A Review of *Sarda sarda* Population Structure Using the Mitochondrial Control Region as a Genetic Marker

Student: David Bartrès Gordils

E-mail: bartresdavid@gmail.com

Double Degree in Biology and Biotechnology

Tutor: Jordi Viñas de Puig

E-mail: jordi.vinas@udg.edu

Date of deposition of memory through TFG platform: 03/07/2024

Acknowledgements

I would like to thank my tutor Jordi Viñas for his continued support and guidance throughout the elaboration of this study as well as the whole LIG for having produced previously all the data that was used in this work and that made it possible in the first place.

I would also like to thank "Peixateria Mercè" for lending me a specimen of Atlantic bonito caught in the Catalan coast for photographic purposes for this work.

Finally, I am grateful to all my friends and family that have supported and encouraged me through this study.

Index

Resum

El bonítol atlàntic *(Sarda sarda*) és un peix marí epipelàgic i nerític que habita la major part de les costes del Mediterrani i l'Atlàntic. És una de les espècies de petits túnids més rellevants en aquestes zones per la seva importància econòmica i nutricional. Tot i el seu interès, es requereix més informació genètica sobre aquesta espècie i la seva estructura poblacional per a una millor comprensió i possibles enfocaments de gestió futura. En aquest estudi s'ha fet una revisió de dades prèvies pertanyents al Laboratori d'Ictiologia Genètica de la Universitat de Girona mitjançant diversos tests genètics i estadístics. La mida de mostra resultant consisteix en gairebé 1000 individus de *S. sarda,* sent fins on sabem l'estudi més extens d'aquesta espècie a l'actualitat, abastant gairebé tota la distribució de l'espècie. Els resultats demostren una alta variabilitat molecular del bonítol atlàntic a les localitzacions estudiades i es correlacionen amb una divisió en dos clades filogenètics altament diferenciats, els quals van ser descrits en estudis anteriors. Aquests clades semblen ser més abundants en algunes localitzacions, el Clade 1 al costat atlàntic de la costa africana i a l'est del Mediterrani i el Clade 2 a l'Atlàntic nord-occidental i a l'oest del Mediterrani. Els haplotips dels individus del Clade 1 són més semblants entre ells però presenten una distribució desigual a les diferents localitzacions, mentre que el Clade 2 experimenta el fenomen oposat. L'estructura poblacional va ser clarament definida per a *S. sarda* com Atlàntic nord-occidental, la costa africana i el mar Mediterrani. L'estructura del Mediterrani (incloent-hi Portugal) es va dividir en Mediterrani occidental, Mediterrani oriental i Malta; aquests resultats, però, no han resultat prou concloents i es requereix més experimentació en aquest camp concret, principalment augmentant la mida mostral al Mediterrani est, especialment al mar de Màrmara. Les proves també van desafiar la hipòtesi de l'Estret de Gibraltar com a barrera biogeogràfica i la deriva genètica cronològica; malgrat això, es van trobar proves a favor de la deriva genètica geogràfica per al bonítol atlàntic. Considerem que aquest treball podria millorar la gestió pesquera en el futur per a *S. sarda* gràcies a la quantitat d'informació genètica i caracterització que proporciona sobre aquesta espècie, juntament amb dades sòlides i robustes per a donar-hi suport.

Resumen

El bonito atlántico (*Sarda sarda*) es un pez marino epipelágico y nerítico que habita la mayor parte de las costas del Mediterráneo y del Atlántico. Es una de las especies de pequeños atunes más importantes en estas zonas debido a su valor económico y nutricional. A pesar de su importancia, se necesita más información genética sobre esta especie y su estructura poblacional para una mejor comprensión y posibles enfoques de gestión futura. Aquí se hizo una revisión de datos previos pertenecientes al "Laboratori d'Ictiologia Genètica" de la Universidad de Girona mediante diversas pruebas genéticas y estadísticas. El tamaño muestral resultante consiste en casi 1000 individuos de *S. sarda*, siendo hasta donde sabemos el estudio más extenso de esta especie hasta la fecha, abarcando casi toda su distribución. Los resultados demuestran una alta variabilidad molecular del bonito atlántico en las localidades estudiadas y se correlacionan con una división en dos clados filogenéticos altamente diferenciados que ya han sido descritos en estudios anteriores. Estos clados parecen ser más abundantes en algunas localidades, el Clado 1 en el lado atlántico de la costa africana y en el este del Mediterráneo y el Clado 2 en el Atlántico noroccidental y en el oeste del Mediterráneo. Los haplotipos de los individuos del Clado 1 son más similares entre si, pero presentan una distribución desigual en las diferentes localidades, mientras que el Clado 2 experimenta el fenómeno opuesto. La estructura poblacional se definió claramente para *S. sarda* como Atlántico noroccidental, costa africana y mar Mediterráneo. La estructura del Mediterráneo (incluyendo Portugal) se dividió en Mediterráneo occidental, Mediterráneo oriental y Malta; estos resultados, sin embargo, no resultaron lo suficientemente concluyentes y se requiere más experimentación en este campo concreto, principalmente aumentando el tamaño de la muestra en el Mediterráneo este, especialmente en el mar de Mármara. Las pruebas también desafiaron la hipótesis del Estrecho de Gibraltar como barrera biogeográfica y la deriva genética cronológica; a pesar de esto, se encontraron pruebas a favor de la deriva genética geográfica para el bonito atlántico. Creemos que este trabajo podría mejorar la gestión pesquera en el futuro para *S. sarda* gracias a la cantidad de información genética y caracterización que proporciona sobre esta especie, junto con datos sólidos y robustos para apoyarla.

Abstract

The Atlantic bonito (*Sarda sarda*) is an epipelagic and neritic marine fish that inhabits most of the Mediterranean and the Atlantic's coasts. It is one of the most important species of small tuna in these areas due to its economical and nutritional value. Despite its importance, more genetic information about it and its population structure is required for a better understanding of this species and possible future management approaches. Here, a review on previous data belonging to the "Laboratori d'Ictiologia Genètica" of the University of Girona was made through different genetic and statistical tests. The resulting sample size consists of almost 1000 individuals of *S. sarda*, being to our knowledge the most extensive study of this species to date, encompassing almost all species distribution. Results demonstrate a high molecular variability for Atlantic bonito throughout studied locations and correlate with a division on two highly differentiated phylogenetic clades that have already been described in previous studies. These clades appear to be more abundant in some locations, Clade 1 in the Atlantic side of the African coast and the eastern Mediterranean and Clade 2 in the North-West Atlantic and the Western Mediterranean. Clade 1 individuals' haplotypes are more similar between them but experience an uneven spread throughout locations, whereas Clade 2 experiences the opposite phenomenon. Population structure was clearly defined for *S. sarda* as North-West Atlantic, African Coast and the Mediterranean Sea. Mediterranean structure (with Portugal included) was divided in the Western Mediterranean, the Eastern Mediterranean, and Malta, these results, however, did not prove conclusive enough and more experimentation in that particular field is required, mainly increasing sample size in the Eastern Mediterranean, specifically the Marmara Sea. Evidence also challenged the hypothesis of the Strait of Gibraltar constituting a biogeographical barrier and chronological genetic drift, in spite of that, evidence in favour of geographical genetic drift was found for Atlantic bonito. To our understanding, this work could improve fishery management in the future for *S. sarda* due to the amount of genetical information and characterization it provides on this species alongside strong and robust evidential data to support it.

Ethics Reflection

With each passing day, fishes increasingly become the centre of investigation in different scientific fields, however, this also highlights the importance of the way they are treated in such studies. In 2016, a revision was made of works published in prestigious journals, and a lot of them did not mention ethical criteria in the handling of the fishes studied (46% out of 250), and out of the ones that did mention them, only an 18% followed official regulations (Bennett *et al.*, 2016).

The CPCSEA (Committee for the Purpose of Control and Supervision of Experiments on Animals [CPCSEA], 2021) is a guideline for fish treatment issued by the government of India, and these regulations greatly improve fish conditions in experiments, however, these regulations can also slow the production of the article or make it more expensive, this combined with the "Publish or Perish" phenomenon can lead to bad practices that worsen the overall study.

In conclusion, experiments on fishes are not as controlled as other animals, and better conditions should be put in place while also guaranteeing some breathing room for investigators.

Sustainability Reflection

In the present study, the Atlantic bonito population worldwide was studied, and this species was subjected to overexploitation in the past, especially in some zones of the Mediterranean Sea (Öztürk, 2009). This causes a great impact in the exploited ecosystem, however, at the same time, a more moderate approach to fishing may make these products more costly, and fewer consumers would be able to purchase them.

Overall, however, environmental sensibilization in this issue would positively affect the discussed ecosystems, allowing them to regenerate during more time and generally making them healthier, that's why overfishing should be more regulated taking into account robust data.

Gender Reflection

Historically, women in research have suffered a lot of inequalities compared to men, and although in genetics this situation has improved recently (Elbardisy and Abedalthagafi, 2021), male investigators still publish 15-20% more than women (Boekhout *et al.*, 2021) which may be due to various factors such as the work environment, the sticky floor phenomenon or social pressures (Handelsmann *et al.*, 2005), this can lead to the famous scissor graphs in which a lot of women start their career in science but few end up in high ranking positions.

In conclusion, women should be empowered in science to improve this field.

1. Introduction

Atlantic Bonito or *Sarda sarda* (Bloch, 1793) is a member of the *Scombridae* family, and a member of the genus Sarda distributed along the coasts of the Atlantic Ocean and the Mediterranean Sea (Collette and Chao, 1975), zones which were studied in this work.

The Atlantic Bonito (Figure 1) is a species of epipelagic and neritic marine fish that can live in a myriad of habitats (Rey *et al.*, 1984), ranging from depths between 80 to 200 meters according to the work of Yoshida (1980), temperatures from 12ºC to 27ºC and salinities ranging from 14 to 39 (Collette and Nauen, 1983). It also has an important ecological role within the ecosystems it inhabits because it feeds on different smaller fishes, squids and shrimps (Bianchi *et al.*, 1999).

S. sarda is one of the most abundant small tuna species in the Mediterranean Sea (Sabatés and Recasens, 2001) and one of the most important worldwide commercially, especially in the Mediterranean and Black Seas (Genç *et al.*, 2018). It has a major nutritional value in the form of omega-3 fatty acids, demand for which is growing (Bimbo, 2013).

Therefore, *S. sarda* constitutes a species of major interest, and yet, despite its economic and nutritive value, there are currently no management measures for this species. However, in 2017, a specific research plan funded by the ICCAT for studying population structure prioritized the Atlantic Bonito (SCRS, 2019). The ICCAT is the International Commission for the Conservation of Atlantic Tunas, an organization that aims to oversee and coordinate research and conservation efforts to guarantee a sustainable management of tuna and other tuna-like species, and it funds many different activities regarding that objective, some of which include scientific research and stock assessments of these fishes, akin to this study.

The study of Rey *et al.* (1984) found that important spawning areas in the Mediterranean were located in the Eastern and Western Mediterranean, other articles, such as Macías *et al.* (2005) arrived at a similar conclusion, however, the exact population structure that *S. sarda* follows within the Mediterranean has been a highly controversial and discussed topic among the years (Roberti *et al.*, 1993; Pujolar *et al.*, 2001; Turan *et al.*, 2016) with no clear conclusion.

This difficulty for assessing a population structure for *S. sarda* may be due to a lack of clear biogeographical barriers, as is the case with the Strait of Gibraltar, discussed in the work of Patarnello *et al.* (2007), and the possible homogenizing impact of oceanic currents (Cowen *et al.*, 2007).

Figure 1. Atlantic Bonito (*S. sarda*) specimen captured in the Catalan coast. Source: Own.

The Strait of Gibraltar has been highly discussed as biogeographical barrier for marine species due to it not posing a major difference for some species, such as Albacore tunas (*Thunnus alalunga*) and Atlantic bluefins (*Thunnus thynnus*), the genetic pools of which are shared between the Mediterranean and the Atlantic equally (Davies *et al.,* 2011; Rodríguez-Ezpeleta *et al.,* 2019).

This lack of a concise population structure or other knowledge of this species obstructs the proposal or implementation of appropriate management measures (Lucena-Frédou *et al.,* 2021) taking into account complex and dynamic population structures that could benefit fisheries for a more sustainable future (Reiss *et al.,* 2009).

The population genetics contribute greatly to the application of regulations and management measures, therefore, the genetic marker for studying the species is also of great importance. The mitochondrial control region is a highly used genetical marker for resolving population structures and phylogenetic relationships in the family Scombridae (Viñas *et al.,* 2004; Viñas *et al.,* 2010; Viñas *et al.,* 2020; Ollé-Vilanova *et al.,* 2024). In spite of that, no global study regarding *Sarda sarda* was made until now.

In the present study, previous collected data using the mtCR as genetic marker for *S. sarda* around its distribution is reviewed. This will improve the knowledge on its population structures and genetic distribution.

2. Objectives

This work mainly aims to:

1. Recollect all data of mtDNA control region belonging to *Sarda sarda* in the Laboratori d'Ictiologia Genètica to construct a correct dataset from this very specific genetic marker that encompasses almost all of the species' distribution.
2. Define the population structure for *S. sarda* between all sampled locations and the Mediterranean alone, thus also checking for genetic drift effects and the role of the Strait of Gibraltar as a biogeographical barrier.
3. Determine how the obtained results correlate with previous studies and propose management measures.

3. Material and Methods

Dataset was constructed out of samples from two sources, the GenBank website (Clark *et al.*, 2016) and the LIG database, however, all the *Sarda sarda* sequences used in this analysis were obtained from the LIG database, they were prior described in the Jordi Viñas *et al.* works from 2004, 2010 and 2020 (Viñas *et al.*, 2004; Viñas *et al.*, 2010; Viñas *et al.*, 2020).

A total of 890 samples of Atlantic bonito (*Sarda sarda*) were genetically analysed, some of which were described and sequenced by Viñas *et al.* (2020): Côte d'Ivoire (n = 122), Senegal (n = 49), Mauritania (n = 48), Morocco (n = 60), Portugal (n = 103), Spain (n = 204), Tunis (n = 49) and Malta (n = 27). Some others were obtained from Viñas *et al.* (2004): Ligurian Sea (n = 58), Ionian Sea (n = 30), Aegean Sea (n = 53), Marmara Sea (n = 20). And finally, Palamós (Spain) (n = 54) and North West Atlantic samples (Atlanta) (n = 13) were obtained from Viñas *et al.* (2010). All locations are displayed in Figure 2.

Palamós and Spain samples are the same location with 8-9 years apart in sample collection (as is displayed in Table 1), therefore, they allow a temporal analysis of the genetic variability.

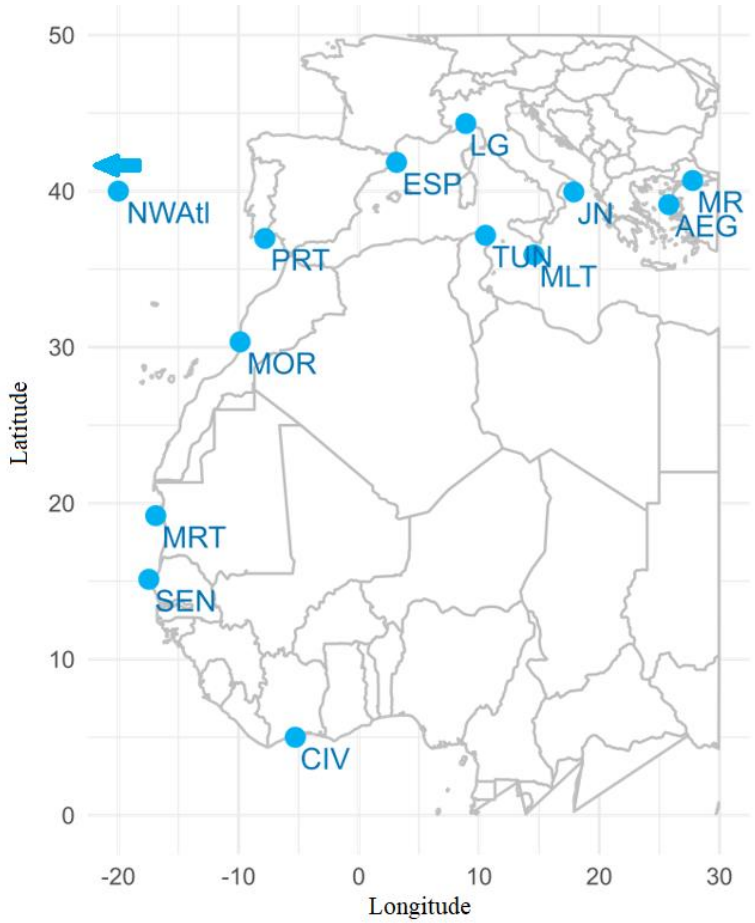

Figure 2. Geographical map of locations analysed. NWAtl's arrow symbolizes its location at the shores of America, in Atlanta. Longitude at x axis and Latitude at y axis.

A Boolean Search was performed in GenBank to find sequences belonging to the mitochondrial control region of *Sarda sarda* that were not already in the LIG's database using the search "*Sarda sarda* mitochondrial control region NOT Vinas", however, none were found. 5 results popped up when using the search "*Sarda sarda* mitochondrial NOT Vinas", however, three of those did not belong to the control region, and the last two were unverified sequences, thus, they were all discarded and only LIG sequences were used.

The LIG sequences were from 2004, 2010 and 2020, they were assembled in BioEdit (Hall, 1999; version 7.7) so they could be modified in order to obtain a consolidation of unique sets among the diverse files of sequences, any gaps that encompassed a certain position for all sequences were also checked and removed. This alignment resulted in 473 sequences (including haplotypes and sequences from single individuals) with a length of 394 nucleotides.

Through DNAsp (Rozas *et al.*, 2017; version 6), identical haplotypes were combined, treating them as the same sequence. Through this criteria, 260 distinct haplotypes were obtained.

Phylogenetic analyses were performed in MEGA11 (Tamura *et al.*, 2021; version 11), in which a phylogenetic tree was constructed using the neighbour-joining (NJ) method due to the large number of haplotypes the file had with 1000 bootstrap replicates. Additionally, a NJ tree was constructed for both the file with sequences combined through DNAsp and the file with no combined sequences whatsoever in order to check if any significant changes in distribution occurred, however, results were completely identical in spite of which method was used, further confirming that the sequences combined through DNAsp were genetically indistinguishable.

Find Best DNA option was used in MEGA to determine the best model to construct the tree, results showed that Tamura-3 parameters model (Tamura, 1992) with a gamma distribution of G = 0.35 was the most optimal for the NJ tree, which was later turned into an unrooted tree through the iTOL website (Letunic and Bork, 2024; version 6.9).

Haplotypes were classified by clades, data which was later processed in RStudio (RStudio Team, 2020) to obtain the different histograms and dispersion diagrams that were analysed, which belongs to R (R Core Team, 2023; version 2023.12.1+402). RStudio was also used to create FST heatmaps and geographical location maps.

R packages used were reshape2 (Wickham, 2007) to transform data for heatmap elaboration, stats and datasets packages (which belong to the R Team cited before) for chi-squared tests and data management, Hmisc (Harrell and Dupont, 2024) to treat heatmap data, RColorBrewer (Neuwirth, 2022) for colouring, tidyverse (Wickham, 2019) for graphical plotting, maps (Becker *et al.*, 2023) for geographical locations plot and readxl (Wickham and Bryan, 2023) to import and read excel files.

Through Arlequin (Excoffier and Lischer, 2010; version 3.5.2.2), different tests were performed; fixation indexes (FST) (Wright, 1949) were obtained for all locations with their corresponding p-values, a multitesting correction of Bonferroni (Holm, 1979) (eq. 1) was done in order to refine the signification value taking into account the number of samples.

$$Bonferroni\ Correction = \frac{pvalue}{n}$$

Equation 1. Formula of the Bonferroni correction, in which pvalue is the original p-value used for this particular study and n is the number of tests done.

Isolation by distance was computed through FST values, which were later linearized (eq. 2) in order to compare genetic distances (FST) to geographic distance and obtain the dispersion diagrams in RStudio.

$$LinFST = \frac{FST}{1 - FST}$$

Equation 2. Formula used for the linearization of FST values in order to compare them to geographic distances.

Arlequin was also used to calculate haplotype (Nei, 1987) and nucleotide diversity (Tajima, 1989) alongside AMOVAs (Excoffier *et al.*, 1992) and Mantel tests (Mantel, 1967) for comparing genetic and geographical distances. Mismatch distribution histograms of observed frequency versus expected frequency of haplotypes were generated through Arlequin and then plotted using RStudio.

Finally, the PopART program (Leigh and Bryant, 2015; version 1.7) was used in order to build a phylogenetic distance tree in which the haplotype frequency was highlighted, an Epsilon factor of 0 was used.

4. Results and discussion

As was specified before, all data was obtained from LIG for the mitochondrial control region. A total of 890 samples of *S. sarda* from 14 different locations around the world resulting in 260 distinct haplotypes with a DNA sequence of 394 bp were analysed. Although data articles were published in 2004, 2010 and 2020, samples were collected in different years, ranging from 1993 to 2021 depending on the location (Table 1). This, to our knowledge, is the most extensive and exhaustive study of this kind ever performed for *S. sarda* or other similar species.

The bonito mtDNA genetic variation was divided in two clades as will be demonstrated later, Clade 1 and Clade 2, molecular diversity analyses were conducted for all locations and results were classified according to clade, being it Clade 1, 2, or the whole sample, labelled as "All".

4.1. Molecular diversity

All localities of *S. sarda* individuals together for both clades show a high haplotype diversity value ($h = 0.955 \pm 0.004$). Clade 1 and Clade 2, whilst also being highly variable, do not exhibit the same degree of variability the whole sample does.

Out of all locations, NWAtl is the one with the biggest haplotype diversity ($h = 1.000 \pm 0.030$) whereas AEG is consistently the site with the lowest haplotype diversity ($h = 0.891 \pm 0.021$). In the case of NWAtl, an h of 1 means that every single individual of that location is a different haplotype from each other.

The majority of values oscillate around 0.9, which signifies a high level of diversity in these locations, however, when taking into consideration all individuals (both clades) for a certain location, haplotype diversity values increase in almost all cases, for example, in Spain's case,

Clade 1 has a haplotype diversity of 0.882 and Clade 2 a value of 0.907, while All has a diversity of 0.948.

The only exceptions to *h* increasing compared to Clades 1 and 2 when observing the all individuals are NWAtl, PAL, JN and MLT. These locations all have one particularity in common, their Clade 2 haplotypes appear to be extremely variable, whereas their Clade 1 haplotypes, comparatively, do not show high diversity. The only exception to this is NWAtl, which shows an *h* value of 1 in both clades, making that entire region hypervariable, however, that may also be due to sample size, since only 13 *S. sarda* individuals for that location were studied.

When considering nucleotide diversity (π), the value of π for the entire location is much higher than any of the two clades by themselves. Furthermore, π is an index that goes from 0 to 0.1, making most values of All samples (which oscillate around 0.07) pretty high.

The only exceptions to this are AEG and the African locations (CIV, SEN, MRT and MOR), which exhibit a mid to low index of nucleotide diversity, making those regions much more similar among themselves, even when considering both clades at once.

In most sites, π of Clade 1 is significantly lower than Clade 2's, except in the cases of NWAtl and SEN. This difference implies that Clade 1 haplotypes, despite being able to be more diverse than Clade 2 haplotypes in some locations, are usually more genetically similar to Clade 2's, which appear to be somewhat more variable than them.

As a whole, Table 1 shows that *S. sarda* has a high variability compared to other Scombridae species, such as is depicted in *Scomber japonicus* (Barahona *et al*., 2017), which used the same mitochondrial control region. These specimens showed a mid to high level of haplotype diversity (h = 0.793 – 0.969) but an extremely low level of nucleotide diversity compared to *S.sarda* (π = 0.004 – 0.008).

Another example of *Scomber japonicus* is the work of Zhu *et al*. (2014) with a mid to high range of haplotype diversity (h = 0.505 - 0.967) and low nucleotide diversity (π = 0.00056 - 0.01042). This study, however, used sequences belonging to the cytochrome b region of the mitochondria instead of the control region, which usually shows lower genetic diversity.

Other Scombridae members such as the *Euthynnus* genus (Ollé-Vilanova *et al.,* 2022) obtained again mid to high *h* values (h = 0.641 – 1.000) and low π values (π = 0.004 – 0.023) and *Scomberomorus cavalla* (Santa Brígida *et al.,* 2007) with mid *h* values (h = 0.545 – 0.794) and very low π values (π = 0.003 – 0.005). These studies used sequences belonging to the mitochondrial control region or D-loop, which is the same genetic marker.

This reveals a consistently high haplotype variability for *S. sarda* compared to other Scombridae species, which tend to range from mid to high values, whereas *S. sarda* displays consistently high values.

Furthermore, nucleotide variability is exceptionally high for *S. sarda* compared to other Scombridae, therefore implying extremely different DNA sequences inside this sample, which exhibits a much higher degree of variation than other studied samples. This phenomenon may be caused by two different lineages which are divergent from each other, these being Clade 1 and Clade 2.

When compared to previous studies done with this data (Viñas *et al.*, 2004; Viñas *et al.*, 2020) molecular diversity indices are similar, however, haplotype diversity values decrease, such as a value of $h = 0.961$ for LG compared to $h = 0.981$ obtained in 2004. Another example is TUN, with a haplotype diversity of $h = 0.927$ compared to h = 0.974 obtained in 2020.

With a bigger sample size than previous articles, haplotypes become less rare and haplotype diversity tends to decrease, although diversity is still high compared to other species due to two divergent clades being studied at the same time.

Nucleotide diversity values remain similar to previous studies, with LG having a global π value of $\pi = 0.072$ compared to $\pi = 0.063$ or AEG with a $\pi = 0.052$ to $\pi = 0.051$ originally from 2004. Other examples are TUN with a $\pi = 0.070$ compared to $\pi = 0.066$ or ESP with $\pi = 0.072$ to $\pi = 0.070$ from the 2020 study.

Although nucleotide diversity values are similar between works, a slight increase in value is observed in this study due to an increase in genetic difference among sequences.

Thus, even with high values compared to other species, these remain consistent with previous studies using the same data albeit with small differences due to a bigger sample being used.

Table 1. Table displaying molecular diversity indices for all studied locations with respective sample collection year, divided by clade, as will be shown in figures 3 and 4 subsequently. N: number of individuals, M: number of haplotypes, h: haplotype diversity, π: nucleotide diversity and SD: Standard Deviation. Sources: A, Viñas *et al.*, 2004; B, Viñas *et al.*, 2010; C, Viñas *et al.*, 2020.

Location	Code	Clade	N	M	h ± SD	π + SD
North-West Atlantic	NWAtl	All	13	13	1.000 ± 0.030	0.078 ± 0.041
(1995)	B	1	4	4	1.000 ± 0.177	0.037 ± 0.026
		2	9	9	1.000 ± 0.052	0.029 ± 0.017
Côte d'Ivoire	CIV	All	122	60	0.932 ± 0.017	0.032 ± 0.016
(2018-2019)	C	1	112	53	0.920 ± 0.020	0.016 ± 0.008
		2	10	7	0.911 ± 0.078	0.025 ± 0.014
Senegal	SEN	All	49	33	0.960 ± 0.018	0.044 ± 0.022
(2018)	C	1	43	28	0.949 ± 0.024	0.024 ± 0.012
		2	6	5	0.933 ± 0.122	0.022 ± 0.013
Mauritania	MRT	All	48	29	0.946 ± 0.020	0.050 ± 0.025
(2019)	C	1	40	23	0.926 ± 0.028	0.023 ± 0.012
		2	8	6	0.893 ± 0.111	0.027 ± 0.016
Morocco	MOR	All	60	31	0.924 ± 0.022	0.056 ± 0.028
(2018)	C	1	45	22	0.875 ± 0.036	0.016 ± 0.009
		2	15	9	0.905 ± 0.054	0.029 ± 0.015
Portugal	PRT	All	103	56	0.960 ± 0.010	0.072 ± 0.035
(2018-2019)	C	1	51	26	0.911 ± 0.031	0.017 ± 0.009
		2	52	30	0.928 ± 0.027	0.029 ± 0.015
Spain	ESP	All	204	87	0.948 ± 0.008	0.072 ± 0.035
(2018-2019)	C	1	94	39	0.882 ± 0.024	0.018 ± 0.009
		2	110	48	0.907 ± 0.022	0.029 ± 0.014
Palamós	PAL	All	54	40	0.978 ± 0.011	0.075 ± 0.037
(2010)	B	1	22	15	0.905 ± 0.057	0.016 ± 0.009
		2	32	25	0.980 ± 0.015	0.037 ± 0.019
Liguria	LIG	All	58	35	0.961 ± 0.013	0.072 ± 0.035
(1993-1994)	A	1	29	18	0.921 ± 0.036	0.015 ± 0.008
		2	29	17	0.919 ± 0.036	0.028 ± 0.015
Ionia	JN	All	30	23	0.968 ± 0.022	0.074 ± 0.037
(1994)	A	1	17	11	0.904 ± 0.057	0.021 ± 0.011
		2	13	12	0.987 ± 0.035	0.029 ± 0.016
Tunisia	TUN	All	49	21	0.927 ± 0.020	0.070 ± 0.035
(2018)	A	1	28	12	0.876 ± 0.048	0.016 ± 0.008
		2	21	9	0.814 ± 0.067	0.030 ± 0.016
Malta	MLT	All	27	17	0.943 ± 0.027	0.070 ± 0.035
(2021)	C	1	9	6	0.833 ± 0.127	0.017 ± 0.010
		2	18	11	0.909 ± 0.051	0.032 ± 0.017
Marmara Sea	MR	All	20	17	0.979 ± 0.025	0.067 ± 0.034
(2001)	A	1	14	11	0.956 ± 0.045	0.024 ± 0.013
		2	6	6	1.000 ± 0.096	0.029 ± 0.018
Aegean Sea	AEG	All	53	14	0.891 ± 0.021	0.052 ± 0.026
(1993)	A	1	41	8	0.827 ± 0.028	0.016 ± 0.008
		2	12	6	0.879 ± 0.060	0.024 ± 0.013
All localities	-	All	890	260	0.955 ± 0.004	0.069 ± 0.033
		1	549	146	0.908 ± 0.009	0.019 ± 0.010
		2	341	114	0.929 ± 0.009	0.030 ± 0.015

4.2. Phylogeny within S. sarda genetic variability

The phylogenetic tree (Fig. 3) obtained from the NJ method clearly divides the *S. sarda* mtDNA sequences in two clades. In Figure 3, Clade 1 was coloured with sky blue and Clade 2 with dark blue. This affirmation is further reinforced by a Bootstrap value of 99%, strongly suggesting that the existence of these 2 clades is extremely well supported, which were already previously detected.

In the studied sample, Clade 1 has a total of 146 different haplotypes corresponding to 549 individuals, whilst Clade 2 possesses a total of 114 different haplotypes which belong to 341 bonito individuals as was shown in Table 1. The whole tree possesses a total SBL (Sum of Branch Lengths) of 1.363, thus resulting in a mean length of $5.242 \cdot 10^{-3}$ per branch taking into account all 260 individuals.

The two clades are separated by a D_A value of 8.74%, superior to D_A value of 8.1% (Viñas *et al.*, 2004). Additionally, Clade 1 mostly possesses a radial expansion shape, which suggests a common ancestor for this clade from which the rest of haplotypes have been recently diverging from, following a sudden population expansion; while Clade 2 has a branched-out shape akin to a bimodal distribution being divided in two main areas, which instead suggests a strong diverging point from which individuals have been gradually differentiating from (Viñas *et al.*, 2004). Furthermore, the phylogeny tree shows that Clade 1 haplotypes are more related to each other than Clade 2 haplotypes, since their phylogenetic branches are closer together, while Clade 2's branches are more spaced out.

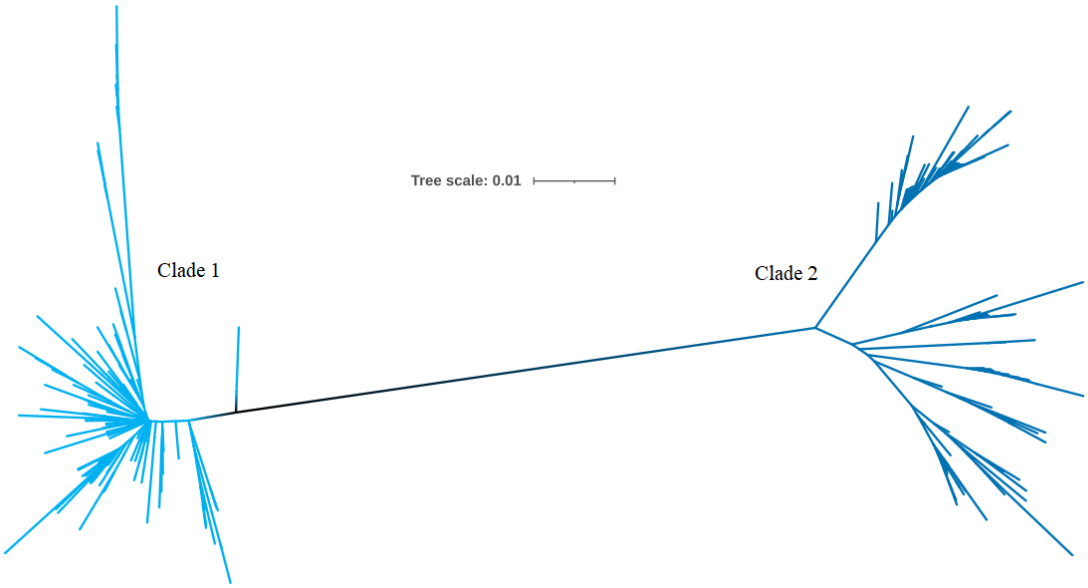

Figure 3. Unrooted phylogenetic tree for all 890 *S. sarda* samples, obtained through the NJ (Neighbor-Joining method) and using the Tamura-3-parameter model with a Gamma distribution of G=0.35.

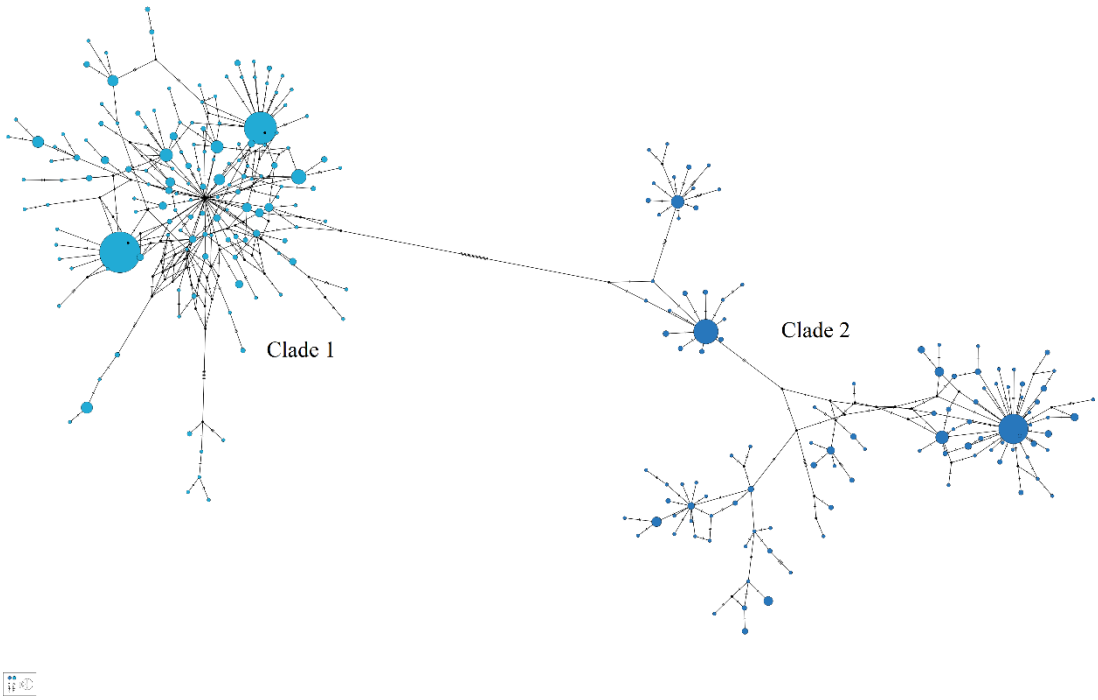

Figure 4. Reticulate phylogenetic tree for all 890 *S. sarda* samples, obtained through PopART with an Epsilon factor of 0 which puts emphasis on haplotype frequency.

A reticulate phylogenetic tree made in PopART (Fig. 4) also shows these two distinct clades, Clade 1 being displayed in sky blue and Clade 2 in a darker shade. However, this reticulate tree displays in a clearer way the different haplotypes and the relationship between each other, alongside their abundancy.

As was observed in Figure 3, haplotypes within Clade 1 are genetically closer to each other than the ones from Clade 2 This is supported by a lower nucleotide diversity (π = 0.019 for Clade 1 and π = 0.030 for Clade 2). Clade 1 not only has a bigger pool of individuals (549 individuals) compared to Clade 2 (341 individuals) but there's two haplotypes that are vastly more frequent than others among specimens of *S. sarda* (hence the bigger circle size), which are also genetically similar due to proximity in the tree; meanwhile, Clade 2 also has two major haplotypes which are further away, thus, showing that they are more diverse genetically than Clade 1's major haplotypes.

These major haplotypes also work as important nodes in the reticulate tree, since a lot of other haplotypes derive from them and are also only related to them genetically.

4.3. Mismatch Distribution Analysis

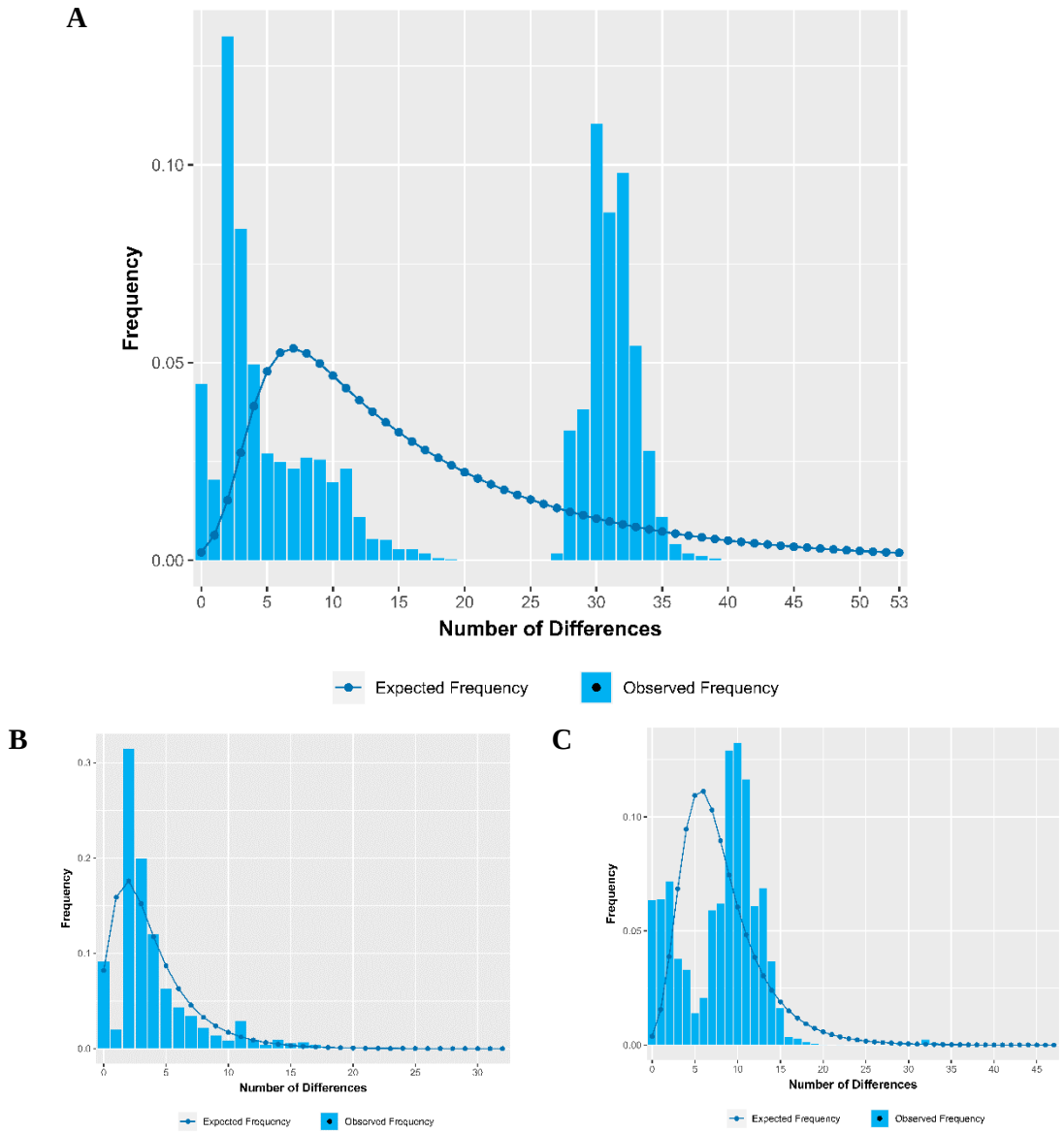

Figure 5. Mismatch graphs for all samples (A), Clade 1 (B) and Clade 2 (C) where the frequency of individuals displaying a certain number or differences in their sequence is displayed, whether it be observed or expected frequency.

The mismatch distribution graph for all samples (Fig. 5A) shows a clear difference between the expected frequency of haplotypes under the sudden expansion model and the observed one, following a bimodal pattern, as expected of two differentiated clades being studied at once. A Tajima's D value of -0.896 but with a p-value>0.1 points that when taking all samples in consideration, there doesn't seem to be a population expansion of any kind. However, a Fu's F statistic (F = -226.548) with negative value and significancy (P < 0.02) suggests that despite Tajima's D value, there has been a recent expansion due to the abundance of multiple rare haplotypes or low-frequency polymorphisms, which is exemplified in the number of individuals with low frequencies in the mismatch graph.

However, Fu's F points at a sudden expansion while Tajima's D, while also pointing at the same conclusion, is not significative. This could be explained by Tajima's D being less sensitive to recent expansions than Fu's F is, as demonstrated in the works of Fu (1997) and Ramos-Onsins & Rozas (2002). Therefore, this would mean that the non-significant Tajima's D suggests that this expansion is recent enough that it hasn't significantly altered the allele frequency spectrum detectable by Tajima's D.

Aside from that, Figure 5A shows two main spikes in observed frequency, one in low difference values and the other in high ones. Therefore, haplotypes and individuals, are either very similar to each other or very different, which correlates to two different clades being studied.

When it comes to Figures 5B and 5C, their observed frequencies match the expected frequency tendency line much more than in Figure 5A's case. This is reinforced by their relative statistics, Clade 1 has a Tajima's D value of $D = -2.289$ with a p-value of $P < 0.01$ and a Fu's F statistic of $F = -32.419$, its p-value being $P < 0.02$, whereas Clade 2 has a Tajima's D value of $D = -1.807$ with an associated p-value of $P < 0.05$ and a Fu's F statistic of $F = -101.670$ with an associated p-value of $P < 0.02$.

All of the aforementioned statistics being significative and pointing towards a sudden expansion in diversity for both Clades 1 and 2 support this idea. However, when observing the graphs, Clade 1 seems to follow the expected frequency line much closer than Clade 2, which is supported by phylogenetic tree observations made in Figures 3 and 4, about Clade 1 having experienced a much more notable sudden expansion, whereas Clade 2's might have been more distant in the past, given that until now, it has proven to be more diverse than Clade 1, since frequencies with an intermediate number of differences are far higher in Figure 5C, suggesting that despite having had a sudden expansion, this clade has been starting to slowly differentiate again.

4.4. Population structure

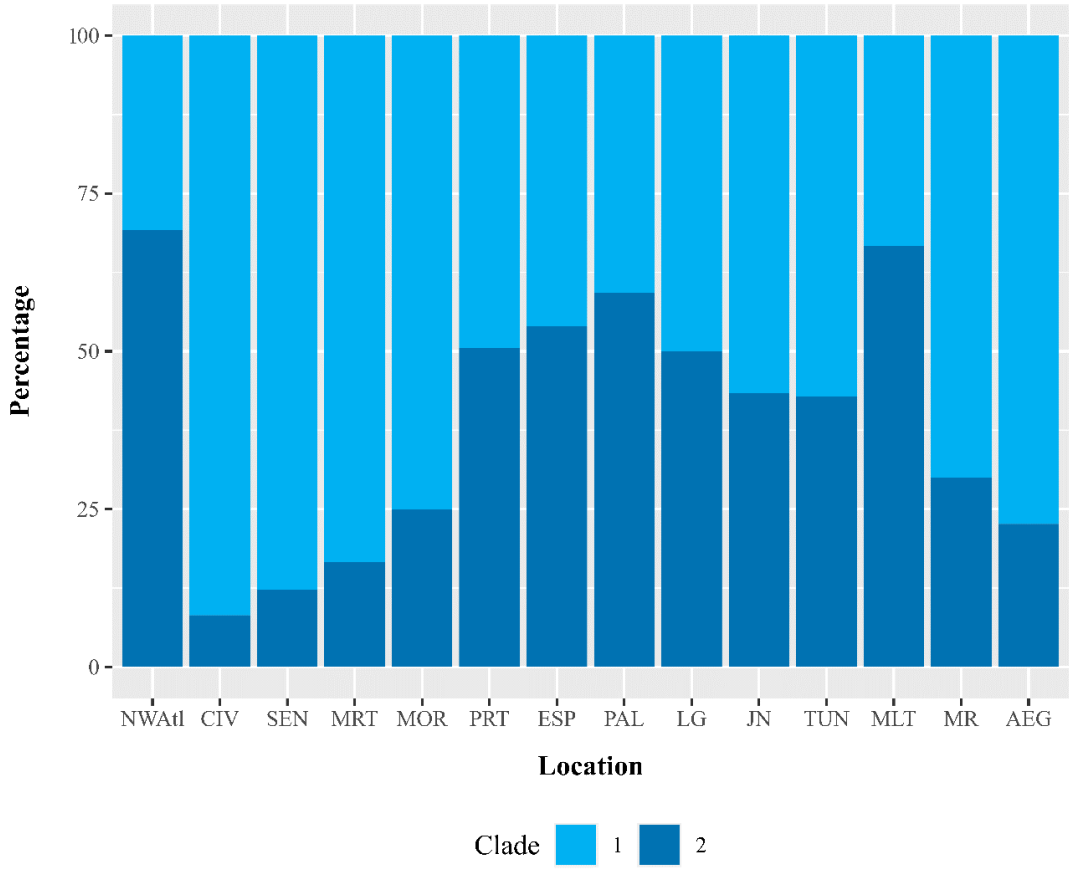

Figure 6. Histogram of all locations studied alongside the percentage in which samples obtained from there belonged to Clade 1 or 2.

Three main zones were analysed, as can be seen in Figure 2, North-American coast, West-African coast and the Mediterranean.

Further analyses showed that this division of clades was unevenly distributed among studied locations, as is shown in Figure 6. To test this hypothesis, a Pearson's Chi-squared test (Pearson, 1900) was conducted. A P-value lower than $2.2 \cdot 10^{-16}$ was obtained, thus rejecting the null hypothesis and confirming that one clade is more abundant than the other depending on the location.

As observed in Figure 6, the histogram takes the shape of a normal distribution with a negative skew, signalling that haplotype clades follow a certain pattern, in which Clade 2 individuals become increasingly abundant until PAL, point in which they start decreasing in number. Exceptions to this are NWAtl and MLT, which possess a high abundance of Clade 2 haplotypes, oddities which could be attributed to the low sample size of these two study sites.

African locations alongside the Marmara and Aegean Seas show a predominance of Clade 1 haplotypes, whilst the rest of locations display a more balanced percentage between the two clades, often oscillating between the 50/50, except for two locations, these being Malta and North-West Atlantic, which have a high abundancy of Clade 2 individuals compared to Clade 1 haplotypes. This distribution is also seen in previous studies (Viñas *et al.*, 2004; Viñas *et al.*, 2020; Ollé-Vilanova *et al.*, 2024).

Despite these disparities potentially being caused by a low sample number of certain locations, it can be inferred that the presence of specific clades is location-dependent.

FST indexes were used to determine genetic differentiation between populations, in these heatmaps (Figure 7), the higher the FST value (and thus, the darker colour) signifies a higher genetic differentiation and therefore, a bigger dissimilarity between populations.

In Figure 7A, all haplotypes were considered. Two main groups exhibiting no genetical difference within them, CIV-SEN-MRT-MOR and PRT-ESP-PAL-LG-JN-TUN-MLT. These two groups appear to be different from each other, NWAtl region and AEG+MR. Current population structure points at an African group and a Mediterranean group, with MR+AEG and NWAtl also separated.

This aggrupation is similar to the work of Ollé-Vilanova *et al.* (2024), which also distinguishes the Mediterranean and African zones for *S. sarda* and the study of Turan (2015), in which the zone of the Aegean Sea appears to be different from all other studied locations. However, that also includes the Marmara Sea, which is not significatively different from AEG (Table 2) in this study, which will be discussed in the Mediterranean population structure section.

The main notable difference figure 7A has towards the other heatmaps (Fig. 7B and Fig. 7C) is that NWAtl doesn't have such a high degree of differentiation towards other locations compared to the rest of the heatmap in figure 7A, however, higher FST values can be seen in this heatmap, due to two distinct clades being studied at once.

Just as clades are unevenly spread out through locations, so are some of their haplotypes, Figures 3 and 4 show that Clade 2 has a higher rate of genetic diversity, having more differentiated haplotypes from each other than Clade 1, however, Figures 7B and 7C show the opposite, with Clade 1 having higher FST values than Clade 2 due to Clade 1's haplotypes varying more from one location to another, whereas Clade 2 haplotypes are more evenly distributed among locations despite being genetically more differentiated.

This also gives insight on the composition of Clade haplotypes of different locations, for example, when taking Malta VS Côte d'Ivoire, a high differentiation is appreciated in Figure 7A despite there not being any notable genetic differentiation in figures 7B and 7C, this means that Clade 1 haplotypes of Malta are extremely similar to Clade 1 haplotypes of Côte d'Ivoire, and the same situation happens with Clade 2 haplotypes, but when both clades are grouped together, high dissimilarities arise.

In this study, PAL and ESP are the same location but apart in time, PAL being Palamós in 2010, and ESP samples, which refer to Spain, were also taken in Palamós, but in 2018-2019 (Table 1).

No significative differences between these two temporal sampling points can be observed by their FST value. This proves that these locations have not varied through time. This also occurs in the work of Turan (2015) where the same location was compared with itself one year apart in time (2011 and 2012) and no genetic variation was found among samples (compared locations included the Black Sea, the Marmara Sea, the Aegean Sea, the Adriatic Sea and the North-Eastern Mediterranean Sea). In this study, samples were taken 8 to 9 years apart. Overfishing and climate change are two events that have a great impact through time in fish populations (Worm *et al.*, 2009; Sumaila and Tai, 2020), and in this particular location (Palamós) these events seem to have had no effect in the Atlantic bonito population, results point at a stabilization of the specimens present in the location.

S. sarda individuals can show a great degree of variation depending on the location and its distance to other locations, and one type of haplotype from a certain clade may predominate in a certain site, but another haplotype may predominate in another site, which are signs of genetic drift. This becomes exemplified with NWAtl being highly different from other locations, and thus, it becomes possible that genetic distances, or fixation indexes in this case, could be directly related to how far away these locations are from each other.

In the AMOVA analysis (Table 3), different groups were constructed in order to test which best explains the source of variation. In all samples, groups formed were NWAtl, Africa and the Mediterranean. This aggrupation yielded the best results. AMOVA results point at this location structure with three distinct groups. However, the aggrupation of AEG with other locations such as JN or LG directly contradicts other studies such as Turan (2015) or Pujolar *et al.* (2001), which will be discussed in the Mediterranean section.

Different sources of variation are accounted for in AMOVA, first is Variation Among Groups, for example comparing values between Africa and the Mediterranean groups as a whole; Among Populations Within Groups compares the populations in a certain group with each other, for example, comparing CIV to SEN, which both belong to the Africa group; and finally, Within

Populations, which compare all locations to each other, for example, TUN to MRT or NWAtl to CIV.

Results show that in all samples, 81.26% of variation is from populations, 15.97% among groups and 2.77% from populations within groups. This means that most of molecular variance in the studied samples comes from different populations between them (81.26%) and that only a small portion of this variation is due to differences between groups, and even smaller between locations from the same group. Therefore, all locations are pretty genetically different between them.

However, these groups are also significantly different from each other, and really similar among them due to their percentage of variation. For example, when examining a random individual from AEG and comparing it to another individual from another population, selected at random, there is an 81% chance that these two individuals will be different. Despite that, if one were to take a random individual from the Africa group and compared it to any individual from the other two groups, there would only be a 16% chance that the two individuals would be different, which indicates that despite having some core differences, groups are relatively similar to each other and most of the variation comes from certain populations.

Finally, if one were to take a random individual from CIV and compare it to a random individual from any of the other African locations, there is only roughly a 3% chance that they would be different, meaning that locations within a given group are extremely similar to each other.

A similar phenomenon occurs with Clades 1 and 2, however, the amount of variation is much lower due to their sum of squares (both having around 2000 total versus almost 12000 from all samples) and their percentage of variation is also distributed differently. In the cases of Clades 1 and 2, a higher amount of differentiation is associated with differences within populations, being the percentages 88.84 and 92.73, respectively. This causes the percentage of variation due to differences among groups to lower significantly compared to all samples.

When studying isolated clades, they are much more similar between them. Therefore, there is a much smaller variation and this is largely caused by isolated populations being very different from other isolated populations. Clade 1 is also the only one with different groups, being these Africa (and Spain), America and the Mediterranean, but this was divided in western and eastern, and even so, the differentiation caused by groups is higher than the one correspondent to Clade 2, which points at the fact that Clade 1's groups are more diverse than Clade 2's.

Results point at three main areas that should be managed independently as fishing stocks: the North-West Atlantic, the African coast and the Mediterranean.

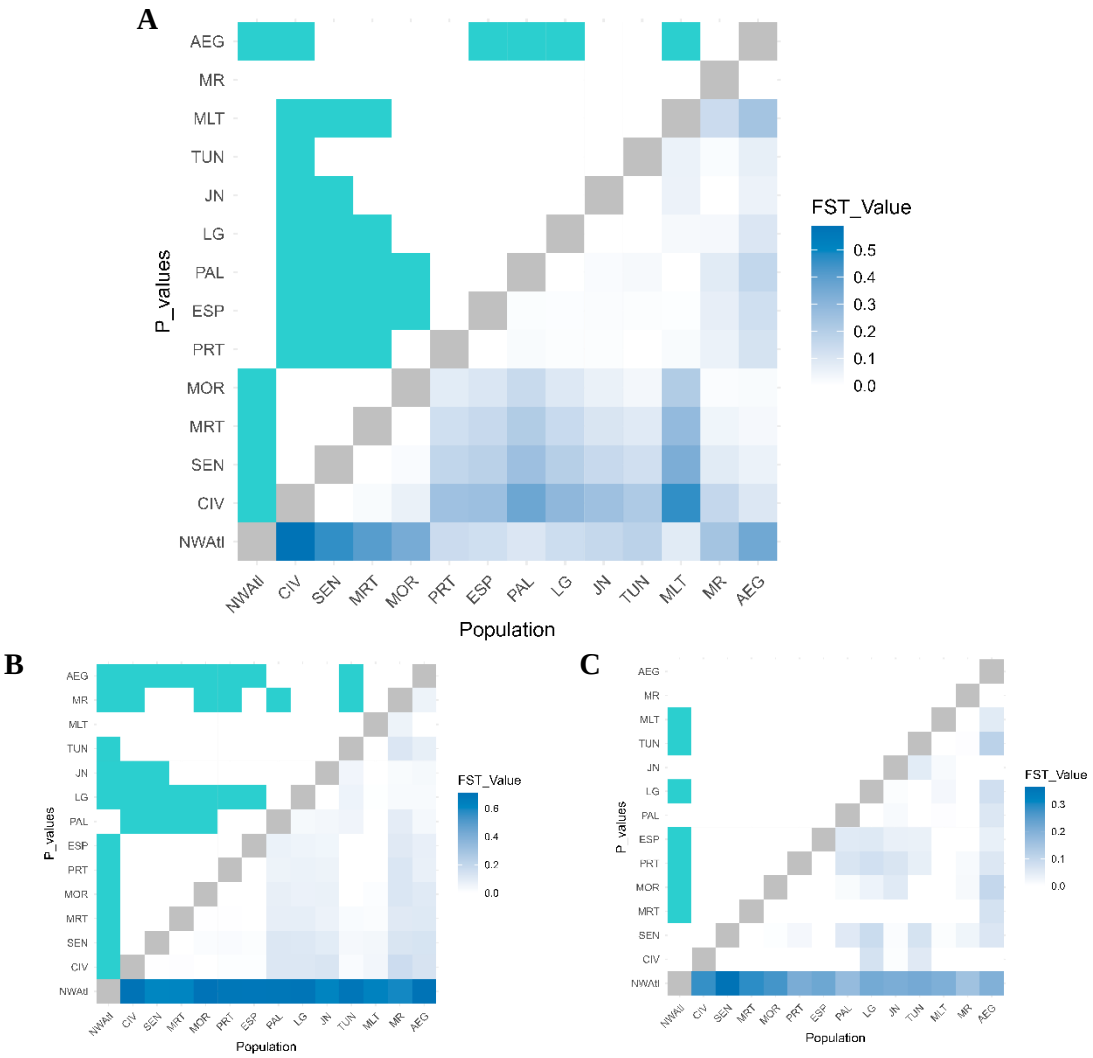

Figure 7. FST Heatmaps of all locations for Clade 1 (B), Clade 2 (C) and both clades when studied as one (A). In the bottom-right corner, FST values are displayed, whilst in top-left corner, significative p-values after Bonferroni correction method are displayed in turquoise.

Table 2. FST index (below diagonal) for each population and their correspondent p-value (above diagonal) considering all data. P-values in bold if significative after Bonferroni correction (P-value < $5.5 \cdot 10^{-4}$).

	NWAtl	**CIV**	**SEN**	**MRT**	**MOR**	**PRT**	**ESP**	**PAL**	**LG**	**JN**	**TUN**	**MLT**	**MR**	**AEG**
NWAtl	--	**0.000**	**0.000**	**0.000**	**0.000**	0.001	0.006	0.019	0.015	0.018	0.004	0.051	0.007	**0.000**
CIV	0.587	--	0.248	0.063	0.004	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.002	**0.000**
SEN	0.461	0.004	--	0.686	0.115	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.002	**0.000**	0.019	0.016
MRT	0.407	0.020	0.000	--	0.348	**0.000**	**0.000**	**0.000**	**0.000**	0.004	0.003	**0.000**	0.091	0.051
MOR	0.346	0.057	0.017	0.000	--	0.001	**0.000**	**0.000**	0.003	0.027	0.032	0.001	0.173	0.085
PRT	0.144	0.253	0.166	0.131	0.079	--	0.769	0.079	0.146	0.206	0.336	0.136	0.057	0.001
ESP	0.131	0.262	0.187	0.153	0.102	0.000	--	0.084	0.102	0.133	0.137	0.201	0.022	**0.000**
PAL	0.101	0.368	0.261	0.211	0.149	0.020	0.013	--	0.320	0.109	0.067	0.371	0.024	**0.000**
LG	0.137	0.294	0.197	0.148	0.092	0.012	0.011	0.000	--	0.554	0.329	0.102	0.132	**0.000**
JN	0.158	0.258	0.154	0.105	0.058	0.012	0.016	0.019	0.000	--	0.424	0.054	0.313	0.039
TUN	0.183	0.216	0.128	0.086	0.036	0.001	0.012	0.027	0.001	0.000	--	0.046	0.188	0.015
MLT	0.082	0.462	0.337	0.283	0.210	0.019	0.007	0.000	0.027	0.054	0.055	--	0.016	**0.000**
MR	0.242	0.161	0.082	0.043	0.014	0.055	0.068	0.081	0.029	0.000	0.017	0.144	--	0.355
AEG	0.359	0.096	0.055	0.030	0.019	0.114	0.131	0.163	0.100	0.054	0.068	0.241	0.000	--

Table 3. AMOVA analysis testing for molecular variance for all locations, Clade 1 and Clade 2. SV: Source of Variation, d.f: degrees of freedom, SS: Sum of Squares, VC: Variance Components and Percentage of variation with associated P-value (in bold if significant)

Clade	Groups	Code	SV	d.f.	SS	VC	%	P-value
All	**America** **Africa** **Mediterranean**	NWAtl CIV+SEN+MRT+MOR ESP+PAL+LG+JN+TUN+MLT+MR+AEG+PRT	Among Groups	2	1015.709	2.367	15.97	**0.000**
			Among Populations within Groups	11	413.668	0.411	2.77	**0.000**
			Within Populations	876	10549.736	12.043	81.26	**0.000**
			Total	889	11979.113	14.821	100	-
1	**America** **Western Mediterranean** **Africa and Spain** **Eastern Mediterranean**	NWAtl PAL+LG+JN+TUN+MLT+PRT CIV+SEN+MRT+MOR+ESP MR+AEG	Among Groups	3	127.395	0.376	9.57	**0.000**
			Among Populations within Groups	10	59.875	0.063	1.60	**0.000**
			Within Populations	535	1868.459	3.492	88.84	**0.000**
			Total	548	2055.729	3.931	100	-
2	**America** **Africa** **Mediterranean**	NWAtl CIV+SEN+MRT+MOR ESP+PAL+LG+JN+TUN+MLT+MR+AEG+PRT	Among Groups	2	41.395	0.316	5.18	**0.008**
			Among Populations within Groups	11	95.207	0.127	2.09	**0.010**
			Within Populations	327	1845.827	5.645	92.73	**0.000**
			Total	340	1982.428	6.087	100	-

From the pairwise FST (Table 4), no genetic differentiation at all can be observed between Portugal (Olhão) and Spain, and no significative differences with the rest of the Mediterranean populations either except for the Aegean Sea (thus Portugal will be grouped with the rest of the Mediterranean). These results challenge the controversy about whether the Strait of Gibraltar constitutes a phylogeographical barrier for marine species (Patarnello *et al.,* 2007) since there doesn't seem to be significative differences from one side of the strait to the other. A Bonferroni correction by multiple testing was also performed on the Mediterranean FST values, this operation yielded a signification value of P-value < 0.0014. Figure 8 reveals a high degree of differentiation from AEG towards other locations, being significantly different from PRT, ESP, PAL, LG and MLT, while most similar to MR with a FST index of 0.000 between the two locations (Table 4), making them genetically identical.

More specific AMOVA tests were made only for the Mediterranean zone to check for genetic differentiation levels between the Aegean Sea, Marmara Sea and the rest of the Mediterranean. Atlantic locations were not included since their high differentiation value towards the Mediterranean could blur the results. Thus, NWAtl, CIV, SEN, MRT and MOR were left out.

AMOVA analyses were conducted to find the population structure of the Mediterranean (Table 5). Through FST Heatmap, 2 groups were made, AEG+MR and the rest of the Mediterranean, since these two locations are identical but different from the rest. However, this aggrupation is not significative. Population structure of the Mediterranean ended up divided in two main areas, the western Mediterranean (PRT+ESP+PAL+LG) and the eastern Mediterranean (TUN+JN+MR+AEG), Malta, however, shows affinity for both groups and can be grouped with whichever while the aggrupation remains significative as shown in Table 5. These results are contradictory with similar works (Turan, 2015; Pujolar *et al.,* 2001) which find AEG to be extremely different to other Mediterranean locations, including the Marmara Sea, which was found to be identical in this study.

This result could be due to a small sample size obtained in the Marmara Sea (n = 20) which may not have shown accurate population structure of *S. sarda* population in that zone. Another possible explanation, although highly hypothetical, would be that overfishing in the Aegean Sea (Öztürk, 2009) in the past caused a bottleneck event in the Aegean bonito population, causing these specimens to have a highly different genetic pool from the rest of the bonito in the Mediterranean.

The Mediterranean structure of *S. sarda* remains unascertained in this study and would need further investigation, increasing and intensifying sampling from the Eastern region of the Mediterranean to better determine a solid population structure. However, current results point at three areas that could be managed independently as fishing stocks: the Western Mediterranean with Portugal included, Malta and the Eastern Mediterranean.

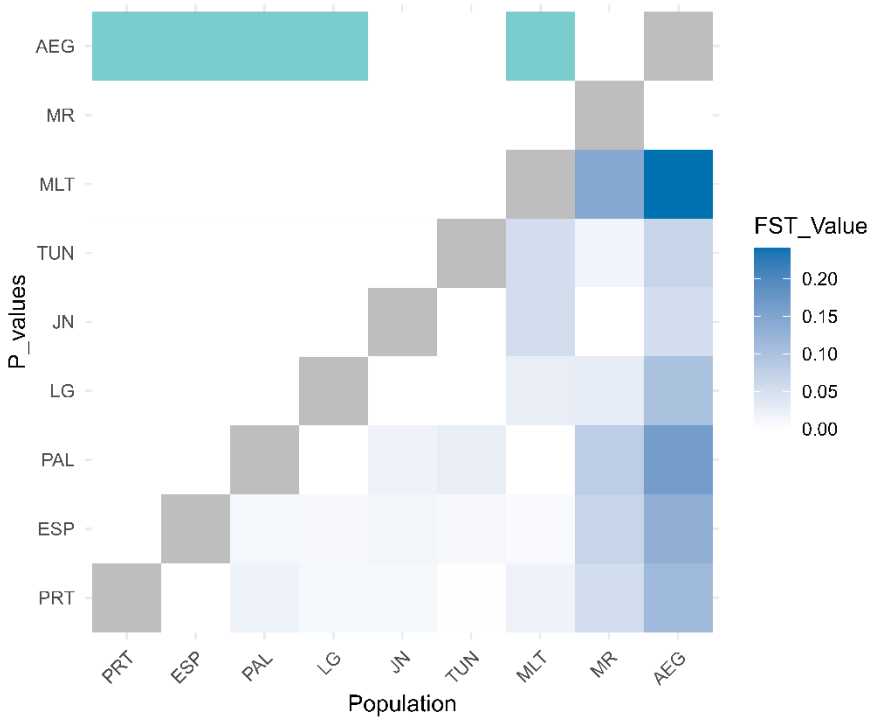

Figure 8. FST Heatmap of Mediterranean locations. In the bottom-right corner, FST values are displayed, whilst in top-left corner, significative p-values after Bonferroni correction method are displayed in turquoise.

Table 4. FST index (below diagonal) for each population within the Mediterranean and their correspondent p-value (above diagonal). P-values in bold if significative after Bonferroni correction (P-value < 0.0014).

	PRT	**ESP**	**PAL**	**LG**	**JN**	**TUN**	**MLT**	**MR**	**AEG**
PRT	--	0.734	0.068	0.144	0.201	0.311	0.124	0.049	**0.000**
ESP	0.000	--	0.113	0.097	0.142	0.115	0.206	0.013	**0.000**
PAL	0.020	0.013	--	0.333	0.141	0.077	0.389	0.035	**0.001**
LG	0.011	0.011	0.000	--	0.540	0.293	0.095	0.150	**0.001**
JN	0.011	0.016	0.019	0.000	--	0.414	0.072	0.339	0.030
TUN	0.001	0.012	0.027	0.001	0.000	--	0.044	0.182	0.019
MLT	0.019	0.007	0.000	0.026	0.054	0.054	--	0.012	**0.000**
MR	0.054	0.068	0.081	0.029	0.000	0.017	0.144	--	0.408
AEG	0.113	0.130	0.162	0.100	0.053	0.067	0.241	0.000	--

Table 5. AMOVA analysis testing for molecular variance for all Mediterranean locations. SV: Source of Variation, d.f: degrees of freedom, SS: Sum of Squares, VC: Variance Components and Percentage of variation with associated P-value (in bold if significative).

Groups	Code	SV	d.f.	SS	VC	%	P-value
Aegean and Marmara **Mediterranean**	AEG+MR ESP+PAL+LG+JN+TUN+MLT+PRT	Among Groups	1	206.974	1.459	9.54	**0.000**
		Among Populations within Groups	7	147.471	0.118	0.77	0.092
		Within Populations	589	8080.991	12.719	89.69	**0.023**
		Total	597	8435.436	15.296	100	-
Western Mediterranean & Malta **Eastern Mediterranean**	PRT+ESP+PAL+LG+MLT AEG+MR+JN+TUN	Among Groups	1	183.630	0.698	4.78	**0.000**
		Among Populations within Groups	7	170.816	0.177	1.21	**0.030**
		Within Populations	589	8080.991	13.719	94.00	**0.004**
		Total	597	8435.436	14.595	100	-
Western Mediterranean **East Mediterranean & Malta**	PRT+ESP+PAL+LG TUN+JN+MR+AEG+MLT	Among Groups	1	107.578	0.273	1.90	**0.000**
		Among Populations within Groups	7	246.867	0.361	2.51	**0.002**
		Within Populations	589	8080.991	13.719	95.58	**0.049**
		Total	597	8435.436	14.354	100	**-**

Figure 9A shows isolation by distance for both clades at once, and, unlike figures 9B and 9C, a gradual increase in genetic distance alongside geographic distance can be appreciated. The slope and shape of the diagram indicate that although faint, there is a correlation between geographic and genetic distance when examining all samples.

Dispersion diagrams show that at low geographic distances, a low genetic distance is to be expected. Furthermore, at the highest geographic distance values, the FST distances spike up by several orders of magnitude, and all these points belong to the NWAtl locations, which implies that the haplotypes of this location are vastly different to other haplotypes, even in the same clade. However, it doesn't seem like individuals within the same clade exhibit much change even in long distances, except for NWAtl, where the differences spike up.

Therefore, it seems that distance between locations does not matter that much when studying the same clade of haplotypes. However, when studying the population of *S. sarda* as a whole, distance does seem to play a factor in how genetically different haplotypes are.

This affirmation is reinforced by different Mantel tests that displayed a significant correlation between linearized FSTs (or genetic distance) and geographic distance for all samples (Fig. 9A) and Clade 1 (Fig. 9B) (both with a p-value = 0.000), the Mantel test realized for Clade 2 (Fig. 9C), however, did not display a significant correlation (p-value = 0.055).

Thus, both graphical analysis and Mantel tests reject the null hypothesis of no correlation between genetic and geographic distances.

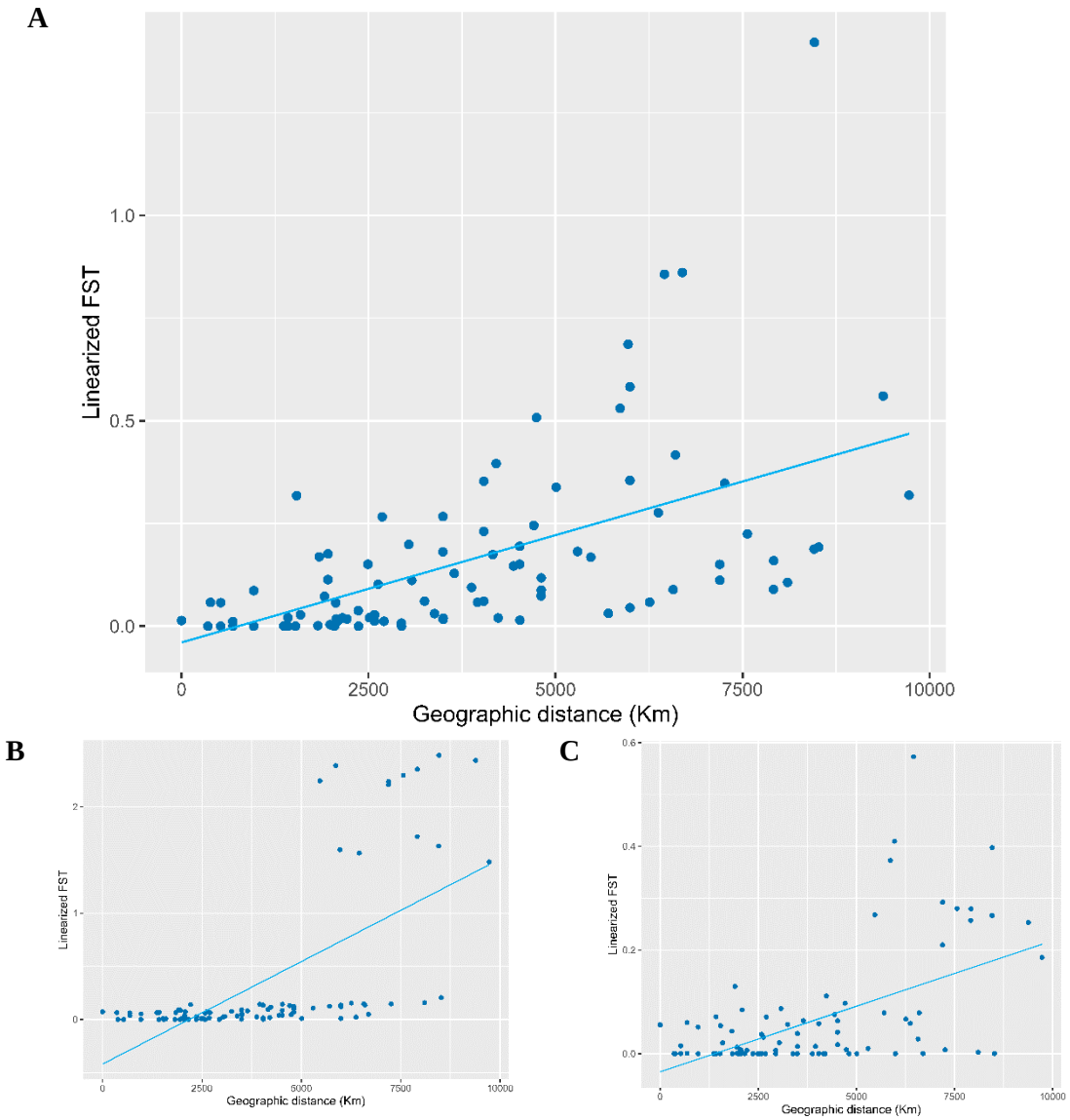

Figure 9. Dispersion diagrams comparing genetic distances (Linearized FST values) and geographic distances (in Km) of Clade 1 (B), Clade 2 (C) and All Samples (A). Regression equation and coefficient are: A -> $y = 5 \cdot 10^{-5}x - 0.0397$ $r^2 = 0.2931$ B -> $y = 0.0002x - 0.4196$ $r^2 = 0.4017$ C -> $y = 3 \cdot 10^{-5}x - 0.0351$ $r^2 = 0.2832$

5. Conclusions

1 - Close to 1000 specimens of *S. sarda* were analysed to construct a solid database made from individuals studied with mitochondrial Control Region DNA, being to our knowledge the most extensive study of this species to date. The database was elaborated successfully and different tests were made to determine variability, phylogeny, and population structure of bonito groups across the Mediterranean and Atlantic.

2 - Atlantic bonito population was divided in two highly differentiated phylogenetic clades, corroborated by previous studies that found the same division. Histogram and population studies show an uneven distribution of clades among populations, Clade 1 being more abundant in the African coast and the East-Mediterranean, and Clade 2 being more frequent in the rest of the Mediterranean and NWAtl.

3 – AMOVA and FST results highlight a clear population structure for *S. sarda* consisting of North-West Atlantic, Africa and the Mediterranean. However, further testing for the Mediterranean region reveals discrepancies towards other articles of the same topic. More investigation would be required and this part of the study remains unascertained. FST indexes reveal no significant differentiation due to chronological reasons in one population after almost a decade apart in sampling. No significant differences were detected between one side of the strait of Gibraltar and the other (PRT and ESP), challenging its role as a phylogeographical barrier. Isolation by distance shows proportional correlation between geographic distance and genetic distances, evidences of genetic drift.

4 - Molecular diversity indexes reveal high haplotype diversity for the studied *S. sarda* population and extremely high nucleotide diversity compared to other Scombridae species. North-West Atlantic region was also found to be hypervariable. However, haplotype diversity decreased compared to previous *S. sarda* studies and nucleotide diversity increased.

5 – Independent management measures were proposed for three different fishing stocks according to population structure; this work has important connotations for fisheries as depicted in the work of Hauser & Carvalho (2008) which could improve their sustainability when taking into account population structure and genetic diversity in order to discern different populations that require separate management strategies or to maintain a high genetic diversity to make populations more resilient to natural selection processes.

This study could help us understand ecological dynamics of bonito to both further our comprehension of it and other similar species, furthermore allowing for better regulations in fisheries for a more efficient management.

References

Barahona, S., Oré-Chávez, D. & Quiroz, R. (2017). High genetic connectivity and Population Expansion of *Scomber japonicus* in the Northern Humboldt Current System revealed by mitochondrial control region sequences. *Revista Peruana de Biología*, 24(2), 163-174. DOI: http://dx.DOI.org/10.15381/rpb.v24i2.12807

Becker, R., Wilks, A., Brownrigg, R., Minka, T. & Deckmyn, A. (2023). maps: Draw Geographical Maps. *CRAN*. DOI: 10.32614/CRAN.package.maps

Bennett, R., Ellender, B., Mäkinen, T., Miya, T., Pattrick, P., Wasserman, R., Woodford, D. & Weyl, O. (2016). Ethical Considerations for field research on fishes. *Koedoe*, 58(1), 1-15. DOI: http://dx.doi.org/10.4102/koedoe.v58i1.1353

Bianchi, G., Carpenter, K., Roux, J., Molloy, F., Boyer, D. & Boyer, H. (1999). Field guide to the living marine resources of Namibia. *FAO Species Identification Guide for Fishery Purposes*, 265, 11. DOI: https://www.cabidigitallibrary.org/DOI/full/10.5555/20013004895

Bimbo, A. (2013). Food Enrichment with Omega-3 Fatty Acids. *Woodhead Publishing Series in Food Science, Technology and Nutrition*, 27-107. DOI: https://DOI.org/10.1533/9780857098863.1.27.

Bloch, M. E. (1793). *Naturgeschichte der ausländischen Fische* (Vol. 7, 123-125). Berlin: Bei dem Verfasser

Boekhout, H., Weijden, I. & Waltman, L. (2021). Gender differences in scientific careers: A large-scale bibliometric analysis. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2106.12624

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Savers, E. W. (2016). GenBank. *Nucleic Acids Research*, 44, D67-D72. DOI: 10.1093/nar/gkv1276

Collette, B. & Chao, L. (1975). Systematica and morphology of the bonitos (Sarda) and their relatives (Scombridae, sardini). *Fishery Bulletin*, 73, 515-625. URL: https://www.biodiversitylibrary.org/part/7964

Collette, B. & Nauen, C. (1983). FAO Species Catalogue. Vol. 2 Scombrids of the World. An annotated and illustrated catalogue of tunas, mackerels, bonitos, and related species known to date. *FAO Fisheries Synopsis*, 125(2), 137. URL: https://www.fao.org/4/ac478e/ac478e00.htm#Contents

Cowen, R., Gawarkiewicz, G., Pineda, J. & Thorrold, S. (2007). Population Connectivity in Marine Systems: An Overview. *Oceanography*, 20(3). DOI: 10.5670/oceanog.2007.26

Davies, C., Gosling, E., Was-Barcz, A. & Brophy, D. (2011). Microsatellite analysis of albacore tuna (*Thunnus alalunga*): Population genetic structure in the North-East Atlantic Ocean and Mediterranean Sea. *Marine Biology*, 158(12), 2727. DOI: 10.1007/s00227-011-1772-x

Department of Animal Husbandry and Dairying. (2021). Guidelines of CPCSEA for Experimentation on Fishes. Committee for the Purpose of Control and Supervision of Experiments on Animals. URL: https://ccsea.gov.in/WriteReadData/userfiles/file/Guidelines%20of%20CPCSEA%20for%20Experimentation%20on%20Fishes-2021.pdf

Elbardisy, H. & Abedalthagafi, M. (2021). The History and Challenges of Women in Genetics: a Focus on Non-Western Women. *Frontiers*, 12. DOI: https://doi.org/10.3389/fgene.2021.759662

Excoffier, L., Smouse, P. & Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491. DOI: https://DOI.org/10.1093/genetics/131.2.479

Excoffier, L. & Lischer, H. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. DOI: 10.1111/j.1755-0998.2010.02847.x

Fu, Y. (1997). Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics*, 147(2), 915-925. DOI: https://DOI.org/10.1093/genetics/147.2.915

Genç, Y., Basçinar, N. & Dagtekin, M. (2019). Feeding hàbits during migration of the Atlantic bonito *Sarda sarda* (Bloch, 1793) to the Black Sea. *Marine Biology Research*, 15(2), 125-136. DOI: https://DOI.org/10.1080/17451000.2019.1596287

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41. 95-98. URL: https://www.academia.edu/2034992/BioEdit_a_user_friendly_biological_sequence_alignment_editor_and_analysis_program_for_Windows_95_98_NT

Handelsman, J., Cantor, N., Carnes, M., Denton, D., Fine, E., Grosz, B., Hinshaw, V., Marrett, C., Rosser, S., Shalala, D. & Sheriddan, J. (2005). More Women in Science. *Science*, 309(5738), 1190-1191. DOI: 10.1126/science.1113252

Harrell, F. & Dupont, C. (2024). Hmisc: Harrell Miscellaneous. *CRAN*. DOI: 10.32614/CRAN.package.Hmisc

Hauser, L. & Carvalho, G. (2008). Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *FISH and FISHERIES*, 9(4), 333-362. DOI: https://DOI.org/10.1111/j.1467-2979.2008.00299.x

Leigh, J. & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods Ecol Evol*, 6(9), 1110-1116. DOI: https://DOI.org/10.1111/2041-210X.12410

Letunic, I. & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* DOI: 10.1093/nar/gkae268

Lucena-Frédou, F., Mourato, B., Fredou, T. & Lino, P. (2021). Review of the life history, fisheries and stock assessment for small tunas in the Atlantic Ocean. *Reviews in Fish Biology and Fisheries*, 31(1), 1-28. DOI: 10.1007/s11160-021-09666-8

Macías, D., Gómez-Vives, M., García, S. & Ortiz de Urbina, J. (2005). Reproductive characteristics of Atlantic bonito (*Sarda sarda*) from the south-western Spanish Mediterranean. *ICCAT Collective Volume of Scientific Papers*, 58(2), 470-483. URL: https://eurekamag.com/research/023/510/023510762.php

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209-220. URL: https://aacrjournals.org/cancerres/article/27/2_Part_1/209/476508/The-Detection-of-Disease-Clustering-and-a

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York Chicheser, West Sussex: Columbia University Press. DOI: https://DOI.org/10.7312/nei-92038

Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes. *CRAN*. DOI: 10.32614/CRAN.package.RColorBrewer

Ollé-Vilanova, J., Vilà-Valls, L., Alvarado-Bremer, J., Cerdenares, G., Duong, T., Hajjej, G., Lino, P., Muñoz-Lechuga, R., Sow, F., Diaha, N., Araguas, R., Sanz, N. & Viñas, J. (2022). Population genetics meets phylogenetics: new insights into the relationships among members of the genus *Euthynnus* (family Scombridae). *Hydrobiologia*, 849, 47-62. DOI: https://DOI.org/10.1007/s10750-021-04707-6

Ollé-Vilanova, J., Hajjej, G., Macias, D., Saber, S., Lino, P., Muñoz-Lechuga, R., Baibbat, S., Sow, F., Diaha, N., Araguas, R., Sanz, N. & Viñas, J. (2024). Atlantic bonito (*Sarda*) genòmic analysis reveals population differentiation across Northeast Atlantic and mediterranean locations: Implications for fishery management. *Marine Environmental Research*, 196, DOI: https://DOI.org/10.1016/j.marenvres.2024.106408

Öztürk, B. (2009). Marine protected areas in the high seas of the Aegean and Eastern Mediterranean Seas, some proposals. *Journal of Black Sea / Mediterranean Environment*, 15(1), 69-82. URL: https://dergipark.org.tr/en/pub/jbme/issue/9844/121875#article_cite

Patarnello, T., Volckaert, F. & Castilho, R. (2007). Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, 16(21), 4426-4444. DOI: https://DOI.org/10.1111/j.1365-294X.2007.03477.x

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157-175. DOI: https://DOI.org/10.1080/14786440009463897

Pujolar, J., Roldán, M. & Pla, C. (2001). Allozyme differentiation of bonito in the Mediterranean Sea. *Journal of Fish Biology*, 59, 169-174. DOI: https://DOI.org/10.1111/j.1095-8649.2001.tb02347.x

Ramos-Onsins, S. & Rozas, J. (2002). Statistical Properties of New Neutrality Tests Against Population Growth. *Molecular Biology and Evolution*, 19(12), 2092-2100. DOI: https://DOI.org/10.1093/oxfordjournals.molbev.a004034

R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL: https://www.R-project.org/

Reiss, H., Hoarau, G., Dickey-Collas, M. & Wolff, W. (2009). Genetic population structure of marine fish: Mismatch between biological and fisheries management units. *Fish and Fisheries*, 10(4), 361-395. DOI: 10.1111/j.1467-2979.2008.00324.x

Roberti, M., Yannopoulos, G., Metrio, G., Ludovico, A., Milella, F., Carone, P., Megalofonou, P., Bingel, F., Unluata, U., Cantatore, P. & Gadaleta, M. (1993). Polimorfismo del DNA mitochondriale in stock di palamita, *Sarda sarda*, del Mar Mediterraneo. *Biol. Mar. Suppl. Notiz. SIBM*, 1, 349-353.

Rodríguez-Ezpeleta, N., Díaz-Arce, N., Walter, J., Richardson, D., Rooker, J., Nøttestad, L., Hanke, A., Franks, J., Deguara, S., Lauretta, M., Addis, P., Varela, J., Fraile, I., Goñi, N., Abid, N., Alemany, F., Oray, I., Quattro, J., Sow, F., Itoh, T., ... (2019). Determining Natal Origin for Improved Management of Atlantic Bluefin Tuna. *Frontiers in Ecology and the Environment*, 17(8), 439-444. URL: https://aquila.usm.edu/fac_pubs/16559

Rozas, J., Ferrer-Mata, A. Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. & Sánchez-Gracia, A. (2017). DnaSP v6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Bio. Evol.*, 34, 3299-3302. DOI: 10.1093/molbev/msx248

RStudio Team. (2020). RStudio: Integrated Development for R. *RStudio, PBC*, Boston. URL: http://www.rstudio.com/

Sabatés, A. & Recasens, L. (2001). Seasonal Distribution and spawning of small tunas (*Auxis rochei* and *Sarda sarda*) in the Northwestern Mediterranean. *Scientia marina*, 65(2), 95-100. DOI: https://DOI.org/10.3989/scimar.2001.65n295

Santa Brígida, E., Cunha, D., Rego, P., Sampaio, I., Schneider, H. & Vallinoto, M. Population analysis of *Scomberomorus cavalla* (Cuvier, 1829) (Perciformes, Scombridae) from the Northern and Northeastern coast of Brazil. *Brazilian Journal of Biology*, 67(4), 919-924. DOI: https://DOI.org/10.1590/S1519-69842007000500016

SCRS. (2019). Report of the 2019 ICCAT small tuna species group intersessional meeting. *ICCAT Collective volume of scientific papers*, 76(7), 1-89. URL: https://www.iccat.int/Documents/CVSP/CV076_2019/n_7/CV76007001.pdf

Sumaila, U. & Tai, T. (2020). End Overfishing and Increase the Resilience of the Ocean to Climate Change. *Frontiers in Marine Science*, 7. DOI: https://doi.org/10.3389/fmars.2020.00523

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585-595. DOI: https://DOI.org/10.1093/genetics/123.3.585

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*, 9(4), 678-687. DOI: https://DOI.org/10.1093/oxfordjournals.molbev.a040752

Tamura, K., Stecher, G. & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*, 38, 3022-3027. URL: https://www.megasoftware.net/citations

Turan, C. (2015). Microsatellite DNA reveals genetically different populations of Atlantic bonito *Sarda sarda* in the Mediterranean Basin. *Biochemical Systematics and Ecology*, 63, 174-182. DOI: https://DOI.org/10.1016/j.bse.2015.10.007

Turan, C., Ivanova, P. & Soldo, A. (2016). Population Structuring and Migration Pathway of Atlantic bonito *Sarda sarda*. *Natural and Engineering Sciences*, 1(3), 56-65. DOI: https://DOI.org/10.28978/nesciences.286366

Viñas, J., Bremer, J. & Pla, C. (2004). Phylogeography of the Atlantic bonito (*Sarda sarda*) in the northern Mediterranean: the combined effects of historical vicariance, population expansion, secondary invasion, and isolation by distance. *Molecular Phylogenetics and Evolution*, 33, 32-42. DOI: 10.1016/j.ympev.2004.04.009

Viñas, J., Bremer, J. & Pla, C. (2010). Phylogeography and phylogeny of the epineritic cosmopolitan bonitos of the genus *Sarda* (Cuvier): inferred patterns of intra- and inter-oceanic connectivity derived from nuclear and mitochondrial DNA data. *Journal of Biogeography*, 37(3), 557-570. DOI: https://DOI.org/10.1111/j.1365-2699.2009.02225.x

Viñas, J., Ollé-Vilanova, J., Hajjej, G., Macias, D., Saber, S., Lino, P., Muñoz-Lechuga, R., Baibbat, S., Habibe, B., Ngom Sow, F., Diaha, N'G & Lucena Frédou, F. (2020). Population genetic of Atlantic bonito in the north east Atlantic and Mediterranean. *Collective Volume of Scientific Papers*, 77(9), 6-12. URL: https://digital.csic.es/bitstream/10261/326603/1/CV077090006.pdf

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20. DOI: http://www.jstatsoft.org/v21/i12/

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V, … (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. DOI: 10.21105/joss.01686

Wickham, H. & Bryan, J. (2023). readxl: Read Excel Files. *CRAN*. URL: https://CRAN.R-project.org/package=readxl

Worm, B., Hilborn, R., Baum, J., Branch, T., Collie, J., Costello, C., Fogarty, M., Fulton, E., Hutchings, J., Jennings, S., Jensen, O., Lotze, Mace, P., Mcclanahan, T., Minto, C., Palumbi, S., Parma, A., Ricard, D., Rosenberg, A., Watson, R… (2009). Rebuilding Global Fisheries. *Science*, 325, 578-585. DOI: 10.1126/science.1173146

Wright, S. (1949). The Genetical Structure of Populations. *Annals of Eugenics*, 15(4), 323-354. DOI: https://DOI.org/10.1111/j.1469-1809.1949.tb02451.x

Zhu, Y., Cheng, Q. & Rogers, S. (2014). Genetic structure of *Scomber japonicus* (Perciformes: Scombridae) along the coast of China revealed by complete mitochondrial cytochrome *b* sequences. *Mitochondrial DNA Part A*, 27(6), 3828-3836. DOI: https://DOI.org/10.3109/19401736.2014.958671