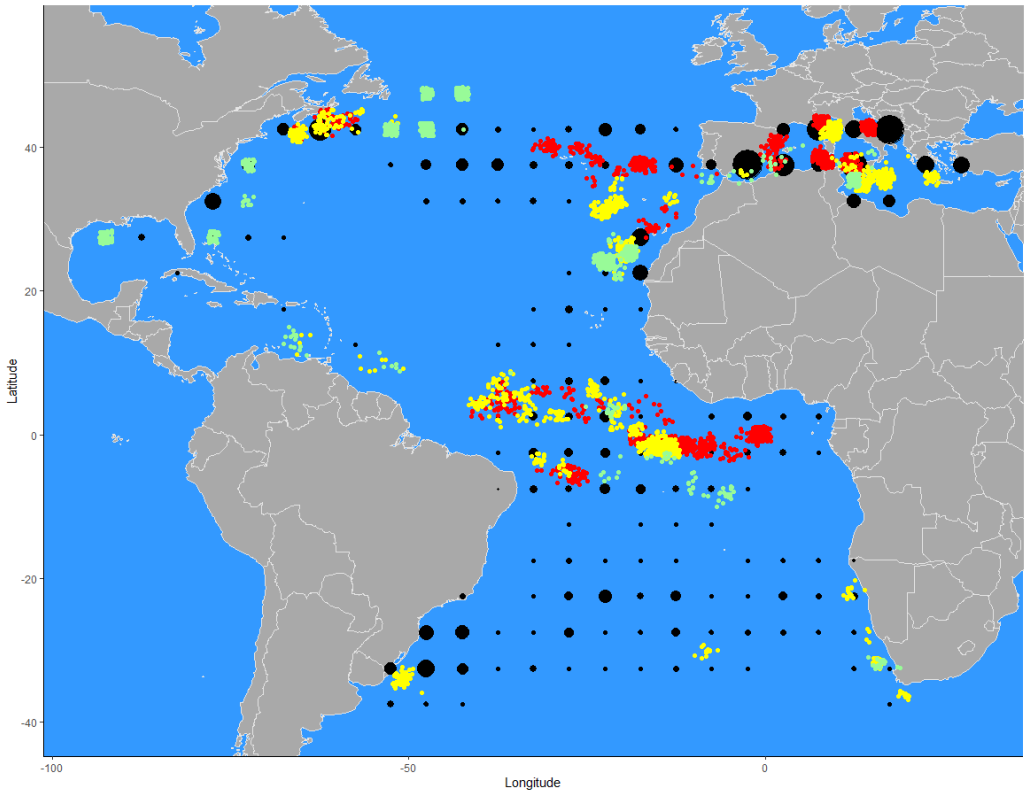


Figure 1. Sampling coverage the North and South Atlantic and Mediterranean. Red dots indicate locations where samples were collected in phase one of the program. Yellow dots indicate sample locations from phase two of the program. Green dots indicate sample locations from phase three of the program. Black dots represent swordfish catch for years 2014-2018 for each 5x5 grid cell area, scaled by contribution to the overall catch. Dots for grid cells contributing to less than 0.1% of the total average catch are omitted. Some sample locations have been obscured to abide by local privacy laws.

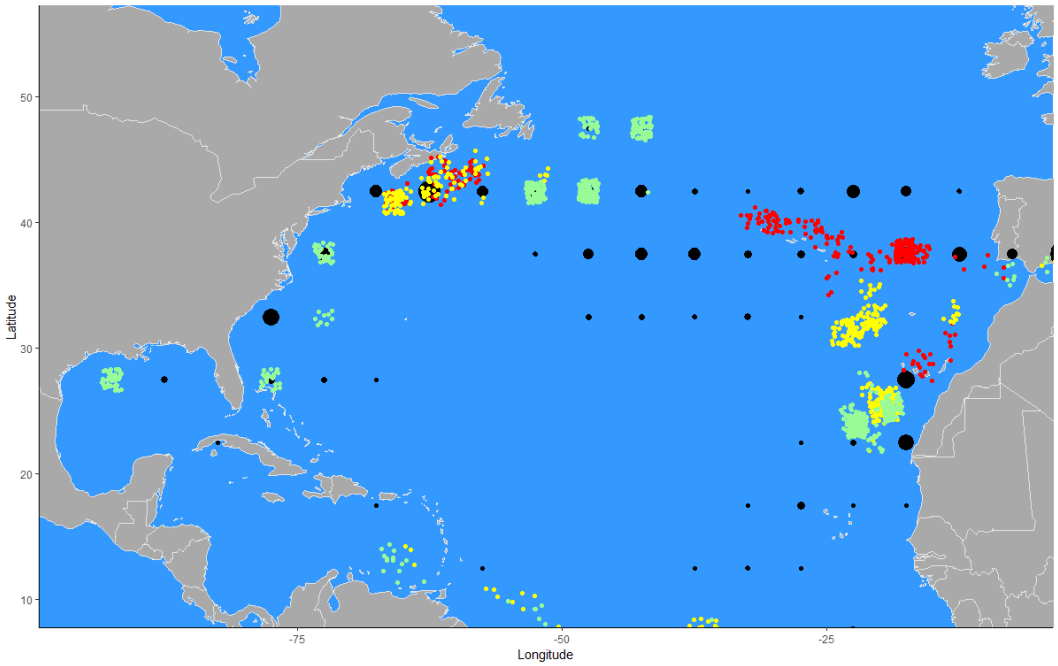


Figure 2. Sampling coverage in the North Atlantic. Red dots indicate locations where samples were collected in phase one of the program. Yellow dots indicate sample locations from phase two of the program. Green dots indicate sample locations from phase three of the program. Black dots represent swordfish catch for years 2014-2018 for each 5x5 grid cell area, scaled by contribution to the overall catch. Dots for grid cells contributing to less than 0.1% of the total average catch are omitted. Some sample locations have been obscured to abide by local privacy laws.

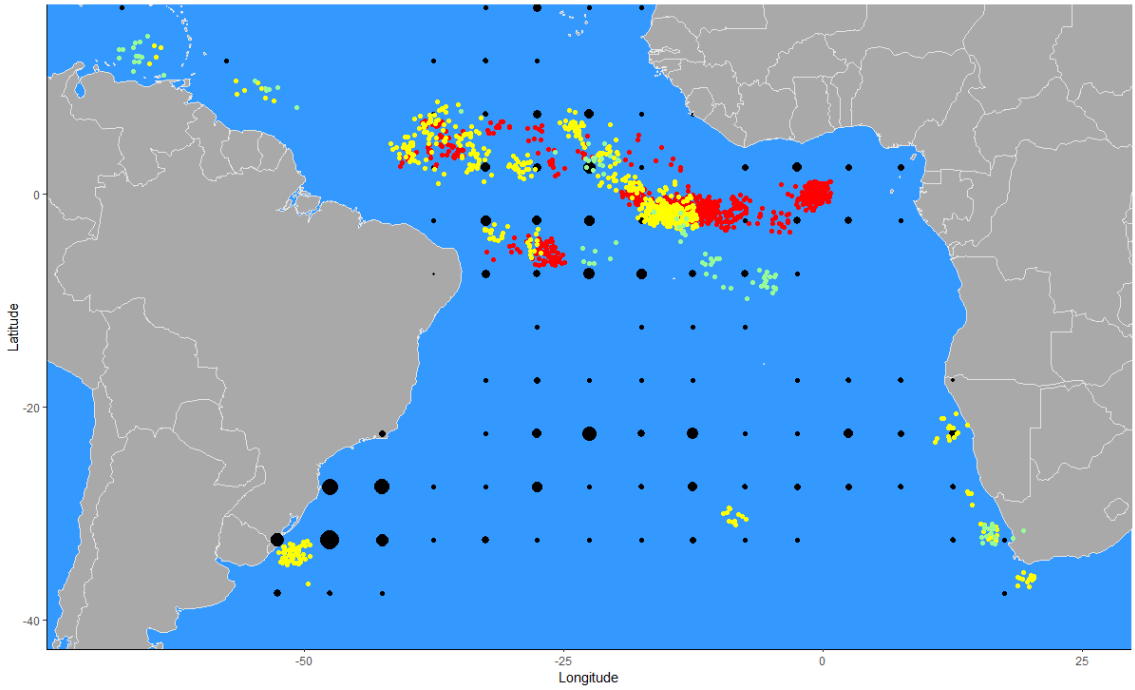


Figure 3. Sampling coverage in the South Atlantic. Red dots indicate locations where samples were collected in phase one of the program. Yellow dots indicate sample locations from phase two of the program Green dots indicate sample locations from phase three of the program. Black dots represent swordfish catch for years 2014-2018 for each 5x5 grid cell area, scaled by contribution to the overall catch. Dots for grid cells contributing to less than 0.1% of the total average catch are omitted. Some sample locations have been obscured to abide by local privacy laws.

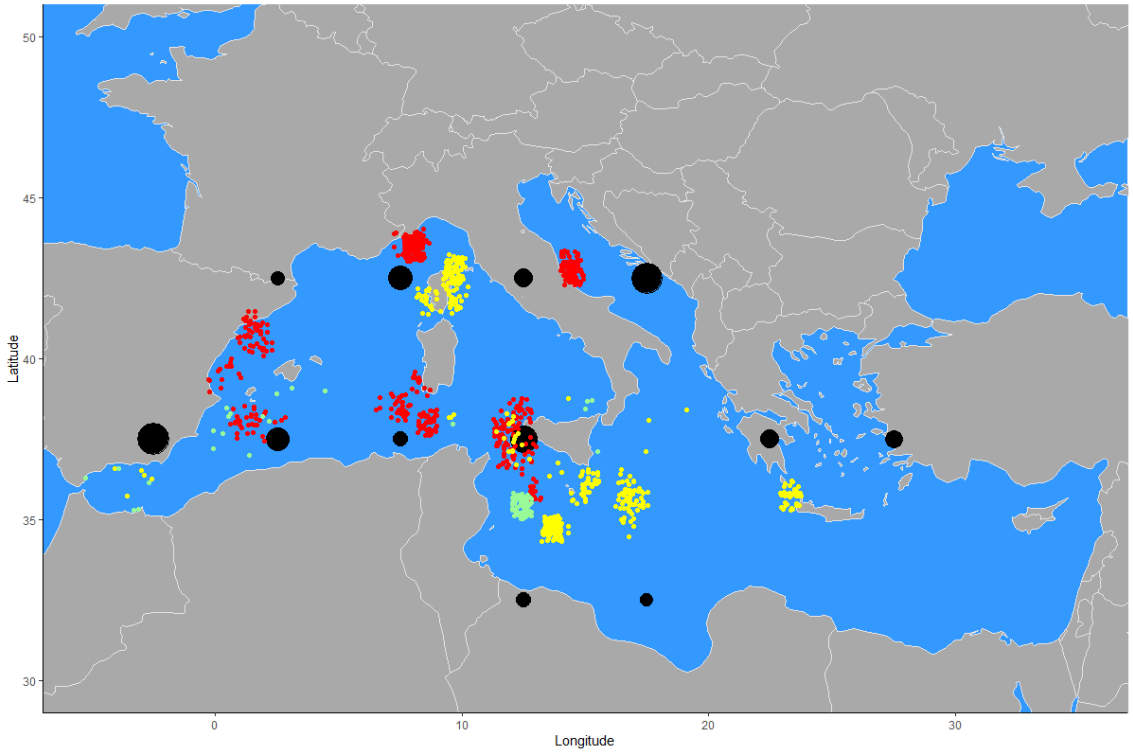


Figure 4. Sampling coverage in the Mediterranean. Red dots indicate locations where samples were collected in phase one of the program. Yellow dots indicate sample locations from phase two of the program. Green dots indicate sample locations from phase three of the program. Black dots represent swordfish catch for years 2014-2018 for each 5x5 grid cell area, scaled by contribution to the overall catch. Dots for grid cells contributing to less than 0.1% of the total average catch are omitted. Some sample locations have been obscured to abide by local privacy laws.

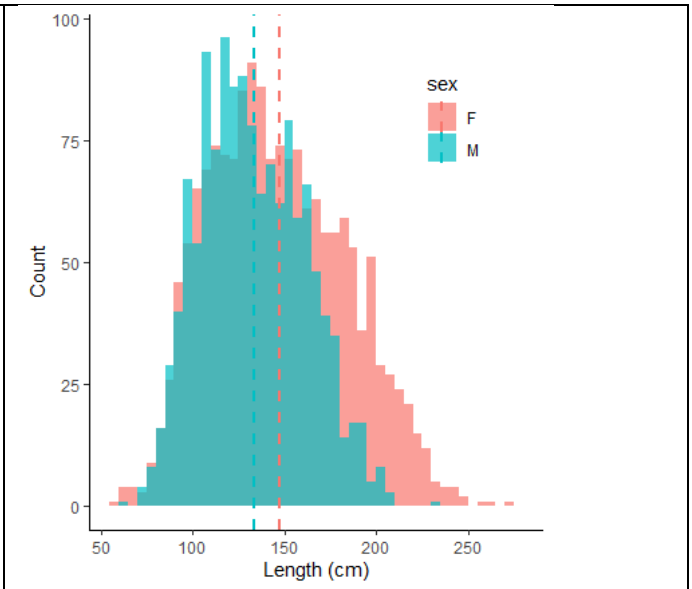
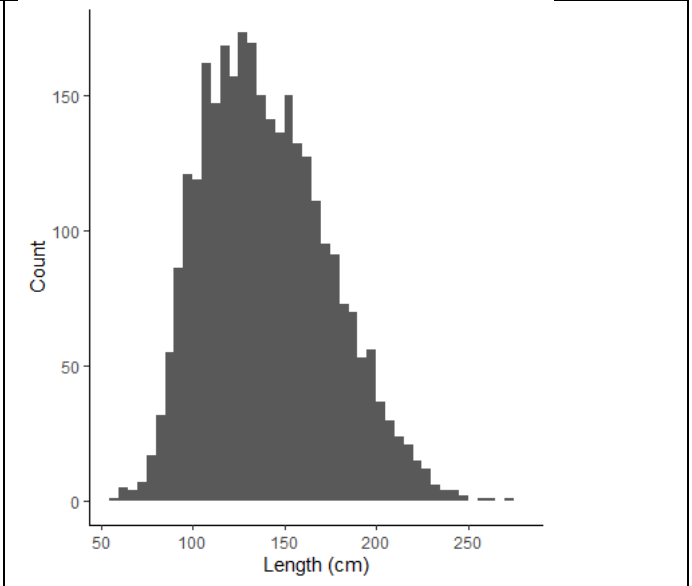
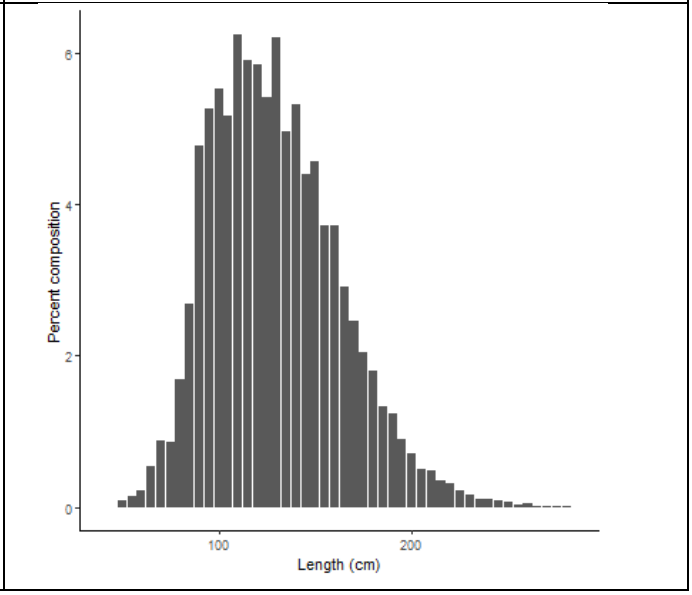
Sample length frequencies, sex disaggregated	
Sample length frequencies	
Catch length frequencies	

Figure 5. Length frequencies for all swordfish sampled in this program disaggregated by sex (top) and non-disaggregated (middle), compared to estimate catch size frequencies from ICCAT task 2 data for 2014-2018 (bottom).

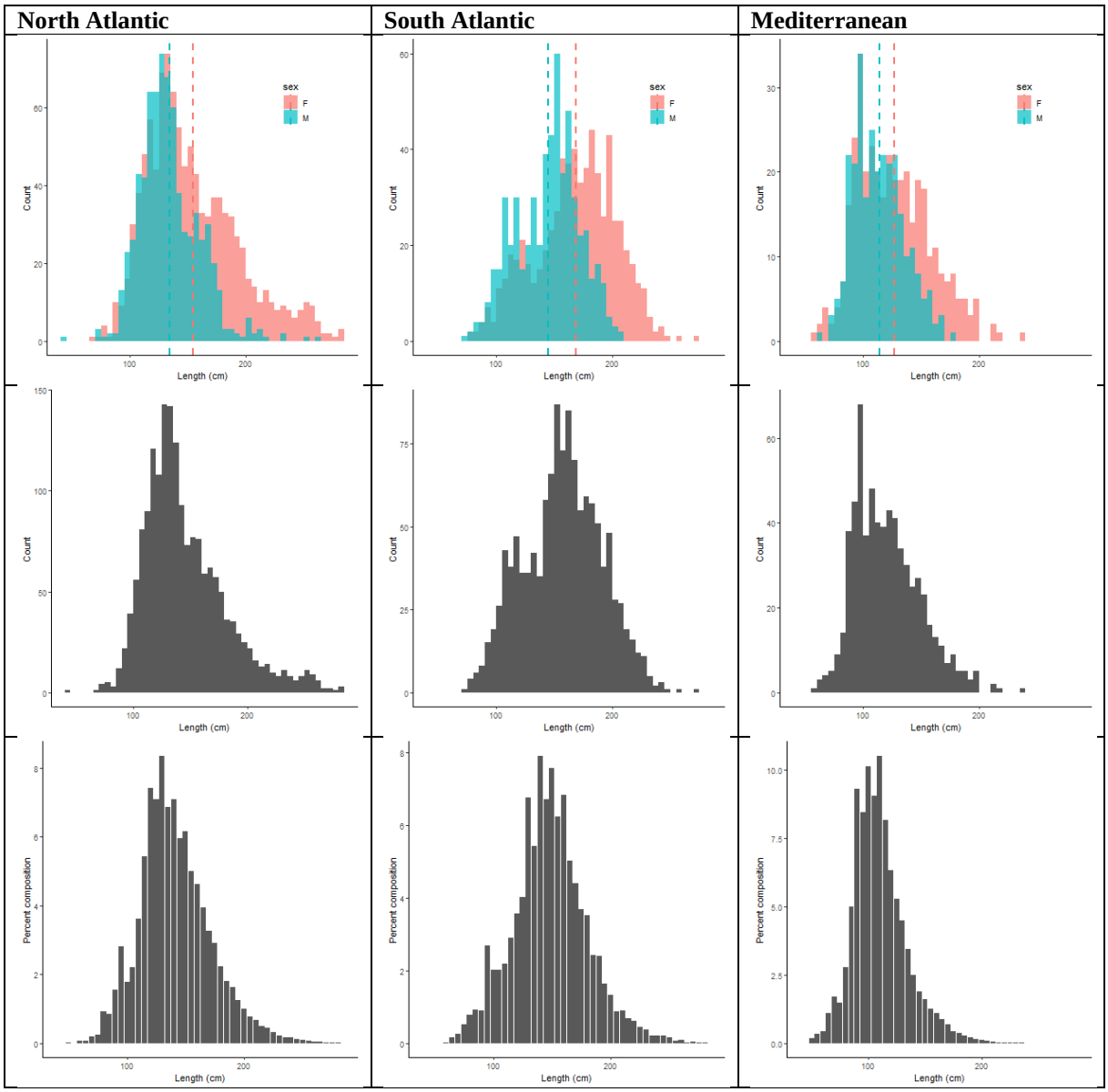


Figure 6. Length frequencies for all swordfish sampled in this program by stock, disaggregated by sex (top) and non-disaggregated (middle), compared to estimate catch size frequencies from ICCAT task 2 data for 2014-2018 (bottom).

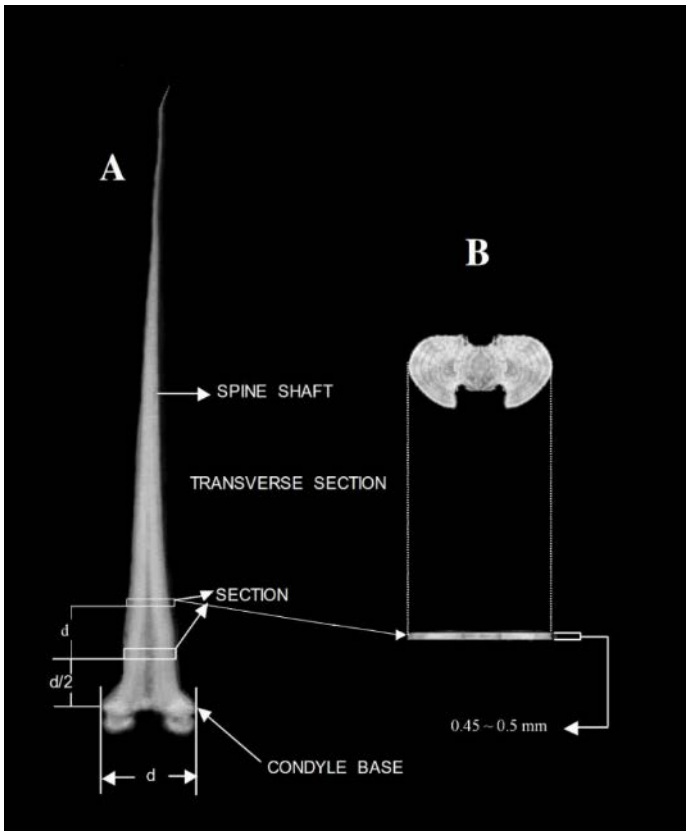


Figure 7. Second anal fin spine of a swordfish, showing A) the condyle base and location of sections at one distance of the width of the condyle base (d) and half distance of the width of the condyle base (d/2) and B) Cross section of the fin spine at distance d (Source: Quelle et al., 2014).

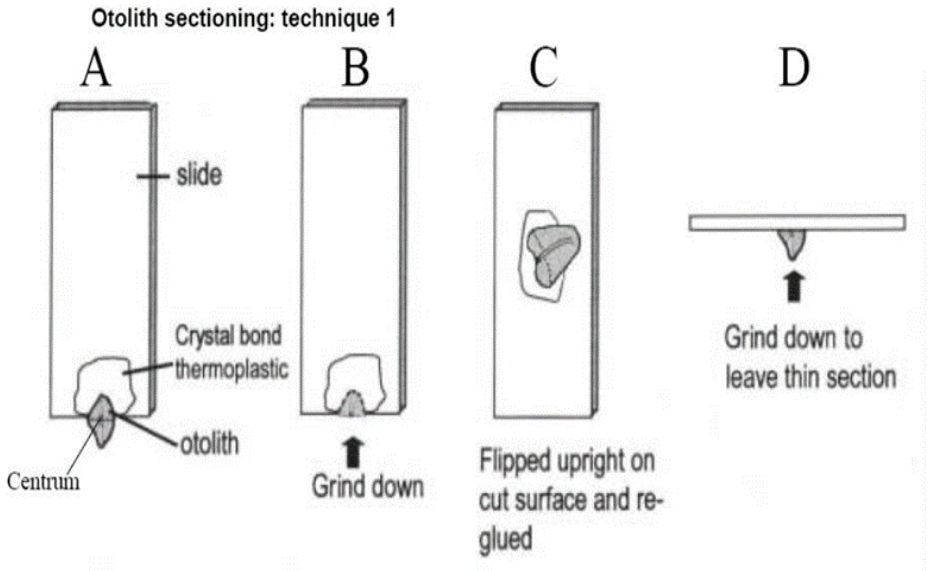


Figure 8. Illustration of the individual grinding process for preparation transverse otolith thin sections (Source: Robbins and Choat, 2002).

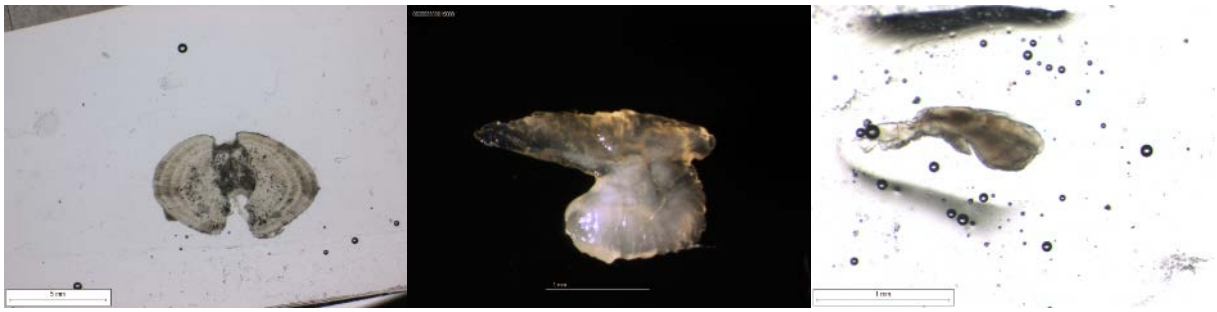


Figure 9. Example of a spine section at one distance from the condyle base (left), a whole otolith (middle), a section otolith for annual ageing (right) for a 139 cm lower jaw fork length male swordfish from the Atlantic Ocean.

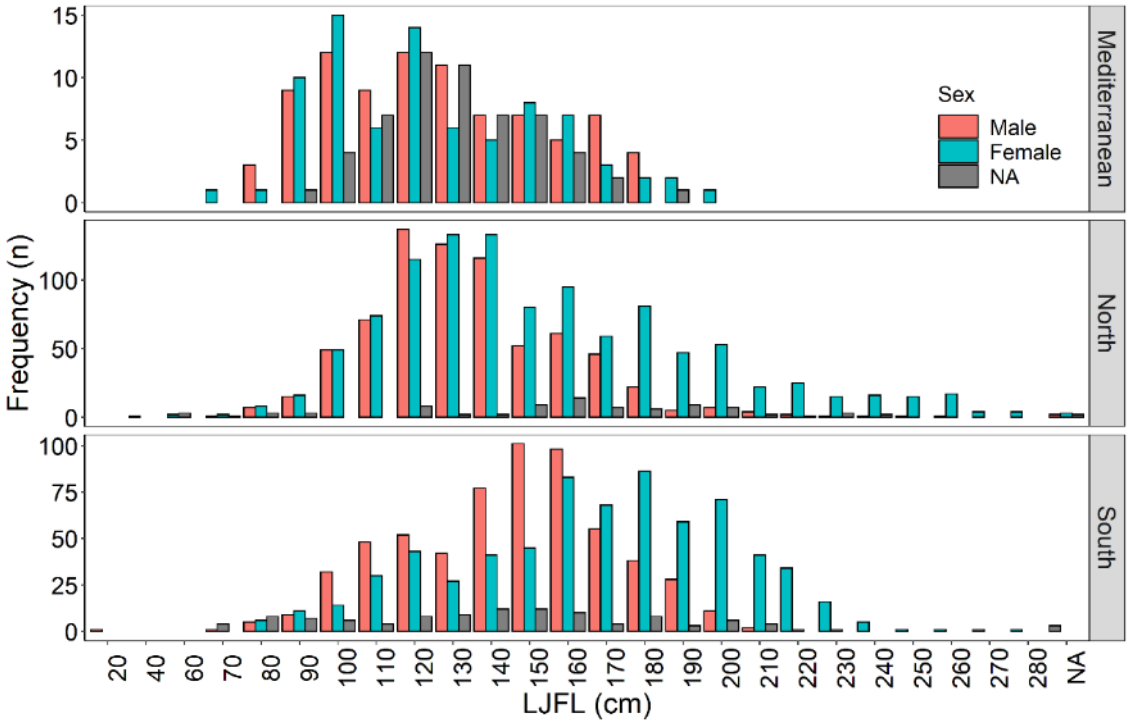


Figure 10. Size (lower jaw fork length, in cm) frequency distribution of swordfish spine samples currently collected for the age and growth study, for the north and south Atlantic (separated at the 5ºN) and Mediterranean Sea.

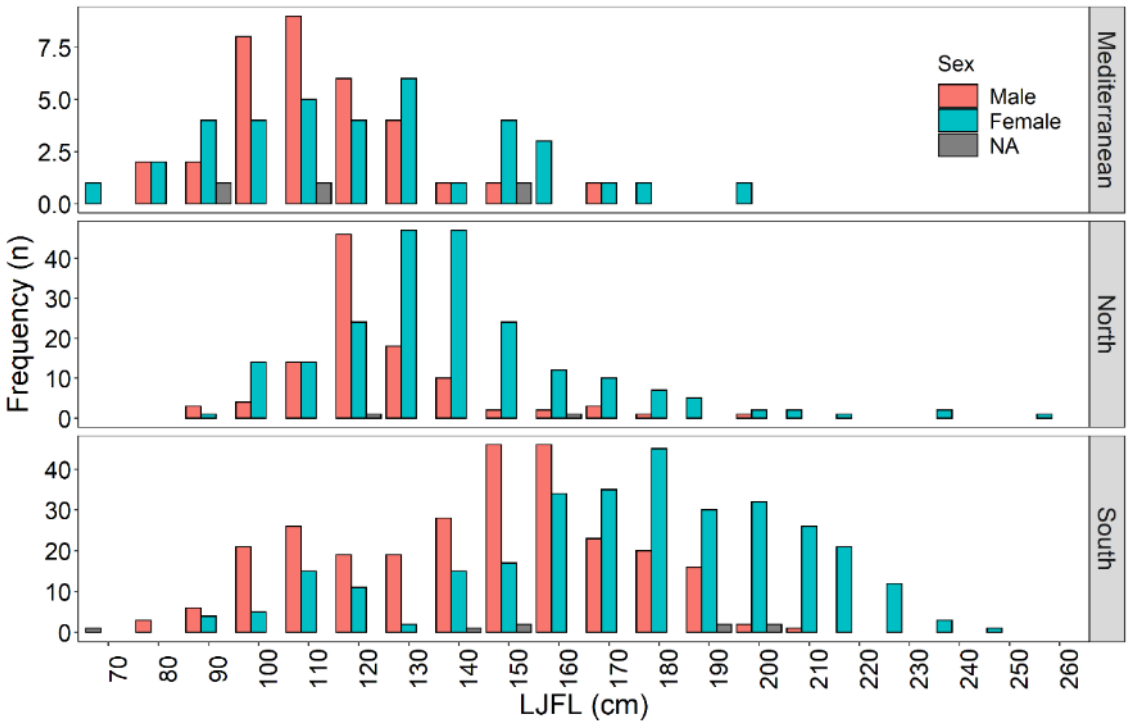


Figure 11. Size (lower jaw fork length, in cm) frequency distribution of swordfish otolith samples currently collected for the age and growth study, for the north and south Atlantic (separated at the 5ºN) and Mediterranean Sea.

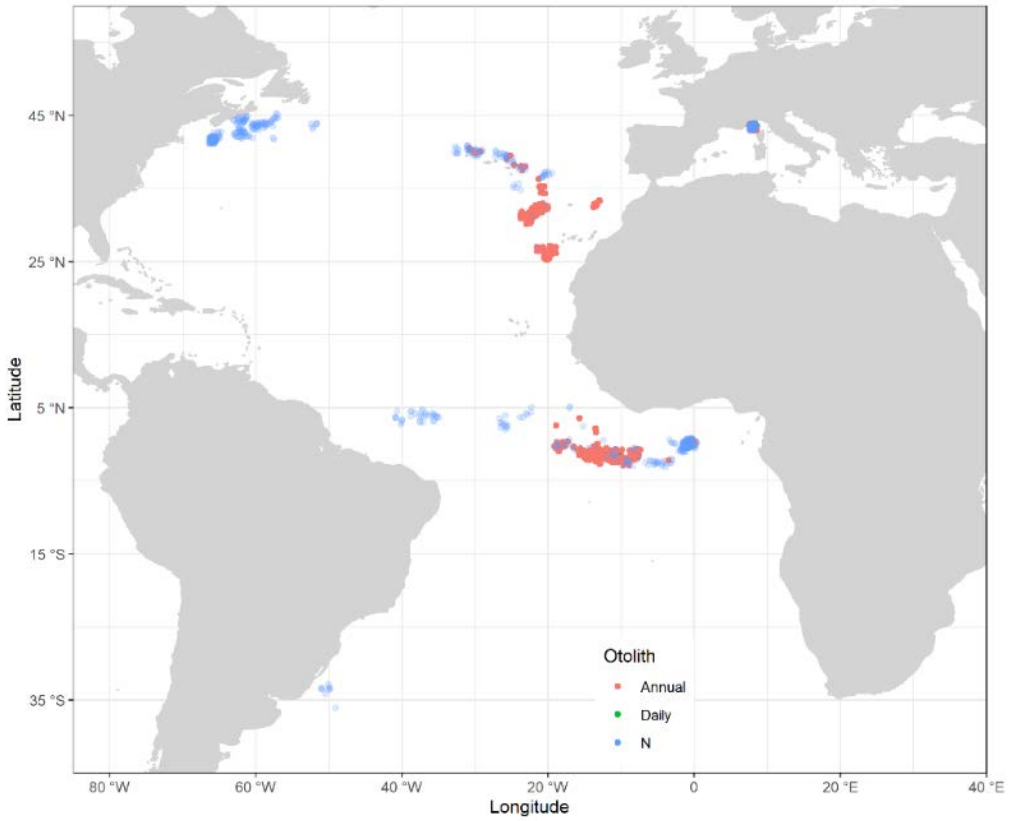


Figure 12. Map with the location of the swordfish (*Xiphias gladius*) ageing samples currently processed or under processing for the age and growth component of the biology programme. Blue circles indicate samples for spines, red circles represent otoliths for annual readings and green circles samples for daily readings.

Appendix 1: Program contributors

Institute	Flag code
AquaBioTech Group	MLT
AquaStudio Research Institute	ITA
Department of Agriculture, Forestry and Fisheries	ZAF
Fisheries and Oceans Canada	CAN
Hellenic Centre for Marine Research	GRC
Institut Français de recherche pour l'exploitation de la mer	FRA
Instituto Español de Oceanografía	ESP
Ministry of Fisheries and Marine Resources	NAM
National Oceanic and Atmospheric Administration	USA
National Taiwan Ocean University	TAI
Oceanis	MLT
Portuguese Institute for the Ocean and Atmosphere	PRT
National Instutute of Marine Sciences and Technologies	TUN
UNIMAR società cooperativa	ITA
Universidad de Oriente	VEN
Universidade Federal do Rio Grande	BRA
Universidade Federal Rural de Pernanbuco	BRA
Università Politecnica delle Marche	ITA
University of Cagliari	ITA
University of Genoa	ITA

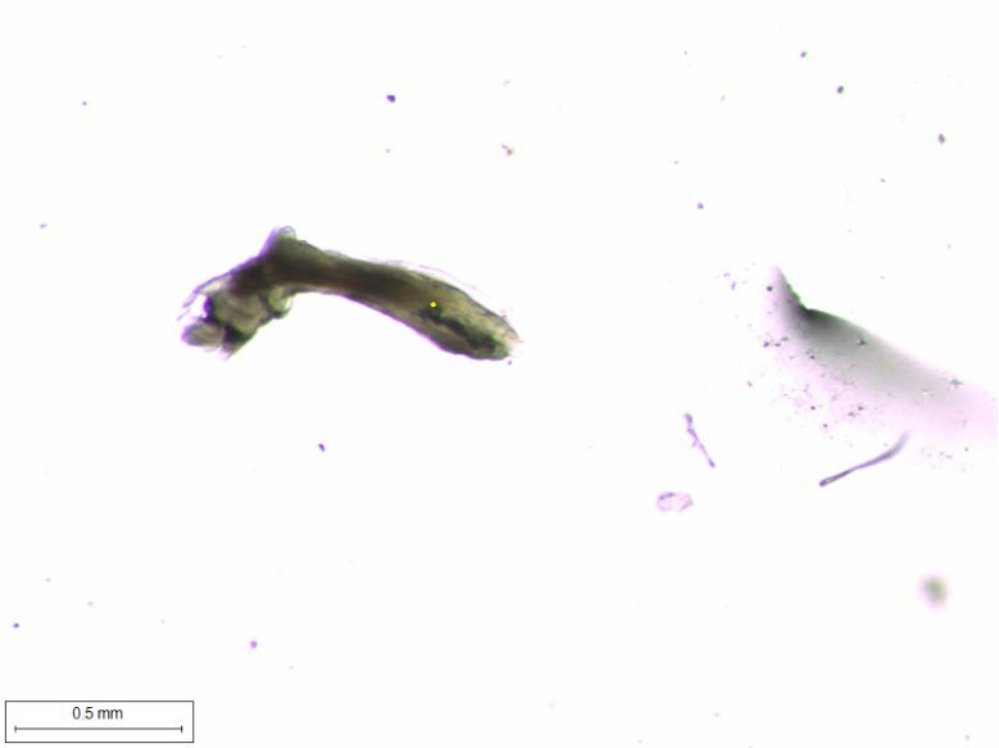
Appendix 2

Discussion on the spines/otoliths pairs images

Two sets of otolith-spine pairs were shown, one composed of 7 individuals with more agreement between reader and a set of large individuals in the exchange exercise.

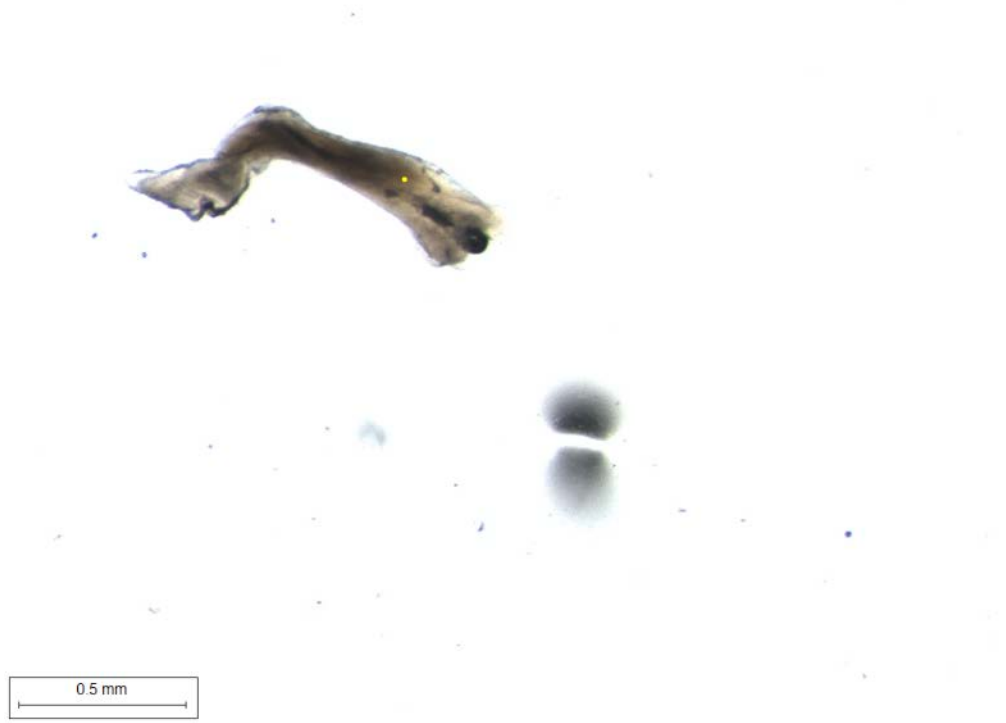
Pairs with more agreement between readers

Pair 1 (99cm male): All readers agreed it was an age 1 (almost age 2), for both spines and otoliths. Kyne suggests the otolith presents an opaque edge but noting it can be very subjective. The spine presents a translucent edge.

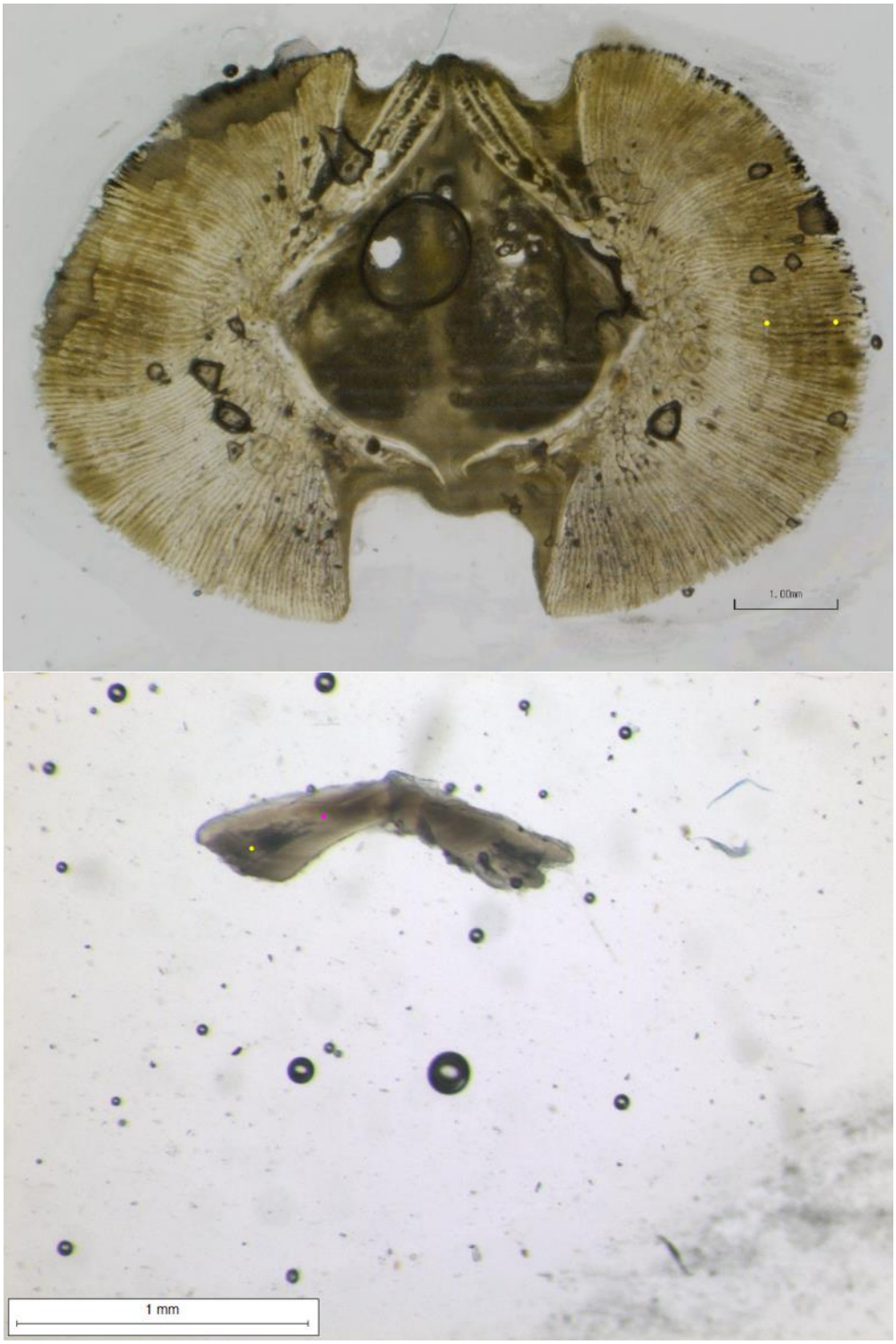


Pair 2 (114 cm female): From the otolith there is some agreement on an age 1, with a wide translucent edge, meaning it would be almost 2 yrs. In that same specimen from the spines it was agreed to be a 2 year old with a translucent edge. Considering previous studies, a female this size could either be age 1 or 2.

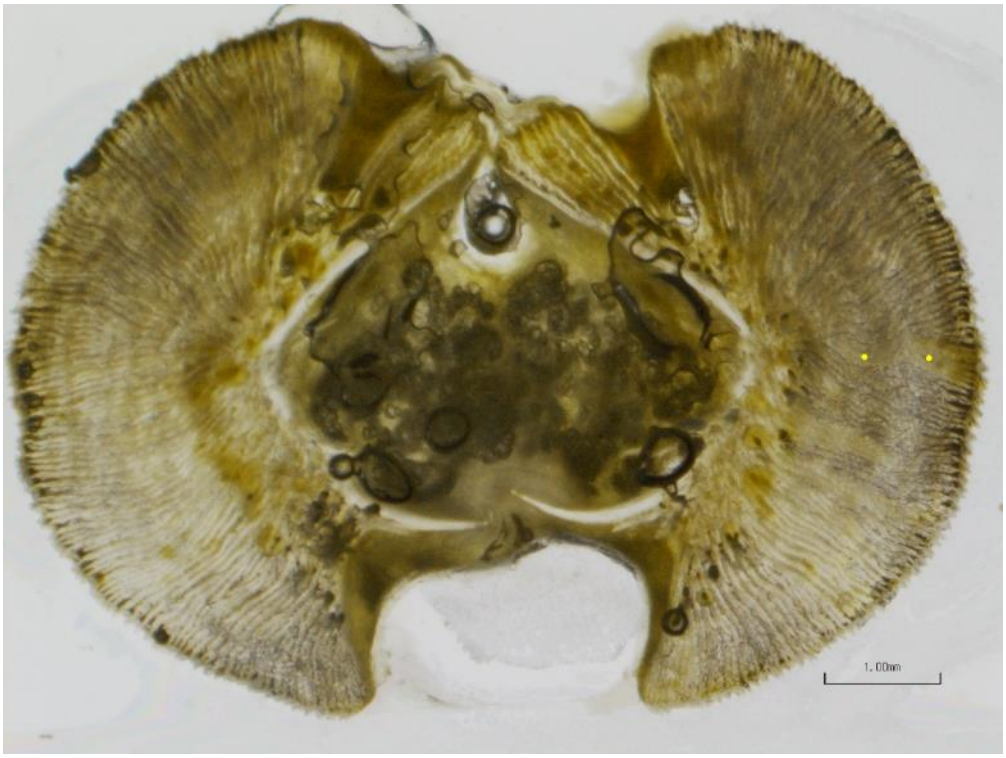
Pair 3 (118 cm female): In the otolith all agreed on age 1 with an opaque edge, noting it could also be a 2 yr old with a narrow translucent. There was less agreement on the spines, however looking to the otolith/spine combination all agreed that it should be an age 1 with a narrow opaque edge. It was noted that is not a clear spine, and it is possible that there is an initial band that could be there given the distance from the focus.



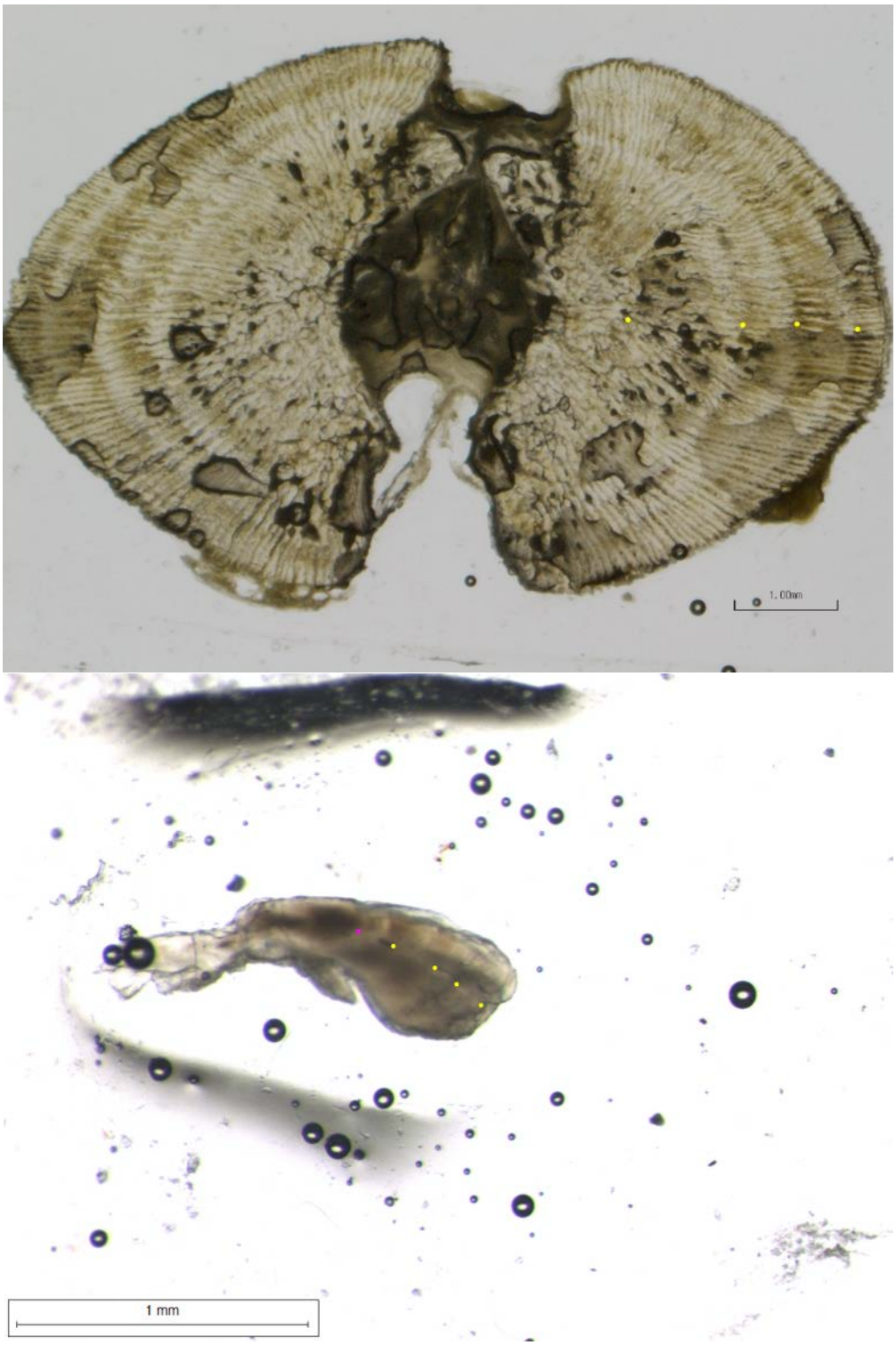
Pair 4 (121 cm female): Some disagreement on the spine ageing as there are some small bands that are very close together, which are likely false bands. But in general, there was an agreement that it is likely an age 2, with a narrow opaque edge. On the same otolith, there was some disagreement between ages 1 and 2.



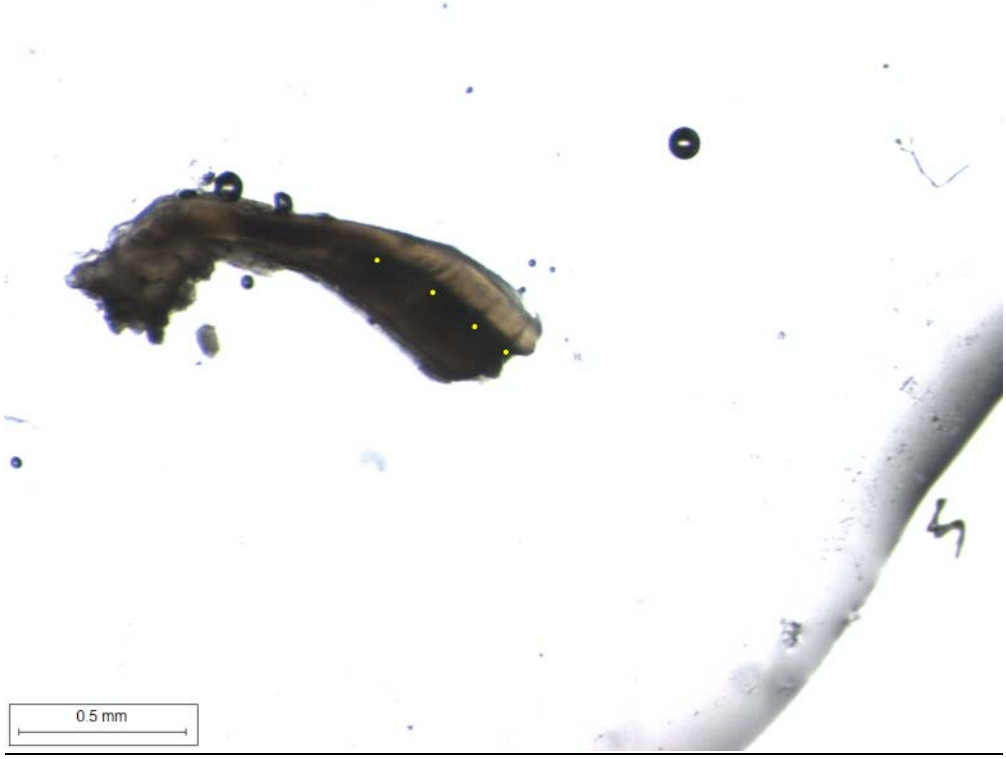
Pair 5 (124 cm male): Some initial disagreement between ages 1 and 2, but at the end agreement it was probably age 2 with a wide opaque edge. For the same otolith, most readers agreed that it was an age 1, however the final agreement was an age 2 if the edge is narrow opaque. It was noted this image is very dark and that digitally enhancing the images could improve readability.



Pair 6 (139 cm male): In the spine there are some differences between ages 3 and 5. The main issue is if there is a 1st ring that has been lost, but it is a small fish so that is uncertain. General agreement at the end that it could be a 4-year-old. There was also some agreement that the spine has a translucent edge. For the otolith most also read age 4, but the first mark could also be considered an increment, in this case having a guideline for the first increment could also help.

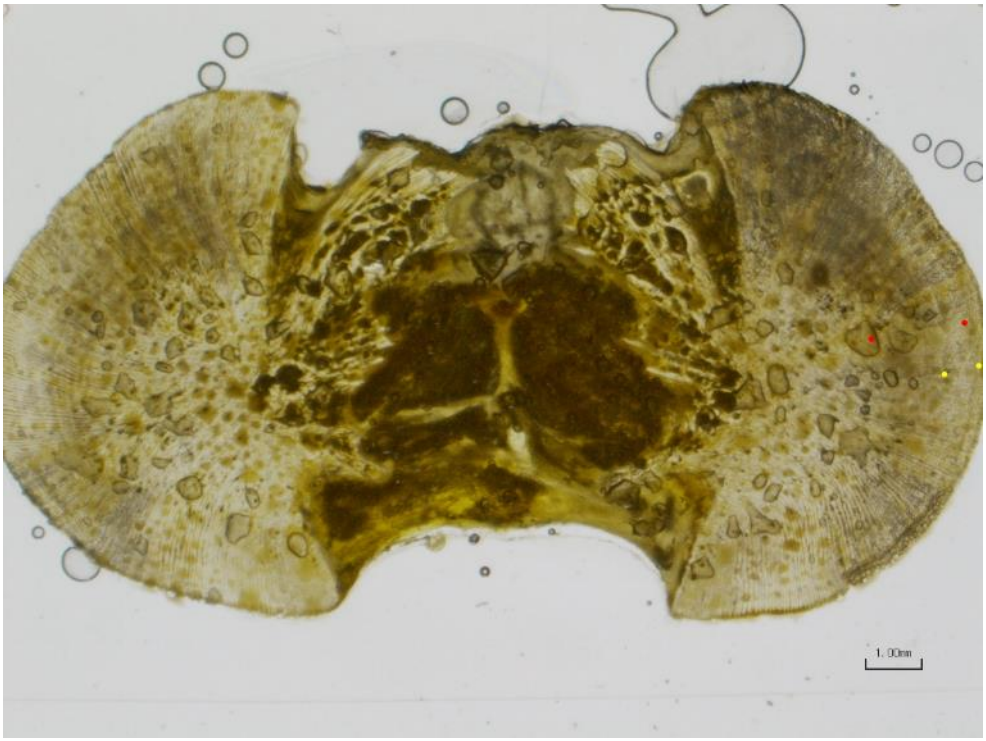


Pair 7 (143 cm male): For the spine there was agreement it was likely an age 3 with a translucent edge. The edge was difficult to decide if it was translucent or narrow opaque, it was agreed it was a translucent edge, so this increment was not counted. For the otolith most agreed it was an age 4, with one reader putting 3 and with the difference being the place of the first band.

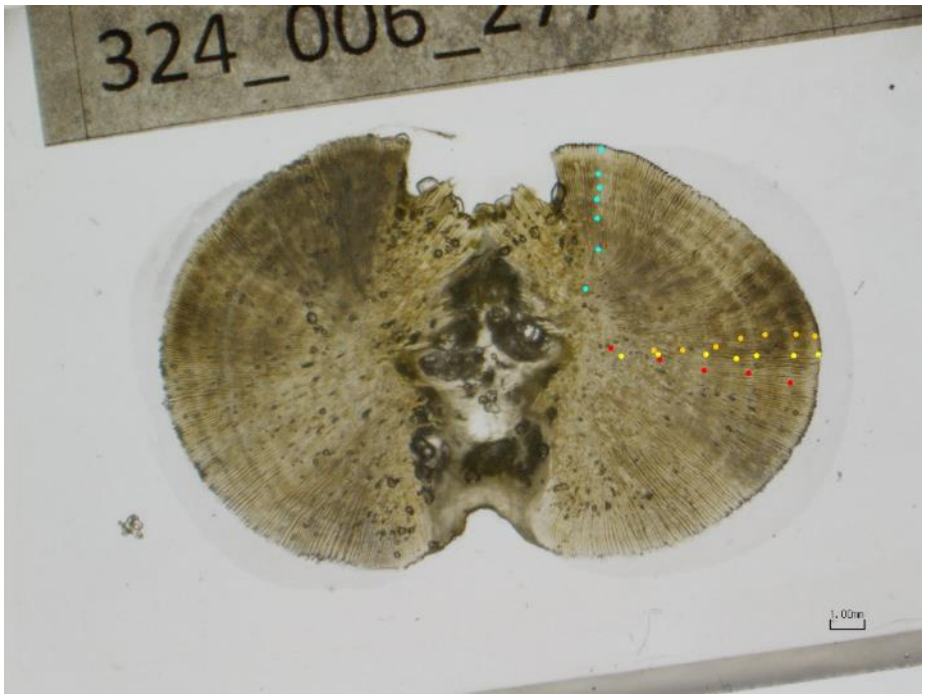


Pairs of large individuals

These pairs were only briefly discussed, with most focus on the first two pairs. For pair 1 (193 cm female) the spine was considered unreadable due to a very large vascularization. It could be possible to estimate ages using back-calculation from the size of the spine. Regarding the otolith there was disagreement between being a 4 or 5 yr old.



Regarding pair 2, (223 cm female), most readers assigned age 7 in for the spine, but this is a case where 1 or 2 initial pairs could be lost in the vascularization area, given the size of the fish. From the same otolith ages were estimated between 9 and 11 (reader 2 revised the reading to 9 increments) and seem to be more in line with the expected age for a female with 223 cm in size. This could be an example where vascularization has obscured the first increments in spines, while the otolith has all increments present.



Appendix 3

Ageing Guidelines for Swordfish

1. Identification of annuli

i. Spines

- Spine sections were photographed at variable magnification depending on spine size using a dissecting microscope illuminated with transmitted light.
- For standardization, if both lobes have a clear pattern, read on the right hand lobe. If there is a great difference between lobes, the clearest one should be used.
- Bands should be marked to the outer edge of each translucent zone (appears light under transmitted light).
- The last translucent band should only be counted if the reader considers that there is at least a small amount of opaque material between the band in question and the spine edge (i.e. being confident that the last translucent band is complete).
- The distance between true annual bands decreases proportionally with age; if the distance between two or three consecutive bands is clearly reduced these might be multiple bands.
- A true band can be tracked along the entire section; false bands or splits will not follow this pattern.
- If possible, provide edge type, where (see **Figure 1**):
    - T = translucent edge (translucent material visible on edge),
    - NO – Narrow opaque edge (opaque material past last translucent band is generally less than 1/3 of previously completed translucent band),
    - WO = wide opaque edge (opaque material past last translucent band is generally greater than 1/3 of previously completed translucent band).~

ii. Otoliths

- Otolith sections were photographed at 40x times magnification using a dissecting microscope illuminated with transmitted light.
- Bands should be marked to the outer edge of each opaque band (appears dark under transmitted light).
- The last opaque band should only be counted if the reader considers that there is at least a small amount of translucent material between the band in question and the otolith edge (i.e. being confident that the last opaque band is complete).
- If possible provide edge type, where (see **Figure 2**):
    - O = opaque edge (opaque visible on edge),
    - NT = Narrow translucent edge (translucent material past last opaque band is generally less than 1/3 of previously completed translucent band),
    - WT = wide translucent edge (translucent material past last opaque band is generally greater than 1/3 of previously completed translucent band)

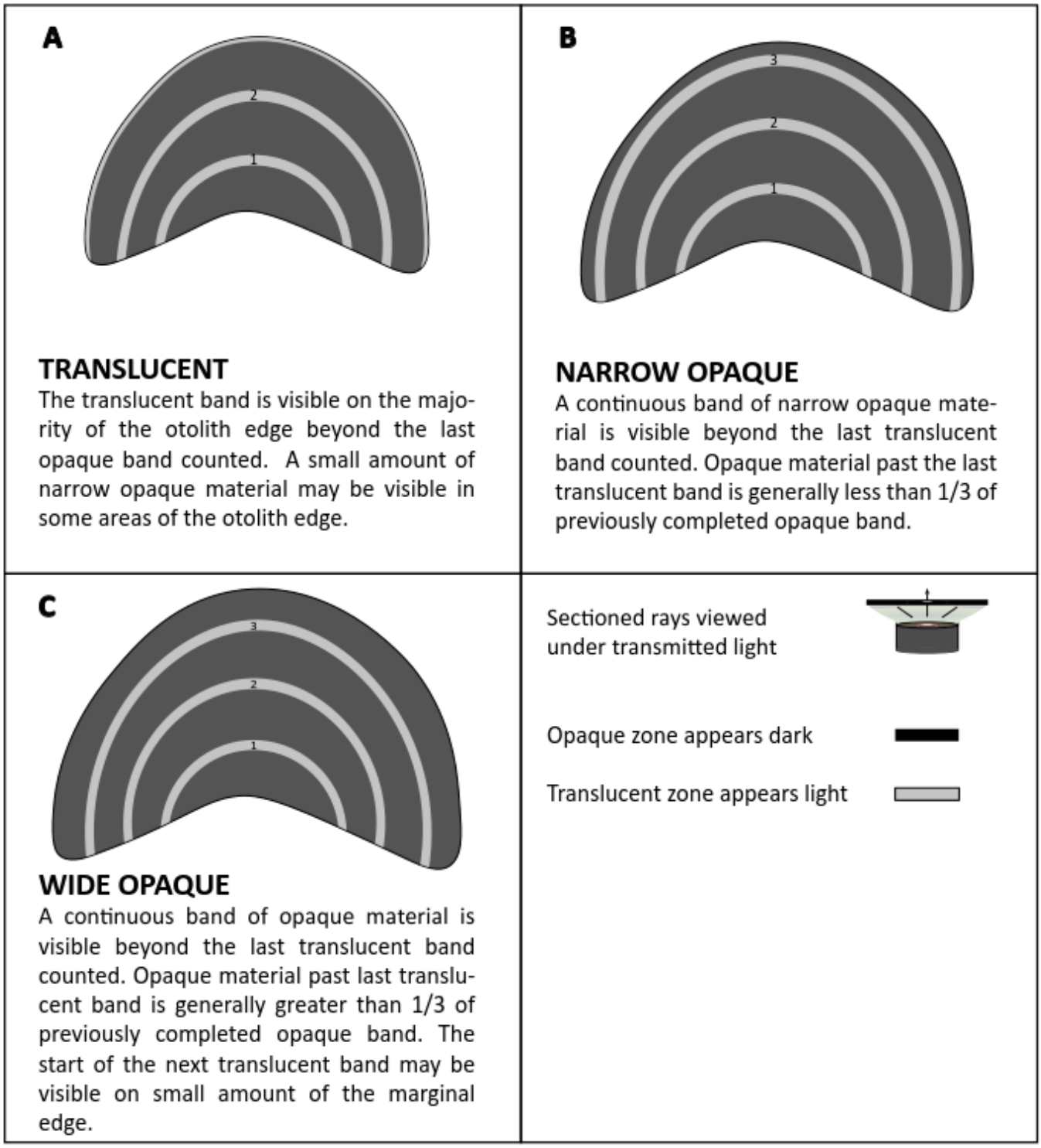


Figure 1. Cycle of band formation and edge classifications for swordfish spines under transmitted light (adapted from Farley et al., 2016).

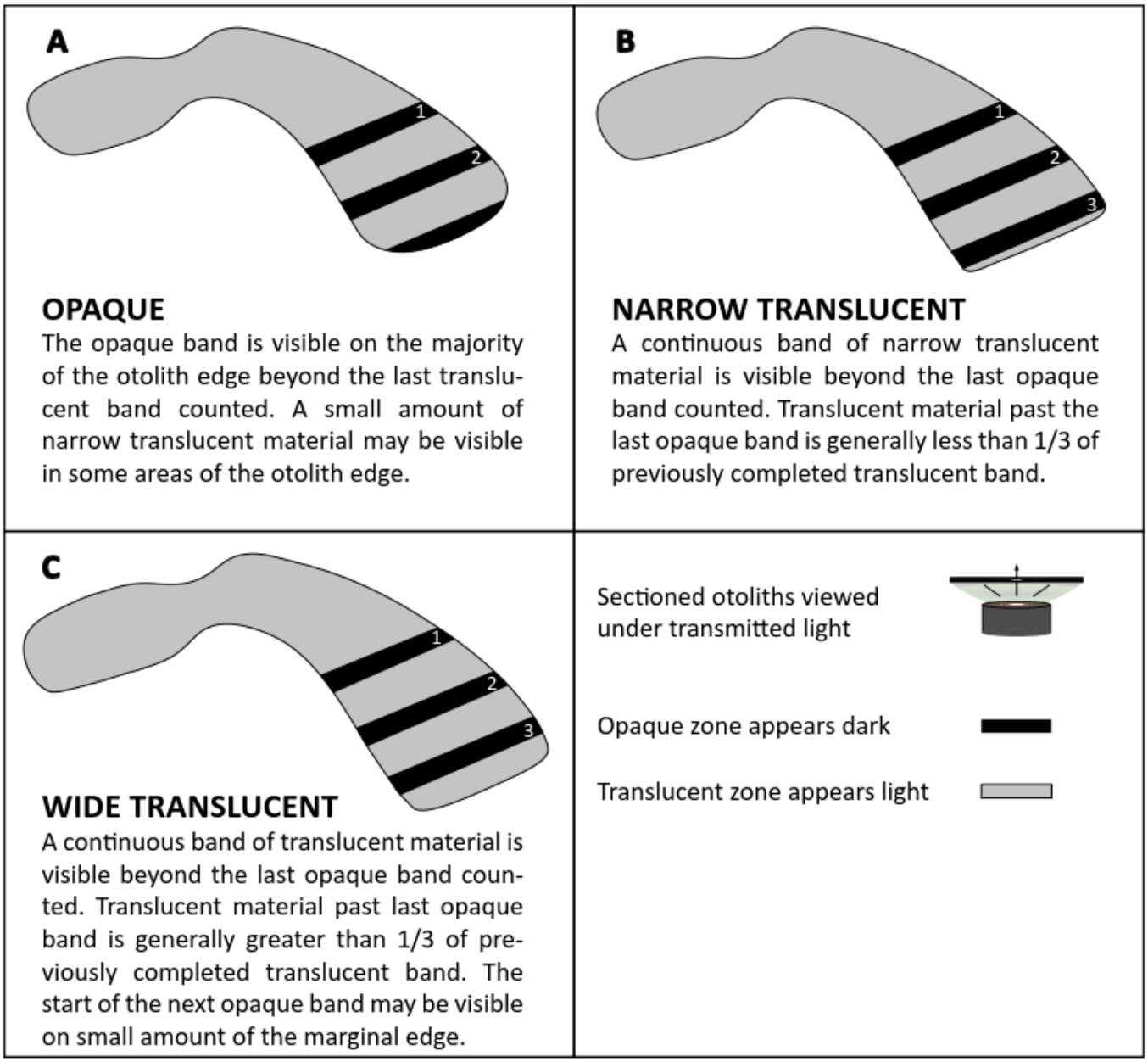


Figure 2. Cycle of band formation and edge classifications for swordfish otoliths under transmitted light (adapted from Farley et al., 2016).

2. Annual band measurements

i. Spines

Annual band measurements in spines will follow Quelle et al. (2014) (see **Figure 3**). A benchmark line, called "Focus line", is drawn, and defined as, the line which connects the two innermost ends of the lobes of the structure. Measurements are taken on a perpendicular line from the focus line to the widest part of the structure. Spine section measurements are obtained as distances between the focus line to the different section structures defined below.

Spine measurements can include:

- Pre-growth structure: Distance between the focus and the area where the growth process starts.
- Inner resorption distance: Distance between the focus and the end of the vascularised area, where the growth tissue starts to be visible.
- Vascularisation: Distance between the inner resorption distance minus the pre-growth structure distance.
- Annulus radius: Distance between the focus and each annulus.
- Adjusted annulus radius: Distance between the annulus radius distance minus the pre-growth structure distance.

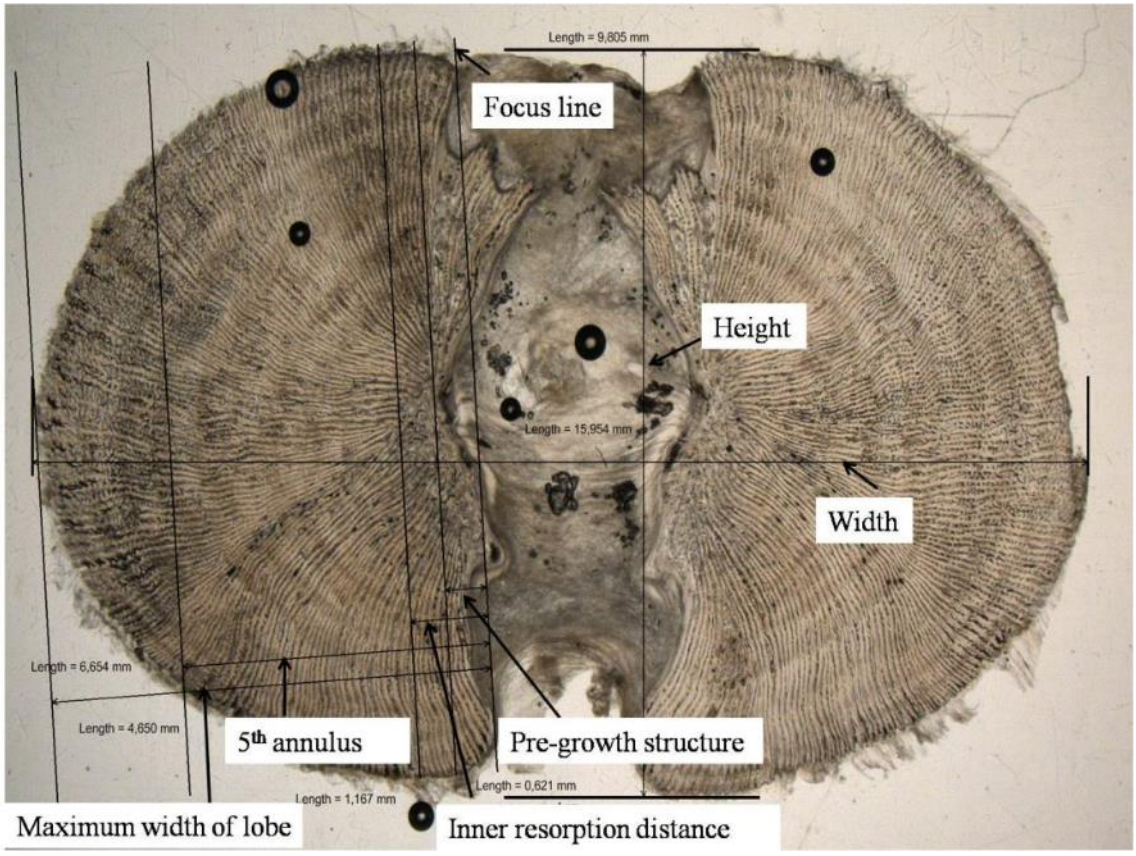


Figure 3 – Swordfish second anal fin spine transverse section and description of measurements (Source: Quelle et al., 2014).

ii. Otoliths

Annual band measurements protocol for swordfish otoliths will be developed and implemented in future phases of the project.

Appendix 4: Summary of analysis for maturity and reproduction in swordfish

A total of 3048 swordfish measured were classified as male (n = 1186), female (n = 1523) and undetermined (n = 338). The sex ratio was calculated as the ratio of females to males.

Six macroscopic maturity stages of gonads were assigned (ICCAT, 2016). Fish were classified as either immature (stage 1) or mature (stages 2 - 5). The L50 was estimated using the macroscopic maturity data. Sample gonads were sent to the coordinator of the reproductive studies in IEO-Málaga (Spain). Microscopic maturity stating of gonads was based on a modification of the criteria of Schaefer (2001) and Farley et al. (2013).

As expected, the analysis of the sex-ratio showed that females were more abundant than males, but it should be verify if the sampling scheme is taking into account both sexes. The estimated L50 for the three stocks was consistently lower than those adopted by the SCRS. However it should be remark that the significant number of histological sections of ovaries examined showed that females microscopically classified as immature were often incorrectly staged as developing (stage 2, mature) when using the macroscopic criteria. Increasing the sampling of swordfish across the Mediterranean Sea and Atlantic Ocean is necessary to collect enough data for the reliable estimation of maturity and other reproductive traits, as is the validation of the macroscopic maturity data using the histological examination of gonads.

Appendix 5: Summary of analysis for swordfish population genetics

Objective 1: Swordfish de novo genome assembly

For the first time, the genome of the swordfish *Xiphias gladius* has been sequenced, assembled and annotated. This initial step paved the way to comprehensive phylogenomic and evolutionary analysis performed by comparing the swordfish genome with high-quality genomes of 19 others fish.

From such analysis, relevant genetic and molecular information regarding the genomic landscape of this iconic species has been discovered.

In particular, we seek to highlight:

A haplotype genome with a size of 687.5 Mb and an average GC content of 40.71%. A total of 4493 scaffolds with an N50 of 9.37 Mb was obtained, with the size of the longest scaffold of 50.25 Mb (**Table1**). By benchmarking the genome against the Metazoa database, a total of 92.5% complete and single-copy, 2.5% complete and duplicated, 1.3% fragmented and 3.5% missing orthologues were found (**Figure3**). This approach provided a measure of genome quality and completeness.

From the functional annotation we obtained that *Xiphias gladius* genome is represented by a total of 43177 genes of which 22575 were assigned to a functional gene coding for a total of 92755 mRNA. Surprisingly, 20602 genes were unassigned.

According to the comparative phylogenomic analysis, we evidenced:

A total of 20602 (47.7%) genes was not assigned to any orthologous group. The great majority of these genes were non-coding genes (**Table2** and **Figure4**). Interestingly, 1514 (3.5%) genes were clustered in swordfish-specific groups, potentially representing a unique evolutionary fingerprint of this species. Among them, a relevant fraction was composed by transposons-related elements, which make the swordfish genome highly dynamic and likely subject to intense remodelling.

Gene families expansion and contraction was investigated. The genes belonging to families in expansion and contraction in the swordfish genome with respect to other species were analysed. Among them, 1021 genes belonging to expanding gene families were found to be associated with 95 biological processes, 95 molecular functions and 93 cellular components. Most of the expanding biological process were related to the immune response (**Figure 5**). Specifically, among the 1021 genes, only 5 were identified to be rapidly expanding and all of them were involved directly in the immune response. These results support the idea that the swordfish population is facing parasite/microbial infection and, generally, marine pollution. The acquirement of ability to physiologically respond to environmental stresses would probably be an investment, from a genetic point of view, to improve and modulate the immune system. At the same time, 3585 genes belonging to contracting gene families were found to be associated with 99 biological processes, 95 molecular functions and 95 cellular components. The main contracting biological processes were those related to DNA integration and carbohydrate and lipids metabolism (**Figure 6**). Among the 3585 genes, up to 57 genes were identified as rapidly contracting and most of them were involved directly in cartilage, nervous system and heart development.

Objective 2: genetic population analysis between the North Atlantic and Mediterranean fish

The double digest restriction-site associated DNA (ddRAD) sequencing technology has been applied to evaluate intra and inter-population differences between 96 Atlantic and Mediterranean specimens (**Figure 7**). In particular, the availability of the assembled genome allowed us to evidence not only how many the differences are (**Figures 8-9**), but also the genes and the gene families the two population are investing in.

According to the genetic population analysis we evidenced:

A total of 1048 genes associated to the genome region where the diversity of the Mediterranean population is higher than that of the Atlantic. Biological processes related to those 1048 genes resulted mainly linked to response to several internal and environmental stimuli (**Figure 10**). On the contrary, 688 genes were associated to genome regions where the diversity of the Atlantic population is higher than that of the Mediterranean. Biological processes related to those 688 genes resulted mainly linked to nervous system, muscular and digestive tract development and pigmentation (**Figure 11**). These results suggested that the Mediterranean swordfish is investing their energy in modulating the molecular machinery involved in the response to several stimuli such as those related to the "closed" and polluted environment of the Mediterranean Sea, if compared to those of the Atlantic Ocean. On the contrary, swordfish belonging to Atlantic population seems to invest energy in processes related to migration and feeding.

ddRAD sequencing technology let as to evaluate a total of 133.166 single nucleotide polymorphisms (SNPs). In particular, 1772 and 109.247 SNPs in North Atlantic and in Mediterranean specimens were found as characterizing, respectively.

25 North Atlantic specimens and 71 Mediterranean specimens were correctly assigned to their respective populations. In particular, Adriatic, Sardinian, Sicilian, Greek and a part of Spanish specimens were assigned to the Mediterranean population, while all Canadian specimens were assigned to the North Atlantic population. However, two specimens from the Spanish area showed an intermediate genotype between the two populations (**Figures 12-13**). This result would suggest the existence of a mixing zone between the North Atlantic Ocean and the Mediterranean Sea, where, probably, both Atlantic and Mediterranean specimens breed.

109.247 SNPs allowed us to identify the Mediterranean population. The Mediterranean population analysed alone showed two cluster of allelic frequencies. Analysing the distribution of these 2 clusters we assumed that in the Mediterranean Sea there are two subpopulations, with different clusters of allelic frequencies. Perhaps, one of this subpopulation could be identified as the Western Mediterranean Sea area (1) and the second population could be identified as the Eastern Mediterranean Sea area (2). A mixing area in Central Mediterranean Sea could be identified, with allelic frequencies between Eastern and Western subpopulations (3) (**Figures 14-15**).

1772 SNPs allowed us to identify the North Atlantic population. North Atlantic population analysed alone, showed 3 cluster of allelic frequencies but all 25 specimens showed all the three different genetic variants suggesting that Atlantic swordfish is a potential panmictic population (**Figures 16-17**).

Objective 3: Genetic population analysis among North and South Atlantic and Mediterranean swordfish focusing on mixing and spawning area

ddRAD analysis will be applied to evaluate genetic population differences among North and South Atlantic and Mediterranean swordfish. In particular, additional 200 samples have been already extracted and will be analysed soon with the ddRAD technique.

The samples have been chosen on the basis of the fishing areas and the gonadal maturity stage in order to identify and investigate mixing and spawning area.

REPORT IN EXTENSO

Genome Sequencing and Assembly

DNA was extracted from a blood sample of a swordfish caught in the Adriatic sea by using HMW DNA Extraction kit. The DNA extracted was sequenced using two different but complementary methods: the first one, based on Nanopore technology (MinION), enabled us to sequence very short fragment of DNA and the second one, based on Illumina technology (NovaSeq 6000), enabled us to sequence longer fragment of DNA. The two sequencing results were used together to obtain the final whole genome assembly. The quality of the Illumina paired-end reads was analyzed using FASTQC [1] and then BBDuk [2] was used to remove low quality bases (minimum

Phred 30), adapters and short sequences (minimum length 35 bp). Genome size estimation was then performed using the high quality reads using BBMap [2]. In **Figure 1** are reported the plots representing: genome fraction coverage (A), GC content distribution (B), Coverage histograms (C) and Mapping quality histograms (D) obtained by Illumina sequencing.

Nanopore reads were corrected with the software NECAT [3] and then assembled with the softwares Canu [4] and wtdbg2 [5] setting an expected genome size of 600 Mbp.

In **Figure 2** are reported the plots illustrating: genome fraction coverage (A), GC content distribution (B), Coverage histograms (C) and Mapping quality histograms (D) obtained by Nanopore sequencing. The raw assemblies were corrected using the Illumina data performing three iterations of NTHITS (options -b 36 -k 40 -t 36 --outbloom --solid) and NTEDIT (-k 40 -z 1000 -d 3). The corrected assemblies were then merged with the software Quickmerge [6] with the option -l 490000. The obtained merged assembly was further polished with five iterations of NTHITS/NTEDIT.

The Illumina reads were mapped against the assembly with minimap2 [7] then Platypus [8] was used to identify mismatches between the assembly and the reads (options --minReads=6 --nCPU=20 --minFlank=10 --trimReadFlank=10 --assemble=1 --assembleBadReads=1 --assemblerKmerSize=55 --assembleBrokenPairs=1 --minMapQual=30 --minBaseQual=30 --minPosterior=30). The high quality variants (filter "PASS" and GT=1/1) were used to further polish the assembly with vcf-consensus [9]. Finally, RNA-seq reads previously obtained in our lab [10] were mapped against the assembly with minimap2 [7] with the option -x splice. The resulting mapping file was processed with Opossum [11] (options --MinFlankEnd 10 --MinFlankStart 10 --SoftClipsExist True --ProperlyPaired False) and then Pilon [12] was used to perform an additional correction of the assembly (options --minqual 30 --minmq 30 --mindepth 6 --flank 0 --threads 20 --duplicates --fix bases --diploid --changes). Two iterations of mapping and Pilon corrections were performed.

The software QUAST [13] was used to obtain the statistics of the assembly (**Table 1**). Read assembly resulted in a haplotype genome with 687.5 Mb in size and an average GC content of 40.71%. A total of 4,493 scaffolds with an N50 score of 9.37 Mb were obtained. The longest scaffold was 50.25 Mb.

Evaluation of the genome completeness using BUSCO v4 [14], used with the datasets Eukaryota, Metazoa and Actinopterygii, confirmed that between 87.4 and 92.5% of complete and single-copy orthologs were found in the assembled genome of *Xiphias gladius* (**Figure 3**). Furthermore, between 1.2% and 2.5% of complete duplicated, between 1.3% and 2.7 fragmented and between 3.5% and 8.2% missing orthologues was reported (**Figure 3**).

Genome Annotation

Repeats were annotated on the final version of the genome using RepeatMasker [15].

The annotation of the genes was performed with multiple steps in order to integrate different sources of information:

Normalized RNA-seq reads [10] were mapped against the genome with STAR version 020201 [16] with the two pass mode enabled. A reference guided transcriptome assembly was performed with Trinity v2.8.6 [17] with the options --SS_lib_type RF --jaccard_clip --full_cleanup --min_kmer_cov 2 --no_normalize_reads --genome_guided_max_intron 100000. The obtained transcripts were filtered to retain only those with a length higher than 500 bp.

The software Maker [18] was run on the assembly providing the RepeatMasker annotation, the Trinity assembly and using zebrafish as model species for Augustus [19] to perform an initial step of annotation. The obtained genes were split into a training and test dataset to train a swordfish Augustus model with etraining and optimize_augustus.pl

The output of STAR obtained by mapping the RNA-seq reads against the assembly was used with GeneMark [20] in order to obtain an HMM model for gene prediction

A final Maker run was performed providing the assembled transcriptome, the RepeatMasker annotation, the Augustus model, the GeneMark model and a dataset of proteins from Lates calcarifer (ASM164080v1)

The obtained GFF3 file was converted to GTF with gffread [19] and then used as input with the RNA-seq BAM file for stringtie in order to annotate splicing isoforms and also annotate non coding RNAs [20]

CDS features were added and final formatting of the GFF3/GTF files were done with the genometools program [21]

The functional annotation of the genes consisted in the prediction of long non-coding RNAs with PLEK [22] whereas descriptions and Gene Ontology annotations were attributed with the software Pannzer2 [23] setting a Minimum query coverage or a minimum sbjct coverage of 0.4 and a Minimum alignment length of 50. KEGG annotations were obtained using the KEGG annotation server and selecting the following as species: hsa, mmu, rno, dre, cel, ecu, nme, bsu, mtu, syn, aae, ape, lcf, npr, lcm, sdu, tru, ipu, srx, xma, onl, cvg.

Comparative Genomics

Comparative genomics analyses were performed to evaluate the conservation of swordfish protein coding genes as well as the expansion/contraction of gene families. The proteomes of the following species were downloaded from Ensembl release 99: *Amphiprion ocellaris, Clupea harengus, Cyprinus carpio, Danio rerio, Esox lucius, Fundulus heteroclitus, Gadus morhua, Ictalurus punctatus, Lepisosteus oculatus, Mola mola, Oreochromis niloticus, Oryzias latipes, Salmo salar, Scophthalmus maximus, Seriola dumerili, Sparus aurata and Tetraodon nigroviridis*. The proteome of the white shark was downloaded from Nicholas et al. [24]. The longest protein per gene was selected from each species and they were used as input for Orthofinder version 2.3.7 [25]. The obtained species tree was converted to an ultrametric tree using Orthofinder with the option -r 450000000 and shown in **Figure 4** and the relative statistics are shown in **Table 2.** A total of 22575 (52.3%) and 20602 (47.7%) swordfish genes were assigned and unassigned to orthologues groups, respectively. A relevant fraction of the unassigned genes belonged to non-coding genes (15587, 75.6%), of which 4974 are coding, 14798 are non-coding, 789 are non-coding low confidence and 41 are unknown. A total of 1514 (3.5%) genes were assigned to swordfish-specific groups and this fraction likely reflect a species-specific evolutionary fingerprint. Indeed, among the most represented and statistically significant biological processes in this species-specific groups, there were DNA-mediated transposition (GO:0006313), DNA integration (GO:0015074) and fat cell differentiation (GO:0045444). In particular, many genes belonging to "*cut and paste*" DNA transposons (i.e. *Tc1*) fell in this group and were highly represented in several swordfish-specific groups (OG0009250; OG0001639). The unique signature of these groups would suggest the involvement of the transposon-related molecular machinery in the swordfish genome evolution and as stress response against environmental challenges (i.e. pollution). However, such a hypothesis needs to be supported and confirmed by further *ad-hoc* analysis and data mining on the swordfish genome.

Gene family expansion and contraction was analysed with the software CAFE [26] using the ultrametric tree and the output from Orthofinder with the option -r 10000. The results from CAFE were filtered in order to keep only families showing a significant expansion/contraction (p-value <=0.05). Results are shown in **Table 3**.

The lists of genes showed to be in expansion/contraction in the swordfish genome as well as the species specific genes were analysed with a Gene Ontology Enrichment Analysis (GOEA) in order to identify the associated functions. The analysis was performed with an in house script [27]. The 1021 genes belonging to expanding gene families were found associated with 95 biological process, 95 molecular function and 93 cellular components. As shown in **Figure 5** the main expanding biological process were those related to immune response such as immunoglobulin production (37 genes), immune response (52 genes), regulation of wound healing (18 genes), negative regulation of I-kappaB kinase/NF-kappaB signalling (7 genes) and neutrophil chemotaxis (6 genes). Interestingly, among 1021 genes belonging to expanding families, only five are rapidly expanding and all of them are involved directly on the immune response. These findings highlight that the swordfish population is facing parasite/microbial infection and, generally, marine pollution. The acquirement of ability to physiologically respond to environmental stresses would probably be an investment, from a genetic point of view, to improve and modulate the immune system. At the same time, the 3585 genes belonging to families in contraction were found associated to 99 biological process, 95 molecular function and 95 cellular components. The contracting biological process were those related to DNA integration and carbohydrate and lipid metabolism (**Figure 6).** Among the 3585 genes belonging to contracting gene families a total of 57 genes are rapidly evolving and most of them are involved directly on cartilage, nervous system and heart development.

The lncRNA transcripts detected in the swordfish genome were mapped against the genomes of the abovementioned species with the software GMAP [28] with the option --cross-species. Then by using in-house scripts the percentages of identity and coverage were calculated for each transcript and shown in **Table 4**.

Double digest restriction-site associated DNA (ddRAD) analysis

DNA was extracted from muscle or fin samples of 96 fish caught in North Atlantic Ocean and Mediterranean Sea (**Figures 7**). Raw reads from 96 ddRAD samples were trimmed with BBDuk [2] setting a minimum base quality of 30 and a minimum read length of 30 bp. The trimmed reads were mapped against the *X. gladius* genome with minimap2 [7] with the option -x sr. Duplicates were removed with sambamba version 0.7.0 [29]. Variant Calling was then performed with Platypus [8] with the options --minReads=4 --minFlank=10 --minMapQual=10 --minBaseQual=30. VCFtools [9] was used to retain only variants labeled as "PASS" and with a MAF >=0.05. A P-distance matrix was also calculated with vcftools as well as a window-based calculation with the following options --window-pi 500000 --window-pi-step 50000. The Mediterranean and the Atlantic populations were compared to identify genomic regions with different P-values. Their distribution was compared with a Wilcoxon-test. The per-window difference in the P-values between the Mediterranean and Atlantic population was calculated, which showed a normal distribution. The Atlantic population showed a significantly ($P<10^{16}$) higher average diversity with respect to the Mediterranean one (**Figure 8**). The mean and standard deviation of the P-distance differences were then calculated and the regions showing a difference higher/lower than the mean plus or minus two standard deviations were considered as significantly different and were shown in **Figure 9**. The associated genes were then extracted and a GOEA was performed. 1048 genes were associated to the genome region where the diversity of the Mediterranean population is higher than the Atlantic one. Biological processes related to those 1048 genes resulted mainly linked to response to several internal and environmental stimuli such as response to hydrogen peroxide, mechanical stimulus, organic cyclic compound, drug inflammatory response, glucose stimulus, cold, response to UV, bacterium and endoplasmic reticulum stress (**Figure 10**). On the contrary, 688 genes were associated to the genome region where the diversity of the Atlantic population is higher than the Mediterranean one. Biological processes related to those 688 resulted mainly linked to nervous system, muscular and digestive tract development and pigmentation (**Figure 11**).

These results suggest that Mediterranean swordfish is investing their energy on modulating the molecular machinery involved in the response to several stimuli probably related to the "closed" and polluted environment of the Mediterranean Sea with the respect to those of the Atlantic Ocean. On the other side, swordfish belonging to Atlantic population seems to invest their energy on process related to migration and feeding.

The association between the different metadata variables associated to the samples was assessed by Fisher test considering a *p*-value less than 0.05 as significant.

Population Genetics

Population genetic analyses was performed to evaluate the structure of the swordfish populations and identify the genetic differences among multiple populations. Sequence data analysis led to the identification of 133.166 single nucleotide polymorphisms (SNPs) in 96 genotypes, A total of 1772 and 109.247 SNPs in North Atlantic specimens and Mediterranean specimens were found, respectively. A Bayesian clustering algorithm elaborated by the program STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to assign 96 specimens to two populations, Mediterranean and Atlantic. This software is designed to explore differences in the distribution of genetic variants by placing samples into groups whose members share similar patterns of variation. It both identifies populations from the data and assigns individuals to that population representing the best fit for the variation patterns found. The analysis was run by setting the predefined number of clusters (K) between 1 and 3 (i.e. the number of sampling areas that we assumed). The "admixture model" was used on the whole dataset with no previous population information and one run with a burnin period of 15000 and 15000 MCMC reps after burnin were performed for the number of clusters predefined.

With K=2, the Bayesian analysis carried out with Structure software revealed coherence in the geographical structuring of populations with the results based on genetic distances. STRUCTURE analysis confirmed that 25 North Atlantic specimens and 70 Mediterranean specimens were correctly assigned to their respective populations

(**Figure 12**). Adriatic, Sardinian, Sicilian, Greek and a part of Spanish specimens were assigned to the Mediterranean population, while all Canadian specimens were assigned to the North Atlantic population. However, two specimens from the Spanish area showed an intermediate genotype between the two populations. This result would suggest that there could be a mixing zone between North Atlantic Ocean and Mediterranean Sea, where, probably, Atlantic and Mediterranean specimens breed (**Figure 13**).

The populations were also analysed individually. First, the sequence data analysis led to the identification of 109.247 SNPs in the Mediterranean population. This finding would indicate evolution forces acting on the genetic variation of population. Indeed, if SNPs change either the functionality of a gene or its expression, and the change provides greater fitness for a population (i.e. a higher capacity to survive and/or reproduce in a given environment), the change will be favoured by natural selection. Therefore, SNPs can be the basis of evolutionary change. STRUCTURE investigated the Mediterranean population, using one run with a burnin period of 15000 and several MCMC reps after burnin of 15000 performed for a number of clusters K=3. Although the analysis was performed with K=3, the software detected only 2 clusters of allelic frequencies. Analysing the distribution of these 2 clusters we assumed that in the Mediterranean Sea there are two subpopulations, with different clusters of allelic frequencies (**Figure 14**). Perhaps, one of this subpopulation could be identified as the Western Mediterranean Sea area (1) and the second population could be identified as the Eastern Mediterranean Sea area (2). A mixing area in Central Mediterranean Sea could be identified, with allelic frequencies between Eastern and Western subpopulations (3) (**Figure 15**).

Secondly, the sequence data analysis led to the identification of 1772 SNPs in the North Atlantic population. This could be related to a low number of specimens or a different evolution forces (genetic drift or bottleneck effect), other than Mediterranean population, acting on population that reduced the frequencies of rare alleles. STRUCTURE investigate the North Atlantic population using one run with a burnin period of 15000 and several MCMC reps after burnin of 15000 performed for a number of clusters K=3. Based on the Bayesian analysis, we assume that the North Atlantic population is potential panmictic population with 3 cluster of allelic frequencies (**Figure 16**). All specimens showed three different genetic variants (**Figure 17**).

In conclusion, the observed clustering indicated the ability of SNPs markers to group together the related genotypes from geographical regions with high level of accuracy. Even though this represents a preliminary study on the swordfish population genetics, our result demonstrated a significant differentiation between Mediterranean and North Atlantic population and a potential mixing area. This result was achieved by using a new sequencing technology which allowed the discovery of a large SNPs data set. Furthermore, the analyses of single populations suggest potential subpopulations in the Mediterranean Sea and three allelic frequencies shared by all specimens of North Atlantic population**.**

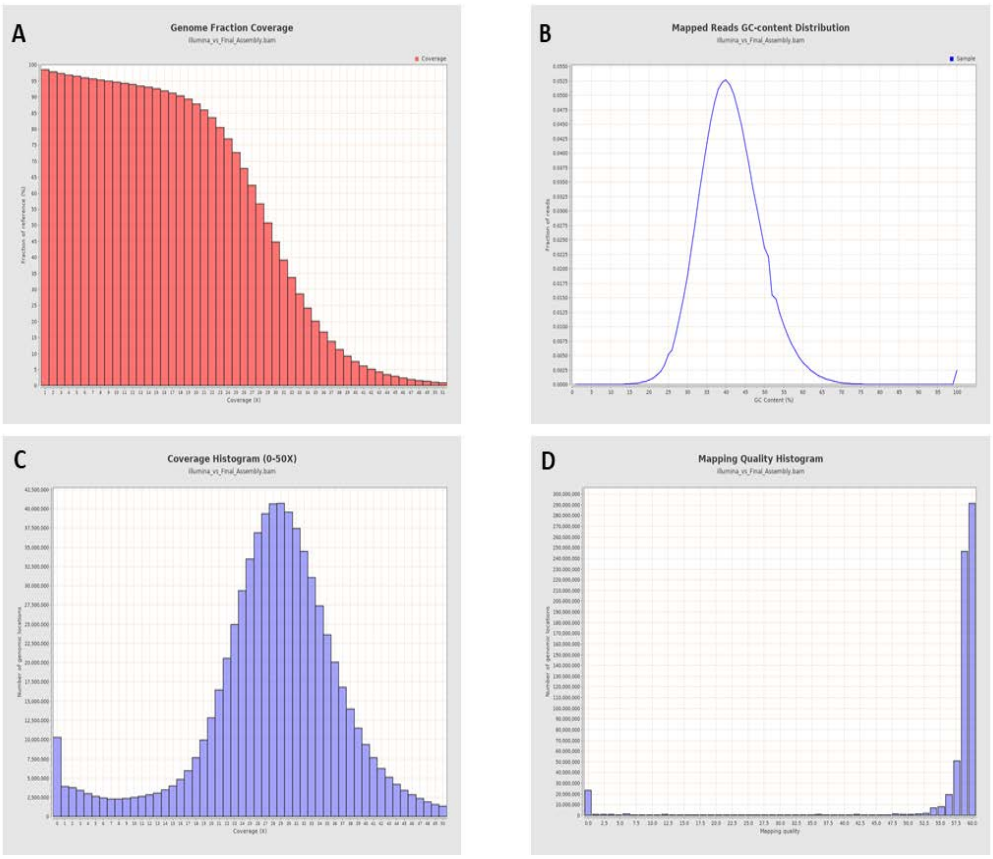


Figure 1: Plots representing: genome fraction coverage (A), GC content distribution (B), coverage histograms (C) and mapping quality histograms (D) obtained by Illumina sequencing.

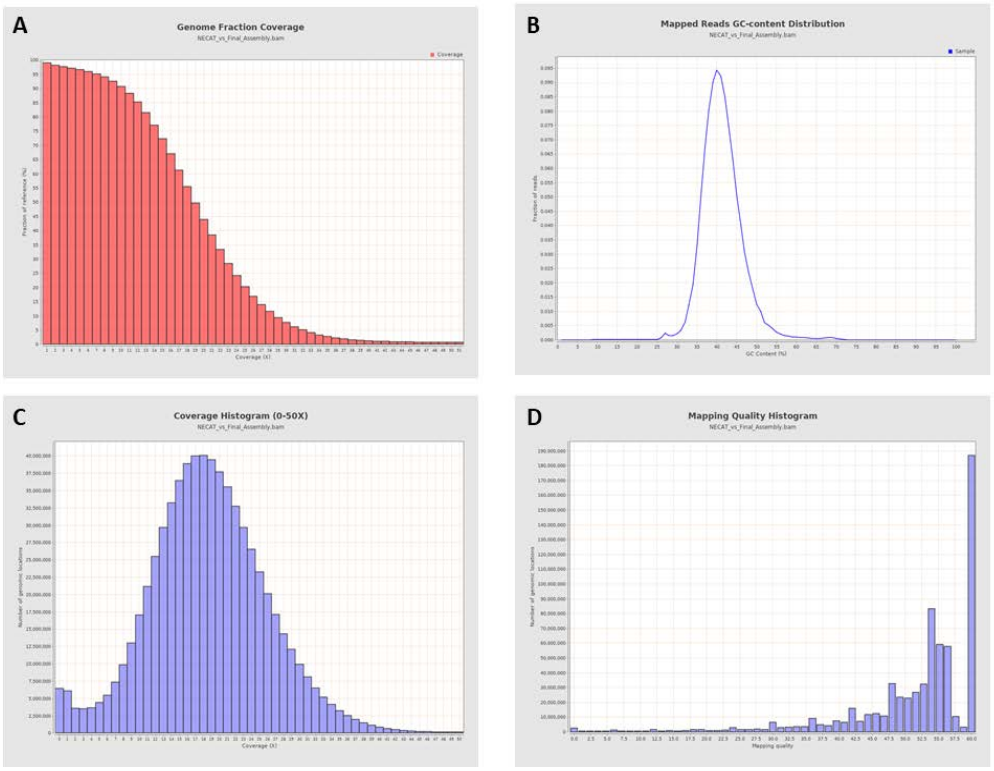


Figure 2: Plots representing: genome fraction coverage (A), GC content distribution (B), coverage histograms (C) and mapping quality histograms (D) obtained by Nanopore sequencing.

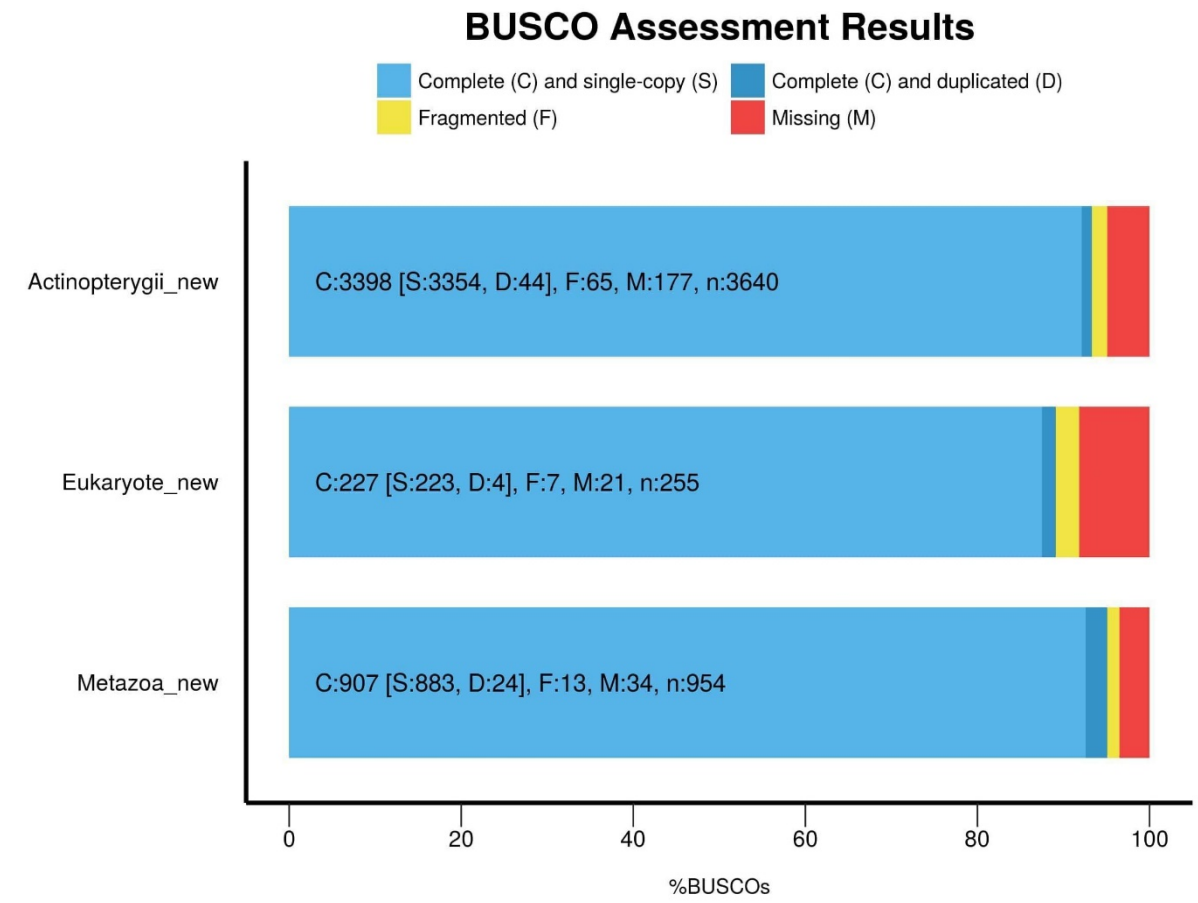


Figure 3: BUSCO assessment results.

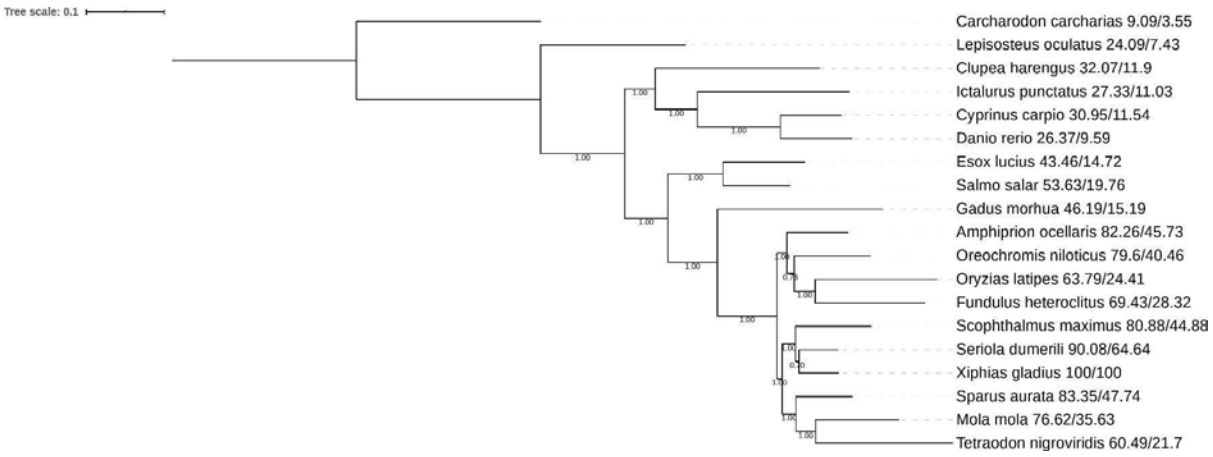


Figure 4. Phylogenomic tree. The first value following the species name is the percentage of conserved protein-coding genes and the second value the percentage of conserved non-coding genes between swordfish and single species.

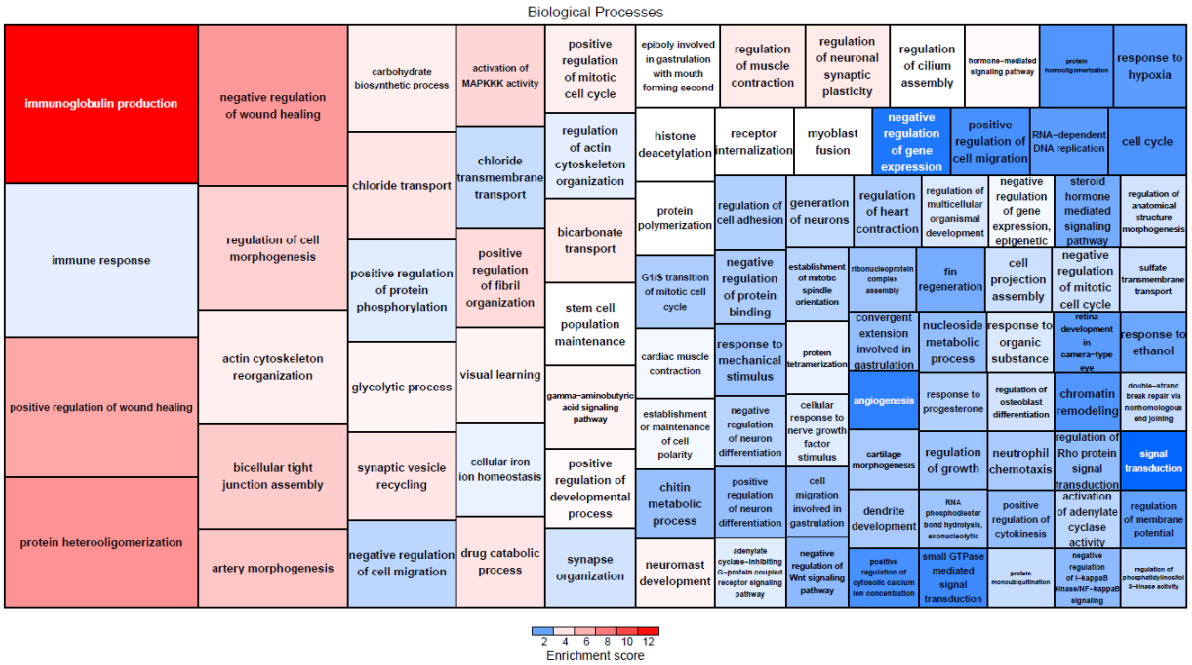


Figure 5: Biological Process of 1021 expanding genes families.

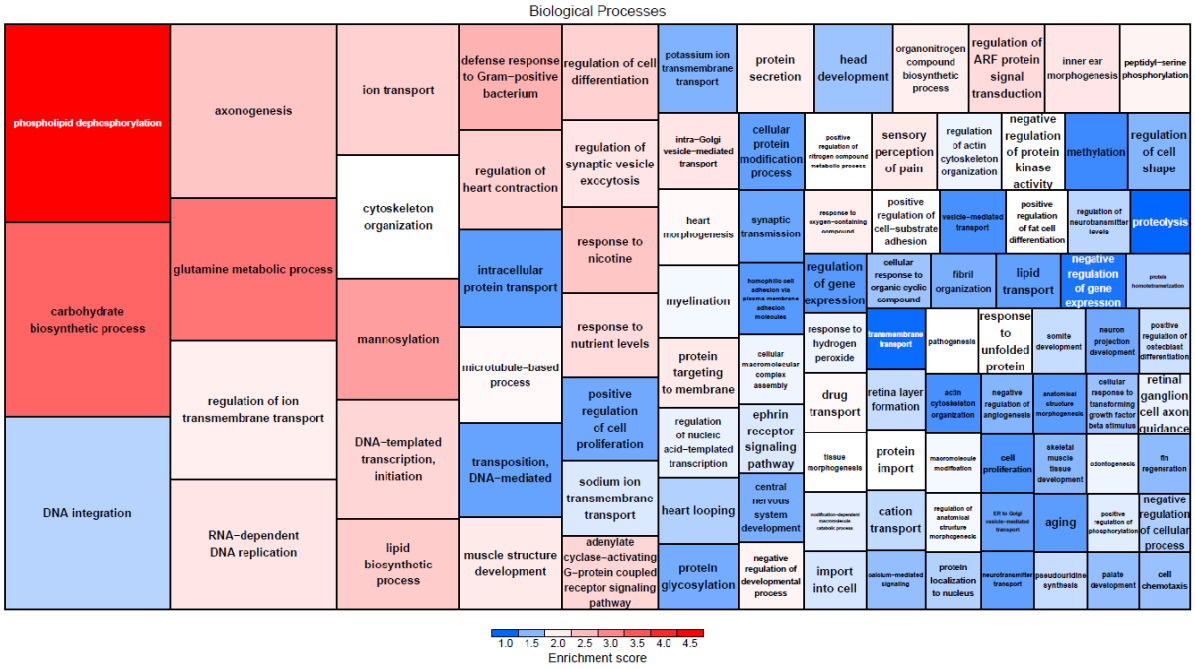


Figure 6: Biological Process of 3585 contracting gene families.

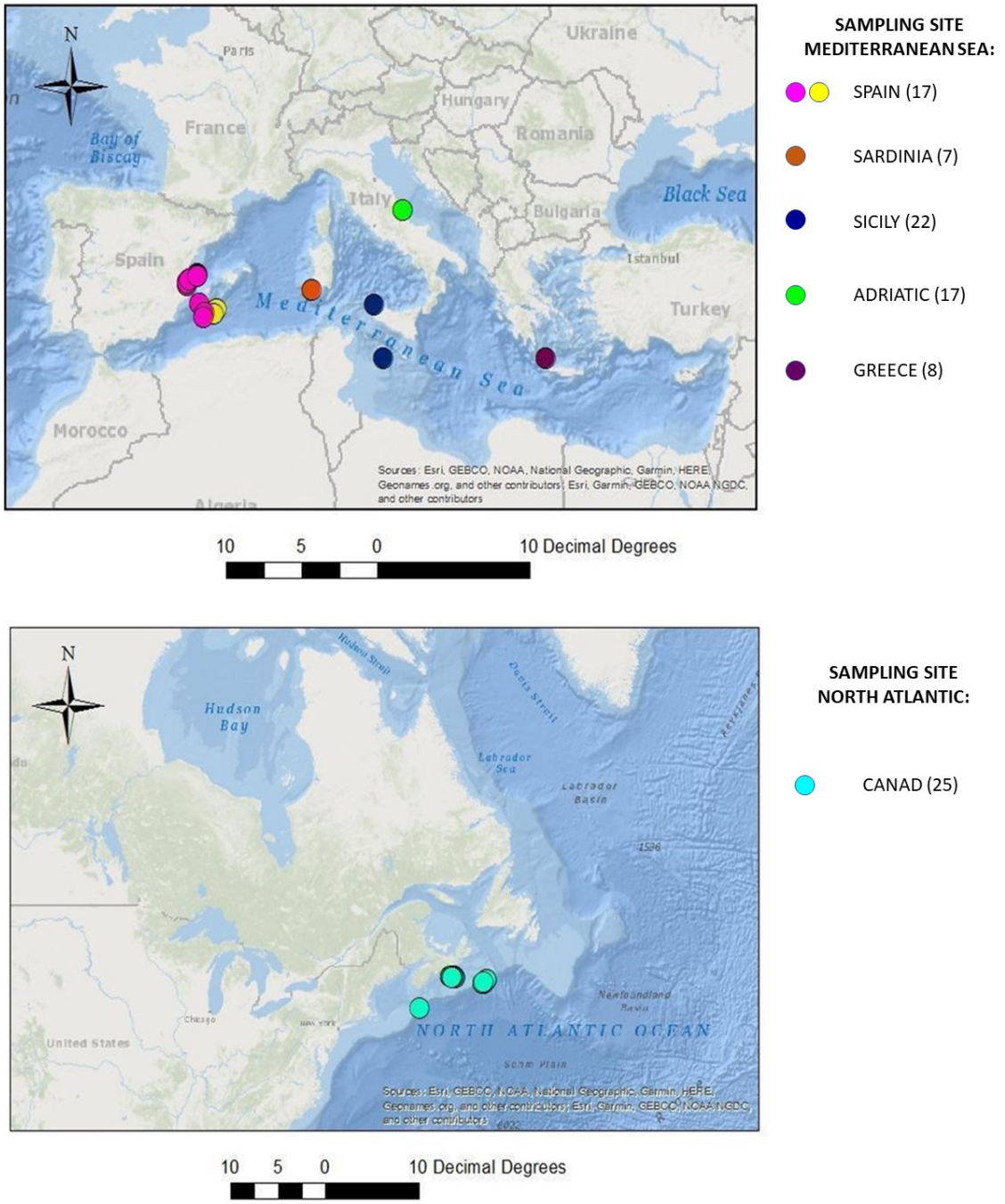


Figure 7. The map of Mediterranean and Atlantic sampling sites. The number of samples per site is shown in brackets.

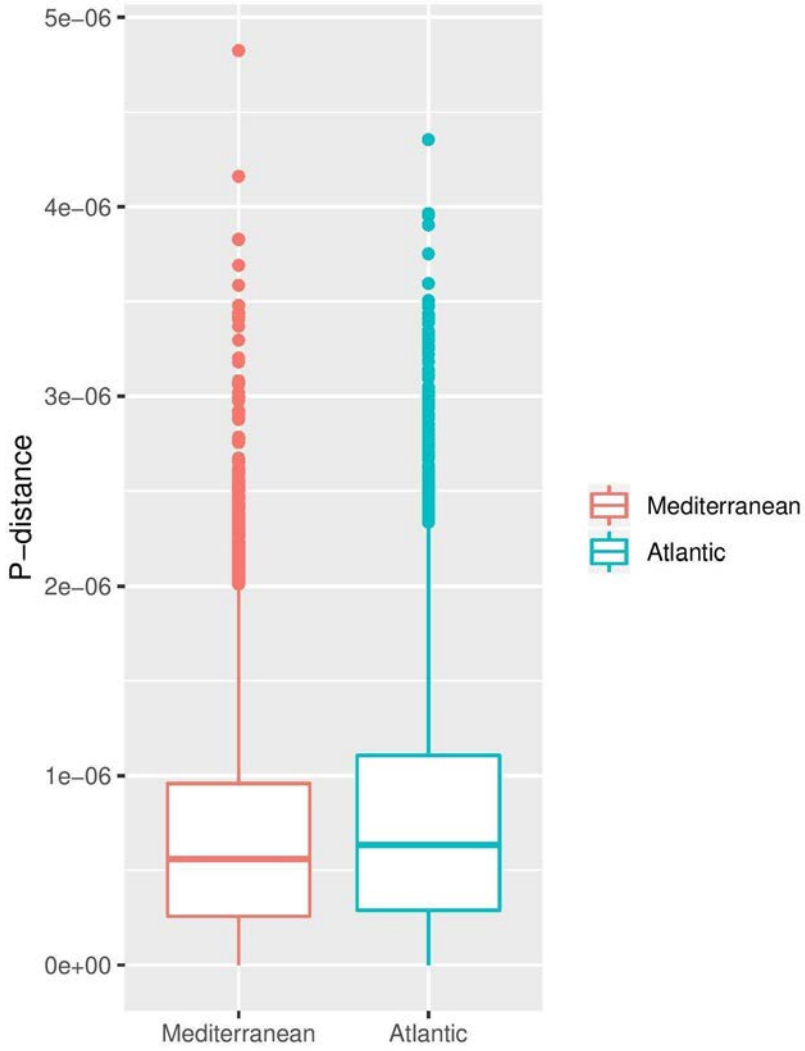


Figure 8. Average genetic diversity between Atlantic and Mediterranean populations

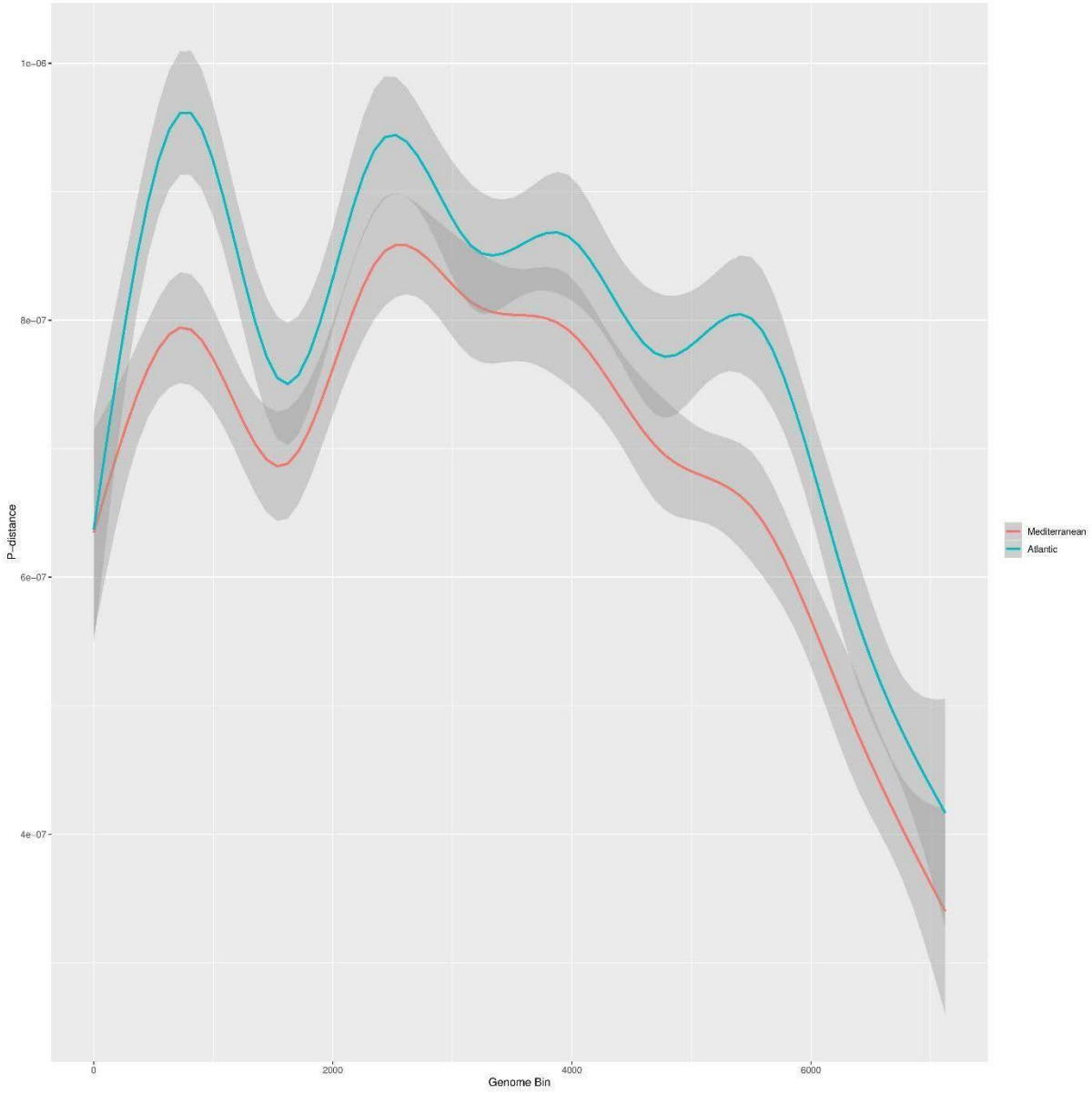


Figure 9. The mean and standard deviation of the P-distance along the swordfish genome

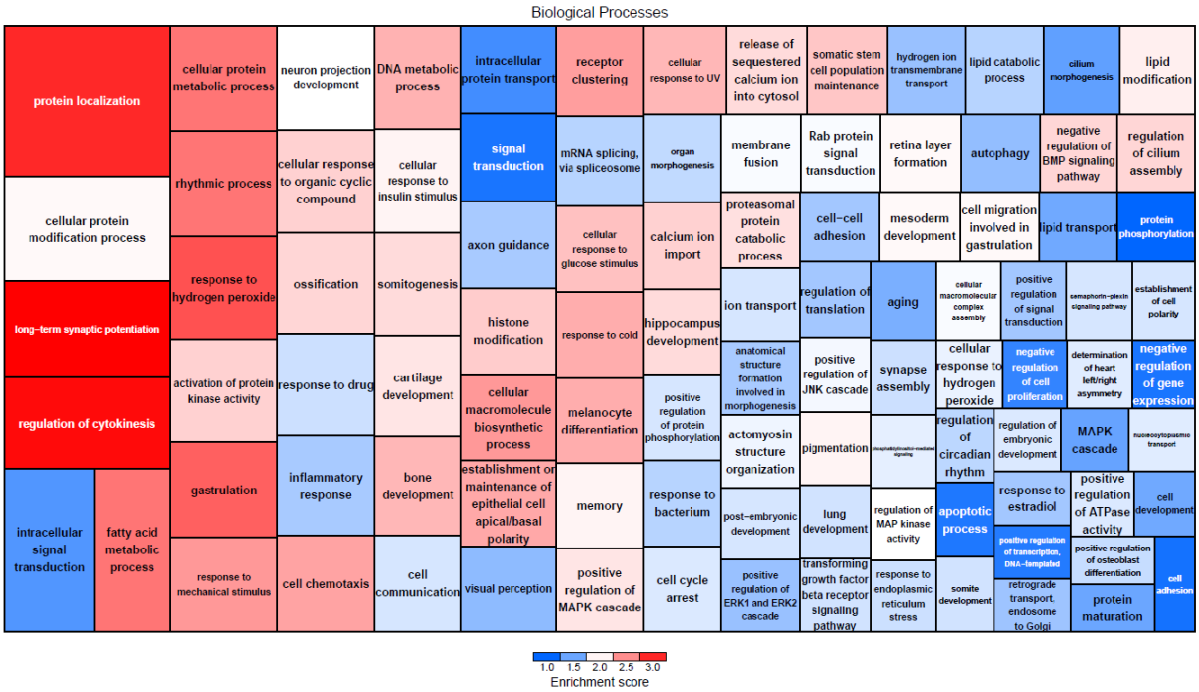


Figure 10: Biological Process of 1048 genes showing increased diversity in Mediterranean population with respect to that of the Atlantic.

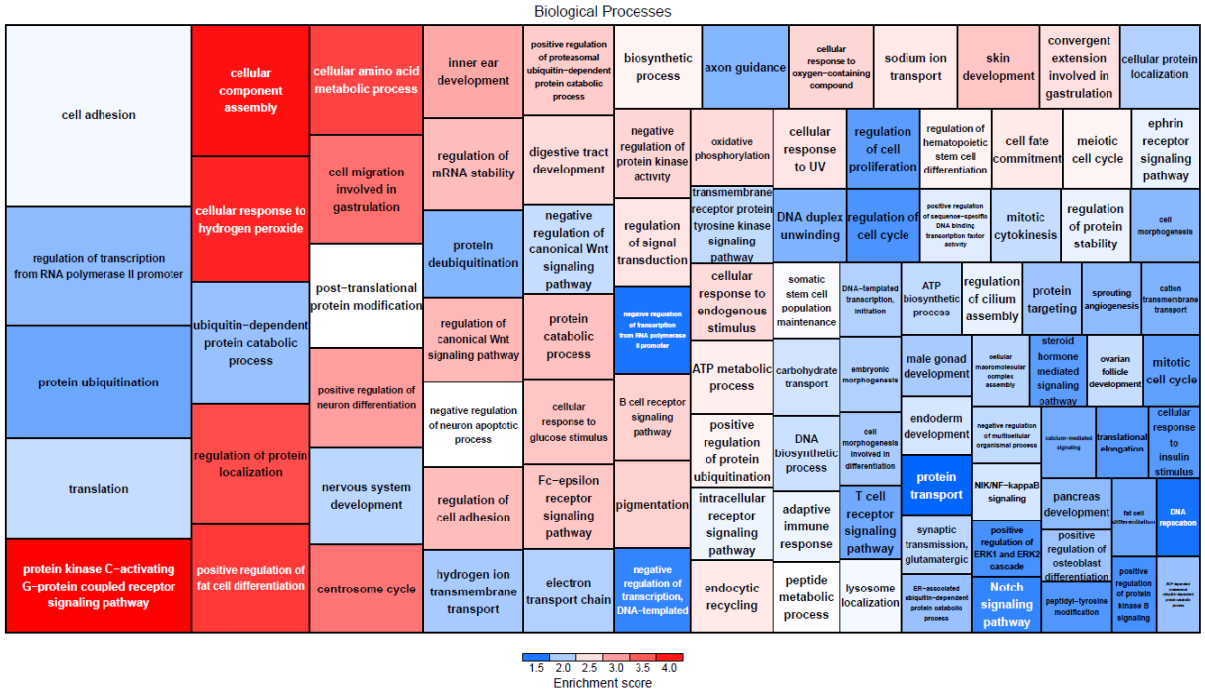


Figure 11: Biological Process of 688 genes showing increased diversity in Atlantic population with respect to that of the Mediterranean one.

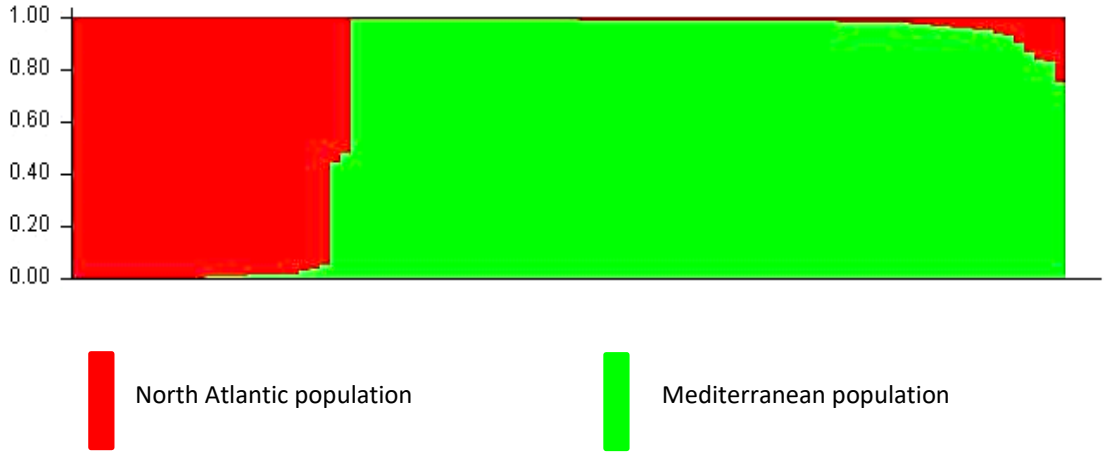


Figure 12. Genetic diversity structure estimated with the Structure software. The two clusters were determined for K=2.

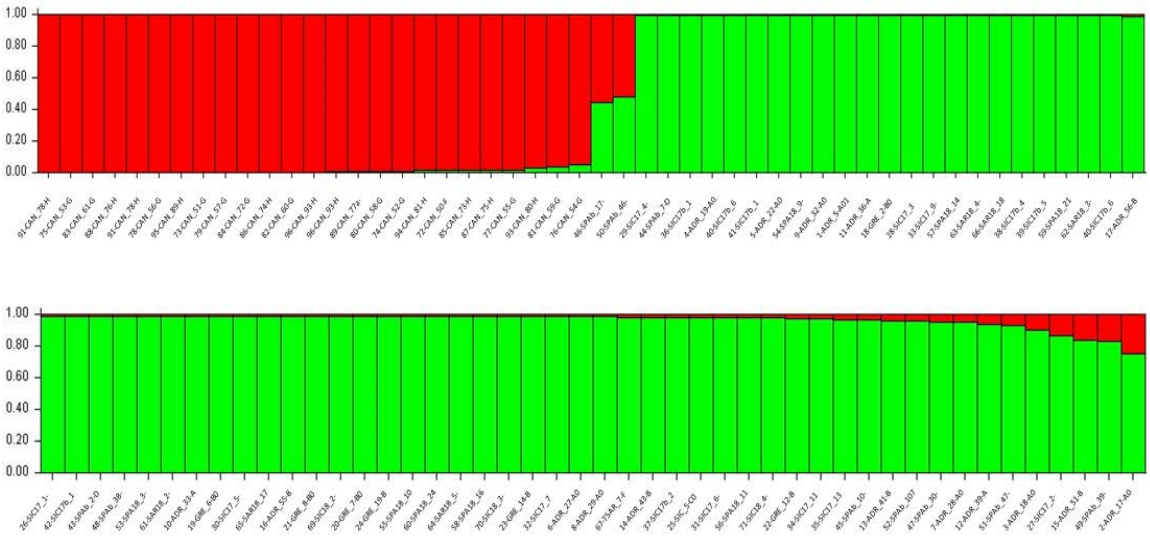


Figure 13. Genetic diversity structure estimated with the Structure software with sort by *Q* in multiple lines.

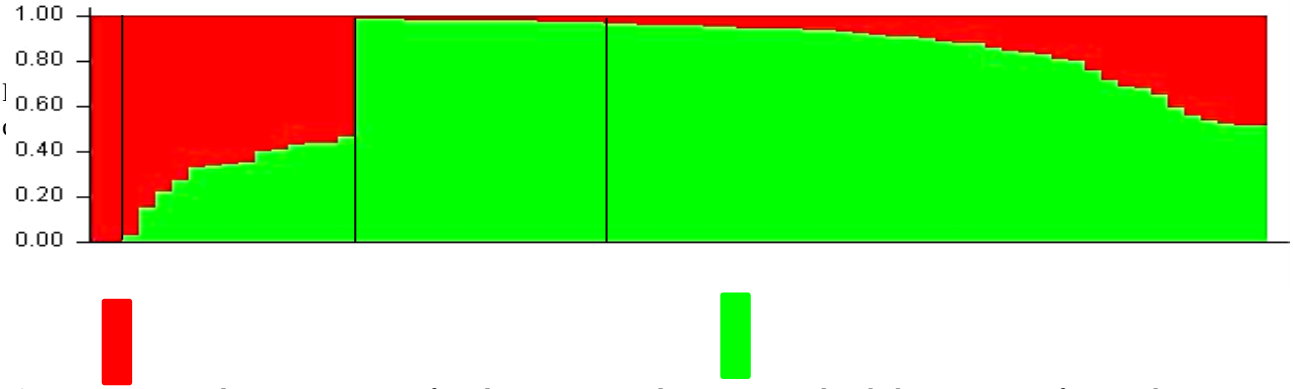


Figure 14. Genetic diversity structure of Mediterranean population estimated with the Structure software. The two clusters were determined for K=3.

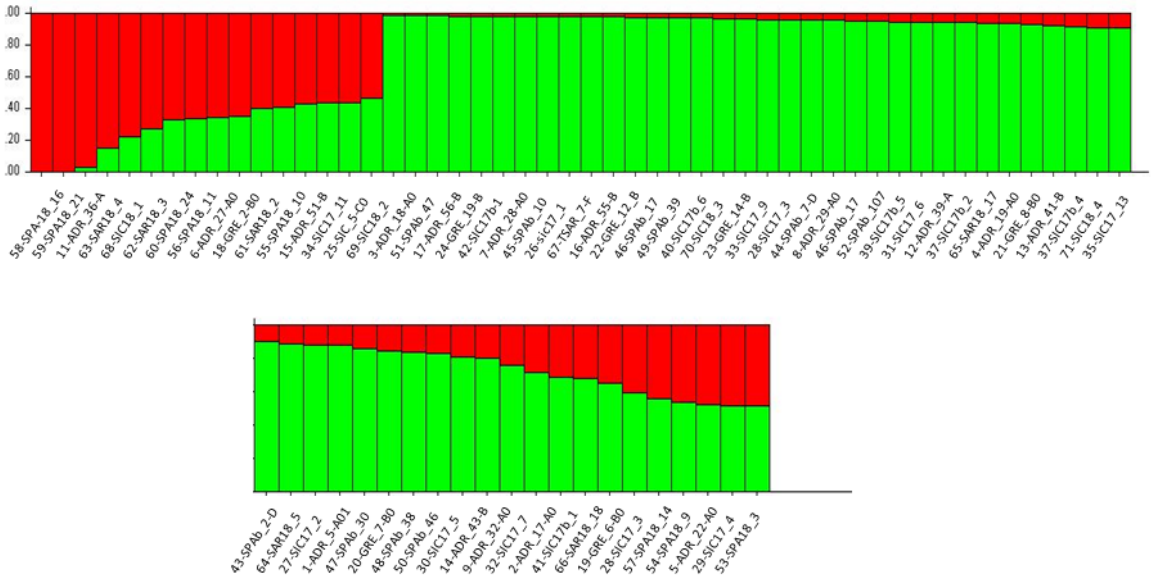


Figure 15. Genetic diversity structure estimated with the Structure software with sort by *Q* in multiple lines.

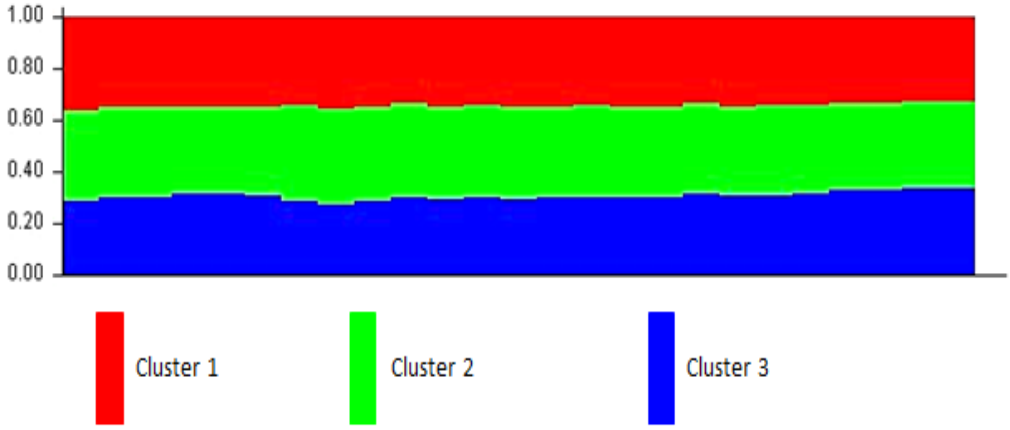


Figure 16. Genetic diversity structure of North Atlantic population estimated with the Structure software. The three clusters were determined for K=3.

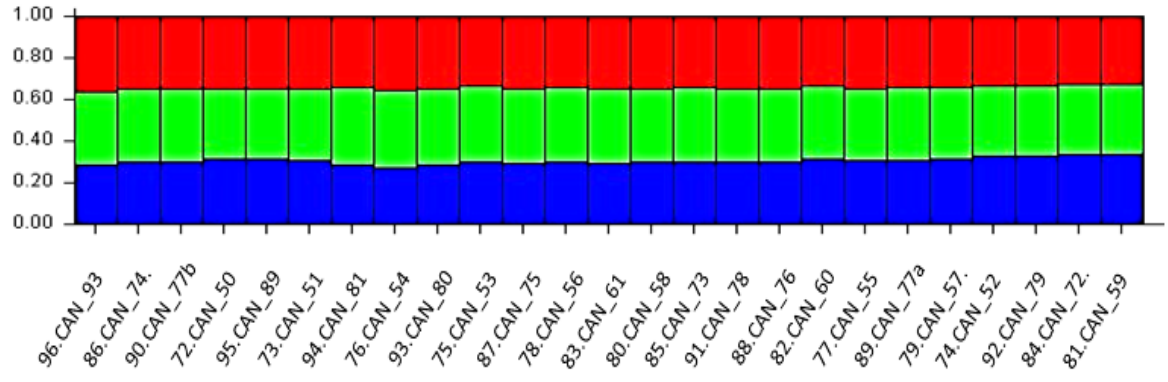


Figure 17. Genetic diversity structure estimated with the Structure software with sort by *Q* in multiple lines.

Table 1. QUAST genome assembly statistics report.

Assembly	**X gladius genome**
# contigs (>= 0 bp)	4493
# contigs (>= 1000 bp)	4492
# contigs (>= 5000 bp)	4114
# contigs (>= 10000 bp)	3405
# contigs (>= 25000 bp)	2292
# contigs (>= 50000 bp)	1729
Total length (>= 0 bp)	687458128
Total length (>= 1000 bp)	687457277
Total length (>= 5000 bp)	686067322
Total length (>= 10000 bp)	680759310
Total length (>= 25000 bp)	663003833
Total length (>= 50000 bp)	642928685
# contigs	4493
Largest contig	5025521
Total length	687458128
GC (%)	40.71
N50	937170
N75	218898
L50	194
L75	578
# N's per 100 kbp	59.86

Table 2: Statistics Per-Species of total number of genes, genes in orthogroups (OG), unassigned genes and genes in species-specific OG resulted from the comparative genomic analysis with OrthoFinder.

	Genes	**Genes in OG**	**Unassigned genes**	**Genes in species-specific OG**
Amphiprion_ocellaris	23592	23155	437	79
Clupea_harengus	24095	23378	717	347
Cyprinus_carpio	44721	43502	1219	614
Danio_rerio	30313	29562	751	1204
Esox_lucius	23954	23373	581	65
Fundulus_heteroclitus	23069	22544	525	162
Gadus_morhua	20095	19284	811	49
Ictalurus_punctatus	23651	23065	586	218
Lepisosteus_oculatus	18341	17966	375	203
Mola_mola	21404	20942	462	40
Oreochromis_niloticus	28189	27397	792	659
Oryzias_latipes	22127	21503	624	246
Salmo_salar	47329	42989	4340	1500
Scophthalmus_maximus	21000	20613	387	4
Seriola_dumerili	23278	22957	321	37
Sparus_aurata	25222	24696	526	218
Tetraodon_nigroviridis	19602	19164	438	158
White_shark_proteins	24520	19893	4627	1957
Xiphias_gladius	**43177**	**22575**	**20602**	**1514**

Table 3. Summary of expanding and contracting gene families according to the CAFE analysis.

Species	Expanded fams	Genes gained	Contracted fams	Genes lost	No change	Avg. Expansion
Oryzias latipes	267 (20)	451	1369 (31)	1519	9853	-0.0929585
Scophthalmus maximus	232 (7)	277	1214 (43)	1449	10043	-0.102011
Fundulus heteroclitus	498 (25)	686	644 (14)	720	10347	-0.0029595
Lepisosteus oculatus	373 (5)	602	4187 (18)	4711	6929	-0.357646
Cyprinus carpio	6723 (70)	9681	293 (5)	322	4473	0.814605
Clupea harengus	742 (38)	1347	1983 (5)	2167	8764	-0.0713726
Oreochromis niloticus	528 (89)	1816	749 (9)	799	10212	0.0885195
Esox lucius	129 (11)	200	1165 (18)	1359	10195	-0.100879
Xiphias gladius	**657 (6)**	**857**	**2007 (60)**	**2464**	**8825**	**-0.139873**
Tetraodon nigroviridis	790 (9)	971	1730 (47)	2007	8969	-0.0901732
Amphiprion ocellaris	778 (25)	1003	689 (8)	738	10022	0.0230655
Daniorerio	597 (27)	948	1272 (40)	1549	9620	-0.0523109
Carcharodon carcharias	1750 (12)	3052	2588 (3)	3154	7151	-0.0088780
Salmo salar	6948 (78)	10958	369 (1)	382	4172	0.920533
Sparus aurata	500 (76)	1628	782 (4)	806	10207	0.0715467
Ictalurus punctatus	376 (20)	764	1527 (10)	1712	9586	-0.0825137
Gadus morhua	323 (13)	596	2608 (36)	3006	8558	-0.209766
Mola mola	405 (14)	505	714 (17)	796	10370	-0.0253286
Seriola dumerili	476 (38)	766	545 (3)	555	10468	0.0183654

Table 4: Statistics of lncRNA analysis per species.

Comparison	Coding	Non-coding	% Coding	% Non-coding	Position Tree
Xiphias_gladius_mRNA_vs_White_shark	5642	1088	9.09	3.55	1
Xiphias_gladius_mRNA_vs_Lepisosteus_oculatus	14946	2276	24.09	7.43	2
Xiphias_gladius_mRNA_vs_Clupea_harengus	19897	3645	32.07	11.9	3
Xiphias_gladius_mRNA_vs_Ictalurus_punctatus	16956	3376	27.33	11.03	4
Xiphias_gladius_mRNA_vs_Cyprinus_carpio	19201	3534	30.95	11.54	5
Xiphias_gladius_mRNA_vs_Danio_rerio	16360	2936	26.37	9.59	6
Xiphias_gladius_mRNA_vs_Esox_lucius	26963	4507	43.46	14.72	7
Xiphias_gladius_mRNA_vs_Salmo_salar	33276	6050	53.63	19.76	8
Xiphias_gladius_mRNA_vs_Gadus_morhua	28657	4650	46.19	15.19	9
Xiphias_gladius_mRNA_vs_AmpOce	51033	14002	82.26	45.73	10
Xiphias_gladius_mRNA_vs_Oreochromis_niloticus	49384	12387	79.6	40.46	11
Xiphias_gladius_mRNA_vs_Oryzias_latipes	39577	7473	63.79	24.41	12
Xiphias_gladius_mRNA_vs_Fundulus_heteroclitus	43073	8672	69.43	28.32	13
Xiphias_gladius_mRNA_vs_Scophthalmus_maximus	50177	13740	80.88	44.88	14
Xiphias_gladius_mRNA_vs_Seriola_dumerili	55885	19791	90.08	64.64	15
Xiphias genome	62042	30618	100	100	16
Xiphias_gladius_mRNA_vs_Sparus_aurata	51712	14616	83.35	47.74	17
Xiphias_gladius_mRNA_vs_Mola_mola	47534	10908	76.62	35.63	18
Xiphias_gladius_mRNA_vs_Tetraodon_nigroviridis	37532	6645	60.49	21.7	19

References

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

https://jgi.doe.gov/data-and-tools/bbtools/

https://www.biorxiv.org/content/10.1101/2020.02.01.930107v1.full.pdf

https://genome.cshlp.org/content/early/2017/03/15/gr.215087.116

https://www.nature.com/articles/s41592-019-0669-3

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5100563/

https://www.ncbi.nlm.nih.gov/pubmed/29750242

https://www.nature.com/articles/ng.3036

https://academic.oup.com/bioinformatics/article/27/15/2156/402296

https://www.nature.com/articles/s41598-019-43872-6

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5322827/

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112963

https://www.ncbi.nlm.nih.gov/pubmed/23422339

https://www.ncbi.nlm.nih.gov/pubmed/31020564

Smit AFA, Hubley R, Green P (1996–2013). RepeatMasker Open-4.0. URL http://www.repeatmasker.org. Last accessed: August 2013.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3571712/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2134774/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160219/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160247/

https://www.ncbi.nlm.nih.gov/pubmed/24091398

https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-311

https://www.ncbi.nlm.nih.gov/pubmed/29741643

https://www.pnas.org/content/116/10/4446

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0721-2

https://academic.oup.com/bioinformatics/article/22/10/1269/237347

https://www.sciencedirect.com/science/article/pii/S2211124719308034

https://academic.oup.com/bioinformatics/article/21/9/1859/409207

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765878/