Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BENI CULTURALI E AMBIENTALI

Ciclo 35

Settore Concorsuale: 05/B1 - ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/05 - ZOOLOGIA

MOVING TOWARD STOCK UNITS IDENTIFICATION BASED ON SPATIAL POPULATION STRUCTURE OF MARINE SPECIES IN THE MEDITERRANEAN SEA AND ADJACENT WATERS THROUGH MULTIDISCIPLINARY AND HOLISTIC APPROACHES FOR THE SUSTAINABILITY OF FISHERIES RESOURCES

Presentata da: Rachele Corti

Coordinatore Dottorato

Donatella Restani

Supervisore

Alessia Cariani

Co-supervisore

Fausto Tinti

Esame finale anno 2023

Table of Contents

Abstract

Investigating stock identity of marine species in a multidisciplinary holistic approach can reveal patterns of complex spatial population structure and signatures of potential local adaptation. The combination of complex life-history traits of marine species with environmental and spatial features generates population structure. Indeed, the integration of molecular and environmental data is crucial to understand selection and demographic processes shaping population structure of marine organisms for management and conservation purposes.

In the first case study, the population structure of common sole (*Solea solea*) in the Mediterranean Sea was delineated using genomic and otolith data, including hundreds of single nucleotide polymorphisms (SNPs) markers, otolith shape and otolith trace element composition data. SNPs were correlated with environmental and spatial variables to evaluate the impact of the selected features on the actual genetic population structure. Specifically, a seascape genetics approach with redundancy (RDA) and genetic-environmental association (GEA) analysis was used to find loci potentially involved in local adaptation. Furthermore, putative functional annotation was investigated to detect genes associated with the described patterns of neutral and adaptive genetic variation. Finally, an integrated holistic approach was applied to combine the tracers with different spatio-temporal scales.

In the second case study, SNPs data was also used to illustrate the population structure of another marine fish targeted by commercial fisheries in Mediterranean Sea, namely European hake (*Merluccius merluccius*). The aim was to identify patterns of neutral and potential adaptive genetic variation by applying seascape genomic framework considering spatial and environmental features. RDA was combined with genetic differentiation to characterize spatial structure and identify factors potentially involved in local adaptation. The area target considered for European hake included distinct sites within the Alboran Sea, extending into the neighboring Mediterranean Sea and Atlantic Ocean.

Results from both genetic and otolith data suggested significant divergence among putative populations of common sole, confirming a clear separation between Western and Eastern Mediterranean Sea, as well as a distinct genetic cluster corresponding to the Adriatic Sea. Evidence of fine scale population structure in the Western Mediterranean Sea was observed at outlier loci level, as well as further differentiation within the Adriatic.

Longitude and salinity variation accounted for the majority of both wide and fine spatial structure. The GEA detected significant associated outlier loci potentially involved in local adaptation, but spatial distribution explained more variance than environmental features in the RDA. Our study not only indicates that separation among Mediterranean sole population is led primarily by neutral processes because of low connectivity due to spatial segregation and limited dispersal, but it also suggests the presence of local adaptation.

The holistic approach by considering the spatio-temporal scales of variation confirmed that the same pattern of separation between these geographical sites is currently occurring and has occurred for many generations. These results should be taken into account to support the application of spatial stock assessment models, including a review and possible redefinition of fishery management units inside dynamic spatial management strategies.

From all analysis performed, a significant genetic differentiation across locations was also found for European hake. Results consistently showed the occurrence of population structure in *Merluccius merluccius* by detecting westward–eastward differentiation among populations, higher levels of structure and potential distinct subgroups at a fine geographical scale using putatively adaptive SNPs. Seascape analysis on European hake revealed that environmental factors better explained the genetic variation at both neutral and putatively adaptive SNPs loci, suggesting their key contribution in shaping the observed genetic structure.

These results enhance the knowledge of the population structure of commercially relevant species, which has a fundamental role in the management of stocks, together with the understanding of which processes and factors are playing a role in the definition of their population structure. The ultimate scope would be to apply this approach in the regular assessments of commercially relevant marine resources to inform the management advice.

Research objectives

In the context of increasing degradation of marine resources, the improvement of understanding of ecological and biological processes of marine resources is crucial to achieve sustainability under fisheries management. Advancing stock identification methods have exposed differences between the spatial structure of biological populations and currently defined stock units. This spatial mismatch between biological population and management unit can affect both the conservation of resources and the subsistence of fisheries communities. Stock identification methods can be valuable in understanding population dynamics. This improvement can make the stock assessment process more robust and thereby enable developing fisheries management strategies to obtain sustainability.

In this perspective, my PhD research project aims to identify stock units based on biological units identification to better understand population structure and stock status of marine fish resources by applying multidisciplinary approaches within a holistic framework. The use of proper management units based on biological evidence under a fisheries sustainable management approach could increase the environmental, economic, and social sustainability of fishing activity. In order to achieve the purpose of my project enhancing the availability of information for each fish stock, I carried out the analysis based on advancing stock identification methods for a better comprehension of population structure of two marine resources in the Mediterranean Sea in a multidisciplinary framework. I focused on common sole (*Solea solea*) and European hake (*Merluccius merluccius*), demersal species of commercial interest in the State of Mediterranean Fisheries. Data used in this project were produced in the framework of the FishPopTrace EU Project and TRANSBORAN Project for *Solea solea* and *Merluccius merluccius* case studies, respectively. For common sole we analysed putative populations inside the Mediterranean Sea while for European hake we targeted distinct sites within the Alboran Sea, extending into the neighboring Mediterranean Sea and Atlantic Ocean.

With the aim of understanding the challenges in the assessment and management of marine fisheries in delineating spatial structure of stock units, available information, advancing methods, and gaps about populations structure and stock units delimitation was summarized in **Chapter 1**. In this chapter, current issues of mismatches between biological population and the management units were discussed, pointing out current strategies that positively affect the development of effective assessment and management strategies. Holistic approaches to assess population structure and connectivity were discussed.

Investigating marine species population structure in a multidisciplinary framework can reveal signatures of potential local adaptation and their consequences for management and conservation. In **Chapter 2**, the population structure of common sole (*Solea solea*) in the Mediterranean Sea was delineated using genomic and otolith data of distinct geographical sites. The impact of the selected features on the actual population structure was evaluated through the correlation of SNPs with environmental and spatial variables. Specifically, a seascape genetics approach was used to identify loci potentially involved in local adaptation.

The power of combining several methods inside a multidisciplinary approach to provide stock delineation can be valuable to define spatial population structure of marine species from an integrated biological perspective. Considering the difficulty of combining methods with different spatial and temporal scales, in **Chapter 3** two integrative analysis (Stock Differentiation Index and integrative multivariate analysis) were performed to combine the results of **Chapter 2**. The assessment and interpretation of the results was discussed for all analysis.

In **Chapter 4**, the population structure of European hake (*Merluccius merluccius*) was investigated using genomic data. The identification of patterns of neutral and potential adaptive genetic variation was achieved by applying seascape genomic framework considering spatial and environmental features. Distinct geographical sites within the Alboran Sea, extending into the neighboring Mediterranean Sea and Atlantic Ocean were targeted to analyse putative populations for European hake. The work was carried out in the TRANSBORAN Project framework.

Finally, the results of the multiple research achieved during my PhD project were discussed in **Chapter 5**. Considering the difficulties and limitations of each case study, the importance of each analysis was discussed reporting the proposed stock units delimitation as conclusions. The discussion comprehended future directions strategies.

Chapter 1
General introduction

1.1 Convergence between biological populations and stock units using stock delineation methods

About most of Mediterranean stocks assessed are currently fished at biologically unsustainable levels (FAO, 2020b). The Common Fisheries Policy (CFP) aims to obtain fishing activity to become environmentally, economically, and socially sustainable in the long-term by providing fisheries resources for the future generations (Reg. (EU) n.1380/2013). Stock units, defined as a group of fishes exploited in a specific area or by a specific method, are used to assess exploited fish species (Smith et al., 1990). Commonly individual stock units are assessed based on single species assumption (Begg et al., 1999). A single species advice for each individual stock unit is provided for fisheries management. For this reason, the availability of information for each fish stock, as those related to stock status and the associated exploitation rates, is crucial to ensure a sustainable exploitation of fisheries resources. Geographical Sub-Areas (GSAs) were established by the General Fisheries Commission for the Mediterranean (GFCM) of the Food and Agriculture Organization of the United Nations (FAO) to assess and manage fish stocks according to geographical boundaries and political features. The reinforcement of information on stock boundaries according to stock identification results have exposed discrepancies between the spatial structure of biological populations and stock units currently defined in stock assessment (Kerr et al., 2017). This spatial mismatch between biological population and management unit can negatively influence stock assessment and therefore the development of sustainable fisheries management, affecting both the conservation of resources and the subsistence of fisheries communities (Kerr et al., 2017; Cope and Punt, 2009). Hence, stock delineation with an evaluation of the spatial structure should be a relevant aspect to include in stock assessment models (Cadrin, 2020). The identification of stock units improves the understanding of population dynamic which is a necessary condition for the application of effective management measures (Berger et al., 2021). The advantage in understanding

population structure and dynamics of marine resources is clear in a conservation prospective also because many factors among which habitat degradation, fishing exploitation and climate change contributed to the variation of spatial distribution of these species (Ciannelli et al., 2013; Cheung et al., 2009).

Over recent years, a large number of methods have been proposed to handle the identification of population structure and settle the issue of the misalignment between stock units and biological populations. Natural tracers such as genetic markers, otolith shape and chemistry, stable isotopes, parasites and morphometrics/meristics data, but also artificial ones (tagging) have been used to reveal insights of population structure and connectivity between marine populations. Much information is located in these tracers (Cadrin et al., 2013). The tracers used to address the structure and connectivity of targeted fish species are briefly described in this chapter.

In recent years, advanced stock identification methods have been widely applied with their results included in scientific advice for management actions. Especially, genomic studies have consistently increased given the power of advanced techniques in discriminating stock units. By changing the way in which sequence data are produced, Next Generation Sequencing (NGS) technologies enabled the transition to population genomic research increasing the number of sequences produced and promoting genetic marker discovery and high-throughput genotyping (Davey et al., 2011; Hemmer-Hansen et al., 2014). The generation of this huge amount of data with a higher genome representation using 'Genotyping By Sequencing' (GBS) approach and an increased statistical power represented an advantage for population genetic studies (Narum et al., 2013). For instance, using Single Nucleotide Polymorphism (SNP) markers, it was possible to describe genetic population structure of marine species. Fine-scale population structure has been detected even by a small amount of SNPs containing highly informative insights (Milano et al., 2014; Randon et al., 2020). As evidenced, genetics tracers provide useful tools for informing fisheries manager in terms of population structure of marine species (Ovenden et al., 2015).

Stock identification methods include, among the others, otolith tracers. The otolith, as the main calcified structure for marine organisms, especially for fish species, is nowadays widely used to identify the stock structure of marine population. Otolith are aragonite structures located in the inner ear with a composition influenced by physical and chemical characteristics of water masses. The otolith grows continuously over time through the life of the fish by daily and seasonal deposition of material without chronological alteration due to environmental variations or changes in fish condition. The information retrieved was experienced by fish during all life (Campana, 1999). As a consequence, the shape of the otolith continues to be used in recent studies as an effective tool in discriminating fish stocks from different geographic areas (Ferhani et al., 2021; Wujdi et al., 2022; Saltalamacchia et al., 2022). Also trace elements are present within an organic matrix in otolith (Campana, 1999) and their composition has been successfully used to tackle issues related to stock identification and fish movements in recent studies (Hüssy et al., 2022; Moreira et al., 2022).

1.2 A holistic approach to assess population structure

The discrimination of stock units is based upon a wide variety of multidisciplinary methods (Cadrin et al., 2013). In the stock identity field, the application of multiple techniques is a widely used strategy. For instance, most recent studies routinely implemented different sources of information originating from the otolith such as stable isotopes, otolith elemental trace, otolith shape or microstructure (Morales-Nin et al., 2022; Singh et al., 2022). As well as other later studies increasingly employed genetics and other natural markers displaying information at a various temporal and spatial scale (Marval-Rodríguez et al., 2022; Randon et al., 2020). For instance, the combination of otolith chemistry with distinct markers allowed the validation of the stock structure signals by providing complementary information towards a more robust stock discrimination (Hüssy et al., 2022; Avigliano, 2022). In a perspective of seascape genomics approach, also environmental features can be combined with population genomics to understand factors shaping population structures (Benestan et al., 2019).

The application of stock identification methods is periodically assessed by Stock Identification Methods Working Group (ICES, 2022). The reviewing and evaluation performed by this ICES expert groups exposed the limitation and consistency of each method in informing on the stock structure and connectivity. The delineation of spatial structure should be disclosed by a holistic approach integrating different stock delimitation methodologies. Hence, the limitations deriving from single technique approaches could be overcome resulting in a consequent corroboration of findings derived from stock structure research (Welch et al., 2015). This strategy is used to inform the proper alignment between population and stock unit mitigating the possible misinterpretations (Cadrin et al., 2014b; Kerr et al., 2017). Furthermore, the consequences of integrating this information into the assessment and management strategies should be evaluated by testing different scenarios (Cadrin et al., 2014b).

In a context of multidisciplinary approach, highly relevant projects have been funded by the European Commission for stock units delineation of the commercially valuable fisheries resources in the Mediterranean Sea. As first project, STOCKMED contributed to stocks identification of demersal and small pelagic species using already available information without adding new data (Fiorentino et al., 2014). To overcome this limitation, MED_UNITs project was launched to identify stock units of relevant demersal species covering all Mediterranean Sea basin to fill information gaps (Spedicato et al., 2022). The challenges in stocks identification are even more prominent in transboundary area, as in the case of Alboran Sea where TRANSBORAN project was launched to identify population structure of fish stocks from Atlantic waters to Mediterranean waters subject to specific oceanographic, ecological, fisheries and management characteristics.

Taking into account the demand of knowledge and in the context of these projects, the ultimate objective of my research is the identification of stock units of two marine species, which are considered important fishery resources in the Mediterranean Sea, following a holistic approach performing multi-tracers stock identification

methods. The project could help gain a better understanding of the actual fish populations of these species in the Mediterranean Sea filling the knowledge gaps on stock structure and promoting the identification of "real" stock units. This approach is expected to reduce the bias associated with the assumption of a given stock units and thus improve sustainable exploitation.

Chapter 2

A multidisciplinary approach to describe population structure of *Solea solea* in the Mediterranean Sea.

2.1 Introduction

The marine realm is often characterised by high dispersal of most species, regarded as the major force maintaining connections between populations (Hauser and Carvalho, 2008). Research efforts have been made to challenge the concept of homogeneous marine populations and to identify population structure in marine fish (Reiss et al., 2009). Marine fish population structure can be defined along a continuum ranging from panmictic (e.g., Japanese Spanish mackerel *Scomberomorus niphonius* (Kitada et al., 2017), and Australian yellowfin bream *Acanthopagrus australis*, (Roberts and Ayre, 2010)) to distinct populations (Haddock *Melanogrammus aeglefinus* (Berg et al., 2021)). Indeed, most of these studies revealed complex spatial patterns (European hake *Merluccius merluccius* (Milano et al., 2014; Westgaard et al., 2017) and horse mackerel *Trachurus trachurus* (Abaunza et al., 2008; Healey et al., 2020)) and eventually helped to define fine-scale population structure. The role of complex sets of biophysical processes, along with environmental features, which are the main factors in connectivity, needs to be understood. Such processes influence population differentiation (Cowen and Sponaugle, 2009), at the same time as interaction among connectivity, population size, and environmental conditions. In some cases, fishing pressure acts as a selective force in local adaptation, which maintains or modifies population structure (Hauser and Carvalho, 2008).

Most of fishery resources in the Mediterranean Sea are over-fished and have been exploited at biologically unsustainable levels (FAO, 2020c). Management regulations based on a proper understanding of the stocks dynamic could help to improve the situation in some areas. Commonly, commercial marine species population status is assessed based on single species and individual stock units assumptions (Begg et al., 1999; FAO,

1999). Stock units in the Mediterranean Sea are generally enclosed within the Geographical Sub-Areas (GSAs), which are statistical units established according to geopolitical features and management needs at the General Fisheries Commission for the Mediterranean (GFCM) of the Food and Agriculture Organization of the United Nations (FAO). Improvements in data availability and stock identification methods have revealed incongruities between the spatial structure of biological populations and currently defined stock units, also in a recent Mediterranean study (Kerr et al., 2017; Spedicato et al., 2022). This spatial mismatch between biological populations and management units can violate stock assessment models assumption, which requires data representative for the entire population. Therefore, defining the fine-scale population structure of commercially exploited marine species is crucial to the development of sustainable management, affecting both the conservation of resources and the subsistence of fisheries communities (Cope and Punt, 2009; Kerr et al., 2017). Assessing and understanding the genetic populations structure of commercially exploited marine fish could provide an empirical solution to tackle these mismatches by providing evidence on the stock boundaries, and their reconciliation with the stock units based on GSAs (Kvamme et al., 2020). Improved and biologically-informed fishery management units are likely to enable more robust stock assessments and enhance effective management and conservation actions (Paris et al., 2018), providing a reliable basis for understanding population dynamics (Casey et al., 2016).

Advanced methods for delineating stocks and assessing population connectivity in marine species include molecular techniques, nowadays enabled by the high throughput potential of Next Generation Sequencing (NGS) technologies (Ovenden et al., 2015). Genomic tools can substantially increase the resolution in defining management units and help to overcome existing limitations on the use of genetic information in fisheries management schemes (Bernatchez et al., 2017; Waples et al., 2008). By changing the way in which sequence data are produced, NGS enabled the transition from population genetics to population genomics research in marine fishes (Willette et al., 2014), offering new possibilities for management of marine resources (Hemmer-Hansen et al., 2014; Valenzuela-Quiñonez, 2016). For instance, the application of Single Nucleotide Polymorphism (SNP) markers revealed a strong differentiation among geographical samples in the European hake (*Merluccius merluccius*) and the further detection of a fine-scale genetic population structure in the Mediterranean Sea (Milano et al., 2014). For both European anchovy (*Engraulis encrasicolus*) (Catanese et al., 2016) and Atlantic mackerel *Scomber scombrus L.* (Rodríguez-Ezpeleta et al., 2016) the complex structure revealed by SNPs suggests that management as independent units would be more a successful strategy.

Among stock identification methods, markers based on the chemical composition and shape (morphometrics) of otoliths have been successfully used to unravel stock identification and fish movements (Campana and Thorrold, 2001; Geffen et al., 2016; Higgins et al., 2010). Recent studies support the efficiency of otolith shape analysis for studying fish population structure (e.g., European sardine *Sardina pilchardus* (Jemaa et al., 2015), European anchovy *Engraulis encrasicolus* (Bacha et al., 2014), blue jack mackerel, *Trachurus*

picturatus (Moreira et al., 2019) and European hake *Merluccius merluccius* (Morales-Nin et al., 2022)). The power of a multidisciplinary approach which combines several data sources and methods to achieve stock delineation has often been demonstrated (Higgins et al., 2010; Spedicato et al., 2022; Pita et al., 2016). As a matter of fact, the integration of outcomes from different methods enhances the robustness of the study (Welch et al., 2015; Izzo et al., 2017) and overcomes drawbacks of single methods by providing a more robust stock discrimination (Hussy et al., 2016; Paris et al., 2018), confirming the value of a multidisciplinary framework in defining spatial structures of marine species (Cadrin, 2020; Cadrin et al., 2014a). By assessing multiple data types, such as genomic and otolith data, a multidisciplinary approach can integrate information on different temporal scales (evolutionary and ecological) to accurately detect the spatial structure of marine populations (Tanner et al., 2016; Abaunza et al., 2008; Welch et al., 2015). This feature is extremely useful to comprehend the temporal disparity between microevolutionary processes, which occur at generational timescales, and ecological-demographic connectivity processes, which become apparent at shorter timescales (Hawkins et al., 2016).

The advances in genomic tools have also enabled the investigation of the influence of environmental factors on genetic variation (Liggins et al., 2019). Over the past few years, seascape genomics studies have explored both adaptive and neutral genetic patterns in marine species, making use of spatial and oceanographic information (Benestan et al., 2016; Segovia et al., 2020). This approach has become increasingly important in detecting patterns of local adaptation in marine fish populations by assessing functional variation of loci associated with environmental features (Liggins et al., 2019).

The present study focuses on common sole (*Solea solea*) (Linnaeus 1758) in the Mediterranean Sea, a demersal species of high commercial interest targeted by large scale (bottom trawlers) and small-scale (mainly gill nets) fisheries over the study area (FAO, 2020b). Common sole habitat spreads over the continental shelf of the entire Mediterranean Sea, especially in areas characterized by soft bottoms and shallow waters (*i.e.* < 100 m). Given its ecological niche, the common sole is unevenly distributed in the Mediterranean Sea, with more than half of the overall fishery production attributable to Adriatic Sea catches (52%), followed by contributions to the harvest from the Aegean Sea (14%) and the Levant Sea (11.5%), according to FAO production data 2007-2017 (FAO, 2020a), aggregated by FAO Major Fishing Areas (FAO, 2022). All other areas account for less than 10% of the Mediterranean harvest of common sole. A formal evaluation of the stock status, at the time of writing, is only available for the Adriatic Sea, and this stock is experiencing a decreasing trend in fishing mortality reflected in remarkable improvements in terms of biomass in recent years (FAO-GFCM, 2021). Because the stock status in all the other FAO-GFCM areas remain unassessed, improved details of population structure could contribute significantly to the design of sustainable management for common sole in the whole Mediterranean Sea (Reiss et al., 2009). The Mediterranean populations of common sole do exhibit some discernible genetic structure, described as a west-to-east genetic differentiation pattern (Rolland et al., 2007) (Guarnieo et al., 2002; Kotoulas et al., 1995). The Adriatic Sea population is also

genetically differentiated from all the other Mediterranean ones. Morphology and composition of otoliths have been used to studied common sole population structure both in the North-Western Mediterranean Sea and the Atlantic (Mérigot et al., 2007; Cuveliers et al., 2010; Morat et al., 2014; AsyMMETR, 2013), in the latter area often in combination with genetic markers (Cuveliers et al., 2010; Delerue-Ricard et al., 2019). Several factors could lead to selection processes in Mediterranean common sole populations. Local selection pressure is likely the result of abiotic factors such as water temperature and river inflow, and the resulting physical gradients across latitudes affecting early life history stages (Vaz et al., 2019), as well as biotic factors such as fish density and the availability of prey (Grati et al., 2013). Many other factors have been studied, for example recruitment success that is independent of nursery ground size which suggests that it is not a density dependent process (Rogers, 1992).

This study assessed the population structure of common sole in the Mediterranean Sea through the simultaneous use of different markers in order to gain a better understanding of the differentiation patterns, and ultimately to fill the knowledge gaps on the species stock structure. The analyses aim at delineating the population units from an integrated perspective, by investigating simultaneously individual genomic profile, otolith shape and trace element composition. We also applied a seascape genomics approach to assess the capacity of spatial distribution and environmental drivers to define neutral and adaptive genetic structure in common sole. Finally, we associated each marker to the predicted functional annotation of the genomic region, and we performed a gene enrichment analysis to identify genes potentially involved in processes of local adaptation leading to fine scale differences within the study area.

2.2 Methods

2.2.1 Molecular and otolith data

Samples and data analysed in this study originated from previous effort carried out in the framework of the FishPopTrace EU Project (`https://fishpoptrace.jrc.ec.europa.eu/`). Population samples were collected over the period from 1999 to 2009 from 13 geographical sites in the Mediterranean Sea (Figure 2.1 and Table 2.1) (Nielsen et al., 2012). Genotype data at 426 SNP markers were obtained as described in previous works (Diopere et al., 2018; Nielsen et al., 2012). Individuals and loci with more than 20% and 10% of missing data, respectively, were removed from the SNP dataset. Monomorphic markers were also discarded prior to the analysis. Details of the evaluation of missing data and filtering procedure are presented in the Supplementary materials 2.5.1.1. The final data set contained 352 individual genotypes at 380 SNPs. Otolith data were produced for population samples from five geographical locations representing the entire Mediterranean basin (Figure 2.1 and Table 2.1), full details of otolith processing and analysis are described in the Supplementary materials 2.5.1.2. In these five sites, the same individuals were processed for

both molecular and otolith data.

Based on geographical position of the 13 sampling sites, we identified nine centroid locations corresponding to a coastal slice defined as the portion of each GSA delimited by the continental shelf boundaries within the depth of 200 metres (m) (Figure 2.1). The sampling sites were associated to the nearest centroids for the seascape analyses. We used the geographical coordinates (latitude and longitude) of the centroid for all analyses (Figure 2.1). R version 3.6.3 was used for all analyses performed within the R environment in this study (R Core Team, 2022).

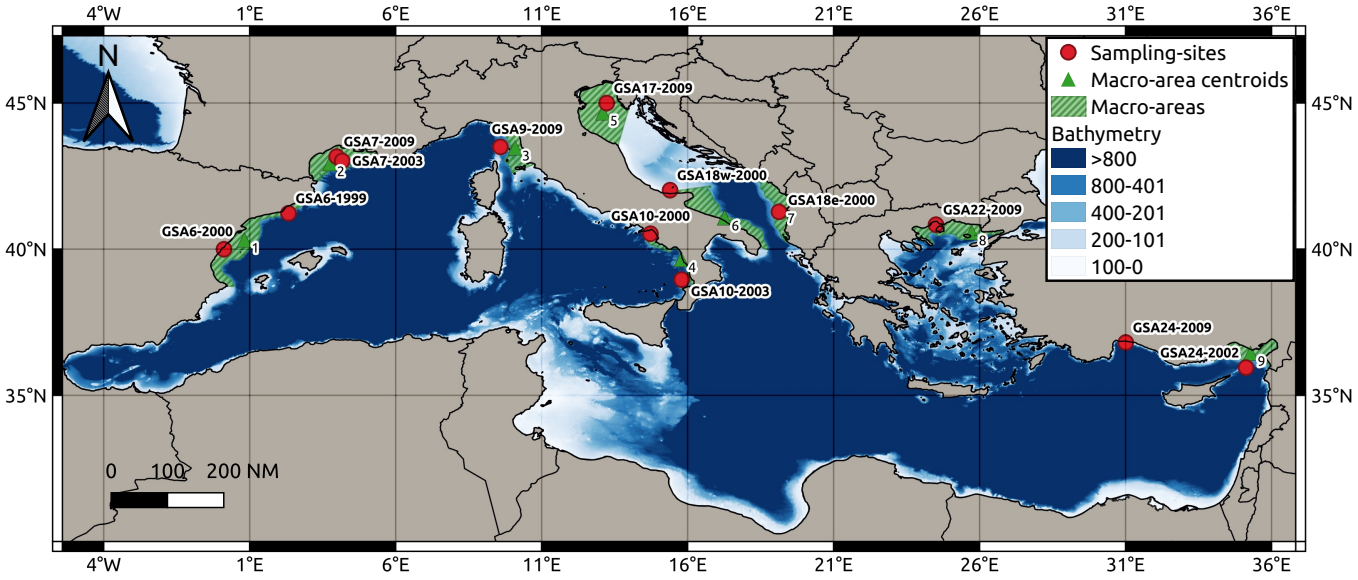

Figure 2.1: Map showing the 13 sampling sites located in the Mediterranean Sea, centroids of nine macro-areas and bathymetry.

Sampling site	Macro-area	GSA	Sampling year	Site ID	N			
					Molecular	Otolith		
						Shape	Trace element composition	
							core	edge
Castellon	1	Northern Spain	2000	GSA6_2000	27			
Barcellona	1	Northern Spain	1999	GSA6_1999	13			
Gulf of Lion	2	Gulf of Lion	2009	GSA7_2009	33	30	27	25
Gulf of Lion	2	Gulf of Lion	2003	GSA7_2003	40			
Viareggio	3	Ligurian Sea and Northern Tyrrhenian Sea	2009	GSA9_2009	36	39	26	30
Gulf of Salerno	4	Southern and Central Tyrrhenian Sea	2000	GSA10_2000	14			
Gulf of Sant'Eufemia	4	Southern and Central Tyrrhenian Sea	2003	GSA10_2003	16			
Chioggia Lagoon	5	Northern Adriatic Sea	2009	GSA17_2009	40	44	26	29
Lesina	6	Southern Adriatic Sea	2000	GSA18w_2000	16			
Albanian coasts	7	Southern Adriatic Sea	2000	GSA18e_2000	11			
Gulf of Kavala	8	Aegean Sea	2009	GSA22_2009	40	47	46	47
Gulf of Antalya	9	Northern Levant Sea	2009	GSA24_2009	27	30	29	26
Antakya coasts	9	Northern Levant Sea	2002	GSA24_2002	39			

Table 2.1: Summary information of sampling site locations and *Solea solea* population samples for genetic and otolith data. Table reports information on geographical areas defined for the analysis and the number of individuals (N) analysed for each technique.

2.2.2 Genetic analysis

To assess genetic variation between samples, observed (Ho) and expected (He) heterozygosity, polymorphic rate, and F_{IS} estimate (Weir and Cockerham, 1984) were calculated using *Genepop v.4.7.5* R packages (Rousset, 2008), respectively (Supplementary materials 2.5.1.1). Global and by locus Hardy-Weinberg (HW) tests were performed using R version of *Genepop v.4.7.5* (Rousset, 2008) (Supplementary materials 2.5.1.1). We estimated global and pairwise values of F_{ST} index (Weir and Cockerham, 1984) and relative significance through 10,000 permutations to assess genetic differentiation among populations using *strataG* R package (Archer et al., 2017) (Supplementary materials 2.5.1.1). We then created a heatmap based on F_{ST} pairwise comparison to inspect population structure. We performed a Discriminant Analysis of Principal Components (DAPC) with *adegenet* R package (Jombart, 2008) to identify the individual-based population structure. We ran the method both including as input the optimal number of clusters determined based on the Bayesian Information Criterion (BIC), as well as providing the sampling sites as prior. To describe the genetic structure, the first option allowed us to identify the optimal number of clusters, while the second one is used to assess the quality of the discrimination (Supplementary materials 2.5.1.1). Outlier loci were identified using four independent approaches namely *pcadapt* R package (Luu et al., 2017), *fsthet* R package (Flanagan and Jones, 2017), Bayescan version 2.1 (Foll and Gaggiotti, 2008) and FDIST approach implemented in *Arlequin v.3.5.2.2* (Excoffier and Lischer, 2010). Throughout, SNPs with MAF lower than 0.05 were removed from the data set before applying the methods (Bekkevold et al., 2020), FDR was set at 5% and 95% CI was utilised to identify loci with *fsthet* (Flanagan and Jones, 2017). The method implemented in *pcadapt* R package relies on PCA to detect population structure, and it allows the identification of genetic markers potentially involved in biological adaptation considering their relationship with the population structure. First, a PCA is used to ascertain population structure and identify the number of principal components K to consider. Then, each SNP is regressed by the K principal components. The results are reported as a vector of z -scores between each SNP, and the first K components obtained with the linear regression. Based on this vector, the outliers are then identified using a multi-dimensional approach. The Mahalanobis distance is used as a statistic to measure the distance between each point and its mean. It calculates the distance between the covariance matrix of the z -scores and the vector of the z -score means. Finally, Mahalanobis distances are transformed in p-values based on the correlations between SNPs and the K principal components to performed multiple hypothesis testing. The q-value procedure is recommended to choose a threshold for p-values FDR approach (Benjamini and Hochberg, 1995). We selected $\alpha = 0.1$ for the threshold of outlier SNPs. We used *fsthet* R package (Flanagan and Jones, 2017) to identify loci with extreme F_{ST} values relative to their heterozygosity H_t. Smoothed quantiles from empirical F_{ST}-H_t distribution were calculated without the need for specifying a population genetics model. Loci lying outside of the quantiles for the 95% confidence level were regarded as outliers. Bayescan relies on differences in allele frequencies between subpopulations to identify candidate

loci under selection. Subpopulation specific F_{ST} coefficients are divided into two components by logistic regression. The population-specific component β is shared by all loci, and the locus-specific component α is shared by all populations. When the locus-specific component α is significantly different from zero means that the component is necessary to explain the diversity pattern and points out a departure from neutrality at the given locus. Specifically, positive values of alpha indicate diversifying selection whereas negative values suggest balancing/purifying selection. For each locus, a posterior probability is computed and SNPs with prior odds (PO) greater than a threshold are considered outliers since PO express how likely the model is with selection compared to neutral model. Here, we ran the program with 50,000 burn-in period, 100,000 number of iterations, 20 pilot runs and thinning interval size set to 50. We set prior odds (PO) to 10 since the data set comprised less than 1,000 SNPs and we set the FDR to 5%. We applied FDIST approach implemented in *Arlequin v.3.5.2.2* (Excoffier and Lischer, 2010) which uses simulations under an infinite island-model to generate the distribution of F_{ST} as a function of heterozygosity. SNPs found in the tails of the distribution are then identified as outliers. We set 50,000 simulation, 100 demes with all geographical samples assigned to only one group, and expected heterozygosity ranging from 0 to 0.5. P-values were corrected through FDR approach (Benjamini and Hochberg, 1995), and SNPs with an associated p-value lower than 0.05 were considered as putative outliers. We then evaluated the overlap in loci to create datasets of neutral and outlier SNPs that could suggest different patterns of genetic divergence. We used Venn diagram to visualize SNPs that were common to each method. To create the outliers dataset we considered the union of all loci identified as outliers by each method, while the remaining markers constituted the neutral dataset. We repeated genetic differentiation and multivariate analysis (PCA, DAPC) using the two resulting datasets to compare potential alternative patterns of divergence. To further investigate population structure, we ran STRUCTURE v.2.3.4 (Pritchard et al., 2000) using the datasets of neutral and outlier SNPs as separate inputs. The software was run assuming admixture, correlated allele frequency and testing the scenarios with locality information (LOCPRIOR). We ran the analysis for K from 1 to 10 with 20 iterations, burn-in period at 20,000 and run length at 100,000. Alpha value was kept as default ($\alpha = 1$). Mean L(K) method was employed to visualise the mean likelihood values of different K partitions and identify the optimal K. R package *pophelper* (Francis, 2017) was utilised to align and merge the multiple runs for each K, as well as to visualise the posterior membership probability of each individual to clusters from 1 to 10. To validate the results, we also performed a spatial principal components analysis (sPCA). sPCA is a multivariate method, implemented by the function *spca* of the *adegenet* R package (Jombart, 2008, 2017). The method is functional to investigate and identify spatial genetic patterns, uncover global and local structure, and evaluate genomic clines due to geographic variation. We performed sPCA to analyse the genetic variation due to spatial variation utilising both neutral and outlier SNP as input data. To incorporate the spatial information, we transformed latitude and longitude of centroids into Cartesian coordinates (xy) (Stanley et al., 2018) using the custom R function *coord_cartesian* found in `https://github.com/rystanley/CartDist/blob/master/CartDistFunction.R`

(Stanley and Jeffery, 2017). The function computes the least-coast distances between each sampling site to account for land barriers. The resulting distance matrix was used to re-project the coordinates into two-dimensional Cartesian space using a non-metric multidimensional scaling (Jeffery et al., 2017; Stanley et al., 2018). We applied the function *jitter* to spread the xy coordinates in 1 km in order to obtain different coordinates for each individual and ran the analysis with the Delaunay triangulation as connection network among geographical sites (Jombart et al., 2008; Lehnert et al., 2019). We ran both the global and local test with 9,999 permutations (Lehnert et al., 2019; Jombart, 2017).

2.2.3 Otolith analysis

We investigated both the variation in otolith shape and the trace element composition. We calculated five shape indices: circularity, roundness, ellipticity, form factor and rectangularity (Tuset et al., 2003). In addition, the otolith shape was described through elliptic Fourier descriptors. We identified outliers with Interquartile Range (IQR) method to remove them before statistical analysis. Otolith contours were extracted from images and Fourier coefficients were calculated using *Momocs* R package (Bonhomme et al., 2014). The detailed procedures may be found in Supplementary material 2.5.1.2.

Shape indices were tested for normality and homogeneity of residuals using Shapiro-Wilk Normality test and Levene's test respectively. Only normally distributed shape indices were retained for further analysis. To avoid redundancy of the data, we calculated Pearson correlations (Mérigot et al., 2007) using *rcorr* function in *Hmisc* R package (Harrell Jr and Harrell Jr, 2019). We investigated whether the relationship between shape indices and total fish length differed significantly for sampling sites with analysis of covariance (ANCOVA) (Keating et al., 2014; Longmore et al., 2010; Rashidabadi et al., 2020) using *lm* R function. If a significant interaction between sampling sites and length was detected, we removed the shape index. Otherwise, if the interaction was not significant and the slopes were equal across sampling sites, the shape index was corrected for size effect based on the linear relationship (Midway et al., 2014; Longmore et al., 2010), using the equation $Y_c = Y - b * L$, where Y is the shape index, b is the common within group slope of the shape-size relationship and L is the measurement of total fish length (mm) (Keating et al., 2014). We used as slope the coefficient b resulting from the linear regression between total fish length and shape index. We repeated ANCOVA analysis to check that the co-variations were effectively treated ($p > 0.01$). We then performed a PERMANOVA analysis to uncover differences between geographical samples using *adonis* function in *vegan* R package (Oksanen et al., 2022). For the analysis, we computed a dissimilarity matrix using *vegdist* function with Mahalanobis method (Rashidabadi et al., 2020) and 1,000 permutations. We investigated the homogeneity of group dispersion with *betadisper* function. Then, we applied *pairwise.adonis* function in *pairwiseAdonis* R package to identify significantly different pairs of geographical sites, applying the Benjamini-Hochberg correction for multiple testing.

The first 40 elliptical Fourier harmonics were obtained from each otolith. Each harmonic n is described by 4 coefficients a_n , b_n , c_n and d_n, a total of 160 elliptical Fourier descriptors (EFDs) were available. We calculated the Fourier power (FP) for each otolith utilising the *calibrate_harmonicpower_efourier* to estimate the number of harmonics required for the elliptical Fourier descriptors analysis. Using the *efourier* function, we computed EFDs and we obtained a collection of coordinates (*Coe* object). Through a PERMANOVA analysis on the EFDs, we assessed differences between geographical sites using *adonis* function in *vegan* R package (Oksanen et al., 2022). For PERMANOVA analysis we followed the same procedure applied to shape indeces. Canonical analysis of principal coordinates (CAP) was performed on both types of morphometric data to visualize the spatial differences in classification accuracy obtained by leave-one-out cross-validation. We used *CAPdiscrim* function in *BiodiversityR* R package to perform the analysis based on Euclidean distances. To determinate the number of PCoA axes (m) to retain in the analysis we follow the general procedure used by (Kindt and Kindt, 2015). We set the argument m from 1 to 100 in the CAP analysis to explore the difference in the proportion of correct assignment of individuals to groups. Thus, we selected the m with the highest percentage and we used it for the CAP analysis. After the PCoA, we checked for the percentage of total variance in the dissimilarity explained by the m PCoA axes. We checked that the axes chosen did not exceed 100% of the variance.

Furthermore, the SIs and the EFDs were combined to evaluate their discriminatory power. We performed canonical analysis of principal coordinates (CAP) following the same procedure described above.

Trace elements composition data were obtained from core region and edge of sole otoliths using laser ablation inductively coupled mass spectrometry (LA-ICPMS) (Chang et al., 2012; Morales-Nin et al., 2014). After assessment of missing values, we replaced concentration of elements for which more than 25% of values were below the LOD with respective LOD value (Mercier et al., 2011). The final data set for both edge and core comprised ^{23}Na, ^{24}Mg, ^{63}Cu, ^{64}Zn, ^{85}Rb, ^{88}Sr, ^{138}Ba, ^{208}Pb. Trace element concentrations were log-transformed for further analysis.

The same statistical analysis was performed for both core and edge elemental data. Before data analysis, we checked for normality and homogeneity of the residuals with Shapiro-Wilk Normality test (Shapiro and Wilk, 1965) and Levene's test (Levene et al., 1960), respectively. We investigated the effect of otolith weight on each element by performing a non-parametric analysis of covariance (ANCOVA) (Moreira et al., 2018; Ferreira et al., 2019). To investigate the effect of otolith weight, we calculated Spearman correlations for each element (Longmore et al., 2010) using *rcorr* function in *Hmisc* R package (Harrell Jr and Harrell Jr, 2019). We performed a non-parametric Kruskal-Wallis statistics to investigate differences in elemental concentrations between geographical sites. We inspected pairwise comparisons between sites for each element through Dunn's test using *dunn.test* function with Bonferroni correction for multiple testing. We performed a Discriminant Analysis of Principal Components (DAPC) with *adegenet* R package (Jombart, 2008) to predict the sampling area membership of each individual using the multivariate element composition assessing

the quality of discrimination. As the optimal number of clusters identified using k-means algorithm did not provide evidence of correspondence with individual groups, we used the sampling sites as a prior.

2.2.4 Environmental analysis

Genetic-environmental association analysis

We represented the spatial structure between geographical sites with distance-based Moran's eigenvector map (dbMEMs) variables obtained through a distance matrix. In order to determine the dbMEMs, we first calculated the spatial distance matrix from geographical coordinates (longitude and latitude of centroids) containing the distances between the locations using the *gcd.hf* function. Then, we computed the distance-based Moran Eigenvector's Maps (dbMEMs) using the *pcnm* function in *vegan* R package (Supplementary materials 2.5.1.3) (Oksanen et al., 2022). In this way, we were able to use spatial variables (dbMEMs) as independent vectors, representing the spatial structure associated with the neighbourhood network across different scales, as explanatory variables in the following analysis (Borcard and Legendre, 2002). Environmental variables were downloaded from E.U. Copernicus Marine Service Information according to their availability over the analysis period from 2000 to 2010 (CMEMS, `http://marine.copernicus.eu/`) (Simoncelli et al., 2019; Teruzzi et al., 2019) and from historical monthly weather data from WorldClim 2.1 (Fick and Hijmans, 2017). From the available products, we selected 17 environmental variables, as reported in Supplementary materials 2.5.1.3, based on their influence on spatial distribution of common sole and their variation between sampling locations (Diopere et al., 2018; Do Prado et al., 2018). To detect multicollinearity, we applied the function *vifcor* in *usdm* R package (Naimi, 2017; Naimi et al., 2014) by which highly correlated variables with correlation values (*th*) greater than 0.65 and with Variance Inflation Factor (VIF) values greater than 2 were excluded. In our seascape genomic analysis we considered as environmental factors the resulting set of four variables: the mean meridional component of the currents (northward) (curM, m/s) using the weighted average value in the water column (Wclm) (curM_Wclm_mean), the range of potential temperature (temp, °C) at the bottom (temp_bottom_range), the mean salinity (sal, psu) at the bottom (sal_bottom_mean) and the mean CO_2 (ocean pCO_2 expresses as carbon dioxide partial pressure) (pco, Pa) (pco_Wclm_mean).

GEA methods

We applied three different genetic–environment association (GEA) methods (Bayenv2, *LFMM* and Samβada) to identify loci potentially involved in local adaptation. First, we used Bayenv2, a Bayesian method that identified the SNPs revealing a significant correlation, expressed by a Bayes factor (BF), between their allele frequencies and at least one of the environmental variables. Starting from neutral SNP loci, the method estimates a covariance matrix. The covariance in allele frequencies is used as null model to test the presence

of correlation between SNPs and each environmental variable (Günther and Coop, 2013; Coop et al., 2010). Then, we used the *cov2cor* function to convert that matrix into a correlation matrix and we checked its correlation with the pairwise F_{st} matrix. SNPs with a BF greater than 10 according to Jeffrey's criterion (Jeffreys, 1961) can be considered potentially under local adaptation due to a specific environmental factor. Secondly, we used a latent factor mixed model (LFMM) that identifies associations among environmental variables and genotypes, while considering the influence of latent factors such as, in our case, the neutral structure of common sole population as produced by DAPC analysis (Caye et al., 2019; Frichot et al., 2013). The correlation between the genotype matrix and the environmental variables is found using hierarchical Bayesian mixed linear models. This method was implemented in *LEA* R package (Frichot and François, 2015). We assessed the significance with a p-value of 0.05 after applying the adjustment method FDR. Then, we used *Samβada*, to detect associations between the genetics data and environmental variables using univariate models that analyse each locus independently. This software uses individual geographical coordinates as a link between environmental and genomics information (Duruz et al., 2019; Stucki et al., 2017). We used the *system* function of Samβada software available at `http://lasig.epfl.ch/sambada` to run Samβada from within the R environment. The program returns log-likelihood ratio and Wald test that after Bonferroni correction can be used to assess the significance of correlations between genetics and the environmental variables ($\alpha = 0.05$).

We determined the dataset of SNPs that resulted as the best candidate loci for local adaptation, by retaining those SNPs that were associated with at least one of the environmental variables in at least one of the methods (*Bayenv2*, *LFMM* and *Samβada*). We then conducted a spatial principal components analysis (sPCA) on the resulting set of markers to investigate and identify spatial genetics structure. The analysis helped to discriminate whether the genetic variation related to candidate SNP loci due to the association with the environmental variables was more relevant than expected considering only the spatial distribution (Benestan et al., 2016; Segovia et al., 2020). We ran linear regressions of the locality scores extracted from the sPCA on each environmental and spatial variable with the purpose of determining which variables significantly explained the genetic variation in the candidate SNPs. Lastly, we computed the R^2_{adj} value to assess the proportion of variance explained.

Redundancy analysis

We evaluated the effect of spatial distribution and/or environmental factors on both neutral and outlier genetic structure at nine macro-areas using a redundancy analysis (RDA), a constrained ordination method used in multivariate analysis (Legendre and Legendre, 2012). We first performed two distinct RDAs using both neutral and outlier SNP datasets as response variable and we combined the spatial factors (dbMEMs) and the environmental factors as explanatory variables. Then, we performed a partial RDA both on neutral and outlier datasets to assess only the influence of environmental factors excluding the effect of the spatial

factors on genetic differentiation and vice versa. All RDAs were performed using *rda* function in *vegan* R package (Oksanen et al., 2022). Following Selmoni et al. (2020) we extracted the first principal component that reaches the 80% of cumulative variance to use them as response variables to compute RDA. The principal components were produced performing a PCA on the standardized matrix using the *prcomp* function, after transforming the matrix of allele frequencies with the Hellinger approach using the *decostand* function. We applied the ANOVA and marginal ANOVAs with 1,000 permutations to assess the significance of the global model and of each variable. We calculated the adjusted coefficient of determination (R^2_{adj}) using the *RSquareAdj* function in *vegan* R package to determine the variation explained by the model (Oksanen et al., 2022). Then, we used the *ordistep* function with 1,000 permutations to perform both forward and backward selection of explanatory variables that best explained the variability of the response variable (optimal model). As previously explained, we evaluated the significance of the model and each variable.

2.2.5 Gene Ontology

We annotated assemblies of representative and unigenes transcript of *Solea solea* downloaded from SoleaDB database (Benzekri et al., 2014) by using EnTAP (Eukaryotic Non-Model Transcriptome Annotation Pipeline) (Hart et al., 2020) in $runP$ mode employing three different databases. Then, a similarity search and best hit selection was performed using DIAMOND (Buchfink et al., 2015). Sequences were mapped against NCBI non redundant protein, Swiss-Prot and TrEMBL databases with expected E-value set at a maximum of 0.000001 and a minimum coverage set at 70%, followed by Gene Ontology (GO) term assignment with eggNOG mapper and InterProScan (Zdobnov and Apweiler, 2001). The flanking regions (120bp) of the 380 SNPs were BLASTed against the *Solea solea* assembly found in SoleaDB (Benzekri et al., 2014). We conducted a selection of results of the annotation based on the matches between reference transcript and the corresponding flanking regions containing the SNPs. At this stage we integrated the results of the representative transcript of *Solea solea* with the unigene ones. For the initial dataset of 380 SNP loci, we kept only sequences with the highest score and the lowest E-value. The same procedure was applied using the *Solea senegalensis* assembly from SoleaDB database (Benzekri et al., 2014) in order to integrate missing results. After the reconstruction of GO term graph obtained by retrieving and merging in single sub-graph all the paths from each term to the respective ontology root (Ashburner et al., 2000), we performed a gene enrichment analysis to understand the over representation of GO terms categories in two subsets: SNP loci associated to potential local adaptation (43 loci) and SNP loci associated with environmental variables (148 loci) compared to the complete dataset of 380 markers. The enrichment analysis based on GO terms was performed with a Fisher's exact test. For the Fisher test, we computed a contingency table for each GO term, counting the number of genes in the SNP loci considered associated or not with the GO term, and the number of genes in the whole dataset associated or not to the GO term to understand the over representation of GO terms categories in the two

subsets. The results were expressed by the odds ratio (OR), defined as the ratio of the proportion of a GO term associated with SNP in the subset to the proportion of this GO term outside this set. The larger the odds ratio, the higher the relative abundance of this GO term compared with the entire data set. GO terms for each subset with $OR > 2$ and $\alpha = 0.05$ were identified as enriched in the three different categories.

2.3 Results

2.3.1 Genetic analysis

We found a decrease in the polymorphic rate of the population samples along the west-to-east geographical axis, ranging from 98.68% to 72.11%. Consistently, there was a significant decrease in average observed (Ho) and expected (He) heterozygosity, while no significant differences were found between Ho and He ($\alpha = 0.05$). After FDR correction only few SNPs per population sample were found significantly out of HW equilibrium. At the population samples level, no significant deviation from the HW equilibrium was found. Accordingly, the estimates of F_{IS} were low and consistent with levels of heterozygosity (Table 2.2).

Sample	Rate polim (%)	He	Ho	F_{IS}
GSA6_2000	98.68	0.3513	0.3433	0.0247
GSA6_1999	96.58	0.3619	0.3546	0.0206
GSA7_2009	98.68	0.3466	0.3685	-0.0629
GSA7_2003	99.21	0.3463	0.3394	0.0199
GSA9_2009	97.89	0.3390	0.3445	-0.0161
GSA10_2000	94.74	0.3541	0.3764	-0.0609
GSA10_2003	95.00	0.3559	0.3954	-0.1100
GSA17_2009	91.84	0.3087	0.3047	0.0131
GSA18w_2000	83.95	0.3384	0.3639	-0.0756
GSA18e_2000	82.11	0.3368	0.3507	-0.0435
GSA22_2009	77.11	0.3038	0.3018	0.0066
GSA24_2009	72.11	0.2999	0.3011	-0.004
GSA24_2002	77.63	0.2775	0.2813	-0.0132

Table 2.2: Genetic diversity statistics for each population samples of the Mediterranean *Solea solea* including percentage of polymorphic SNPs rate (Rate polim), mean observed (Ho) and expected (He) heterozygosity, and F_{IS} estimation. Code of population samples are given in Table 2.1

Pairwise F_{ST} values ranged from -0.003 to 0.172, with an overall significant F_{ST} of 0.069 ($p = 0.001$). After

Benjamini–Yekutieli correction, all pairwise F_{ST} remained significant ($\alpha = 0.05$) except for the comparisons involving samples from GSA6, GSA7, GSA10, GSA17 and GSA18 (Suppl. Figure 2.7). A heatmap of pairwise F_{ST} pointed to three main areas with significant levels ($\alpha = 0.05$) of genetic divergence corresponding to the Western Mediterranean, Adriatic Sea, and Levantine Sea (Figure 2.2).

The clustering pattern of population samples resulting from DAPC analysis was concordant with the structure suggested by the genetic pairwise F_{ST}. Based on k-means algorithm and the lowest BIC score, three clusters were identified as corresponding to three main areas: Western Mediterranean, Adriatic Sea and Eastern Mediterranean (Figure 2.3). Western and eastern samples were separated along the first discriminant axis. The second discriminant axis showed a separation between samples from the Aegean and Levantine areas when using prior information about the sampling sites. Moreover, the GSA18e_2000 sample separated from the rest of Adriatic samples and resulted in an intermediate position between samples from other Adriatic sites and Eastern Mediterranean area. In the DAPC analysis performed with prior knowledge of subgroups we retained 125 PCs and five discriminant functions. The proportion of conserved variance for the first two discriminant functions of whole dataset was 83.52% and the proportion of overall correct assignment was 84.38% (Figure 2.3A). Samples GSA10_2000, GSA10_2003, GSA18w_2000, GSA18e_2000 and GSA22_2009 showed a higher correct assignment rate with respect to the other population samples (Suppl. Figure 2.9A). Methods for outlier detection identified different numbers of SNPs as outliers (Suppl. Figure 2.8): *pcadapt* identified eight outlier SNPs, *fsthet* identified 25, *Bayescan* identified 40 and *Arlequin v.3.5.2.2* identified 33. The number of SNPs identified by pcadpat was not sufficient to reveal the population structure, thus its results were removed from further analysis. However, six of the eight loci resulting from *pcadpat* were also found by other methods. The union of markers identified by all methods resulted in a outlier dataset of 43 SNP loci. Accordingly, the neutral dataset included the 337 markers which were not detected as outlier by any of the methods (Suppl. Table 2.5). DAPC analysis performed on the neutral dataset corroborated the result obtained with the whole SNPs dataset, indicating a broad structure characterised by a separation of samples along the longitudinal axes of the Mediterranean. Two clusters were identified by the k-means algorithm and the lowest BIC score. In the DAPC analysis performed with prior knowledge of subgroups we retained 100 PCs and three discriminant functions. The proportion of conserved variance for the first two discriminant functions of neutral dataset was 77.51% and the proportion of overall correct assignment was 59.70% (Figure 2.3B, Suppl. Figure 2.9B). DAPC analysis performed on the outlier dataset revealed higher levels of structure detecting potential distinct subgroups within Western Mediterranean area in respect to whole and neutral datasets. The second discriminant axis showed a deeper separation between samples from Tyrrhenian and western basin areas. Four clusters were identified by the k-means algorithm and the lowest BIC score. In the DAPC analysis performed with prior knowledge of subgroups we retained 25 PCs and three discriminant functions. The proportion of conserved variance for the first two discriminant functions of outlier dataset was 87.37% and the proportion of overall correct assignment was 47.44% (Figure 2.3C, Suppl.

Figure 2.9C). In addition, we obtained concordant results from *STRUCTURE*, provided in Supplementary materials (Suppl. Figure 2.11,2.10), they were highly consistent with the pattern revealed by DAPC analysis.

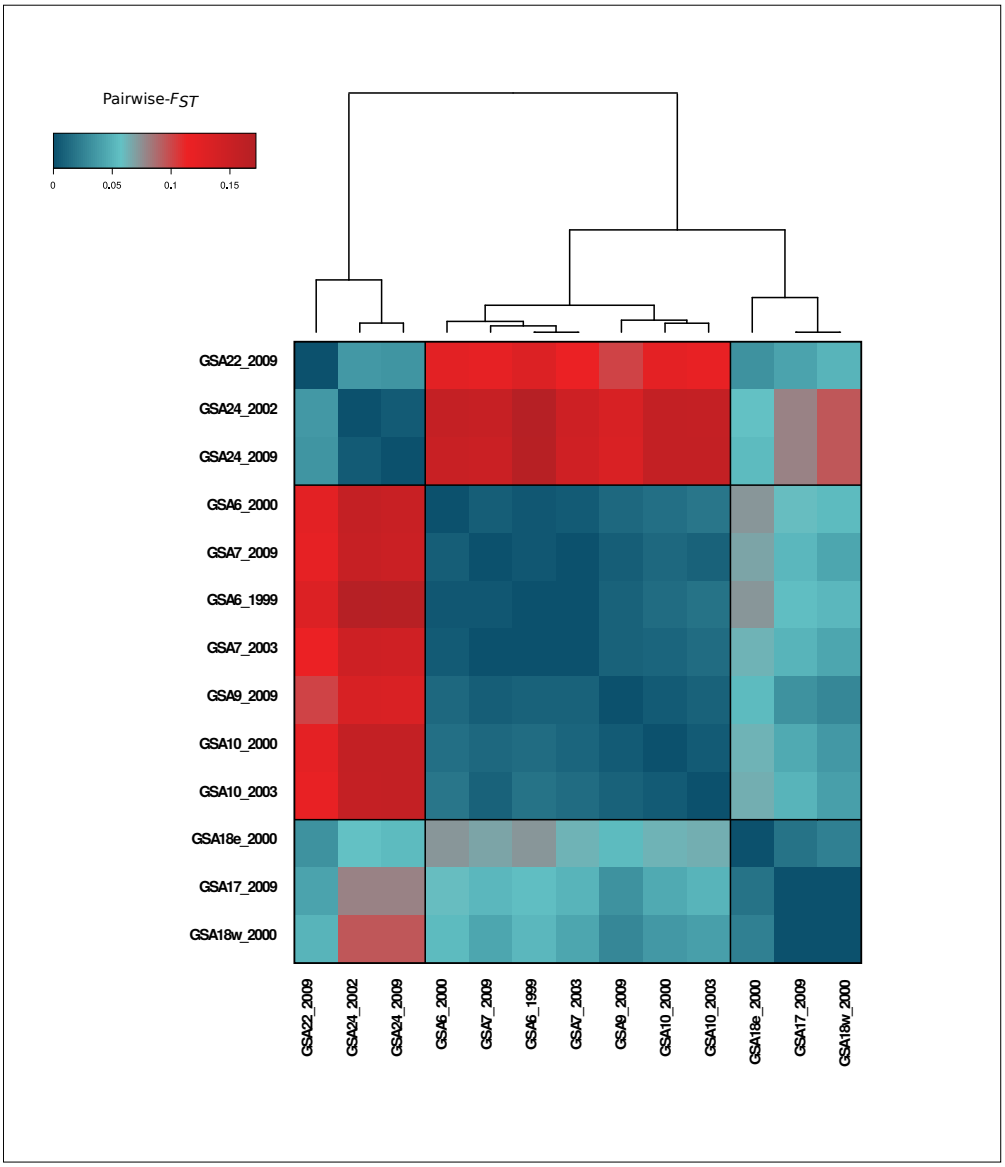

Figure 2.2: Heatmap of pairwise-F_{ST} values among population samples of *Solea solea* computed on 380 SNP loci. Code of population samples are given in Table 2.1

The west-to-east differentiation pattern detected by previous analyses was also confirmed by sPCA results which highlighted a spatial cline in the allele frequencies across longitude for neutral and outlier SNP loci (Suppl. Figure 2.12, 2.13). In the sPCA analysis, the global test was significant (p = 0.0001) for both datasets, reinforcing the evidence of large scale population structure at neutral level. However, the local test was not significant ($p > 0.05$), indicating a lack of fine-scale structure. We inspected the allele loadings of

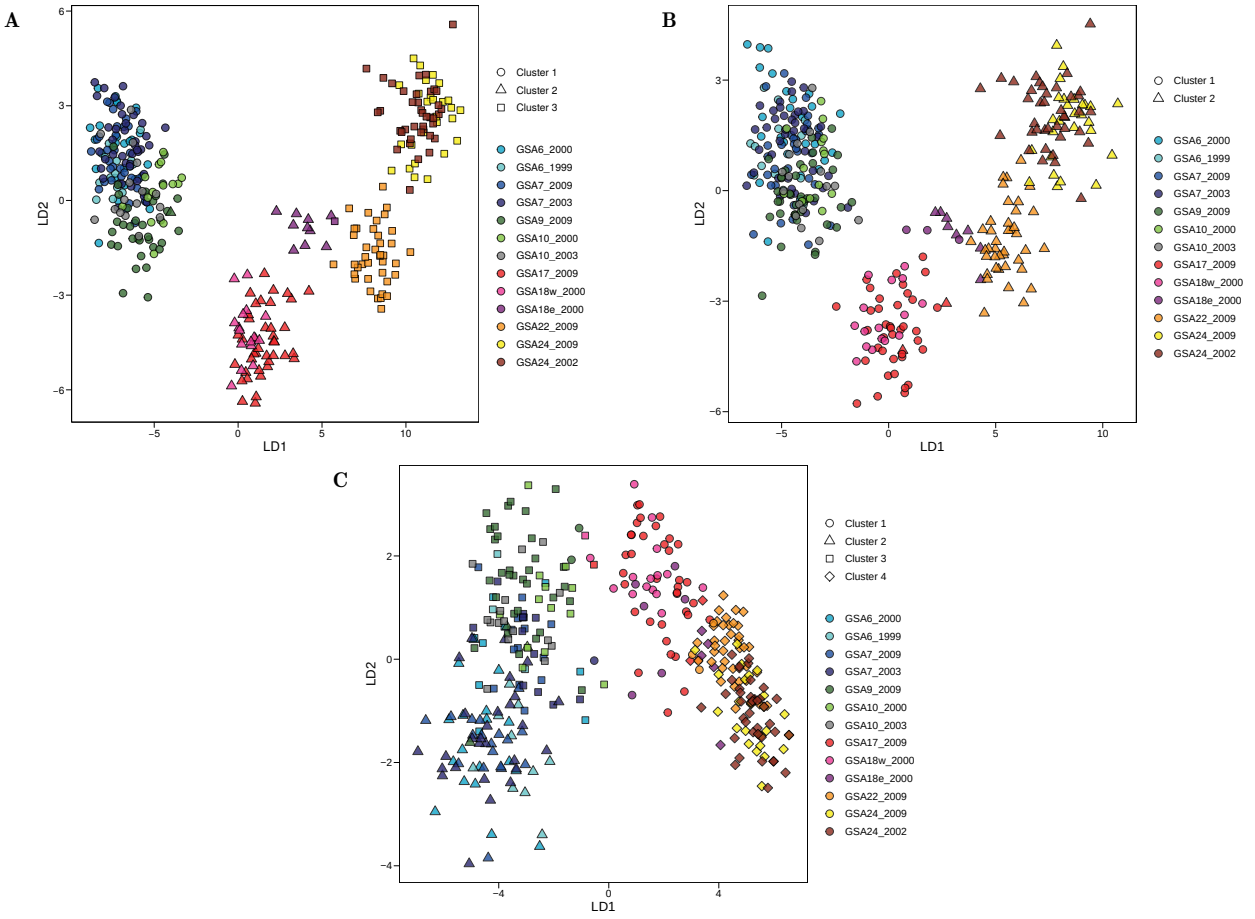

Figure 2.3: Discriminant Analysis of Principal Components (DAPC) plots of the 13 population samples of *Solea solea* for whole (A, 380 SNP loci), neutral (B, 337 SNP loci) and outlier (C, 43 SNP loci) dataset. Geographical population samples are colour coded according to sampling site, with Site ID reported as in Table 2.1. The subdivision of individuals according to identified clusters is shown by assigning to data points different geometric shapes.

the sPCA to assess the contribution of alleles to both neutral and outlier structure. We found in total that alleles of 19 SNPs contribute most to the first axis of the sPCA (Suppl. Table 2.6).

2.3.2 Otolith analysis

A total of eight individuals were detected as outlier data points and thus removed from shape index dataset. Roundness, ellipticity and rectangularity were normally distributed ($p < 0.01$), thus retained for further analysis. Levene's test for homogeneity of variance was significant ($p<2.2 \times 10^{-16}$, $\alpha = 0.01$). We found a significant negative correlation between roundness and ellipticity ($r = -0.802, p < 0.001$). Nonetheless, roundness, ellipticity and rectangularity were included in subsequent analysis because they were not correlated with more than one other index ($p < 0.001$). None of the shape indices was correlated with total length of fish, and only ellipticity showed a significant association ($p = 0.008$). There was no significant interaction between total fish length and sampling sites for $\alpha = 0.01$, and thus the slope across sampling sites was equal. To remove the effect of total length on ellipticity, we corrected the values by division with the ellipticity/fish length slope. The *adonis* test in the PERMANOVA analysis found that site was a significant explanatory variable at $\alpha = 0.01$. All pairwise site comparisons were significantly different ($p < 0.05$), as reported in Supplementary Table 2.7, except for GSA9_2009-vs-GSA17_2009, GSA17_2009-vs-GSA22_2009 and GSA17_2009-vs-GSA24_2009. The result of dispersion test was not significant ($p = 0.159, \alpha = 0.01$) validating the *adonis* test results. CAP analysis highlighted a separation of samples from the Gulf of Lion with 75% correct reclassification compared to lower values for the other locations (Figure 2.4A and Suppl. Table 2.8).

The otolith shape was summarized by 13 harmonics (thus 52 coefficients) as they reached 99% of the cumulative Fourier power (Suppl. Figure 2.14). All pairwise comparisons among sampling sites resulted in significant differences in the PERMANOVA analysis ($p < 0.05$) (Suppl. Table 2.7). In contrast, the dispersion test resulted significant with a p-value of 0.001 ($\alpha = 0.01$) making the *adonis* test results less reliable. The CAP on Fourier descriptors using 13 harmonics identified Aegean and Turkish samples as separate groups (Figure 2.4B). In both cases Aegean samples (95.74%) showed a higher reclassification success compared to Turkish samples (68.97%) (Suppl. Table 2.8). The CAP on integrated morphometric data identified the samples from Aegean Sea and Gulf of Lion as separate groups (Figure 2.4C). Integrating the two data sets provided higher overall reclassification success, with the Aegean samples showing the highest value of correct reclassification (Suppl. Table 2.8).

From the analysis of elemental data, only Zn ($p = 0.02$) fulfilled the assumption of normality and homogeneity ($p < 0.01$) for both core and edge data. ANCOVA analysis suggested the influence of otolith weight only on elemental concentration of Rb for $\alpha = 0.01$ for core data, while for edge data the influence of otolith weight on elemental concentrations of Ba, Rb and Mg was significant ($\alpha = 0.01$). None of the elements at either otolith core and edge were correlated with otolith weight ($\alpha = 0.01$), based on Spearmans correlation. Significant differences ($\alpha = 0.05$) in the elemental concentration were found between sites for all the elements of both core and edge data (Figure 2.5). The sampling sites were weakly identified by DAPC methods for

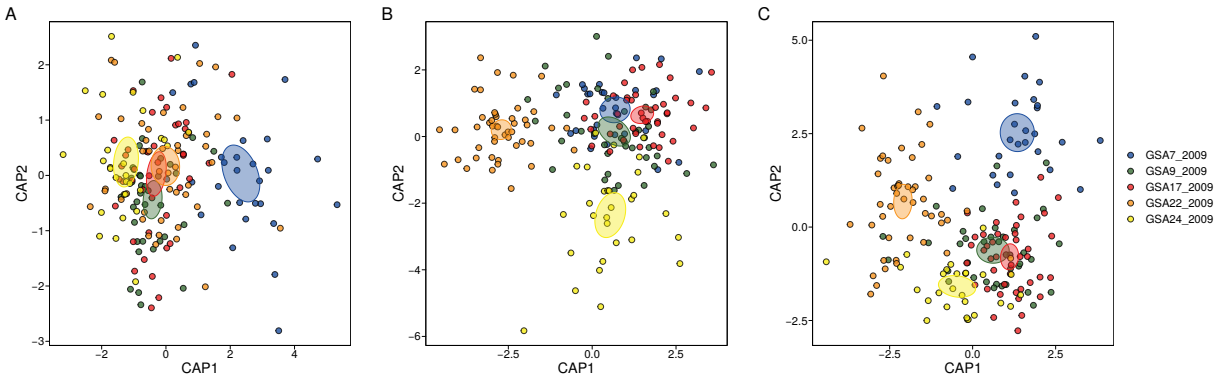

Figure 2.4: CAP (Canonical analysis of principal coordinates) plots for otolith shape indices (SIs) (A), elliptic Fourier descriptors (EFDs) (B), and integrated morphometric data (SIs and EFDs) (C) from five Mediterranean population samples of *Solea solea* (GSA7_2009, GSA_2009, GSA17_2009, GSA22_2009 and GSA24_2009.

core data, except for GSA22_2009, while the correct assignment of individual in sampling sites was higher for edge data (Suppl. Figure 2.15). The elements that most contributed to the discrimination of geographical sites are showed in Supplementary Figure 2.16.

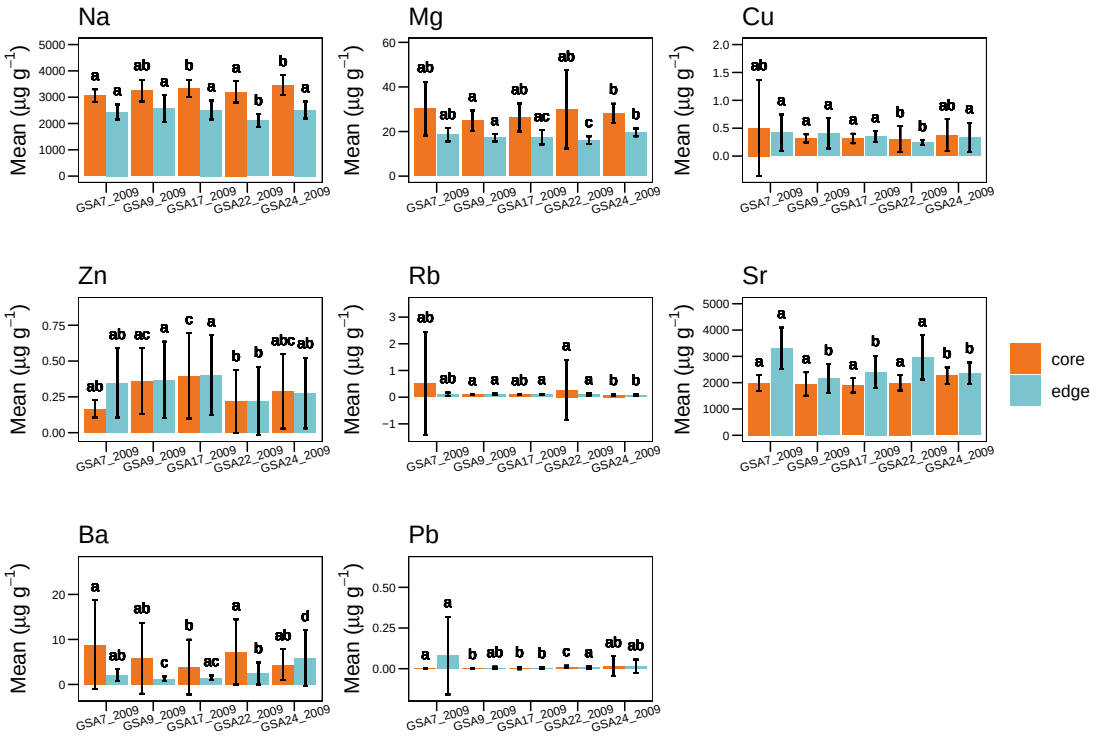

Figure 2.5: Mean±SD elemental concentration of each element in otolith core and edge per sampling sites (GSA7_2009, GSA9_2009, GSA17_2009, GSA22_2009 and GSA24_2009). Significant differences among sampling sites obtained from Dunn's test are shown in the plot. The bars with different lowercase letters corresponding to significant differences ($\alpha = 0.05$) between sample sites.

2.3.3 Genetic-environmental association analysis

We identified 13, 55 and 106 SNPs from *Bayenv2*, *LFMM* and *Samβada* respectively, which showed at least one significant association with an environmental variable. The intersection of the SNPs identified among the three GEA analysis (*Bayenv2*, *LFMM* and *Samβada*) resulted in six loci (Suppl. Figure 2.18, Suppl. Table 2.5) and the union of all SNPs identified by GEA methods resulted in 148 SNPs (Suppl. Table 2.5). The regression analysis of locality scores from the sPCA performed on the GEA union dataset of 148 SNPs returned only dbMEM1, dbMEM3 and dbMEM5 as significant spatial variables, while dbMEM2 and dbMEM4 were not ($p > 0.05$) (Table 2.3). All the environmental variables were significantly associated with the SNPs, sal_bottom_mean was the best predictor among environmental variables, while curM_Wclm_mean, temp_bottom_range and pco_Wclm_mean were less strong, but still significant, predictors. Longitude was the best predictor of locality scores and latitude was also a good predictor (Table 2.3).

Variables	R^2_{adj}	p-values	Significance
dbMEM1	0.768	$< 2 \times 10^{-16}$	*
dbMEM2	-3.71×10^{-5}	0.321	
dbMEM3	0.119	1.58×10^{-11}	*
dbMEM4	0.001	0.253	
dbMEM5	0.047	2.45×10^{-5}	*
sal_bottom_mean	0.628	$< 2 \times 10^{-16}$	*
curM_Wclm_mean	0.088	8.87×10^{-9}	*
temp_bottom_range	0.172	2.89×10^{-16}	*
pco_Wclm_mean	0.046	2.75×10^{-5}	*
longitude	0.897	$< 2 \times 10^{-16}$	*
latitude	0.354	$< 2 \times 10^{-16}$	*

Significant variables are indicated with the following symbol:

* $p < 0.05$

Table 2.3: Regression analysis of locality scores from the sPCA on the 148 SNPs detected by GAE methods across the Mediterranean population samples of *Solea solea*.

2.3.4 Redundancy analysis

The global RDA model performed on both outlier and neutral SNPs identified significant patterns of variation ($p < 0.05$). The percentage of genetic variation explained was 38.13% for the outlier dataset of loci and 9.97% for the neutral ones, with the first two axes accounting respectively for 35.61% and 7.82% of the explained

variation. The forward and backward procedure using *ordistep* function selected different spatial variables for the two datasets of SNPs, while temp_bottom_range, sal_Wclm_mean and pco_Wclm_mean were identified for both RDA on neutral and outlier SNPs (Figure 2.6, Suppl. Table 2.9). From marginal ANOVAs, spatial and environmental variables selected were all significant predictors of neutral and outlier genetic variation ($p < 0.05$). The results of variation partitioning based on neutral SNPs indicated that spatial variables (1.9%) explained approximately the same individual fraction of variance as the environmental variables (1.5%). Shared variance (6.3%) resulted in the highest amount of explained variance (Suppl. Figure 2.20). In contrast, the spatial variables explained more of the individual fraction of variance in the outlier SNPs (7.0%), whereas environmental variables explained 5.5%. As for neutral SNPs, the highest amount of explained variance was represented by shared variance (25.6%) (Suppl. Figure 2.21). Partial RDA performed with environmental variables as condition, confirmed that most of the genetic variation both for neutral and outlier structure was explained by spatial variables (Suppl. Figure 2.22, 2.23, Suppl. Table 2.9).

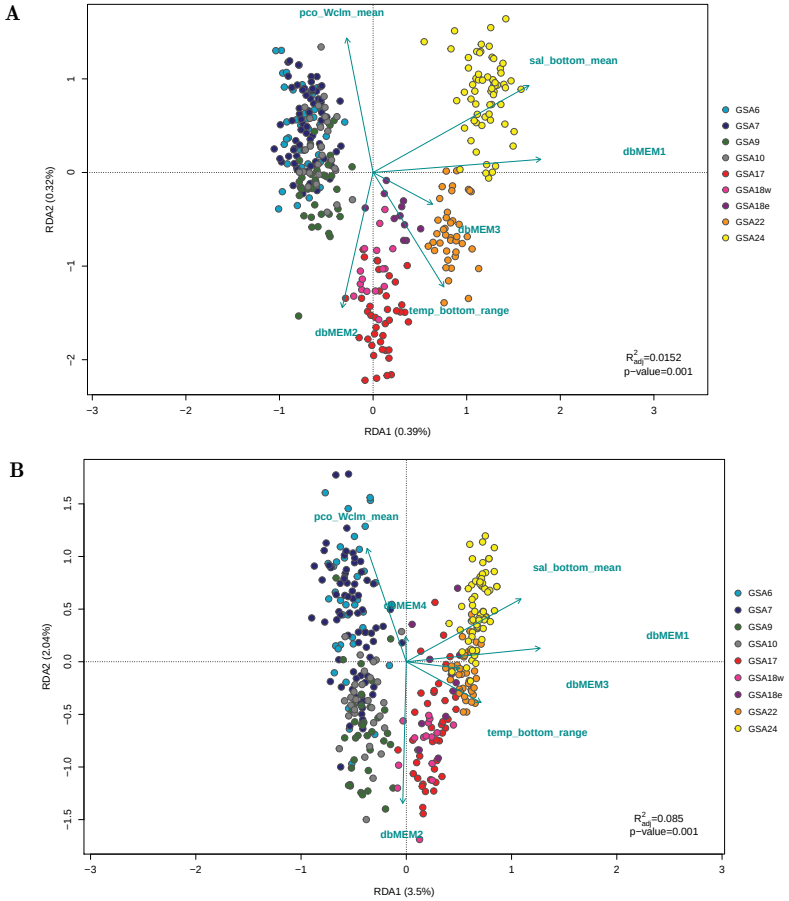

Figure 2.6: Redundancy analysis (RDA) performed on neutral (A) and outlier (B) SNPs datasets from population samples of *Solea solea* following the forward and backward selection of variables using *ordistep* function. Spatial (dbMEMs) and environmental variables found as significant predictors are shown with arrows.

2.3.5 Gene Ontology

Based on the outcomes of transcript annotation and BLAST search, association with known genes was retrieved for 360 out of 380 loci. In details, 36 out of 43 outlier loci and 132 out of 148 loci associated with environmental variables were related with functional annotation. A total of 20 annotated markers were in common between the two datasets. Enrichment analysis revealed that the vast majority of significant results was due to GO terms connected to the outlier dataset, while only a negligible number of significantly enriched terms was ascribed to the environmental dataset. GO terms were significantly over represented in all three aspects of gene ontology (Suppl. Figure 2.24, 2.25 and 2.26). Enrichment analysis highlighted 93 biological processes, and the top three predominantly enriched GO terms were regulation of nucleobase-containing compound metabolic process, cell migration involved in gastrulation and embryo development. For the 19 molecular functions, the top three predominantly enriched GO terms were nucleic acid binding, RNA and DNA binding. Finally, 12 enriched GO terms were found for the cellular components. When retrieving SNPs associated to the enriched GO terms, a total of 23 loci were associated with GO terms significantly over represented, and belong to 20 successfully annotated genes (listed in Suppl. Table 2.10).

2.4 Discussion

Evidence for significant population structure

Understanding patterns of population structure may contribute to marine fisheries conservation because it can identify existing misalignment of current management units with the biological units (Kerr et al., 2017). Through this study we examined population structure of common sole in the Mediterranean Sea while also investigating the efficacy of different tracers providing complementary information and identifying factors contributing to genetic differentiation. A multidisciplinary approach provides the opportunity to reveal fine population structure often missed with single markers. In particular, molecular approaches were used to identify structure at the microevolutionary time scale, while otolith shape and elemental composition analyses for a subset of sampling sites described the structure at a life span scale (Tanner et al., 2016). The current work confirmed the presence of a high level of differentiation in common sole populations across the Mediterranean area.

A strong signal of a well-defined spatial structure was found inside the Mediterranean basin at different time scales. Genetic analysis revealed substantial differences and clustered Northern Mediterranean population samples of *Solea solea* into three main groups occupying three main areas (Western Mediterranean, Adriatic, Eastern Mediterranean), with higher levels of heterogeneity in the western area compared to the eastern one. In addition, genetic results confirmed a further differentiation pattern within the Adriatic Sea,

with Northern Adriatic and South-Western Adriatic population samples (GSA17_2009 and GSA18w_2000) separating from South-Eastern Adriatic (GSA18e_2000). This pattern has been already suggested in previous studies based only on mitochondrial markers (Sabatini et al., 2018; Guarnieo et al., 2002). However, stock assessment has historically been carried out using only the GSA17 (Northern Adriatic Sea) as management unit since the landings of common sole in the western part of the GSA18 are negligible (Sabatini et al., 2018). From the analysis on genetic diversity, common sole samples from Eastern Mediterranean showed the lowest values of observed heterozygosity compared to those from other locations. A decreasing polymorphic rate could be indicative of demographic processes as bottleneck events, potentially linked to high mortality, which causes may deserve further studies (Pinsky and Palumbi, 2014). As suggested with other species such as European hake (*Merluccius merluccius*) (Morales-Nin et al., 2014), the genetic differences observed between population samples of *Solea solea* into the three main areas in the Mediterranean basin are probably led by neutral drift, because of limited connectivity likely due to restricted larval dispersal across such a wide spatial scale, surface water-masses circulation in the Mediterranean with several cyclonic gyres (El-Geziry and Bryden, 2010) as well as by the life-cycle traits of common sole, with early developmental stages spending several months in lagunar and estuarine habitat (Primo et al., 2013) and biogeographical barriers. For instance, the Strait of Sicily has been considered a geographical boundary separating the population structure across Eastern and Western Mediterranean at phylogenetic levels (Patarnello et al., 2007). The pattern of population structure was confirmed by all methods applied. In particular, DAPC analysis separated western from eastern population samples along the first discriminant axis and samples from Aegean and Levantine areas along the second discriminant function axis. The use of different approaches for population differentiation helps to address both demographic and adaptive processes, indicated by variation at the neutral and outlier SNP loci, respectively. The presence of three main clusters was in fact revealed by using a complete genetic dataset as well as a selected dataset of only neutral markers. Genetic differentiation at a finer spatial scale was highlighted by the analysis conducted on a separate dataset of outlier SNPs. Consistency between neutral and outlier SNP loci signal was well supported, with a greater resolution of outlier SNP loci in revealing fine scale segregation patterns compared to neutral markers. In details, results based on 43 outlier loci discriminated the common sole samples from the Tyrrhenian Sea (GSA9 and 10) from the other Western Mediterranean samples. These insights of subtle genetic clustering based on outlier markers, putatively under selection, in the Tyrrhenian area suggest that local selection could play detectable role in shaping the genetic structure of common sole. The results obtained from a recent study conducted by (Diopere et al., 2018) on this demersal flatfish in the North-Eastern Atlantic Ocean, using the same panel of genomic markers, is in line with our results, indicating that the genetic structure of common sole is shaped by a combination of neutral and adaptive processes. It appears that neutral genetic structure is due to dispersal limitation of this species throughout its distribution, while adaptive processes such as environmental and ecological features play as key factors in local adaptation. Furthermore, outlier detection demonstrated

to be a useful tool in revealing fine population structure in areas in which connectivity and gene flow are limited by geographical boundaries among others factors as in our study, as well as in Atlantic area where connectivity is more favored (Diopere et al., 2018).

Environmental effects on genetic variation

Another goal of the study was to investigate the influence of environmental variables on the population structure within a seascape analysis framework. Studies have shown that the use of constrained ordination methods such as RDA are effective tools in detecting the genetic basis of local adaptation (Forester et al., 2018). Seascape analysis confirmed the distinction between common sole population samples from the Western and Eastern Mediterranean areas, with the longitude as the best predictor. Both neutral and outlier SNPs datasets identified clusters, however a larger proportion of the genetic variation was explained by clusters in the outlier loci dataset, confirming the higher power of resolution of outlier markers than neutral ones to detect fine scale population structure. In this respect, RDA models, which included spatial distribution and environmental factors, explained 38.13% of the outlier genetic structure while only 9.97% of the neutral one. The majority of variation remains unexplained in RDA models probably because it can be explained by additional factors. In fact, neutral genetic variation could be due to random processes such as genetic drift or mutation, but also other factors can contribute to demographic isolation (Gagnaire et al., 2015).

The analysis of environmental data and their association with genetic variation suggested that the population structure of Mediterranean common sole is likely to be partially due to environmental factors such as temperature or salinity and to spatial distribution. As a result of high level of unexplained variance and up to 25.6% of shared variance explained by the two types of variables, the environmental ones exhibit a spatial structure, that it is also captured by the spatial variables. This means that both spatial and environmental factors might significantly impact genetic structure. Especially for the outlier dataset, the residual influence of spatial variables was predominant compared to environmental ones explaining 7% of variance meaning that other spatial structure related to other environmental variables, geomorphological or connectivity features act in shaping genetic differentiation among population samples at multiple scales. This result seems to be consistent with the life-cycle traits of this demersal flatfish species. On the other side, environmental variables also explained 5.5% of variance in the outlier dataset not explained by any spatial structure, even though there are significant salinity and temperature gradients across the Mediterranean Sea (Skliris, 2014; Spedicato et al., 2022). Particularly, several outlier SNPs were also associated with environmental variables in the GEA analysis, suggesting the potential involvement of these loci in local adaptation.

Concordance between otolith and genetic variation

Regarding shape analysis of the otolith outlines, the results showed statistically significant differences among the five population samples, hence otolith shape analysis is confirmed a valuable discriminating method in marine fishes. The analysis performed on shape indices showed significantly different dimensions for GSA7_2009 compared to the other samples, with a higher correct assignment to the original locations. This separation could suggest low levels of variability within samples from Gulf of Lion (GSA7_2009) and higher differentiation between GSA7_2009 samples and samples from the other locations. Individual variability might reduce the power of the classification method, and may reflect local environmental conditions (Morales-Nin et al., 2022). In addition, even with correction for fish length effects, some otolith dimensions vary with age and growth conditions which could contribute to variability within areas (Burke et al., 2008). On the contrary, elliptical Fourier descriptors (EFDs) demonstrated to be more successful than shape indices in discriminating among individuals from different sampling locations. The results obtained from analysis of EFDs clearly separated Eastern Mediterranean population samples from the other sampling locations with a h percentage of allocation to the original locations. Finally, overall reclassification success was higher when morphometric data were integrated, meaning that the combination of shape indices and EFDs improved the power of discrimination between different groups.

Although at a lower resolution because of the reduced sampling extent compared to genetic data, integrated morphometric data were effective in confirming significant divergence between Western and Eastern Mediterranean population samples separated from the rest of the common sole population samples, a pattern already suggested from genetics analysis.

Many factors can influence the trace element concentrations in otoliths, including environmental factors such as temperature, salinity and chemical composition of water masses, fish physiology and ontogenetic change in habitat occupation (Morales-Nin et al., 2022). In relation to the otolith trace element composition, the correct population assignments depends on the otolith material deposited in a localised geographical area. The otolith core forms during embryo development and early life history and can represent the pelagic life stage, including the spawning area and possibly movement to nursery areas. The otolith edge represents the most recent resident area of that individual. In our case the discrimination methods detected stronger signals in the core elemental composition data, with population samples from the Aegean (GSA22_2009) more clearly differentiated. Whether this is limited by the number of samples in each population, or a reflection of the chemistry of the water masses, is difficult to say (Vasconcelos et al., 2007). In fact, there is an improvement in the discrimination power of core data compared to edge data, but not with perfect correspondence because of fish movement after their early life stages. Multiple factors might be limiting the power of discrimination methods, such as the low number of elements used in the analysis, low residence time in the location due to habitat shift in such a complex life cycle as the one of common sole. The core and

edge mean concentrations measured in the single-element analysis (Figure 2.5) were in line with the values assessed by other studies (Chang and Geffen, 2013; Cuveliers et al., 2010). The presence of Ba as one of the microelements that most contributes to the discrimination between groups in the core data is consistent with the fact that Ba is known as a proxy of freshwater influence occurring in nursery area (Suppl. Figure 2.16) (Morat et al., 2014), as well as areas of high productivity (Chang and Geffen, 2013). Among other elements, Cu, Zn and Pb seem to be the microelements that most contribute to the discrimination between groups both in core and edge data (Suppl. Figure 2.16). Although waters masses near coast and shelf are naturally enriched with heavy metals, high concentrations of Cu, Zn and Pb can indicate industrial pollution in that area, originating from agriculture activities (Cu, Zn) and from extensive mining for the latter (Pb) (Chang and Geffen, 2013; Constenla et al., 2021; Geffen et al., 2003). As reported in (Constenla et al., 2021), common sole could be used also as a sentinel species for possible changes in marine water due to environmental pollution directly affecting marine resources. Otolith concentration of Sr is sensitive to both salinity and water mass composition. This may explain why GSA22_2009 was so clearly differentiated from the other sampling sites based on the otolith core data. The salinity regime of the Eastern Mediterranean basin is substantially higher compared to those existing in the other parts of the Mediterranean (Suppl. Figure 2.17) (Gačić et al., 2011). Moreover, this pattern is confirmed by RDA analysis in which, among environmental variables, salinity significantly contributed to the genetic differentiation of Eastern Mediterranean common sole population samples, although at a lower degree with respect to spatial variables. Also, among SNPs loci associated with environmental factors, salinity resulted as the best predictor among environmental variables with respect to the regression analysis on the loading score in sPCA analysis.

Looking for candidate adaptive genes

In order to validate putative SNPs under selection, research studies aim to link outlier markers to functional genes that can be involved in adaptive processes, by associating the function of known genes near to the identified locus to the variation of key environmental parameters. Several studies have successfully identified as outlier SNPs and SNPs associated with environmental variables found in genes involved in adaptation processes, using various methods of genome scan analysis (Benestan et al., 2016; Bernatchez et al., 2019; Selmoni et al., 2020). Different methods to detect outlier markers and loci associated to environmental factors, such as outlier detection methods and genetic-environment association (GEA), could be combined to obtain more information and enhance the possibility to identify loci potentially under selection (de Villemereuil et al., 2014; Rellstab et al., 2015). The use of additional information such as environmental variables improves the power of GEA analysis in identifying loci, compared to outlier detection analysis, and reveals signals of selection. As suggested by Forester et al. (2018), the overlapping approach which combines results from different methods can be effective in detecting adaptation while maintaining a low false-positive rate.

As a result, both list of putative loci under selection (43 SNPs) and/or influenced by environmental factors (148 SNPs) were used to perform Gene Ontology (GO) enrichment analysis. Our aim in performing the GO enrichment tests was to explore whether the functions of known genes were under or over-represented in sets of genes associated with both environmental variables or outlier SNPs. Most of the loci highlighted in the enrichment analysis, which are mainly loci also included in the environmental dataset, were associated with different gene product functions and all of them were significantly over represented in different aspects of gene ontology. However, the successfully annotated genes connected with these loci in most cases did not directly correspond to gene expression regulation processes generally connected with our environmental variables. Nevertheless, these loci could have an impact on overall gene expression because they are connected with genes involved in RNA and DNA binding, molecular interaction and metabolic processes among others. The exception were a few genes that are involved in muscle contraction and encoded proteins that support mitochondrial metabolic demand that were expressed in response to high pCO2 (Cline et al., 2020). Outlier SNPs 1023 and 464 that are located into these annotated genes are associated with pCO_2 environmental variables (pco_Wclm_mean) in at least one GEA method and they are also included in the loci that most contribute to the pattern identified by sPCA. Overall, based on our results, further work including more complete genomic information is necessary to expose the influence of environmental features on the microevolutionary process shaping the population structure of this species, with special emphasis on the assessment of putative functional local adaptation processes, which ultimately could lead to differences in life history traits.

Conclusion and future impacts

This study has produced a major advance in the knowledge-based management of common sole in the Mediterranean Sea, defining and assessing the population structure of this relevant demersal resource with multidisciplinary analyses. In our study, the identified west-to-east differentiation pattern observed is fully coherent with the population structure described by Rolland et al. (2007). Using whole and neutral datasets, we identified three areas (Western Mediterranean, Adriatic, Eastern Mediterranean) that largely approximate to the FAO division of the Mediterranean Sea in the Major Fishing Areas (MFAs) 37.1, 37.2 and 37.3. In addition, the outlier dataset revealed a more subtle division within the western area (37.1). A very large portion of MFA 37.2 is the Ionian sea (including the Libic coast), and our sampling did not cover the whole Mediterranean area. Even if more samples from the Southern and insular Mediterranean Sea are needed to achieve a complete pattern of common sole population units in the Basin, our findings could provide a new perspective for management decisions, taking into account the current stock assessment scenario. Following a multidisciplinary approach, we also tried to elucidate the selection forces that influence the observed pattern of population structure. The identified genetic clusters agreed with the subdivision in MFA areas, with

population samples reassigned with high success rate by DAPC analysis on genomics data as well as CAP analysis on otolith data. Moreover, the power of outlier loci in detecting finer population structure helps to reveal metapopulations, and it should be considered in the assessment and management of fisheries plans. Both genomics and otolith data provided adequate discrimination power to delineate population structure in the present study, but the temporal and spatial scales of our sampling design could be a limitation for the identification of proper stock units for management purposes. The multidisciplinary approach in our work is perfectly in line with the GFCM 2030 Strategy to focus on subregional level processes (FAO, 2021). However, our study relied only on samples from the European areas, that are not fully representative of the species distribution in the basin. The collection of samples from geographical sites in the southern part (i.e., the African coast and Sicily channel) would enhance the usefulness of the findings and the power to discriminate substructure at a finer geographical scale within the three main areas. In addition, comparison with recent samples would allow an evaluation of the genetic diversity of the species following two decades of management measures that greatly reduced the fishing pressure on this stock (FAO-GFCM, 2021). In conclusion, our multidisciplinary approach was reliable in revealing the population structure of common sole, and providing new insights into the importance of local adaptations that produce finer scale structure in this highly valuable marine resource. These findings represent a valuable asset to implement an effective integrated approach combining comprehensive quantitative results in order to delineate proper stock management units for this species.

2.5 Supplementary materials

2.5.1 Methods

2.5.1.1 Genetic analysis

Assessment of missing data and data filtering

Preparation and processing of SNP data can produce misleading artefacts which can generate erroneous patterns or considered hint of population structure and adaptive processes (O'Leary et al., 2018). In this scenario, O'Leary et al. (2018) stresses the importance of creating a rigorous and documented filtering process to minimize or remove errors. In particular, filtering SNP data is crucial to remove genotyping errors and prevent wrong interpretations of standing-out signals (Hendricks et al., 2018; Meirmans, 2015). However, the filtering procedure is contingent on the analysis and type of data, and the development of a standardized method is not always feasible.

In our study, we applied several filtering methods to understand how to handle missing data. Specially, we tried to respond to the following questions. How different amounts of missing data affect the data?

Which filtering approach is more suitable to obtain reliable data avoiding loss of information? Removing all variables or samples with missing data, imputing missing data, filtering for different thresholds of missing data or combining the last two methods are some ways to account for missing data. Among those, filtering for different values of acceptable missing data seems the more flexible method, despite the arbitrary choice of filtering thresholds.

In our study, the percentage of missing data in the original data set was 7.56%. Among the geographical samples, $GSA10\_2000$, $GSA10\_2003$ and $GSA18e\_2000$ were particularly affected by missing data. We assessed the amount of missing data at two different levels: individuals and SNPs. We applied two main filtering methods defined as fixed and sequential filtering approaches. Through the fixed filtering approach, we applied different coupled thresholds of allowed missing data per loci and individuals on the original data set. Whereas, the sequential methods used the same thresholds having as input of each step the data set obtained from the previous filtering stage. We tried also this second method due to the high level of missing data in the data set to minimize the effect of tight thresholds, and retain as many loci and individuals as possible. For this same reason we started filtering from loci since samples sizes were uneven and the three above-mentioned populations were particularly affected by missing genotypes. The thresholds were chosen based on previous studies (Diopere et al., 2018; Nielsen et al., 2012) and trials. The thresholds employed were the following:

- for loci: 20, 18, 16, 14, 12 and 10%
- for individuals: 40, 36, 32, 28, 24 and 20%

The final data set should represent the real population genetic differentiation, preserving a reasonable number of individuals for each of the 13 geographical sampling sites without being affected by the filtering approach. Since most of the metrics used in populations genetics (i.e., F_{ST}) are contingent on the allele frequency, we decided to use MAF (minor allele frequency) as a metric to choose the final data set (Berner, 2019). For this purpose, we investigated MAF distribution of different filtering steps for the entire data set (populations pooled) and for each separate geographical sample. We used Kolmogorov-Smirnov test to determine statistically significant differences in MAF distributions at each filtering step with respect to the original distribution ($\alpha = 0.05$). Moreover, we inspected the data set without the three populations more affected by missing data. We assessed the results of the filtering steps for this data set through Principal Component Analysis (PCA). We then excluded monomorphic SNPs from the final data set prior to the analysis.

Heterozygosity, F_{IS} and Hardy-Weinberg equilibrium

We used *Genepop* program (R version 4.7.5) to compute the observed and expected heterozygosity, estimate F_{IS} and test for Hardy-Weinberg equilibrium. We used *genedivFis* function to calculate the observed and

expected heterozygosity and F_{IS} values per sample. Then, we obtained a global estimate of F_{IS} (argument `pairs=FALSE`) with *Fst* function. We used *test_HW* function to test three alternative hypotheses: probability test (exact test), heterozygote deficiency and excess (score test). We used the complete enumeration method since the individuals were less than 1000 and the distinct alleles less than 5 in the data set. Then, we used the same function to test each sample (multi-locus versions). Finally, the function also returned a global result for all SNPs and all samples.

F-statistics (F_{ST})

As input file we used the *genind* object, which we transformed a *gtype* object by using *genind2gtypes* function. We used the *overallTest* function to calculate global F_{ST} and the *pairwiseTest* function to calculate the pairwise F_{ST}. For both functions we set `stats = "fst"` and `nrep = 10,000`, and we then applied Benjamini and Yekutieli p-values correction (Benjamini and Yekutieli, 2001). We considered significant pairwise comparisons with p-value lower than $\alpha = 0.05$.

Discriminant Analysis of Principal Components (DAPC)

Discriminant Analysis of Principal Components (DAPC) is a method to identify genetic clusters and describe the relationships between them. This method relies on the computation of new variables as linear combinations of the alleles. These new variables, referred to as discriminant functions, reflect the genetic variation among individuals and groups of individuals minimising the variation within clusters. However, DAPC needs the definition of prior groups, and this requirement is often difficult to meet. Alternatively, the method allows to infer the groups by clustering algorithm such as *k*-means. This multivariate approach combines PCA, *k*-means clustering and Discriminant Analysis (DA) to group individuals into a pre-determined number of clusters that better describes the data. First, data are transformed by PCA to reduce the number of variables, then *k*-means clustering with increasing number of K is performed. Finally, relationships between the different clusters are described by DA. This latter method leads to the identification of discriminant functions, which are variables that optimize the variance between groups while minimizing the variance within them. Moreover, it allows to identify probability memberships of individuals to each group, that can be interpreted as proximity to the clusters.

The first step relies on the *find.clusters* function. This function performs *k*-means clustering algorithm over a range of *k* to identify clusters while maximising the variation between them. We ran *find.clusters* function on the *genind* object for *k* from 1 to 20, setting 25 starting points for each run (`n.start`) and 10 iterations for each *k* (`n.iter`). First, the function transforms the data using PCA, at this step we retained 400 PCs in the iteractive mode looking at the cumulative distribution of variance. Indeed, *k*-means algorithm benefits from retaining all the variance, thus reducing the number of PCs at this step is only useful in terms

of computation. Then, we used the Bayesian Information Criterion (BIC) to compare clustering solutions obtained. Secondly, *dapc* function was run to provide transformation of the data by PCA prior to the discriminant analysis (DA). Given the optimal number of clusters identified by *k*-means, we ran the *dapc function* for the range of k from 2 to 6. We retained 170 PCs and two discriminant functions since the k values were lower than 10.

In the second step we performed the DAPC analysis with prior knowledge of the subgroups. Using prior populations, it is possible to assess the quality of the discrimination by looking at the re-assignment of each individual to its group of origin Figure (2.9). We used the 13 geographical sampling locations as `grp` argument in the *dapc* function. Moreover, when prior groups are defined you can use the α-score criterion or the cross-validation procedure to find the best number of PCs to retain. We ran the cross-validation method starting with a low number of replicates (`n.rep=30`). From those results, we chose a smaller range of PCs and we repeated the cross-validation with higher repetitions (`n.rep=1,000`). The number of PCs with the lowest error was 125. So, we ran the DAPC with this optimal number of PCs and, having a number of k greater than 10, we retained 5 discriminant functions. Here we performed a DAPC analysis with prior knowledge of the population samples on all datasets integrating the clusters information of the probability memberships of individuals from find.cluster to the results obtained with prior knowledge in one graph.

2.5.1.2 Otolith analysis

For otolith analysis we used the data produced by FishPopTrace project (`https://fishpoptrace.jrc.ec.europa.eu/`) as described below. Both the right and left otolith of the same individuals were extracted, cleaned, and photographed in the project. In our study only images of left otoliths were used for morphometrics analysis. For otolith shape analysis we used ImageJ-Fiji (`https://imagej.net/software/fiji/`) to automatically detect the otolith outline and calculate the shape indices (SIs) based on otolith morphometric parameters (area, perimeter, length, and width). The images were convert to silhouettes and EFDs extracted using the program SHAPE (Iwata and Ukai, 2002) and the *Momocs* R package (Bonhomme et al., 2014). All the analysis on Elliptical Fourier descriptors were performed using *Momocs* R package (Bonhomme et al., 2014). We used *import_jpg* function to import otolith outlines from otolith images. Starting from closed outlines, we then created the *Out* object with *Out* function. This object is a list of closed outlines, for each outline there are a list matrices of coordinates and a factor specifying the sampling sites. After outlines inspection, we decided to normalize the outlines aligning the shapes using as control points two landmarks chosen on each shapes (*def_ldk* function). The shapes were then centered on the origin with *coo_center* function, scaled by the centroid size with *coo_scale* function and aligned with a full generalized Procruster alignment using *fgProcruster* function.

2.5.1.3 Environmental analysis

Monthly mean value of physical and biochemistry environmental variables were downloaded at a resolution of 12 km^2 from E.U. Copernicus Marine Service Information (`https://marine.copernicus.eu/`). We selected the following variables considering their availability over the analysis period from 2000 to 2010. Meridional component of the currents (northward) (curM, m/s), zonal component of the currents (eastward) (curZ, m/s), salinity (sal, psu) and potential temperature (temp, °C) data were extracted from Mediterranean Sea Physical Reanalysis (`https://doi.org/10.25423/MEDSEA_REANALYSIS_PHYS_006_004`) on a grid with 1/16° x 1/16° horizontal resolution (approximatively 6 km) and 72 vertical levels of thickness. Chlorophyll (chl, milligram m^{-3}), dissolved oxygen (dox, millimol m^{-3}), nitrate (NO_3) (nit, millimol m^{-3}), CO_2 (ocean pCO_2 expresses as carbon dioxide partial pressure) (pco, Pa), pH (ph, PH), phosphate (PO_4) (pho, millimol m^{-3}), phytoplankton carbon biomass (phyto, mol m^{-3}), net primary production (tendency of mole concentration of particulate organic matter expressed as carbon in sea water due to net primary production) (ppn, mol m^{-3} s^{-1}) were extracted from Mediterranean Sea Biogeochemical Reanalysis (`https://doi.org/10.25423/MEDSEA_REANALYSIS_BIO_006_008`) on a grid with 1/16° x 1/16° horizontal resolution (approximatively 6 km) and 72 vertical levels of thickness. For those variables in each macro-area, we calculated average, range, minimum and maximum values according to the variables features. In order to account for differences between bottom and surface, we selected both bottom and the weighted average value in the water column (Wclm). Also, daily mean values of turbidity (Diffuse Attenuation Coefficient at 490nm (Kd_490)) (turb, m^{-1}) were downloaded from E.U. Copernicus Marine Service Information (https://doi.org/10.48670/moi-00299) at a resolution of 1 km_2. In addition, monthly mean values of total precipitation (mm) as proxy of freshwater input were downloaded from WorldClim 2.1 (`https://worldclim.org/data/monthlywth.html`) at a resolution of 441 km_2. We selected 17 environmental variables reported in Table 2.4 based on their influence in common sole distribution.

From the *pcnm* function in *vegan* R package we obtained 5 dbMEMs eigenvectors (Oksanen et al., 2022). According to the scales of the pattern showed by dbMEMs represented in Figure 2.19, dbMEM1 and dbMEM2 were arbitrarily defined as broad scale variables, and dbMEM3, dbMEM4 and dbMEM5 as fine-scale variable. Visual inspection of dbMEM1 suggested a broad scale west-to-east gradient. These eigenvectors were used as spatial variables in the RDA analysis.

2.5.2 Results

2.5.2.1 Genetic analysis

Environmental variables
curM_Wclm_mean
curZ_Wclm_mean
sal_bottom_mean
sal_bottom_range
temp_bottom_mean
temp_bottom_range
mean_turb
mean_prec
min_prec
chl_Wclm_mean
dox_Wclm_mean
nit_Wclm_mean
pco_Wclm_mean
ph_Wclm_mean
pho_Wclm_mean
phyto_Wclm_mean
ppn_Wclm_mean

Table 2.4: Summary of 17 environmental variables.

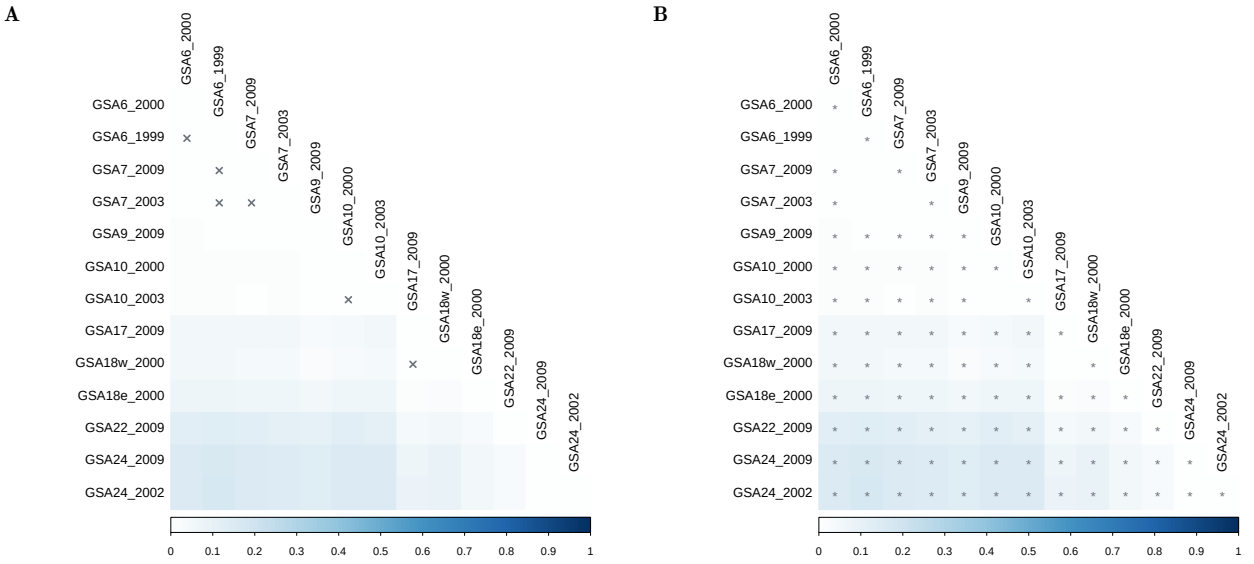

Figure 2.7: Matrix of pairwise F_{ST} after Benjamini–Yekutieli correction for strataG indicating non significant comparison (A) and significant comparison (B) for ($\alpha = 0.05$).

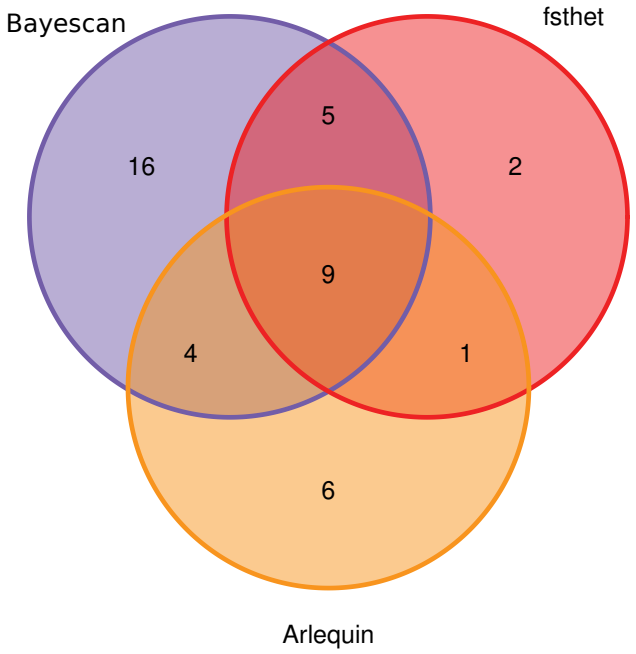

Figure 2.8: Venn diagram of SNPs identified by *fsthet*, *Bayescan* and *Arlequin* in outlier detection analysis.

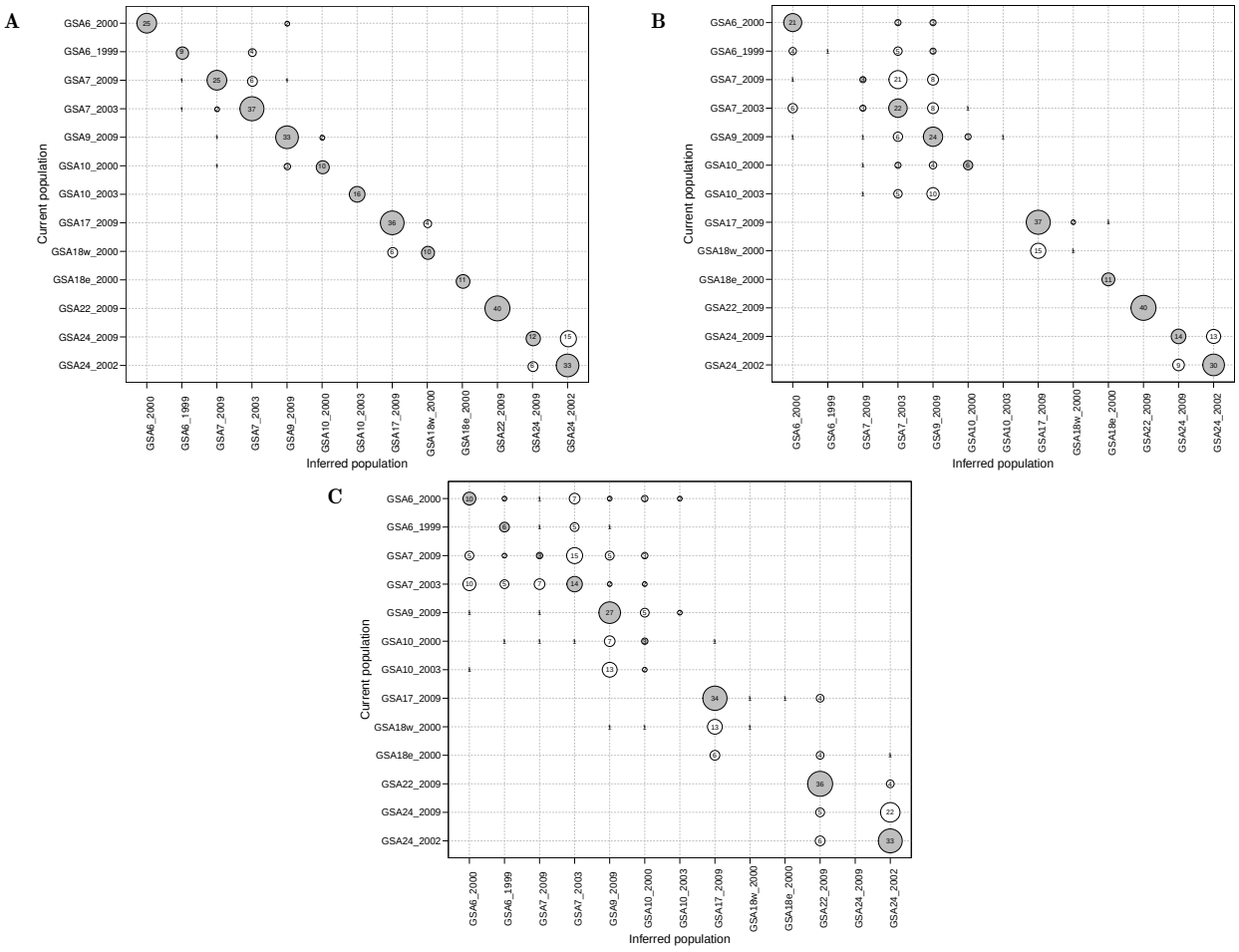

Figure 2.9: Comparison of individual assignment between inferred groups and actual samples corresponding to sampling sites resulted by DAPC analysis for whole (A), neutral (B) and outlier (C) dataset.

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP1001	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1002	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1004	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1008	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1010	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
SNP1012	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
SNP1014	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1017	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1018	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1022	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1023	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
SNP1025	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1029	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1030	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
SNP1031	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1033	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1035	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1038	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP104	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1040	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1046	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1047	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1049	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1052	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1054	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1058	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
SNP106	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1060	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1062	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP1063	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1067	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1068	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1070	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1074	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1077	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1079	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1084	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1089	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1091	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1094	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP1105	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1110	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1111	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1113	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1114	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1118	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1120	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1125	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1127	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1128	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1129	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1133	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1137	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1139	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1146	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1147	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
SNP1154	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
SNP1156	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP1159	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP116	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1160	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1163	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1166	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1169	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1182	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE
SNP1183	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1184	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1190	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1191	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1193	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1199	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP120	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1200	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1202	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1203	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1205	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP1206	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1208	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1213	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1214	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1220	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1231	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1232	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1234	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1235	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1236	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1238	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP1240	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1246	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1250	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1251	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1259	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1260	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
SNP1261	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1262	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1264	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1268	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1269	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP127	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1282	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1283	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1284	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1286	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1293	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1294	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1301	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1302	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1305	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1307	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1308	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP131	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1310	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1314	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1317	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1319	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1320	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP1328	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1333	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1335	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1337	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP134	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1340	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
SNP1343	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1346	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1347	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1349	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP135	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1353	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
SNP1354	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
SNP1355	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1359	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1370	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1373	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1376	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1379	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1380	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1388	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1400	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1402	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1407	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1408	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1411	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1416	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1417	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1427	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP1432	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1436	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1439	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP1445	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1451	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1455	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1456	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1466	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP147	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1479	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1484	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1489	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
SNP1491	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1492	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1493	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1496	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1501	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP1502	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1512	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1515	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1516	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1519	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
SNP1531	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP1536	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP1546	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP1554	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP158	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP164	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP167	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP170	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP182	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP184	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP188	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP191	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP192	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP199	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
SNP2	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP200	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP201	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP206	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP220	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP225	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP228	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP232	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP235	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP246	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP25	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP250	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP256	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP267	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP275	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP276	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP278	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP284	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP35	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP350	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP355	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP357	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP36	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP360	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP37	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP376	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP383	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP386	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP388	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP391	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP392	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP393	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP394	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP395	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP396	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP398	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
SNP399	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP410	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP411	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP418	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP42	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP422	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP424	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP431	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
SNP432	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP435	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP44	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP450	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP451	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP452	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP455	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP461	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP463	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP464	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
SNP465	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP466	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP469	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP476	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP488	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP490	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP497	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP503	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP504	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP516	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP520	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP528	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP530	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP533	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP543	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP550	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP563	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP567	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP569	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP573	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP577	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP58	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP590	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
SNP597	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
SNP599	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP60	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP600	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP601	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP605	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP609	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP614	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP626	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP633	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP637	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
SNP638	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP640	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP642	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP645	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP647	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP652	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP664	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP666	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP674	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP708	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP709	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP726	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP73	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
SNP737	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP739	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP747	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP749	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP750	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP752	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP757	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP766	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP767	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP768	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP769	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP771	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP773	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP776	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP779	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP780	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP781	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP782	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP788	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP789	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
SNP790	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP80	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP800	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP803	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP809	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP810	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP811	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP812	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP813	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP814	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP815	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP817	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE
SNP821	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP826	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP831	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP834	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP835	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP840	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP844	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP845	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP848	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP851	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP853	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP854	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP855	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP861	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP864	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP871	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP875	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
SNP877	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
SNP879	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP88	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP883	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP884	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP885	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP890	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP891	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP898	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP899	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP90	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
SNP914	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
SNP915	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP919	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP920	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP923	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP928	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP930	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP933	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP935	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP943	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP944	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP945	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP946	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP948	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP949	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP951	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP953	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP955	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
SNP956	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP961	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP962	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP963	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP965	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP968	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP970	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP971	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP972	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP973	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP975	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP977	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP986	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP987	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP988	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP989	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
SNP990	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 2.5: Details of loci identified as neutral (337 SNP loci) or outlier (43 SNP loci), SNPs identified by *Samβada*, *LFMM* and *Bayenv* associated with the environmental variables (curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean), and SNPs identified by at least one GAE methods (148 SNP loci).

SNPs name	Neutral	Outlier	GAE methods	*Samβada*	*LFMM*	*Bayenv*
SNP992	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
SNP995	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
SNP996	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE

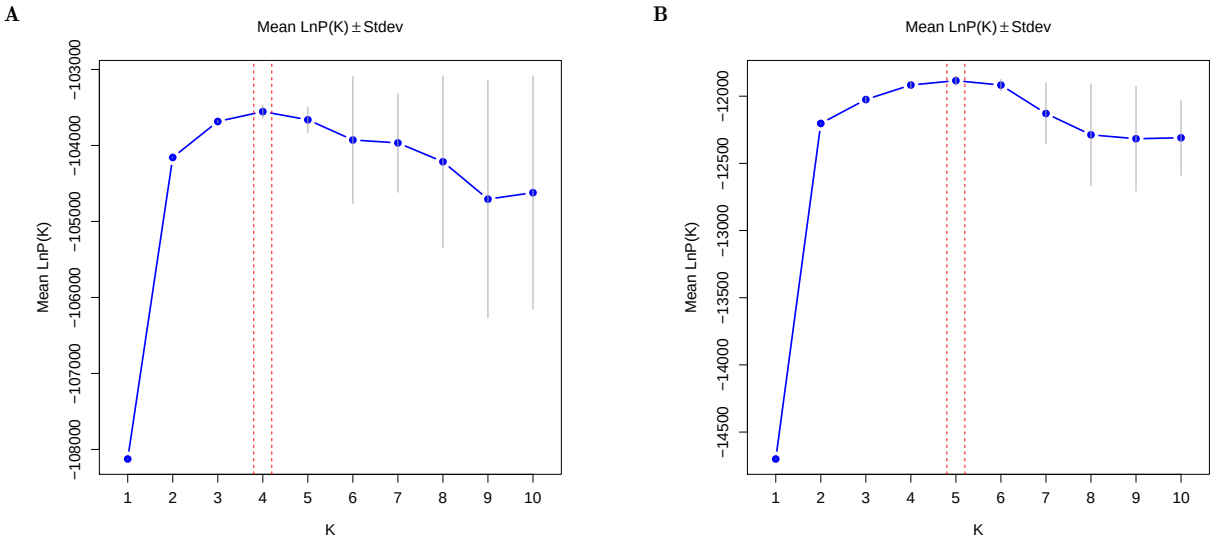

Figure 2.10: The mean L(K) over 20 iterations for each K. The highest mean L(K) value corresponds to the optimal K and it is showed in red dashed lines

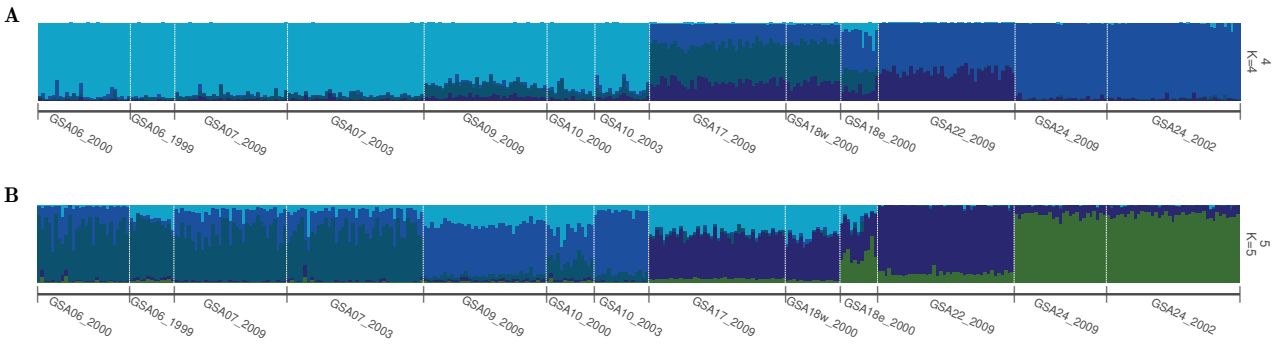

Figure 2.11: Graphical representation of the Bayesian clustering analysis of STRUCTURE for neutral (A) and outlier (B) SNPs datasets.

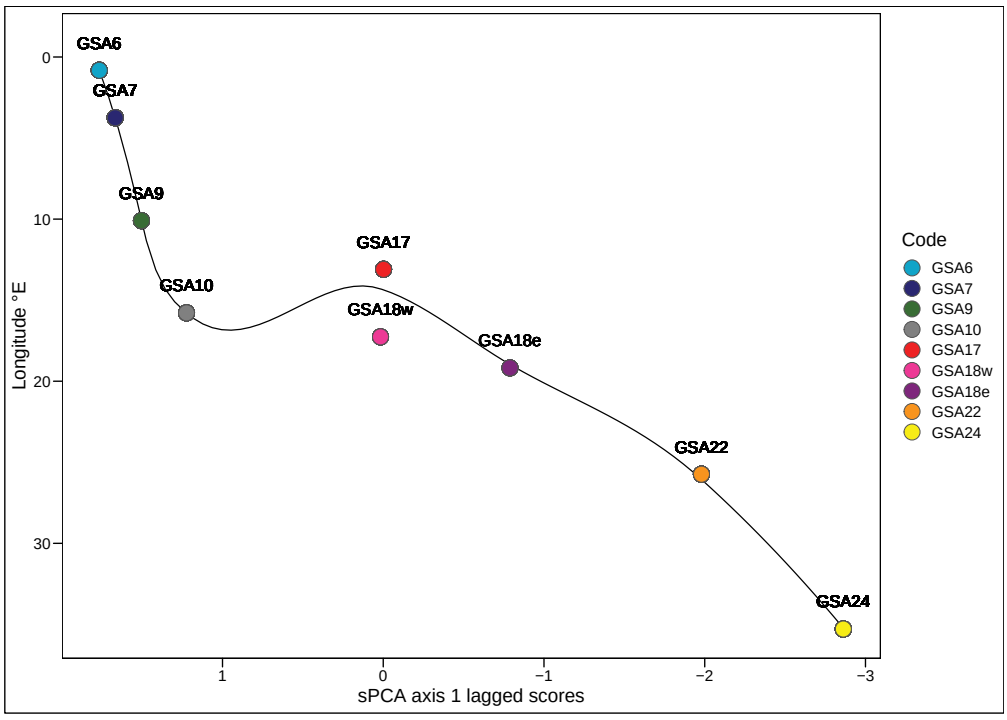

Figure 2.12: The graph shows the longitude genetic cline for neutral SNPs loci plotting sPCA axis 1 scores against longitude of sampling sites.

MARKERS	Neutral	Outlier
SNP1030	NO	YES
SNP464	NO	YES
SNP184	YES	NO
SNP930	YES	NO
SNP853	YES	NO
SNP1337	YES	NO
SNP990	YES	NO
SNP664	YES	NO
SNP638	YES	NO
SNP1018	YES	NO
SNP1436	YES	NO
SNP809	YES	NO
SNP1074	YES	NO
SNP898	YES	NO
SNP1184	YES	NO
SNP1546	YES	NO
SNP199	YES	NO
SNP971	YES	NO
SNP1063	YES	NO

Table 2.6: SNPs identified by loadingplot function imposing a threshold value above which alleles are annotated in both neutral and outlier datasets. We selected the 5% of alleles that contribute the most to the first sPCA axis.

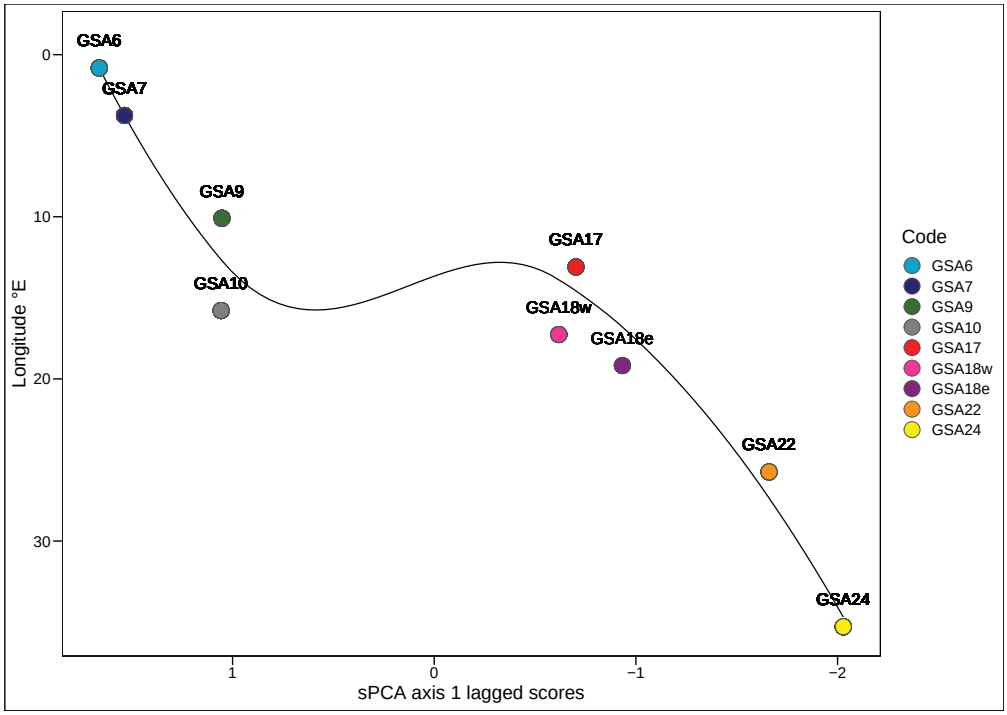

Figure 2.13: The graph shows the longitude genetic cline for outlier SNPs loci plotting sPCA axis 1 scores against longitude of sampling sites.

2.5.2.2 Otolith analysis

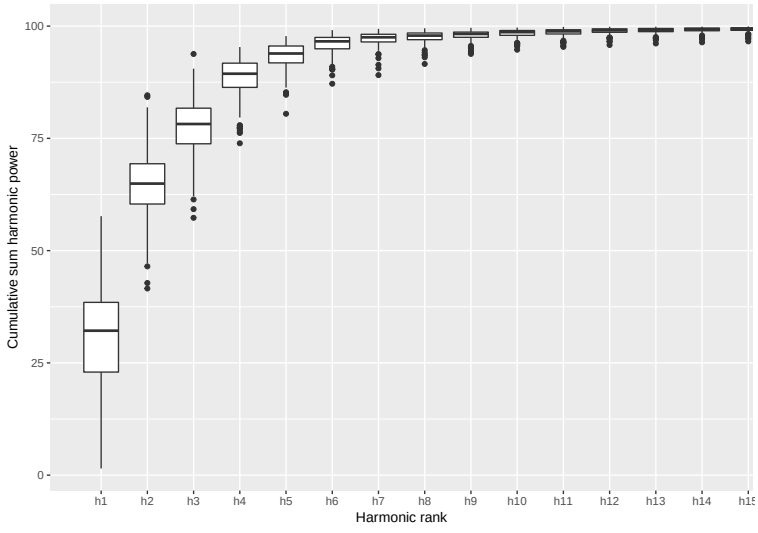

Figure 2.14: Cumulative Fourier power in relation to the number of harmonics describing the otolith shape. Each harmonic rank is composed by 3 harmonics.

	Pairwaise comparison	F model	p-adjusted	Significance
SIs	GSA7_2009 vs GSA9_2009	23.6415	0.0025	*
	GSA7_2009 vs GSA17_2009	15.0566	0.0025	*
	GSA7_2009 vs GSA22_2009	13.7214	0.0025	*
	GSA7_2009 vs GSA24_2009	31.4544	0.0025	*
	GSA9_2009 vs GSA17_2009	2.0840	0.1277	
	GSA9_2009 vs GSA22_2009	3.9083	0.0100	*
	GSA9_2009 vs GSA24_2009	5.9418	0.0040	*
	GSA17_2009 vs GSA22_2009	0.4822	0.6593	
	GSA17_2009 vs GSA24_2009	4.0592	0.0125	.
	GSA22_2009 vs GSA24_2009	5.0402	0.0050	*
EFDs	GSA7_2009 vs GSA9_2009	1.5526	0.0012	*
	GSA7_2009 vs GSA17_2009	1.9799	0.0012	*
	GSA7_2009 vs GSA22_2009	2.6622	0.0012	*
	GSA7_2009 vs GSA24_2009	1.7576	0.0012	*
	GSA9_2009 vs GSA17_2009	1.5848	0.0033	*
	GSA9_2009 vs GSA22_2009	2.3969	0.0012	*
	GSA9_2009 vs GSA24_2009	1.3777	0.0070	*
	GSA17_2009 vs GSA22_2009	3.4781	0.0012	*
	GSA17_2009 vs GSA24_2009	2.0571	0.0012	*
	GSA22_2009 vs GSA24_2009	2.3830	0.0012	*

Significant pairs are indicated with the following symbols:

. $p < 0.1$

* $p < 0.05$

Table 2.7: Results of PERMANOVA analysis using otolith shape: shape indices (SIs), elliptical Fourier descriptors (EFDs).

		Predicted sites						
	Original sites	GSA7_2009	GSA9_2009	GSA17_2009	GSA22_2009	GSA24_2009	Total	%correct
SIs	GSA7_2009	21	0	0	7	0	28	75.00
	GSA9_2009	2	13	11	11	1	38	34.21
	GSA17_2009	1	12	4	19	6	42	9.52
	GSA22_2009	6	10	4	18	7	45	40.00
	GSA24_2009	0	6	4	6	13	28	44.83
	Total							37.91
EFDs	GSA7_2009	20	4	5	0	1	28	66.67
	GSA9_2009	3	17	8	7	4	38	43.59
	GSA17_2009	3	12	29	0	0	42	65.91
	GSA22_2009	2	0	1	44	0	45	93.62
	GSA24_2009	2	4	3	0	20	28	68.97
	Total							68.78
SIs + EFDs	GSA7_2009	22	0	2	3	0	28	82.14
	GSA9_2009	2	18	11	4	3	38	47.37
	GSA17_2009	2	9	31	0	0	42	73.81
	GSA22_2009	0	2	0	42	1	45	93.33
	GSA24_2009	0	3	5	2	18	28	64.29
	Total							72.93

Table 2.8: Cross-validation classification matrix from CAP analysis using otolith shape: shape indices (SIs), elliptical Fourier descriptors (EFDs) and both (SIs + EFDs).

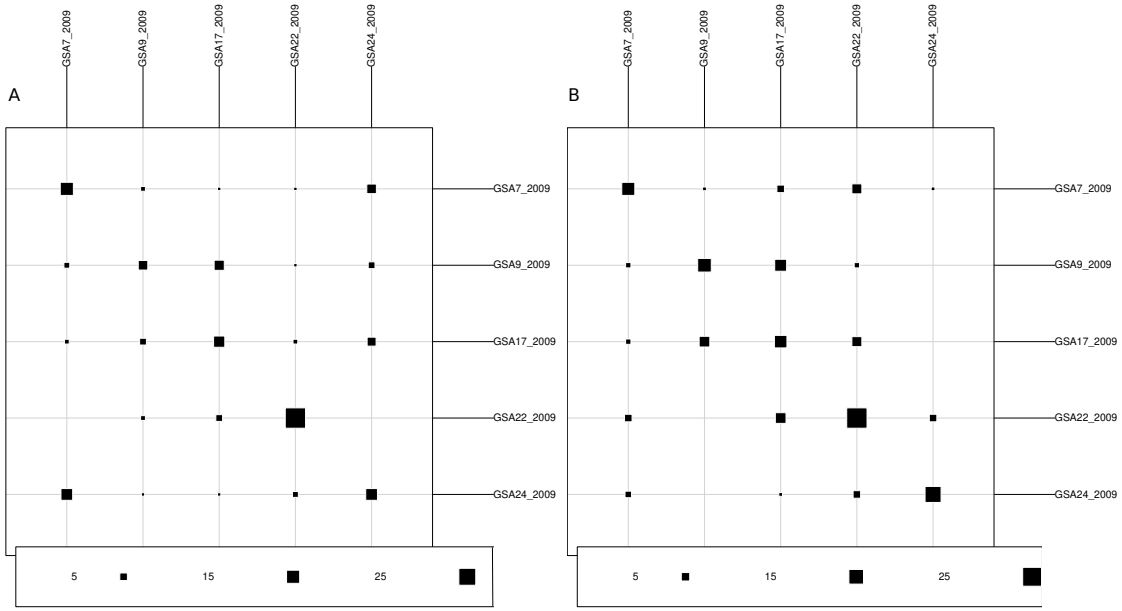

Figure 2.15: Membership counts plot showing the number of corrected assignments to the group of origin for core (A) and edge (B) data respectively. Rows correspond to actual groups, while columns correspond to inferred groups.

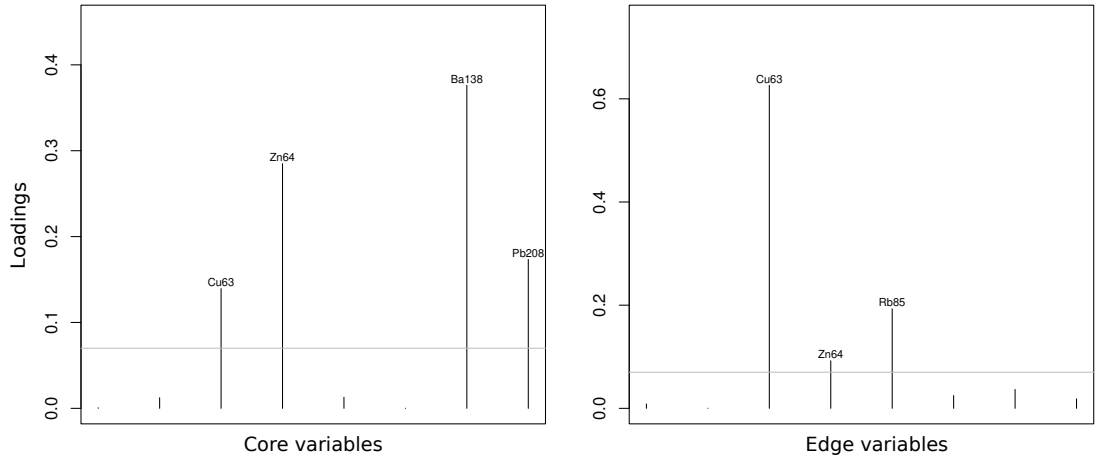

Figure 2.16: Loadings of elements for the first and second DAPC components for core and edge data. The plot shows the weight of each variable (loadings) in a given analysis.

2.5.2.3 Environmental analysis

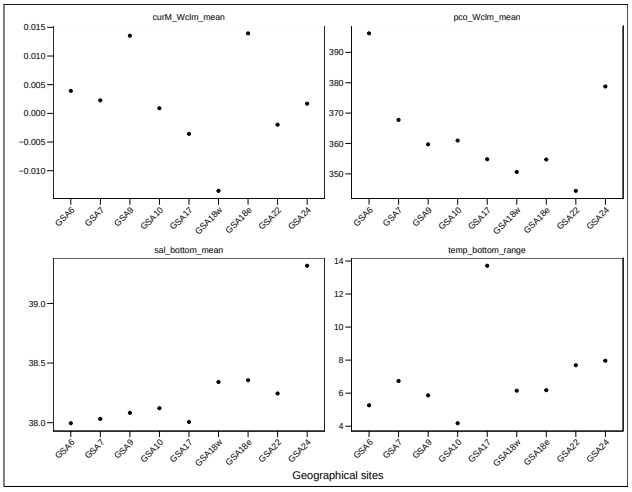

Figure 2.17: Plot represents the environmental values with their corresponding measures used in this study for macro-area reported as the corresponding GSA: the mean meridional component of the currents (northward) (curM, m/s) using the weighted average value in the water column (Wclm) (curM_Wclm_mean), the range of potential temperature (temp, °C) at the bottom (temp_bottom_range), the mean salinity (sal, psu) at the bottom (sal_bottom_mean) and the mean CO_2 (ocean pCO_2 expresses as carbon dioxide partial pressure) (pco, Pa) (pco_Wclm_mean).

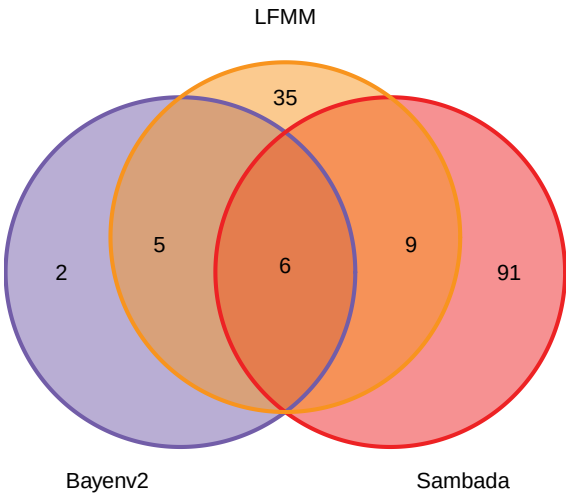

Figure 2.18: Venn diagram of candidate SNPs identified by *Sambada*, *LFMM* and *Bayenv* associated with curM_Wclm_mean, temp_bottom_range, sal_bottom_mean and pco_Wclm_mean.

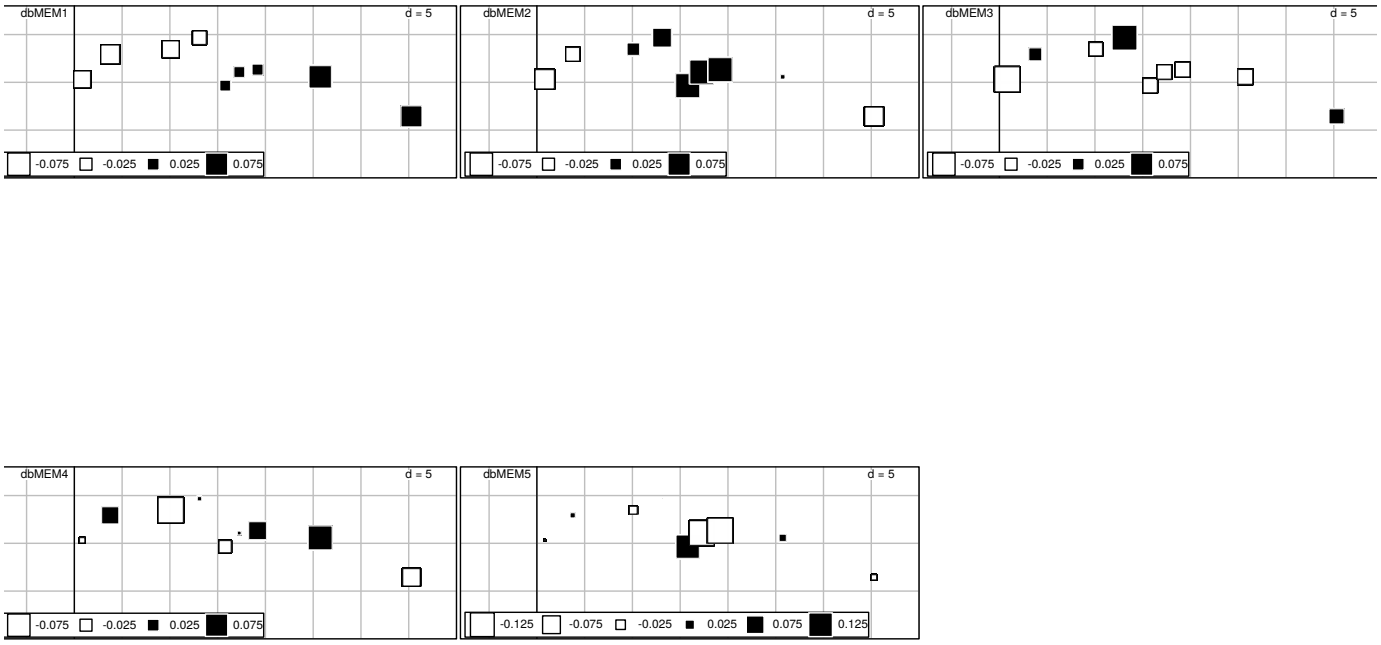

Figure 2.19: Maps showing 5 dbMEMs eigenvectors used as spatial variables with positive and negative eigenvalues represented by black and white squares, respectively.

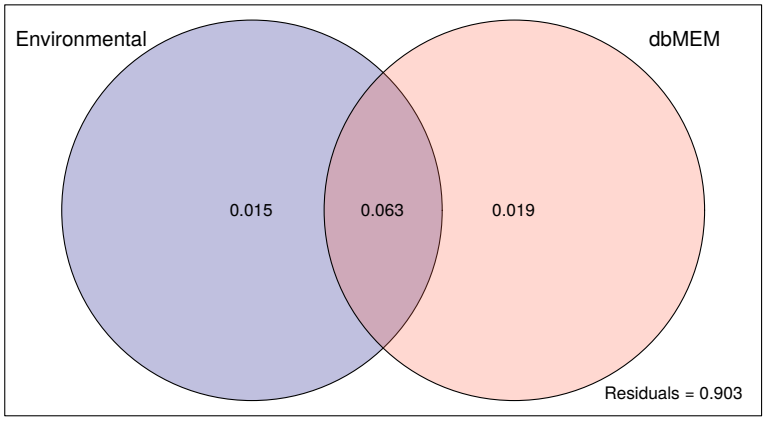

Figure 2.20: Results from variation partitioning analysis on neutral SNPs expressed by the adjusted coefficient of determination (R^2_{adj}). Venn's diagram shows the independent and shared variance explained by environmental and spatial variables.

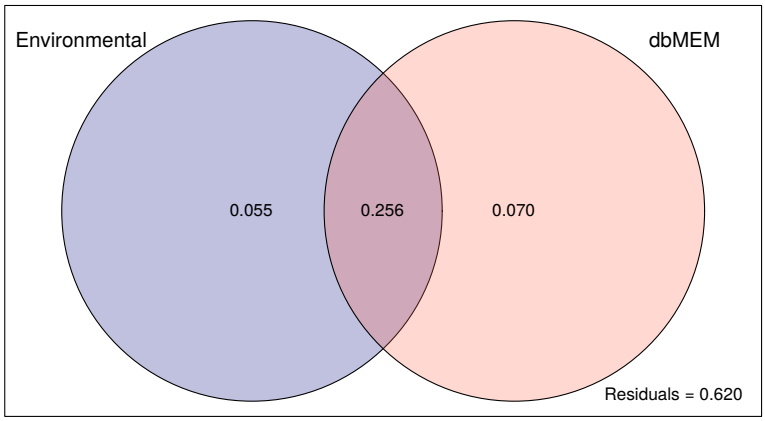

Figure 2.21: Results from variation partitioning analysis on outlier SNPs expressed by the adjusted coefficient of determination (R^2_{adj}). Venn's diagram shows the independent and shared variance explained by environmental and spatial variables.

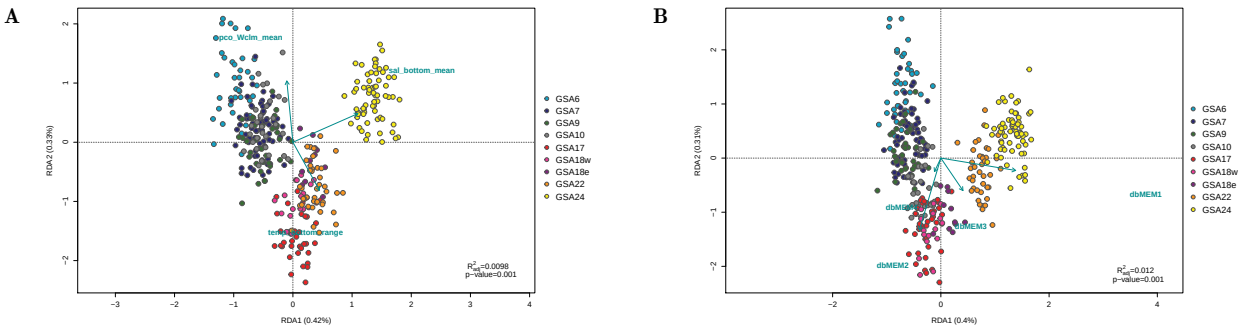

Figure 2.22: Partial redundancy analysis (RDA) conditioned on spatial (A) and environmental (B) variables performed on neutral SNPs. Environmental and spatial variables are shown with arrows.

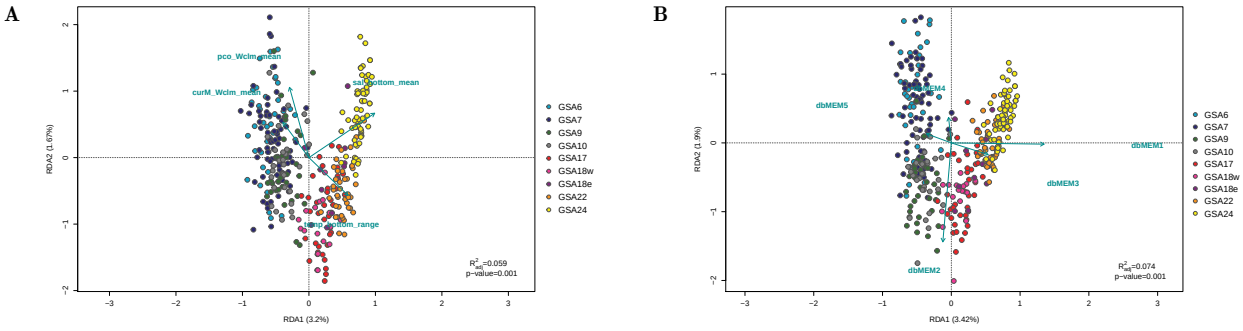

Figure 2.23: Partial redundancy analysis (RDA) conditioned on spatial (A) and environmental (B) variables performed on outlier SNPs. Environmental and spatial variables are shown with arrows.

		Selected variables (ordistep function)			
SNPs	Analysis	Environmental	Spatial	p	R^2_{adj}
Neutral	RDA			0.0010	0.0152
		sal_Wclm_mean*			
		temp_bottom_range***			
		pco_Wclm_mean*			
			dbMEM1*		
			dbMEM2***		
			dbMEM3**		
	partial RDA			0.0010	0.0098
		pho_Wclm_mean***			
		temp_bottom_range***			
		dox_Wclm_mean***			
	partial RDA			0.0010	0.0120
			dbMEM1***		
			dbMEM2***		
			dbMEM3***		
			dbMEM4***		
Outlier	RDA			0.0010	0.0850
		sal_Wclm_mean**			
		temp_bottom_range***			
		pco_Wclm_mean***			
			dbMEM1*		
			dbMEM2**		
			dbMEM3***		
			dbMEM4***		
	partial RDA			0.0010	0.0590
		sal_Wclm_mean***			
		temp_bottom_range***			
		pco_Wclm_mean***			
		curM_Wclm_mean***			
	partial RDA			0.0010	0.0740
			dbMEM1***		
			dbMEM2***		
			dbMEM3***		
			dbMEM4***		
			dbMEM5***		

Significant explanatory variables are indicated with the following symbols:

* $p < 0.05$

** $p < 0.01$

*** $p = 0.001$

Table 2.9: Results from RDA and partial RDA for each response variables divided in neutral and outlier SNPs

2.5.2.4 Gene Ontology

Figure 2.24: GO terms enriched in the biological process aspect. Names of GO terms founded significantly enriched ($p < 0.05$) are reported along the vertical axes.

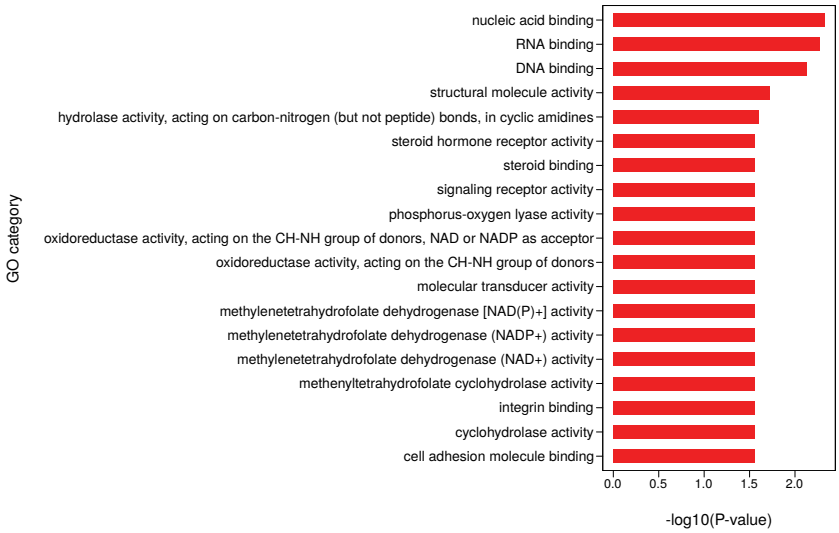

Figure 2.25: GO terms enriched in the molecular function aspect. Names of GO terms founded significantly enriched ($p < 0.05$) are reported along the vertical axes.

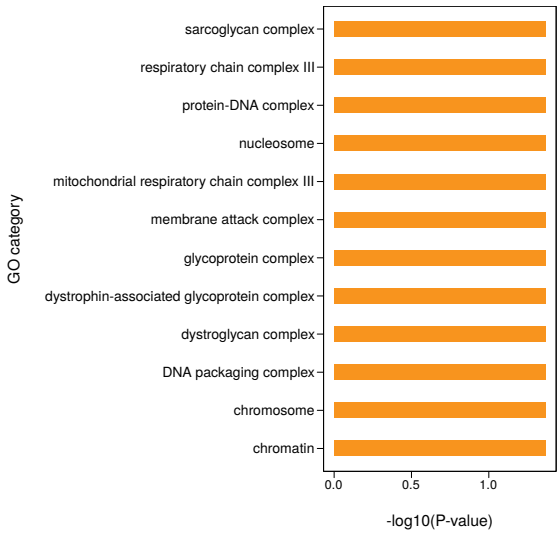

Figure 2.26: GO terms enriched in the cellular component aspect. Names of GO terms founded significantly enriched ($p < 0.05$) are reported along the vertical axes.

SNPs	Gene symbol	Gene description	Aspects
SNP1012*	sgcb	sarcoglycan, beta (dystrophin-associated glycoprotein)	CC
SNP1023*	LOC104962262	titin-like	CC, MF, BP
SNP206	LOC119503422	cytochrome b-c1 complex subunit 7	CC, BP
SNP398*	c7a	complement component 7a	CC, BP
SNP1293	fn1a	fibroectin 1a	MF, BP
SNP1531	Er-beta		MF, BP
SNP943	mthfd2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase	MF, BP
SNP250*	ybx1	Y box binding protein 1	MF, BP
SNP1058*	ampd1	adenosine monophosphate deaminase 1 (isoform M)	MF, BP
SNP528	znf106a-202	zinc finger protein 106	MF
SNP530	znf106a-202	zinc finger protein 106	MF
SNP569	cirbpa	cold inducible RNA binding protein a	MF, BP
SNP90*	fxr1	fragile X mental retardation, autosomal homolog 1	MF, BP
SNP1353*	LOC111650759	keratin, type I cytoskeletal 13-like	MF, BP
SNP1354*	LOC111650759	keratin, type I cytoskeletal 13-like	MF, BP
SNP73*	LOC103378706	60S acidic ribosomal protein P2-like	MF, BP
SNP854	cd74	CD74 molecule, major histocompatibility complex, class II invariant chain a	BP
SNP1439*	ldha	lactate dehydrogenase A4	BP
SNP398*	selenop	complement component 7a	BP
SNP464*	LOC118300284	heat shock protein, alpha-crystallin-related, b11	BP
SNP877*	selenop	selenoprotein P	BP
SNP228	LOC103362009	troponin C, skeletal musclelike	BP
SNP476	tcap	titin-cap (telethonin)	BP

Table 2.10: SNPs associated with GO terms significantly enriched ($p < 0.05$) in at least one of the three aspects of Gene Ontology, biological processes (BP), molecular functions (MF), cellular components (CC). These 23 markers are all included in the outlier dataset, with 13 (*) of them also included in the environmental dataset.

Chapter 3

A holistic study of population structure with multi-tracers for the common sole (*Solea solea*) of the Mediterranean Sea.

3.1 Introduction

The delineation of marine populations is made more challenging due to the structural complexity of the populations in terms of marine fish connectivity through the exchange of individuals, but also the complexity of marine realm in terms of abiotic and biotic factors that interact with marine species distribution (Ciannelli et al., 2013).

In fact, the populations of almost all marine species resulted spatially structured to various degrees of connectivity over different temporal and spatial scales, resulting in different levels of aggregation from distinct population structures, through metapopulation to panmictic populations (Berryman, 2002; Ciannelli et al., 2013). Furthermore, other factors such as habitat fragmentation, fishing exploitation, invasive species and environmental alterations due to climate change have been acting as individual or cumulative forces that contribute to the variation and reduction of spatial distribution patterns making the definition of populations structure more complicated (Carroll et al., 2007). For example, fishing exploitation, by targeting older age classes, and by reducing mean size of individuals and the spatial heterogeneity of the populations, can lead to potentially decreased recruitment or increased population variability affecting the resilience of the population in facing environmental fluctuations (Hsieh et al., 2010; Perry et al., 2010).

The interaction between fishing and climate is substantial in the modification of the exploited marine system. In particular, fishing, by acting on individuals, also affects population, community and ecosystem levels making the status of the whole marine system more sensitive to climate forcing (Perry et al., 2010). Another anthropogenic factor that should be consider is the impact of oil and gas infrastructure in maintaining or

disrupting the resilience of marine species, although the effects of these structures on seascape connectivity have been suggested but not completely demonstrated (McLean et al., 2022). As a consequence of all these drivers, the variation of marine species in their abundance and distribution disrupts the interaction among and between species inside the marine ecosystem leading to the collapse of the fisheries stocks (Ciannelli et al., 2013). Therefore, the precise picture of spatial extent, demographic parameters and structure of marine populations is essential for a correct evaluation of stock in a fisheries management prospective in order to achieve sustainability and conservation of the resources. Commonly stock units are assessed based on single species assumption according to geographical and political features (Begg et al., 1999; Smith et al., 1990; FAO, 2020b). The mismatch between the biological population and the delimitation of management stocks leads to a biased assessment and potential mismanagement of the stocks (Kerr et al., 2017). An accurate stock units definition that leads to a robust assessment performance can be obtained properly by describing the spatial structure of population and the connectivity among populations (Berger et al., 2021). Hence fishery exploitation adapted to the underlined population structure becomes necessary in a conservative and sustainable framework (Kerr et al., 2017).

In fact, a variety of methods have been developed to investigate and define stock structure. In this direction, the International Council for the Exploration of the Sea Stock Identification Methods Working Group (ICES-SIMWG) periodically reported on updated applications of stock identification methods that accurately detect significant patterns of population structure (ICES, 2021). To overcome the challenges described above, integrated approaches are proposed to reveal the current stock structure with the aim of recognizing exact management units (Welch et al., 2015). In this respect, all available information on the characteristics of the population obtained from multiple methods and tracers should be gathered into a comprehensive approach (Kerr et al., 2017). The chance to detect actual spatial structure and an increased confidence of stock evaluation can be accomplished with a methodology that can combine different sources of information. One of the practices considered as the best in order to achieve powerful stock structure outcome is a holistic approach to comprehend spatial structure (Waldman, 1999; Abaunza et al., 2008; Cadrin et al., 2014b). In fact, the actual configuration of population is produced by ecological, demographic and evolutionary processes that act on spatial variations with different time scales (Carroll et al., 2007; Cadrin et al., 2014b). In fact, the advantage of these approaches is the combination of tracers with different spatio-temporal scales meaning that the method strengthens the power of one tracer to reveal an indication of separation where another one can fail avoiding their limitations (Abaunza et al., 2008; Pita et al., 2016; Welch et al., 2015).

The stock structure of common sole (*Solea solea*) inside Mediterranean basin was investigated in a recent study using mainly two tracers, molecular markers and otolith shape and elemental composition (Corti et al., 2022 *in prep.*). Different methods were implemented to identify spatial differences without applying an integrative methodology to fully elucidate the population structure of common sole. The molecular

approach investigated potential segregation throughout the stock using Single Nucleotide Polymorphism (SNPs) markers over an evolutionary time scale (Ovenden et al., 2015). Previous findings that described a west-to-east genetic differentiation pattern (Rolland et al., 2007; Guarnieo et al., 2002; Kotoulas et al., 1995) of common sole in Mediterranean Sea was reinforced by recent results of a strong genetic differentiation between three main sub-regions in the Mediterranean Sea (Western Mediterranean Sea, Adriatic Sea and Eastern Mediterranean Sea) detected by neutral genetic analysis and a less strong but significant genetic differentiation within each sub-region detected by outlier analysis (Corti et al., 2022 *in prep.*). The otolith approach also investigated the potential spatial variations across the stock using shape descriptors and elemental concentrations combining ontogenetic, environmental and genetic influences over the fish life span (Geffen et al., 2016). A signal of population structure was also detected using otolith tracers with significant differences between geographical sites which provide more directly information connected to demographic implications on potential segregation within the life cycle of the fish (Corti et al., 2022 *in prep.*). In fact, the information provided by each tracer was not redundant considering that each analysis offered stock structure delineation insight over different time and spatial scales. The integration of various sources of information is essential in a sustainable management framework especially for marine species targeted by commercial fisheries (Cadrin, 2020).

Considering the previous knowledge on recent past research, the current study aims to summarize the available information to provide a delineated stock structure describing the main mechanisms shaping the stock differentiation that could be suitable for stock assessment management strategies. To achieve this, we applied an integrated holistic approach to combine the tracers with different spatio-temporal scales that already presented insight of stock structure in common sole. We selected two complementary methods previously used because of their capacity to synthesize the information provided for these methods but also to clearly delineate similarities and differences between methods in terms of the most plausible stock structure: a semi-quantitative stock differentiation index (SDI) and a quantitative integrative multivariate analysis. With this approach, we expected to fill the gaps in knowledge concerning the spatial structure of stocks for the management of fisheries resources by exposing different structuring processes operating at contrasting scales.

3.2 Materials and Methods

In this study two easy user methodologies, SDI and multivariate analysis, were applied to integrate the results of population structure of 13 common sole population samples in the Mediterranean Sea obtained in a recent study from different tracers (Figure 3.27, Table 3.11). The multivariate procedure was also applied to the five geographical sites having available data for each tracer (Supplementary materials). The results across these two approaches were compared and combined to provide the most plausible stock structure of common

sole inside Mediterranean Sea. The approach applied in the previous study involved different tracers with different spatio-temporal scales (Tanner et al., 2016) and several methods including genetic analysis on both neutral and outlier SNPs markers and otolith analysis. Specifically, the sources of data applied to the study were: (1) neutral genetic; (2) outlier genetic; (3) otolith shape; (4) otolith elemental concentration core; and (5) otolith elemental concentration edge. Because of data availability, not all methods were applied on the same locations. Analyses were performed on genetics data for all 13 sampling sites, while otolith analyses were carried out only for five sampling sites out of 13 (Table 3.12). Previous study contained the details of the sampling locations, a description of the methods used in the analysis of genetic and otolith data and the results for each method revealing differences between locations independently for each methodology.

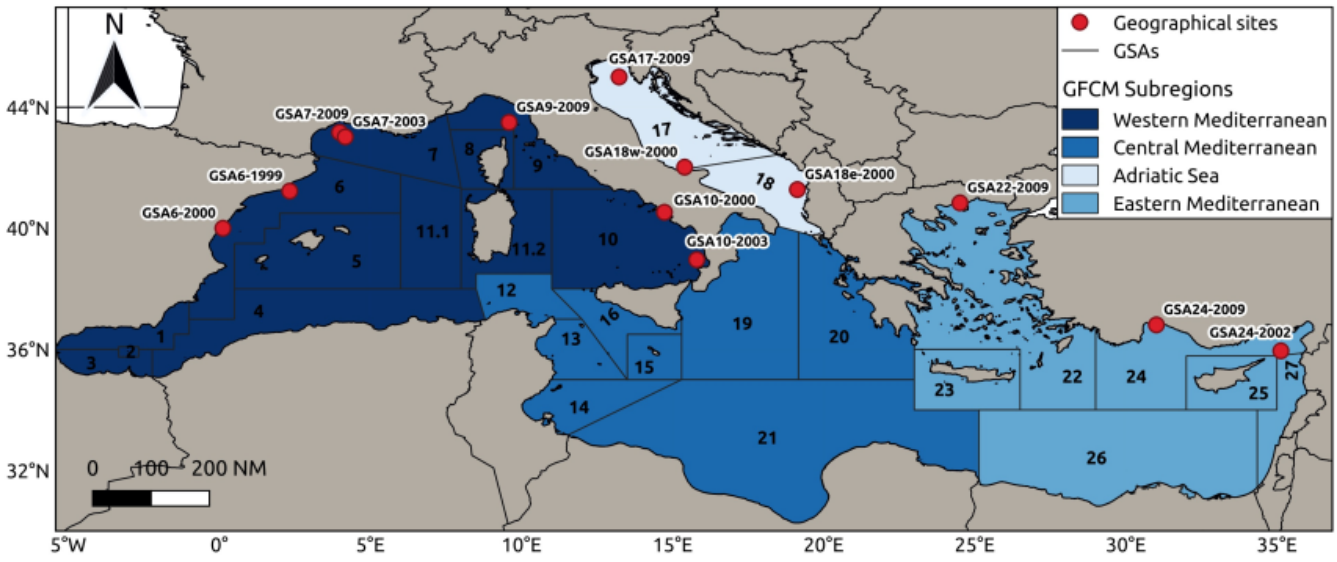

Figure 3.27: Map showing the 13 geographical sites located in the Mediterranean Sea. Black lines show the Geographical SubAreas (GSAs) and colors show the subregions as established by the General Fisheries Commission for the Mediterranean (GFCM).

3.2.1 Stock Differentiation Index (SDI)

Stock Differentiation Index is a semi-quantitative approach that offers the opportunity to combine tracers with different spatial and temporal resolutions succeeding into revealing stock structure (Welch et al., 2015; Izzo et al., 2017). The method assigns difference values (DV) to each pairwise comparison for each method based on the results obtained in the previous analysis. To assign the difference values for the genetic tracers we used the results of the pairwise F_{ST} computed by *StrataG* that identified differences in pairwise significant comparisons for $\alpha = 0.05$ both for neutral and outlier datasets as they provide information of contrasting

Sampling sites	GSAs	Sampling year	Sites ID	Number of methods	GFCM subregions
Castellon	Northern Spain (GSA6)	2000	GSA6_2000	2	Western Mediterranean
Barcellona	Northern Spain (GSA6)	1999	GSA6_1999	2	Western Mediterranean
Gulf of Lion	Gulf of Lion (GSA7)	2009	GSA7_2009	5	Western Mediterranean
Gulf of Lion	Gulf of Lion (GSA7)	2003	GSA7_2003	2	Western Mediterranean
Viareggio	Ligurian Sea and Northern Tyrrhenian Sea(GSA9)	2009	GSA9_2009	5	Western Mediterranean
Gulf of Salerno	Southern and Central Tyrrhenian Sea (GSA10)	2000	GSA10_2000	2	Western Mediterranean
Gulf of Sant'Eufemia	Southern and Central Tyrrhenian Sea (GSA10)	2003	GSA10_2003	2	Western Mediterranean
Chioggia Lagoon	Northern Adriatic Sea (GSA17)	2009	GSA17_2009	5	Adriatic Sea
Lesina	Southern Adriatic Sea (GSA18)	2000	GSA18w_2000	2	Adriatic Sea
Albanian coasts	Southern Adriatic Sea (GSA18)	2000	GSA18e_2000	2	Adriatic Sea
Gulf of Kavala	Aegean Sea (GSA22)	2009	GSA22_2009	5	Eastern Mediterranean
Gulf of Antalya	Northern Levant Sea (GSA24)	2009	GSA24_2009	5	Eastern Mediterranean
Antakya coasts	Northern Levant Sea (GSA24)	2002	GSA24_2002	2	Eastern Mediterranean

Table 3.11: Summary information of geographical sites and number of methods used in the integrative analysis. The last column reports geographical subregions defined by GFCM.

Type of tracers	Number of geographical sites	Spatial scale (GFCM subregions)	Temporal scale
Genetic neutral	13	Western Mediterranean Adriatic Sea Eastern Mediterranean	Evolutionary
Genetic outlier	13	Western Mediterranean Adriatic Sea Eastern Mediterranean	Some generations
Otolith shape	5	Western Mediterranean Adriatic Sea Eastern Mediterranean	Life span
Otolith elemental concentration core	5	Western Mediterranean Adriatic Sea Eastern Mediterranean	Life span (Larval stage)
Otolith elemental concentration edge	5	Western Mediterranean Adriatic Sea Eastern Mediterranean	Life span (Adult stage)

Table 3.12: Tracers used to deal with the stock structure of the common sole. The resolution of each tracer is displayed with the corresponding spatial and temporal scales and the number of geographical sites.

stock structure mechanisms. For otolith shape tracers the results of PERMANOVA analysis computed using *adonis* function in *vegan* R package were used to understand which pairs of sites were significantly different using *pairwise.adonis* function in *pairwiseAdonis* R package. The resulting pairwise significance was used to assign the difference values. For otolith elemental concentration tracers, the results of non-parametric Kruskal-Wallis statistics that investigate the differences in elemental concentrations between geographical sites were used to perform Dunn's test using *dunn.test* function resulting in multiple pairwise comparison between sites that were used to assign the difference values. As a general rule, a difference value of 1 ($DV = 1$) is assigned to a pairwise comparison when the method detects a significant comparison for that tracer. When the method detects a non significant comparison, a difference value of 0 ($DV = 0$) is assigned. Then, SDI values for each pair are calculated as:

$$SDI = \sum DV \,/\, Count\, DV$$

Where $\sum DV$ represents the sum of DVs between tracers for one pair of geographical sites and $CountDV$ is the total number of tracers used. SDI values spread between 0 and 1. High SDI values implies considerable spatial structure while low SDI values suggest weak or no spatial differences between geographic sites. An overall SDI value evaluates the spatial structure across the stock. The evaluation of the results is performed following the thresholds defined by Izzo et al. (2017). A weak spatial structure is expected with $SDI < 0.33$, moderate indication of spatial structure is imparted with SDI values between 0.33 and 0.66, and a strong spatial structure is denoted by $SDI > 0.66$.

3.2.2 Integrative multivariate analysis

To apply this second approach a dataset was produced containing the two first principal components (PCs) of the PCA performed for each tracer using *dudi.pca* function of *ade4* R package (Thioulouse et al., 2018). Thus, the dataset consists of two columns for each tracer with the first and the second PCs values for each geographical site. Specifically, the dataset is composed by the PCs from the following source of data: neutral genetic (Gen.neut1, Gen.neut2); outlier genetic (Gen.out1, Gen.out2); otolith shape (Oto.shape1, Oto.shape2); otolith elemental concentration core (Oto.core1, Oto.core2); and otolith elemental concentration edge (Oto.edge1, Oto.edge2). First, a PERMANOVA analysis was performed on the multivariate dataset to test the significance of differences between geographical sites. As input for the analysis, we created a dissimilarity matrix using *vegdist* function. Then, the PERMANOVA was performed using *adonis* function in *vegan* R package with Euclidean method and 1,000 permutations (Oksanen et al., 2022). *FactoMineR* R package (Husson et al., 2017) was used to perform multivariate analysis on these datasets in the contest of principal components methods. And specifically, *PCA* function of *FactoMineR* R package was applied on multivariate dataset to compute principal component analysis. This analysis allows to extract the information

contained in these multiple quantitative variables by identifying the dimension across which the variation in the data is maximal. Outputs provided by the analysis were extracted and visualized with *factoextra* package (Kassambara and Mundt, 2017) that allows for an easy and effective interpretation of the results. Thus, R functions provided in the *factoextra* R package enable to extract results for geographical sites but also for variables accounting the contributions of each variable on the variability of a given principal component (Kassambara, 2017).

3.3 Results

3.3.1 Stock Differentiation Index (SDI)

Pairwise SDI were calculated between each pair of geographical sites. For the five tracers to define stock structure that we used, 78 pairwise comparisons were made using 13 geographical sites. Based on determined thresholds (Suppl. Table 3.15), among pairs 58 show strong evidence of separation (74%), 14 show moderate evidence of separation (18%) and only 6 show no difference in spatial segregation compared to total number of pairwise comparison (Suppl. Table 3.16). As a general result, the stock of common sole indicates strong evidence of spatial structure inside Mediterranean Sea with an overall $SDI = 0.81$ across the 13 geographical sites. Considering only the five geographical sites for which all available tracers data are available, the overall $SDI = 0.72$ is decreased compared to the one with all geographical sites, but it still indicates strong evidence of spatial separation between geographical sites unless with a lower power. Grouping the geographical sites by subregions, an overall $SDI = 0.75$ indicates strong evidence of spatial structure considering the comparisons among and within Adriatic Sea and Eastern Mediterranean Sea subregions. Whereas, for the geographical sites grouped in Western Mediterranean Sea subregion the overall comparison expressed a lower degree of spatial structure, but moderate with $SDI = 0.41$, evidencing still a complex structure within this region. In general, geographical sites located at great distances expressed high SDI values, whereas contiguous sites expressed SDI values lower than 0.66 with few exceptions (Table 3.13). Eastern Mediterranean samples resulted strongly separated from all other sites, except for GSA24_2009 from GSA7_2009, GSA9_2009, GSA17_2009 ($SDI = 0.60$) and for geographical sites within GSA24 ($SDI = 0.60$), evidencing a moderate separation (Table 3.13). Regarding the central area, all Adriatic geographical sites are differentiated from the Western and the Eastern Mediterranean ones, except for the comparison GSA17_2009 vs GSA9_2009 ($SDI = 0.60$). Among Adriatic geographical sites there is a pattern of segregation, even within geographical sites included in GSA18. Indeed, inside the Adriatic area there is evidence to suggest that south-east sites are separated from north and south-west sites. The north and south-west coast of Adriatic Sea locations are similar, and both clearly separate from the south-east site, suggesting an independent dynamic in the Adriatic Sea. SDI highlighted a clear separation of Western Mediterranean geographical sites from Central and

Eastern Mediterranean areas. Regarding the Western Mediterranean area, geographical sites that are very close appear to be connected, with continuous sites expressing moderate to minimal or even null evidence of separation (Table 3.13). Moreover, geographical sites comprised within the same GSAs in the Western Mediterranean area showed null values ($SDI = 0$) indicating no evidence of segregation probably due to the fact that are temporal replicates within the same GSAs area.

	GSA6_2000	GSA6_1999	GSA7_2009	GSA7_2003	GSA9_2009	GSA10_2000	GSA10_2003	GSA17_2009	GSA18w_2000	GSA18e_2000	GSA22_2009	GSA24_2009	GSA24_2002
GSA6_2000													
GSA6_1999	0												
GSA7_2009	**0.5***	0											
GSA7_2003	**0.5***	0	0										
GSA9_2009	**1**	**0.5***	**0.4***	**0.5***									
GSA10_2000	**1**	**0.5***	**1**	**1**	**0.5***								
GSA10_2003	**1**	**1**	**0.5***	**1**	**0.5***	0							
GSA17_2009	**1**	**1**	**0.8**	**1**	**0.6***	**1**	**1**						
GSA18w_2000	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0					
GSA18e_2000	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**				
GSA22_2009	**1**	**1**	**0.8**	**1**	**0.8**	**1**	**1**	**1**	**1**	**1**			
GSA24_2009	**1**	**1**	**0.6***	**1**	**0.6***	**1**	**1**	**0.6***	**1**	**1**	**1**		
GSA24_2002	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0.5***	

Pairwise comparison with SDI values in bold indicates strong evidence of separation between geographical sites (SDI values > 0.66) and SDI values in bold with "*" indicates moderate evidence of separation between samples (0.33 < SDI values < 0.66). While other SDI values (SDI < 0.33) indicates minimal evidence of separation.

Table 3.13: Matrix of pairwise comparisons among geographical sites reporting SDI values for each pair calculated following the described approach. Underlined SDI values represents pairwise comparisons among the five geographical sites having available data for all tracer.

The situation described above is confirmed by grouping the locations by GSAs. Furthermore, there was strong evidence of separation of Tyrrhenian sites from Western Mediterranean geographical sites, except for GSA7 vs GSA9 (0.45) and a moderate separation in between Tyrrhenian sites (Table 3.14). Considering only the SDI values of the five geographical sites containing all tracers, the general pattern described is the same accounting for moderate to strong evidence of separation of the three main geographical area: Eastern Mediterranean, Adriatic Sea, and Western Mediterranean subregions. There is also a moderate separation among Eastern Mediterranean and Western Mediterranean geographical sites (Table 3.13).

3.3.2 Integrative multivariate analysis

From PERMANOVA analysis, a p-value of 0.001 was obtained from the *adonis* test, thus the geographical sites can be considered significantly different for $\alpha = 0.01$ in the multivariate dataset, while multi-level pairwise comparison among geographical sites was not performed due to the low number of geographical

GSAs	6	7	9	10	17	18	22	24
6	0							
7	0.25	0						
9	**0.75**	**0.45***	-					
10	**0.88**	**0.88**	**0.5***	0				
17	**1**	**0.9**	**1**	**1**	-			
18	**1**	**1**	**1**	**1**	**0.5***	**1**		
22	**1**	**0.9**	**0.8**	**1**	**1**	**1**	-	
24	**1**	**0.9**	**0.8**	**1**	**0.8**	**1**	**1**	**0.5***

Pairwise comparison with SDI values in bold indicates strong evidence of separation between GSAs (SDI values > 0.66) and SDI values in bold with "*" indicates moderate evidence of separation between samples (0.33 < SDI values < 0.66). While other SDI values (SDI < 0.33) indicates minimal evidence of separation.

Table 3.14: Matrix of pairwise comparisons among GSAs reporting SDI values for each GSA calculated following the described approach. GSAs containing only one geographical sites per GSAs do not express any SDI values in the diagonal.

sites per GSAa. From PCA, most of the information contained in the data is retained by the first three principal components that explained 81.42% of the variation (Suppl. Figure 3.35). In general, all variables expressed a good quality of representation on the first three principal component as reported by cos2 (square cosine, squared coordinates) value that represents the quality of representation of variables (Figure 3.28). In particular, variables (Oto.edge1, Oto.core1, Oto.core2, Oto.edge2 and Gen.neut1) have a strong effect on first and second principal components compared to other variables (Figure 3.28A). The quality of representation is considerably lower for those variables on second and third principal component, but variables still maintain a good representation, especially Gen.out1, Oto.core2, Oto.edge1 and Oto.shape2 meaning that also third principal component contains a source of information for these variables (Figure 3.28B). Thus, all three principal components of the multivariate analysis contain different information, each principal component represents specific variables.

Regarding the contributions of each variable to principal components, a variable with a contribution larger than 10% could be considered more substantial in contributing to the results. Mainly otolith elemental concentration data but also genetics data both neutral and outlier contributed most to the first two principal components, that correspond to different scale of segregating mechanisms. Whereas otolith shape data could be meaningful in contributing to the third principal components (Figure 3.29 and Suppl. Figure 3.36).

Regarding PCA results for geographical sites, the first principal component which is the one with the highest variance (33.3%) display a western-eastern general patterns, whereas the second principal component

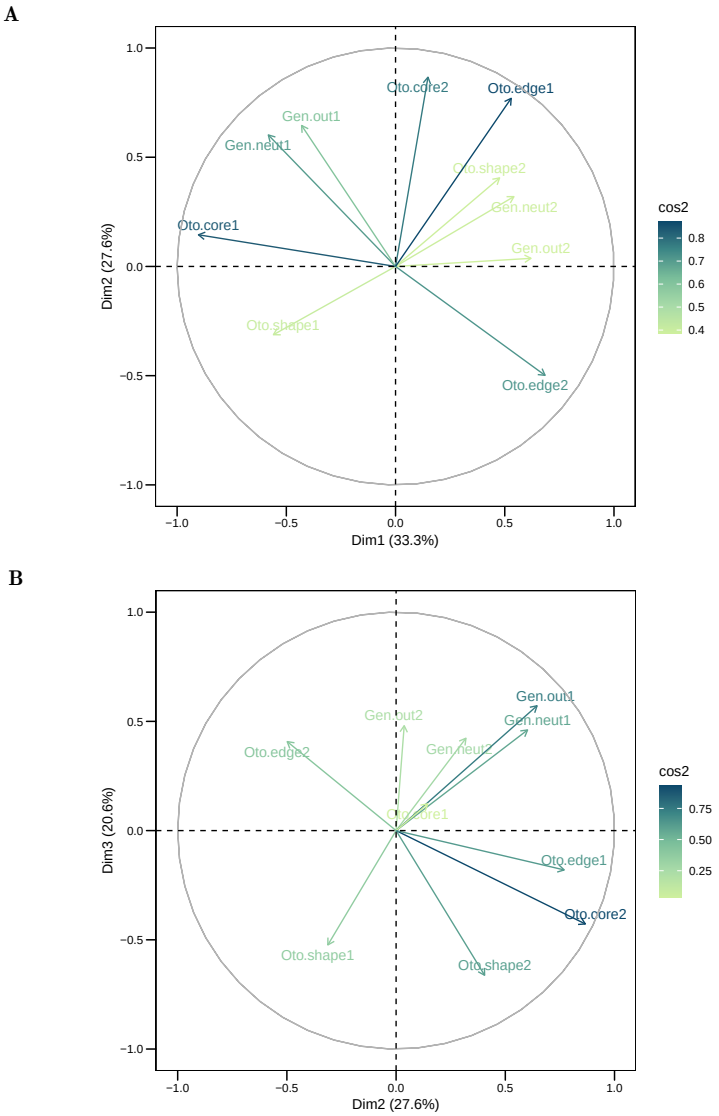

Figure 3.28: Variable correlation plot of PCA analysis showing the relationship between variables with the corresponding cos2 values that express the quality of representation of each variable on the principal component. The plots present axes 1 and 2 (A), and axes 2 and 3 (B) representing different combination of the first three principal components.

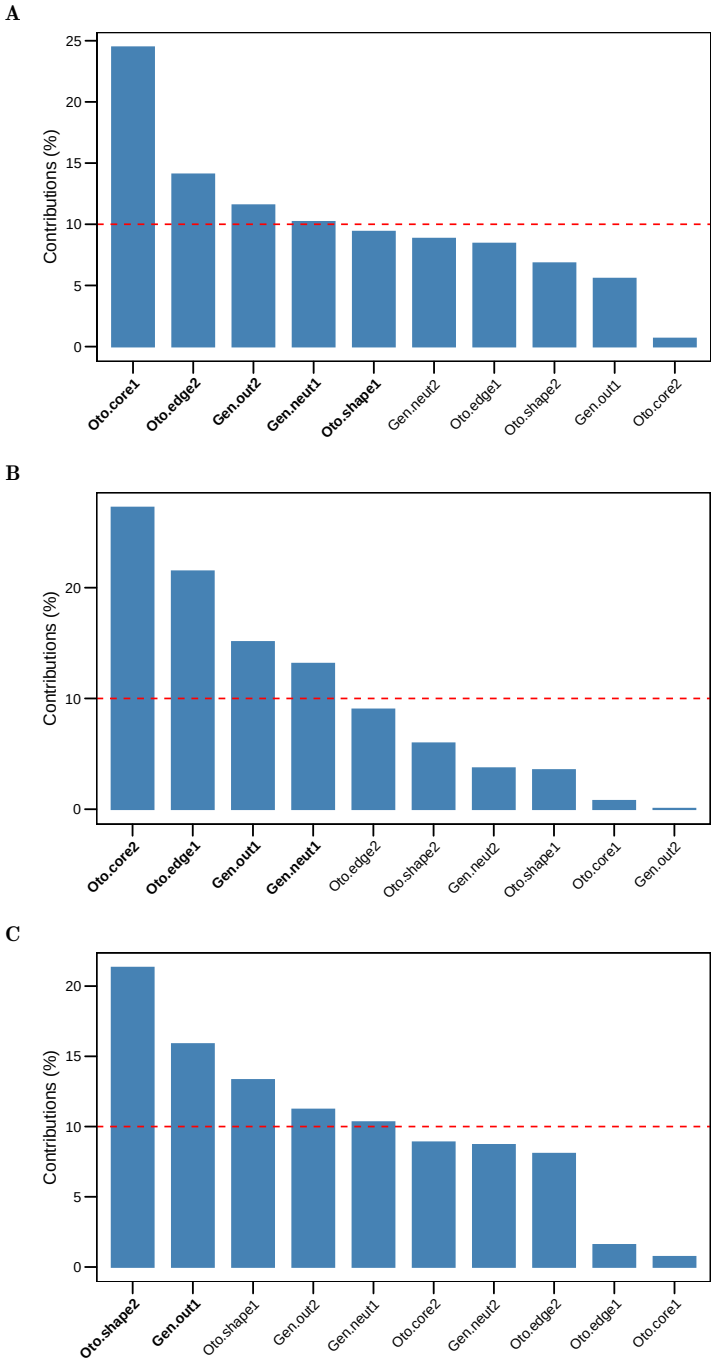

Figure 3.29: Barplots of variable contributions showing the total contribution of each variable to PC1 (A), PC2 (B) and PC3 (C), respectively. The red dashed line represents the expected average contribution with a cutoff of 10% because we consider a uniform contribution of the 10 variables. The most significantly associated variables (p-value < 0.05) with each principal component identified with the *dimdesc* function of *FactoMineR* R package are reported in bold.

(27.6%) indicates a local gradient showing differences between GSA22 and GSA24, GSA17 and GSA18, and also among Western Mediterranean geographical sites (Figure 3.31). Finally, the third principal component indicated differences between GSA6-GSA7 and GSA9, but also of GSA6-GSA7 from GSA10. A more specific picture was exhibited at geographical sites level (Figure 3.30).

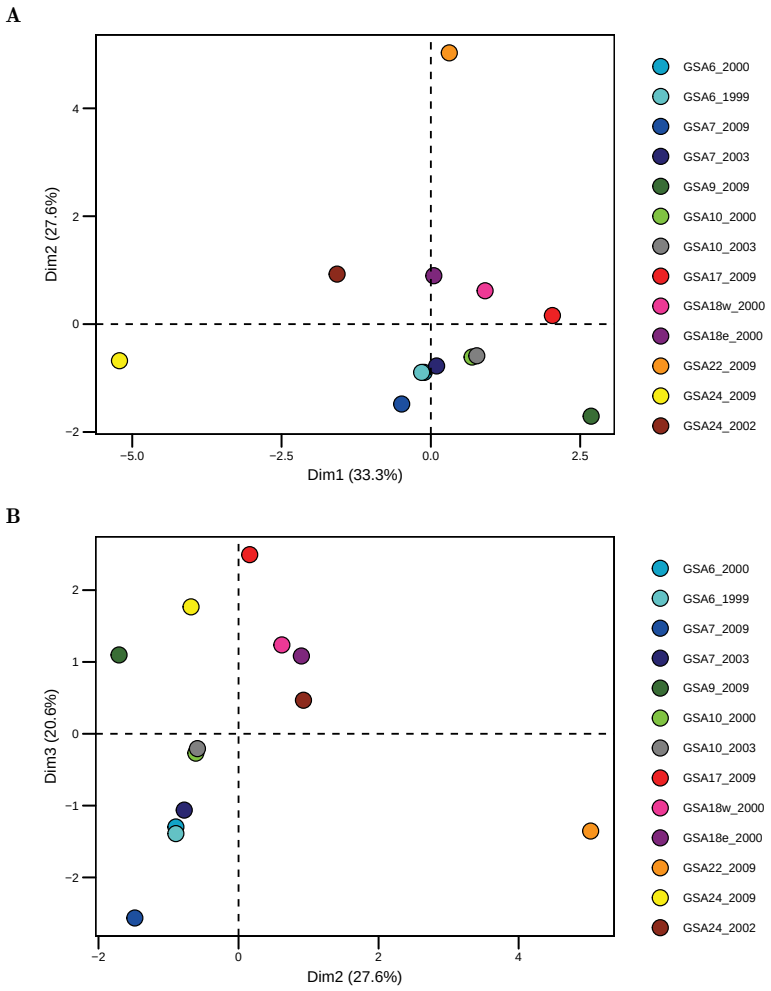

Figure 3.30: Graph visualization of 13 geographical sites of PCA output for PC1-PC2 (A) and to PC2-PC3 (B), respectively.

Oto.core1 and Oto.edge2 are the variables that most account for the general western-eastern pattern of separation between geographical sites (Figure 3.32A and Figure 3.29A). Oto.core2 and Oto.edge1 are the variables that most account for the local pattern of separation between geographical sites inside Eastern Mediterranean and Adriatic Sea (Figure 3.32A, 3.29B). For the genetic tracers, Gen.neut1 and Gen.out2 are the variables that account for the separation between Western Mediterranean, Adriatic Sea, and Eastern Mediterranean (Figure 3.32A, 3.29A and Suppl. Figure 3.36A), while Gen.neut1 and Gen.out1 influenced

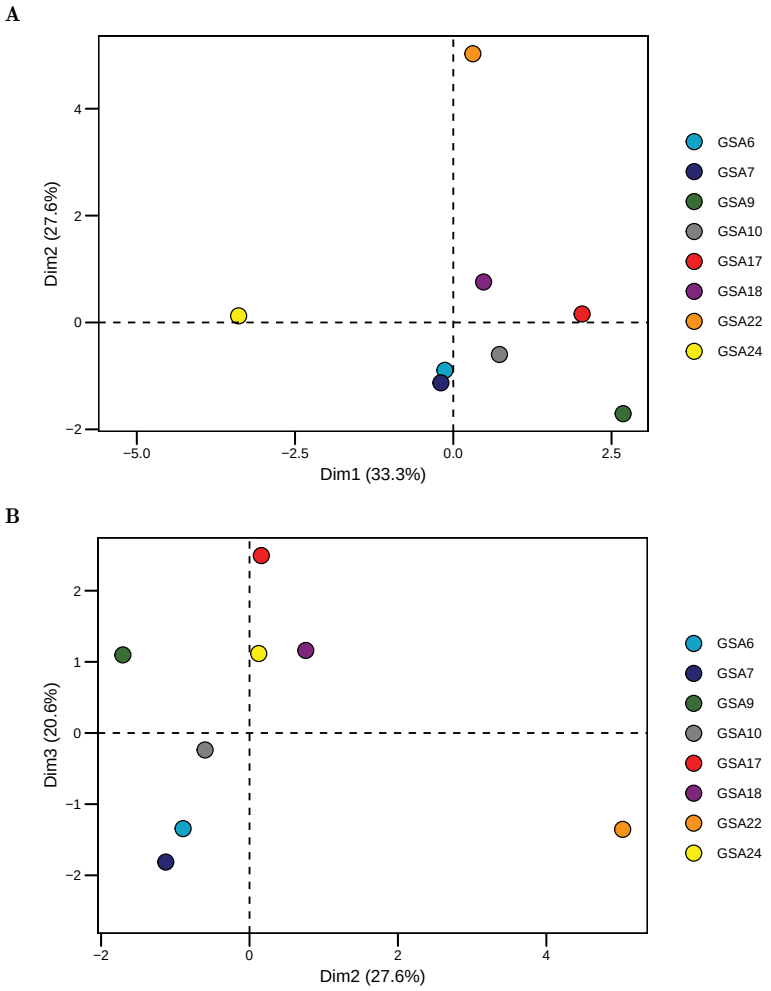

Figure 3.31: Graph visualization of PCA output of geographical sites grouped by GSAs of PC1-PC2 (A) and to PC2-PC3 (B), respectively.

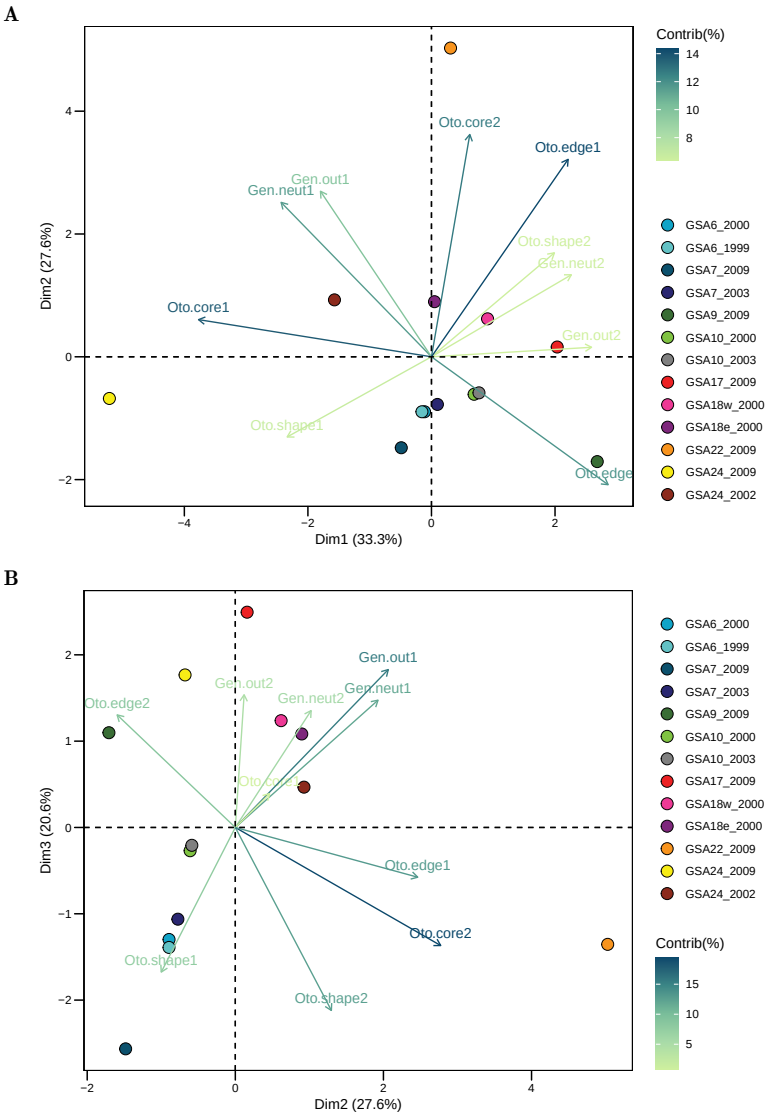

Figure 3.32: Biplot of PCA output displaying differences between geographical sites and tracer variables contributions to PC1-PC2 (A) and PC2-PC3 (B) showed with arrows. The contribution of each tracer is expressed by percentage (%).

the separation within the three main geographical area (Figure 3.32A, 3.29B and Suppl. Figure 3.36A). Oto.shape2 and Gen.out1 are involved in the separation between Western Mediterranean geographical sites and in their separation from others (Figure 3.32B, 3.29C). Also, the spatial segregation within Eastern Mediterranean and Adriatic geographical sites is mainly described by Oto.shape2 and Gen.out1. The same pattern is displayed grouping the geographical sites by GSAs (Suppl. Figure 3.37). The general pattern of separation between subregions is described in (Figure 3.33, Suppl. Figure 3.38) where there is a clear separation between Western Mediterranean, Adriatic Sea, and Eastern Mediterranean.

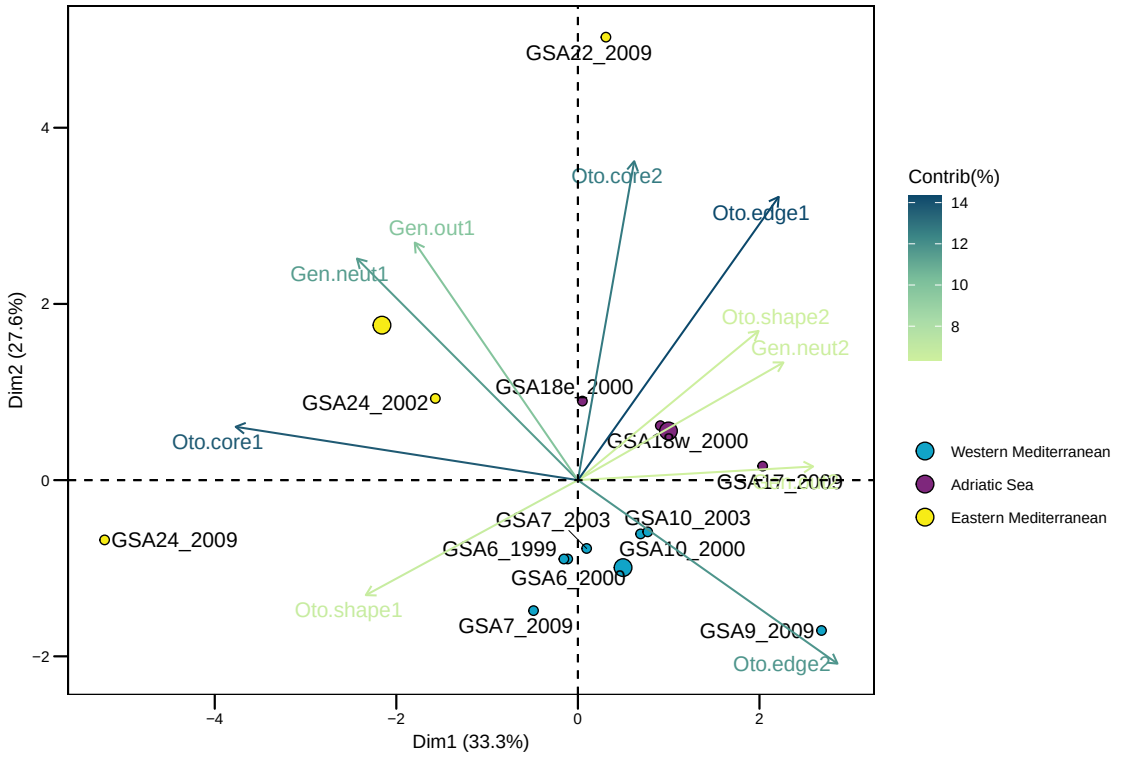

Figure 3.33: Biplot of PCA results on PCs values extracted from integrative multivariate analysis on different type of data of geographical sites displaying both differences between geographical sites grouped by GFCM subregions and variable contributions to PC1-PC2 expressed by percentage. Larger points represents the barycentre for each GFCM subregion.

From results on five geographical sites, the first two principal components of PCA explain 74.9% of the variation (Suppl. Figure 3.39). Considering only the five geographical sites, variables expressed slightly more quality of representation (cos2 values) than whole geographical sites on the first three principal component (Suppl. Figure 3.40). Regarding the contributions of each variable to principal components in five geographical sites situation, the contribution is almost the same as with whole geographical sites (Suppl. Figure 3.41). Also, the results of PCA described the same pattern previously explained (Suppl. Figure 3.42).

3.4 Discussion

We applied a holistic approach, summarising all the available information, aiming to bridge the gap in the understanding of population structure of common sole in the Mediterranean Sea. Although it is necessary to combine the results of different techniques merging their divergent resolutions, there is not a unique procedure for obtaining an exact definition of stocks.

The SDI approach applied by Izzo et al. (2017); Randon et al. (2021); Welch et al. (2015) has proved to be a powerful tool that is capable of integrating current information while successfully tackling the issues related to stock structure identification. In SDI approach the results of all pairwise comparisons among different geographical sites for the corresponding techniques/methods were tabulated defining a semi-quantitative measure of separation hence identifying where significant differences were found among geographical sites. SDI value thresholds established by Izzo et al. (2017) were followed to have enough power in discriminating sub-populations. The SDI values and the number of tracers reported for each pairwise comparison of geographic sites in an easy-to-read table (Suppl. Table 3.16) allowing to evaluate the weight of results. The general division pattern resulted in our study is even maintained by looking only at the five locations containing data for all tracers. Furthermore, adjacent sample locations were merged by GSAs to provide a summarized stock structure overview.

As an integrative multivariate analysis, we applied a PCA to emphasize variations among tracers and reveal mechanisms and scales of stock segregation in the observed dataset. The analysis exhibited where differences between geographical sites were detected and by which tracers. Furthermore, the results can be pooled by GSA and subregions to extract an overview of stock structure at broadest possible spatial scale evaluating the contribution provided by each tracer considering their temporal scale. Inside a holistic framework the integrative analysis combined in this study in association with different timescales, offered the basic groundwork for the interpretation of joint results and identification of management units. In fact, in this study we demonstrated that both approaches can identify differences between geographical locations combining spatio-temporal scales.

Previous studies suggested that common sole population is well structured within the Mediterranean Sea. Our study confirmed that, based on strong separation pattern indicated by SDI values, the population of common sole appears to be consisting of at least three very likely large stocks at regional level inside Mediterranean basin: one Eastern Mediterranean, one inside Adriatic Sea and one in Western Mediterranean (Table 3.14). While looking at subregional level (Table 3.13), according to SDI values there is moderate evidence of separation between geographical sites inside the same subregion suggesting evidence of potential complex stocks, which is also supported by the multivariate analyses. In other words, under a scenario of homogeneous populations within large region low SDI would be observed, which is not the case. More precisely there is indication of separation of Western Mediterranean area into two main separate stocks:

Northern Spain/Gulf of Lion and Tyrrhenian Sea. The same moderate separation is detected within Adriatic Sea from Northern to Southern geographical sites and within Eastern Mediterranean with Aegean Sea divided from Northern Levant Sea. Nevertheless, we may acknowledge that lower SDI values of some exceptions to leading pattern (Table 3.14, 3.13) were probably due to the differences in data availability of tracers (Table 3.12). One of the advantages of this approach is the application of semi-quantitative values to assess differences between locations and associated separation sighted by multiple tracers. Indeed, the application of SDI approach as a quantitative method for pooling multi-disciplinary data could be a good tool for holistic studies of stock structure of benthonic species as demonstrated by Randon et al. (2021), while it may need to be supported by complementary approach that capture quantitative gradients. The threshold established in SDI approach allowed a better definition of the groups based on numerical values. The moderate evidence of separation is related to separation among adjacent groups at subregional scale while strong evidence of separation is linked to the isolation of groups separated by wide geographical distances at regional scale. This confirmed the suggestion that spatial distance might be the leading driver in population structure of common sole given the lowly migratory nature of flatfish. Our findings of segregation at subregional scale is in line with recent studies that have confirmed a spatial stock structure of common sole in the Eastern English Channel (EEC) (Du Pontavice et al., 2018; Randon et al., 2018). The life-history traits of this species have been proved to have a major role in local segregation, in fact there is demonstration of low connectivity among common sole sub-populations within the ECC (Lecomte et al., 2020). Furthermore, the existence of metapopulation structure of sole in the EEC (Randon et al., 2020) was successfully proved by using complementary tracers to investigate marine species.

In order to combine the results obtained from tracers, the second approach used in this study, multivariate analysis, might boost the identification of management units by understanding the contribution of spatially and temporally distinct techniques. The multivariate approach reveals patterns in datasets by generally removing the redundancy in the data (Kassambara, 2017). Similarly based on individual representation of geographical sites from integrative multivariate analysis the population of common sole appears to be consisting of at least the three stocks inside Mediterranean basin previously explained by SDI approach as indicated in Figure 3.33. Looking specifically to more adjunct geographical sites at subregional level there is accordance to SDI results. The results of multivariate analysis confirmed the separation inside Western Mediterranean with a division between North Spain and Tyrrhenian Sea. Inside Adriatic Sea there is a separation between Northern and Southern geographical sites with a transition towards Northern Levant Sea that is divided from Aegean Sea (Figure 3.30). As a result, the integrative multivariate approach has proved to be a valuable tool in this context because it has the ability to extract and visualize differences between geographical sites by taking into account the contribution of each tracer and to group them attending the spatio-temporal scales of variation. In fact, the type of temporal information should be considered in the identification of stock boundaries based on significant differences between geographical sites because of their

importance in understanding the main mechanisms shaping the stock differentiation. Looking in details, the information given by each independent tracers are connected to two main temporal scales, a long term corresponding to evolutionary time scale and short term corresponding to individual life span. The general western-eastern pattern of segregation divided the geographical sites into three main area by the first principal component, as well as the more subregional pattern of segregation by the second principal component. This pattern of segregation was associated to genetics information but also to elemental concentration in both core and edge (Figure 3.32). Looking at the variables contribution in the multivariate analysis the genetics seem to drive the general eastern-western pattern of separation that was probably related to genetic drift and/or isolation by distance at broad spatial scales. While the segregation at subregion level that were associated to otolith elemental concentration and otolith shape, were probably related to recent event of isolation and adaptation to local environments (Figure 3.29). In particular, the pattern of segregation within Western Mediterranean geographical sites associated mainly to genetic outlier information was probably related to local adaptation at short term temporal scale. Thus, our results are in accordance with the spatio-temporal scales of the tracers (Figure 3.34). Furthermore, SDI approach results, in which few pairwise comparison, such as GSA7vsGSA17, GSA7vsGSA22, GSA9vsGSA22, GSA17vsGSA22 and GSA22vsGSA24, detected significant differences for almost all tracers ranging from small to large temporal scale confirmed previous findings, meaning that the separation between these geographical sites is currently occurring and has been occurred for many generations. The processes occurring at the evolutionary time scale are far longer than the mechanism existing at the ecological time scales that are those of primary importance in terms of stock assessment or fishery management prospective.

Although in this study we reported some relevant insights originated from the results that enhance the knowledge of the stock structure of common sole in the Mediterranean Sea, the identification of proper spatial management units can not be limited to three large units (i.e., Western, Central and Eastern Mediterranean) corresponding to subregions. Indeed, there is evidence for more complex structures, while whether they emerge from spatial heterogeneity or potential metapopulation structures require further research as it can bias assessment outcomes and derived references points triggering a potential mismanagement of the species (Goethel and Berger, 2017). Both complementary tracer and/or populations dynamics simulations may help to understand the demographic extent and implications of alternative population structure scenarios of the Mediterranean populations (e.g., (Goethel et al., 2021)). Implementing the strategy reported in this study in a contest with high degree of concordance among tracers and data availability could provide certainty on the stock status. The incorporation of both biological population structure and bioeconomic models by performing simulated scenery of different management plans has been demonstrated to be a successful way of handling demersal stocks (Cadrin, 2020; Spedicato et al., 2022). In this sense, dynamic spatial management would be preferable to adopt as a management approach. Nevertheless, the conclusion of both approaches could be limited by the fact that geographical sites do not cover the whole distribution of common sole

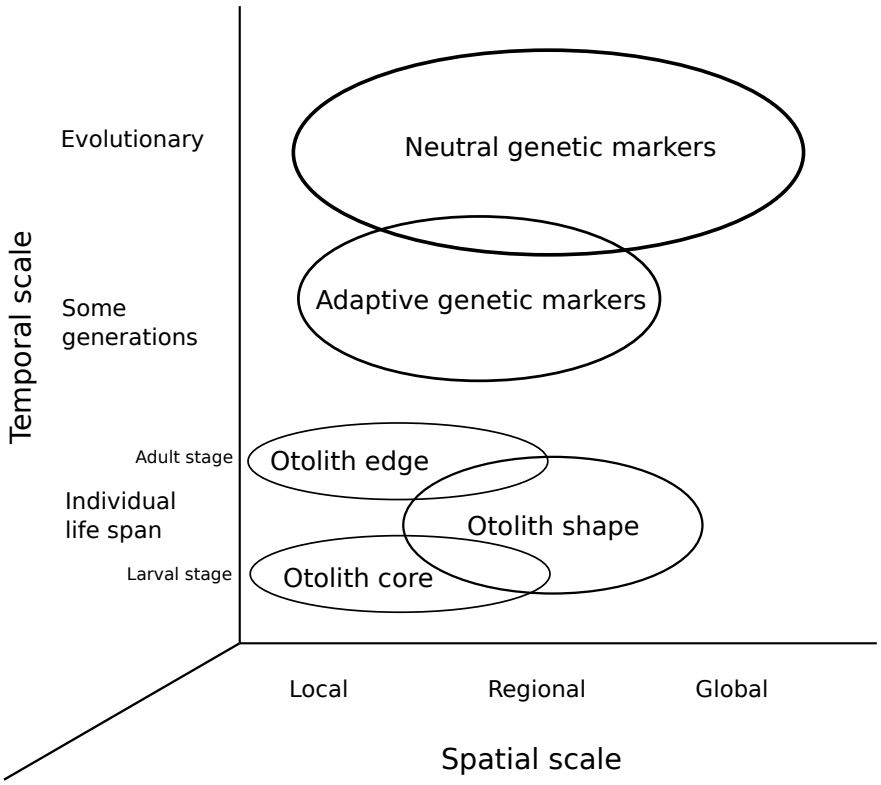

Figure 3.34: Summary of spatio-temporal scales represented by different markers applied in the current stock delineation approach. The spatial scale covers a distance between local to global areas. The temporal scale ranging from individual life span stages to evolutionary time.

populations of the Mediterranean basin. In particular, the Central Mediterranean subregion and, in general, the southern part of the Mediterranean Sea (i.e., the African coast and Sicily channel) is not represented in this study, but the remaining subregions where the species occurs more and from where you have the majority of the landings are nearly well represented. Therefore, our results will help to better design the sampling collection for future studies attending to applied question. However, due to this limitation, the whole structure of common sole inside Mediterranean Sea could be potentially incomplete due to the lack of data in terms of geographical locations, but also tracers (e.g., oceanographic dispersion modelling and fisheries data). Specifically, this constraint did not ensure a detailed interpretation of the spatio-temporal patterns of common sole stocks. The use of more tracers could help delineating a more precise stock structure of common sole inside the Mediterranean Sea. Considering these aspects, further investigations both in terms of stock structure and connectivity are needed to give an overall and clear picture of the status of this stock spread inside the Mediterranean Basin. In conclusion, this holistic approach could provide a meaningful tool in stock identification process to clearly integrate different dataset and comprehend results in a multi-technique stock structure work in order to spread outcomes in an easily legible way to fisheries managers and other stakeholders (Cadrin, 2020). Our findings in terms of segregation of geographical sites between and within subregions confirmed a high level of structure of common sole inside Mediterranean area with evidence of low connectivity between adjacent areas and presence of sub-populations that implies consequence for fisheries management and set the scientific basis for the future application of spatial stock assessment models.

3.5 Supplementary materials

3.5.1 Stock Differentiation Index (SDI)

SDI value	Evaluation of evidence for separation between geographical sites
SDI > 0.66	Strong
$0.33 \leq SDI \geq 0.66$	Moderate
SDI < 0.33	Weak

Table 3.15: SDI thresholds for evaluating patterns of stock structure in the common sole following Izzo et al. (2017) approach.

Table 3.16: Summary table reports difference values between each pair of geographical sites of the five available tracers applied to the common sole ($DV = 1$ for significant difference, $DV = 0$ for no significant difference). Number of tracers and SDI values are indicated for each pair of geographic sites. Established SDI thresholds are reported in Table 3.15. Pairwise comparison in which the number of tracers is maximum (5) are represented in bold.

	Tracers						
Pairwise	Genetic neutral	Genetic outlier	Otolith shape	Otolith core	Otolith edge	Number of methods	Pairwise SDI values
GSA6_2000 vs GSA6_1999	0	0				2	0
GSA6_2000 vs GSA7_2009	1	0				2	0.5
GSA6_2000 vs GSA7_2003	1	0				2	0.5
GSA6_2000 vs GSA9_2009	1	1				2	1
GSA6_2000 vs GSA10_2000	1	1				2	1
GSA6_2000 vs GSA10_2003	1	1				2	1
GSA6_2000 vs GSA17_2009	1	1				2	1
GSA6_2000 vs GSA18w_2000	1	1				2	1
GSA6_2000 vs GSA18e_2000	1	1				2	1
GSA6_2000 vs GSA22_2009	1	1				2	1
GSA6_2000 vs GSA24_2009	1	1				2	1
GSA6_2000 vs GSA24_2002	1	1				2	1
GSA6_1999 vs GSA7_2009	0	0				2	0
GSA6_1999 vs GSA7_2003	0	0				2	0
GSA6_1999 vs GSA9_2009	0	1				2	0.5
GSA6_1999 vs GSA10_2000	0	1				2	0.5
GSA6_1999 vs GSA10_2003	1	1				2	1
GSA6_1999 vs GSA17_2009	1	1				2	1
GSA6_1999 vs GSA18w_2000	1	1				2	1
GSA6_1999 vs GSA18e_2000	1	1				2	1
GSA6_1999 vs GSA22_2009	1	1				2	1
GSA6_1999 vs GSA24_2009	1	1				2	1
GSA6_1999 vs GSA24_2002	1	1				2	1

Continued on next page

Table 3.16 – continued from previous page

Pairwise	Methods						
	Genetic neutral	Genetic outlier	Otolith shape	Otolith core	Otolith edge	Number of methods	Pairwise SDI values
GSA7_2009 vs GSA7_2003	0	0				2	0
GSA7_2009 vs **GSA9_2009**	0	1	1	0	0	5	0.4
GSA7_2009 vs GSA10_2000	1	1				2	1
GSA7_2009 vs GSA10_2003	0	1				2	0.5
GSA7_2009 vs **GSA17_2009**	1	1	1	1	0	5	0.8
GSA7_2009 vs GSA18w_2000	1	1				2	1
GSA7_2009 vs GSA18e_2000	1	1				2	1
GSA7_2009 vs **GSA22_2009**	1	1	1	0	1	5	0.8
GSA7_2009 vs **GSA24_2009**	1	1	1	0	0	5	0.6
GSA7_2009 vs GSA24_2002	1	1				2	1
GSA7_2003 vs GSA9_2009	0	1				2	0.5
GSA7_2003 vs GSA10_2000	1	1				2	1
GSA7_2003 vs GSA10_2003	1	1				2	1
GSA7_2003 vs GSA17_2009	1	1				2	1
GSA7_2003 vs GSA18w_2000	1	1				2	1
GSA7_2003 vs GSA18e_2000	1	1				2	1
GSA7_2003 vs GSA22_2009	1	1				2	1
GSA7_2003 vs GSA24_2009	1	1				2	1
GSA7_2003 vs GSA24_2002	1	1				2	1
GSA9_2009 vs GSA10_2000	0	1				2	0.5
GSA9_2009 vs GSA10_2003	1	0				2	0.5
GSA9_2009 vs **GSA17_2009**	1	1	1	0	0	5	0.6
GSA9_2009 vs GSA18w_2000	1	1				2	1
GSA9_2009 vs GSA18e_2000	1	1				2	1
GSA9_2009 vs **GSA22_2009**	1	1	1	0	1	5	0.8
GSA9_2009 vs **GSA24_2009**	1	1	1	0	0	5	0.6
GSA9_2009 vs GSA24_2002	1	1				2	1
GSA10_2000 vs GSA10_2003	0	0				2	0

Continued on next page

Table 3.16 – continued from previous page

Pairwise	Methods						
	Genetic neutral	Genetic outlier	Otolith shape	Otolith core	Otolith edge	Number of methods	Pairwise SDI values
GSA10_2000 vs GSA17_2009	1	1				2	1
GSA10_2000 vs GSA18w_2000	1	1				2	1
GSA10_2000 vs GSA18e_2000	1	1				2	1
GSA10_2000 vs GSA22_2009	1	1				2	1
GSA10_2000 vs GSA24_2009	1	1				2	1
GSA10_2000 vs GSA24_2002	1	1				2	1
GSA10_2003 vs GSA17_2009	1	1				2	1
GSA10_2003 vs GSA18w_2000	1	1				2	1
GSA10_2003 vs GSA18e_2000	1	1				2	1
GSA10_2003 vs GSA22_2009	1	1				2	1
GSA10_2003 vs GSA24_2009	1	1				2	1
GSA10_2003 vs GSA24_2002	1	1				2	1
GSA17_2009 vs GSA18w_2000	0	0				2	0
GSA17_2009 vs GSA18e_2000	1	1				2	1
GSA17_2009 vs **GSA22_2009**	1	1	1	1	1	5	1
GSA17_2009 vs **GSA24_2009**	1	1	1	0	0	5	0.6
GSA17_2009 vs GSA24_2002	1	1				2	1
GSA18w_2000 vs GSA18e_2000	1	1				2	1
GSA18w_2000 vs GSA22_2009	1	1				2	1
GSA18w_2000 vs GSA24_2009	1	1				2	1
GSA18w_2000 vs GSA24_2002	1	1				2	1
GSA18e_2000 vs GSA22_2009	1	1				2	1
GSA18e_2000 vs GSA24_2009	1	1				2	1
GSA18e_2000 vs GSA24_2002	1	1				2	1
GSA22_2009 vs **GSA24_2009**	1	1	1	1	1	5	1
GSA22_2009 vs GSA24_2002	1	1				2	1
GSA24_2009 vs GSA24_2002	1	0				2	0.5

3.5.2 Integrative multivariate analysis

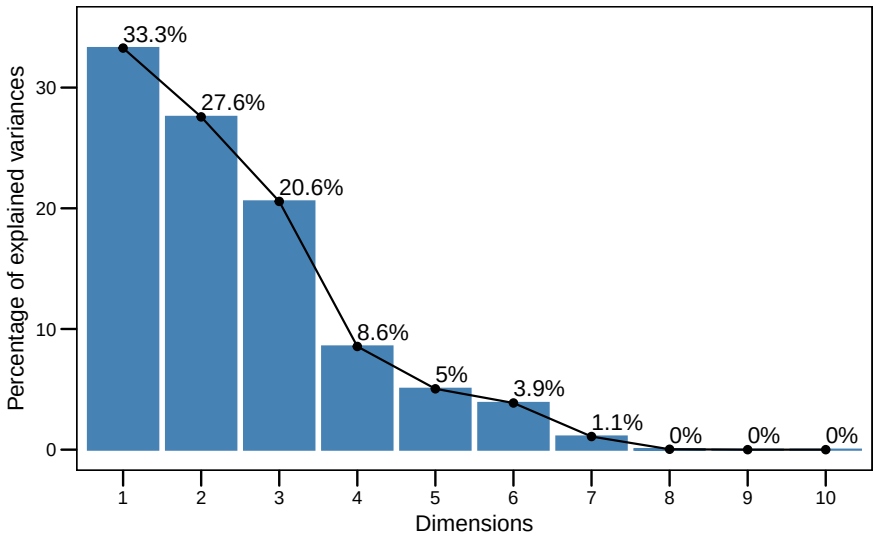

Figure 3.35: Scree plot shows the eigenvalues extracted from PCA analysis measuring the amount of variation retained by each principal component.

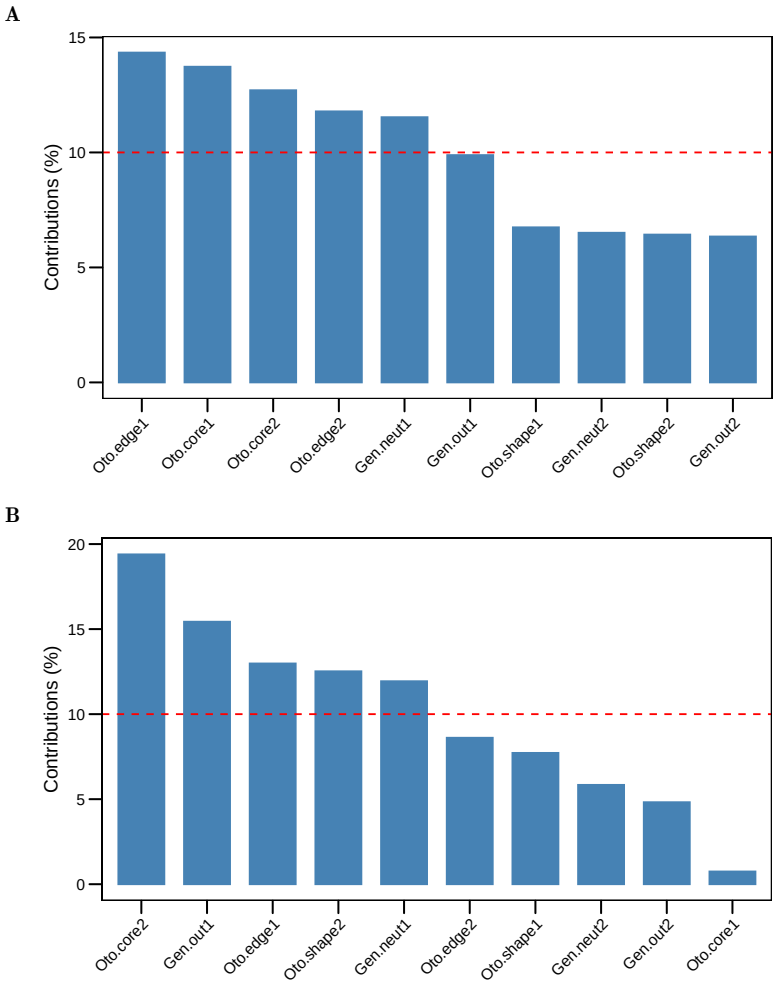

Figure 3.36: Barplots of variable contributions showing the total contribution of each variable to PC1-PC2 (A) and to PC2-PC3 (B), respectively. The red dashed line represents the expected average contribution with a cutoff of 10% considering a uniform contribution of the 10 variables.

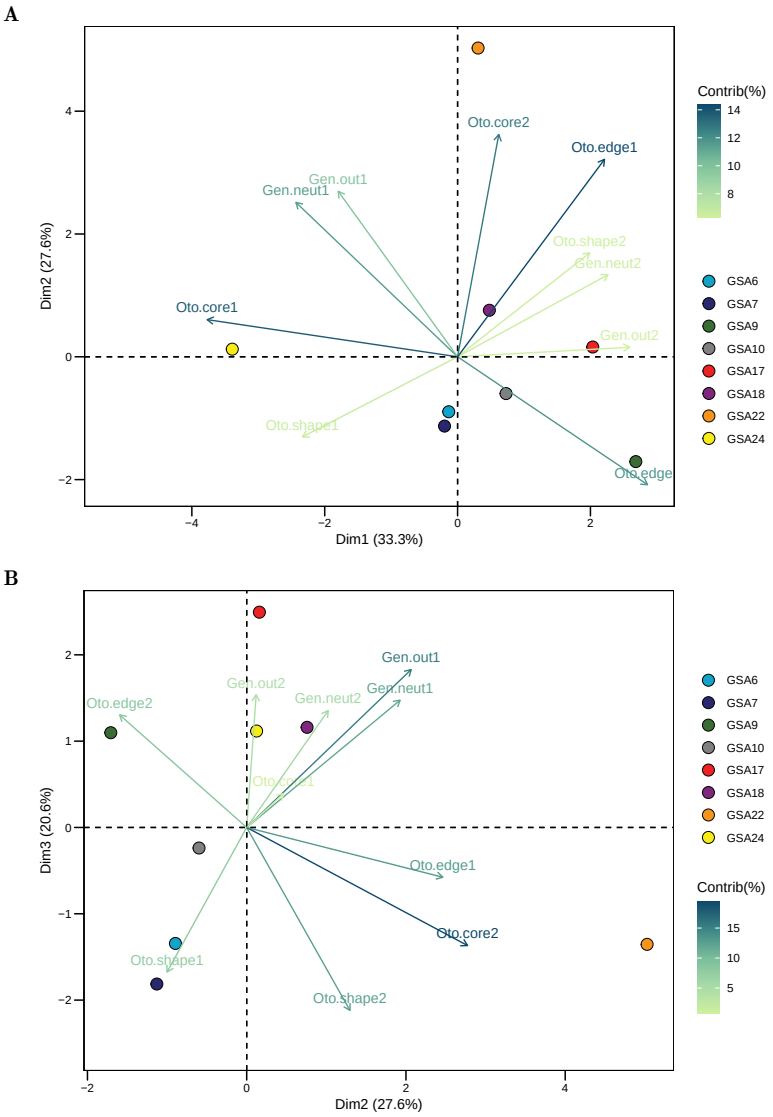

Figure 3.37: Biplot of PCA output displaying differences between GSAs and tracer variables contributions to PC1-PC2 (A) and PC2-PC3 (B) showed with arrows. The contribution of each tracer is expressed by percentage (%).

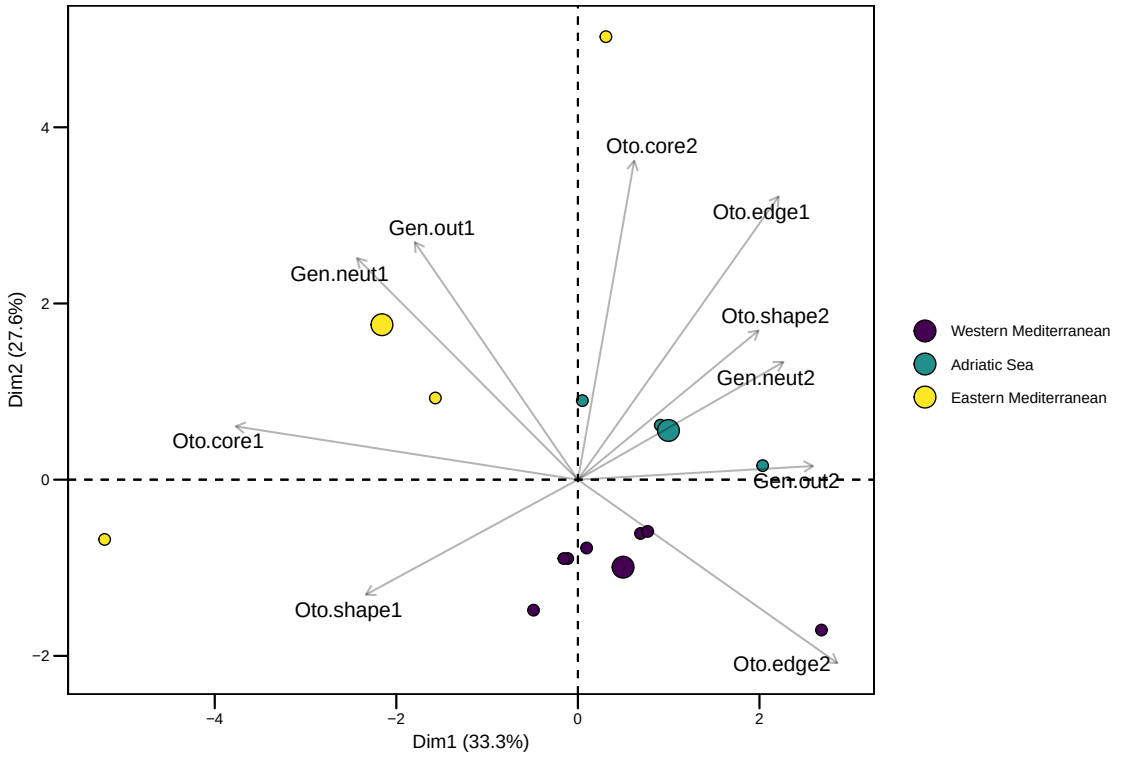

Figure 3.38: Biplot of PCA output displaying differences between geographical sites grouped by GFCM subregions and tracer variables contributions to PC1-PC2 showed with arrows.

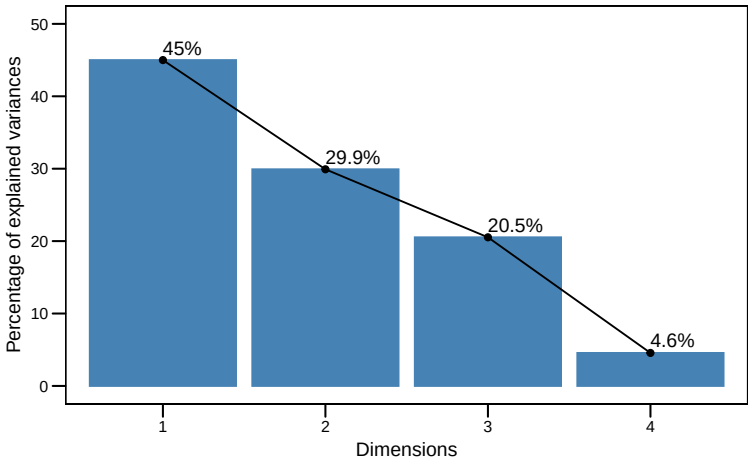

Figure 3.39: Scree plot shows the eigenvalues extracted from PCA analysis performed on five geographical sites measuring the amount of variation retained by each principal component.

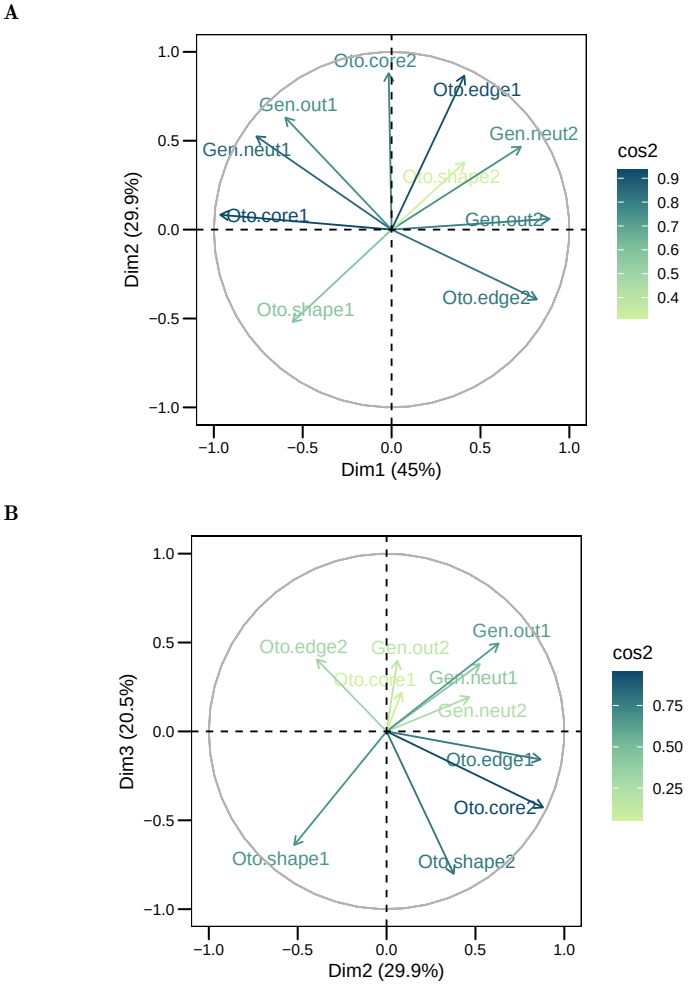

Figure 3.40: Variable correlation plot of PCA analysis performed on five geographical sites showing the relationship between variables with the corresponding cos2 values that expressed the quality of representation of each variable on the principal component. The plots present axes 1 and 2 (A), and axes 2 and 3 (B) representing different combination of the first three principal components.

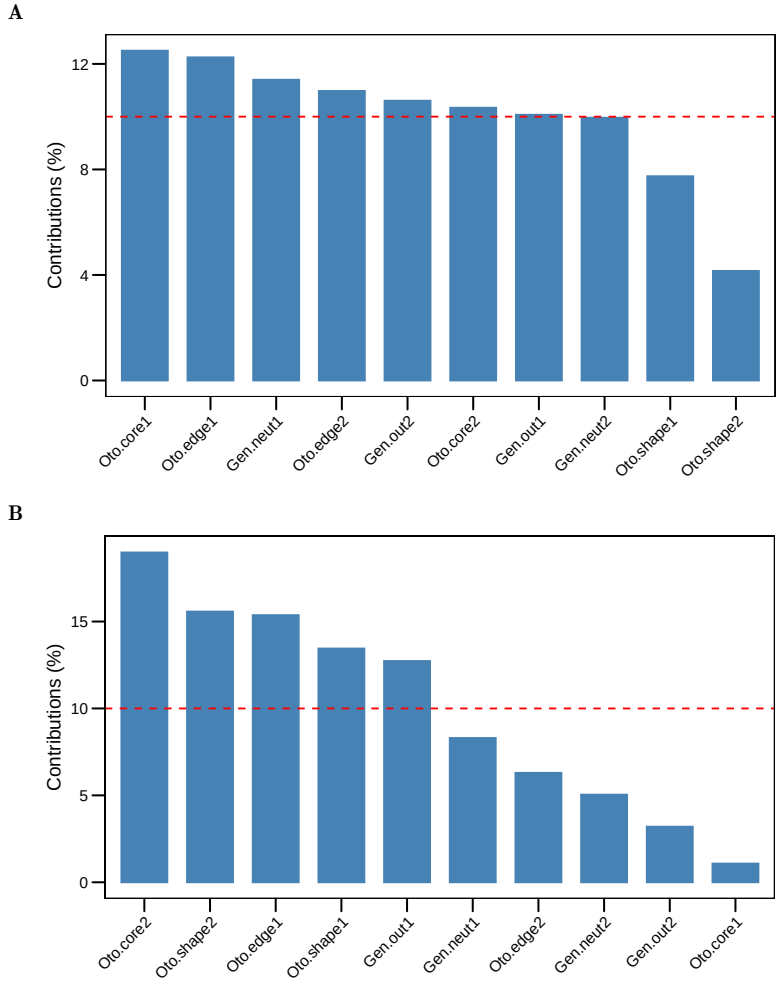

Figure 3.41: Barplots of variable contributions obtained in the PCA analysis using five geographical sites showing the total contribution of each variable to PC1-PC2 (A) and to PC2-PC3 (B), respectively. The red dashed line represents the expected average contribution with a cutoff of 10% considering a uniform contribution of the 10 variables.

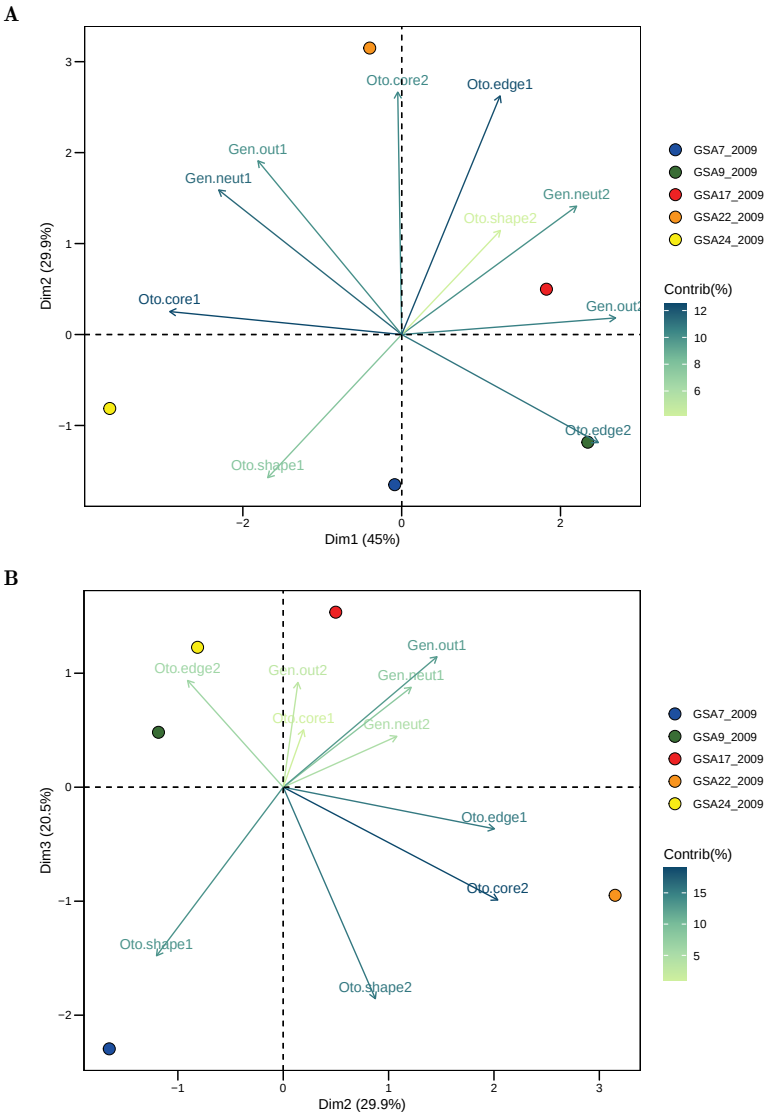

Figure 3.42: Biplot of PCA output displaying differences between five geographical sites and tracer variables contributions to PC1-PC2 (A) and PC2-PC3 (B) showed with arrows. The contribution of each tracer is expressed by percentage (%).

Chapter 4

Seascape genomics approach to describe population structure of marine species: *Merluccius merluccius* case study.

4.1 Introduction

Defining the spatial structure and boundaries of commercial marine species is essential for the sustainable management of fisheries (Cadrin, 2020). Stock units delimitation has been mainly based on economic, social, and political basis instead of biological reasons, which has leaded to a mismatch between the spatial structure of biological populations and currently defined stock units (Reiss et al., 2009). Consequently, stock assessment modelers based the evaluation of stock status on unbiased stock boundaries leading to ineffective management for fisheries resources (Kerr et al., 2017). The consequences of failures in fisheries management could be a depletion of the stocks. Understanding population dynamics and defining the fine-scale population structure of commercially exploited marine species is crucial to enable more robust stock assessments and achieve a successful management for the conservation of resources (Cadrin, 2020; Paris et al., 2018). Molecular techniques, among other methods, have been demonstrated to be a valuable tool for the spatial delineation of stocks and for the assessment of population connectivity in marine species (Ovenden et al., 2015). The distribution of marine species has been adversely affected by climate change (Pinsky et al., 2018). In this respect, the investigation of the influence of environmental factors and spatial extent on genetic variation is a key issue to understand the ability of the species to adapt to environmental changes. Seascape genomics approaches detected both adaptive and neutral genetic patterns in marine species (Liggins et al., 2019).

The present study focuses on European hake *Merluccius merluccius* (Linnaeus, 1758), a demersal species of high interest from a commercial point of view (FAO, 2020b; ICES, 2021). European hake is widely distributed in the Eastern Atlantic Ocean (from Norway to Mauritania) and in the entire Mediterranean Sea

from coastal waters to deeper waters, usually around 200-500 m depths depending on area and life stages (Korta et al., 2015). A formal evaluation of the stock status is available for almost all its habitat distribution. International Council for the Exploration of the Sea (ICES) is responsible for the assessment and management of European hake in the Northeast Atlantic Ocean by assuming two stocks (ICES, 2021). In Southeast Atlantic Ocean, Fishery Committee for the Eastern Central Atlantic (CECAF) and Morocco assessed European hake as one stock. General Fisheries Commission for the Mediterranean (GFCM) of the Food and Agriculture Organization of the United Nations (FAO) managed European hake in the Mediterranean Sea through geographical subareas (GSAs). The assessment of European hake stocks covers a large number of GSAs, aggregating multiple GSAs within benchmark approach (FAO, 2019). Stocks of European hake in the Mediterranean Sea were actually fished at biologically unsustainable levels resulting in an overexploited status (FAO, 2019, 2020b). Because the stock identity of hake inside Mediterranean and Atlantic waters is uncertain without biological evidence, detailed description of population structure could contribute significantly to the definition of sustainable management for European hake. The necessity of defining new assessment and management units for the European hake has been highlighted in various genetic studies. In the Northeast Atlantic Ocean, stock status is actually assessed considering the northern and the southern stock as separated by Capbreton Canyon without biological evidence and with no agreement on the population structure of the European hake (Lundy et al., 1999; Castillo et al., 2005; Pita et al., 2011, 2014). Recent studies applying High Throughput Sequencing technologies suggested that the delimitation of these stock units should be revised (Milano et al., 2014; Westgaard et al., 2017; Leone et al., 2019). Despite several genetic studies provided strong evidence of a separation between European hake populations of Atlantic Ocean and Mediterranean Sea (Cimmaruta et al., 2005; Pita et al., 2010, 2014; Leone et al., 2019), within Mediterranean Sea only outlier loci revealed a strong differentiation among Western, Central and Eastern Mediterranean locations (Milano et al., 2014). In fact, stock units of hake within the Mediterranean is still unclear with limited information for some GSAs (Fiorentino et al., 2014). Recent projects are expected to fill these gaps. Among the others, TRANSBORAN project aimed at investigating the spatial population structure of sardine, European hake, and blackspot seabream in the Alboran Sea and adjacent waters following a multidisciplinary approach. The aim of the project is the identification of stock units to determine if the current GSA boundaries are appropriate spatial scale for the assessment and the management of these species. In this context, the aim of the present research, as part of TRANSBORAN project, is to describe the population structure of European hake within the Alboran Sea and adjacent waters using SNP markers. Furthermore, we also applied a seascape genomics approach to assess the contribution of spatial distribution and environmental drivers in defining neutral and adaptive genetic structure in European hake.

4.2 Materials and Methods

Data and analysis reported in this chapter are carried out in the framework of the TRANSBORAN Project. Population samples were collected from 15 geographical sites within the Alboran Sea, and neighbouring Mediterranean waters and from adjacent Atlantic Ocean in a period of eight weeks between October-November 2018 (Figure 4.43 and Table 4.17). Genotype data at 573 SNP markers were obtained as detailed in the Supplementary materials 4.5.1. Loci and individuals with a call rate below 80% were removed. Monomorphic loci and loci with a minor allele frequency (MAF) lower than 1% were identified using *adegenet* R package (Jombart, 2008) and removed. The final dataset contained 590 individual genotypes at 453 SNPs. Based on geographical position of the 15 sampling sites, we identified seven centroid locations corresponding to a coastal slice defined as the portion of each FAO and GSA area following the continental shelf boundaries within the depth of 500 metres (m) in the Mediterranean Sea and Atlantic Ocean (Casey and Pereiro, 1995; Recasens et al., 1998) (Figure 4.43). The sampling sites were associated to the nearest centroids for the seascape analyses. We used the geographical coordinates (latitude and longitude) of the centroid for all analysis (Figure 4.43).

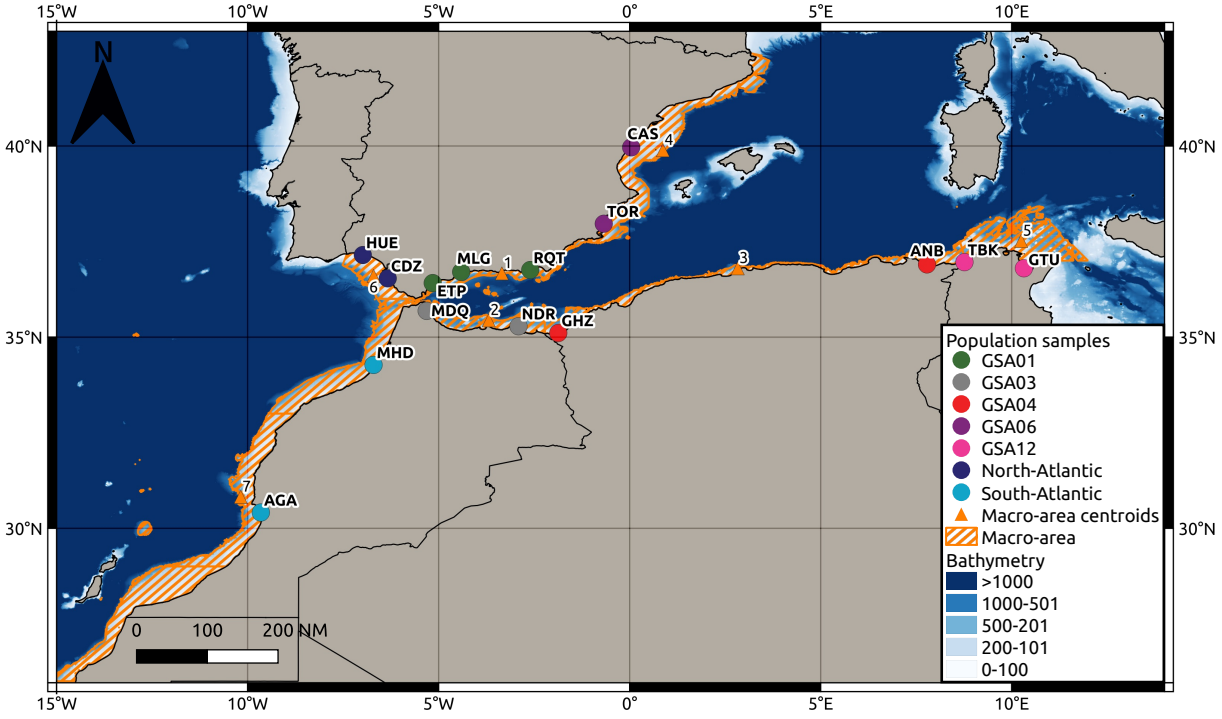

Figure 4.43: Map showing the 15 sampling sites located in the Alboran Sea and adjacent waters with the corresponding GSAs, the identified seven macro-areas, their corresponding centroids and bathymetry.

Sampling site	Site ID	FAO Major Fishing Areas	Macro-area	N
Agadir (Morocco)	AGA	Morocco Coastal (Division 34.1.1)	7 (South Atlantic)	40
Mehdia (Morocco)	MHD	Morocco Coastal (Division 34.1.1)	7 (South Atlantic)	40
Huelva (Spain)	HUE	Portuguese Waters - East (Division 27.9.a)	6 (North Atlantic)	40
Cadiz (Spain)	CDZ	Portuguese Waters - East (Division 27.9.a)	6 (North Atlantic)	40
Estepona (Spain)	ETP	Balearic (Division 37.1.1)	1 (GSA01)	40
Malaga (Spain)	MLG	Balearic (Division 37.1.1)	1 (GSA01)	33
Roquetas (Spain)	RQT	Balearic (Division 37.1.1)	1 (GSA01)	38
M'Diq (Morocco)	MDQ	Balearic (Division 37.1.1)	2 (GSA03)	37
Nador (Morocco)	NDR	Balearic (Division 37.1.1)	2 (GSA03)	40
Ghazaouet (Algeria)	GHZ	Balearic (Division 37.1.1)	3 (GSA04)	41
Annaba (Algeria)	ANB	Balearic (Division 37.1.1)	3 (GSA04)	41
Tabarka (Tunisia)	TBK	Balearic (Division 37.1.1)	5 (GSA12)	40
Gulf of Tunis (Tunisia)	GTU	Balearic (Division 37.1.1)	5 (GSA12)	40
Torrevieja (Spain)	TOR	Balearic (Division 37.1.1)	4 (GSA06)	40
Castellon (Spain)	CAS	Balearic (Division 37.1.1)	4 (GSA06)	40

Table 4.17: Summary information of sampling site locations of *Merluccius merluccius* population samples. Table reports information on geographical areas defined for the analysis and the number of individuals (N) analysed.

Linkage disequilibrium was tested in each sample using the probability test implemented in GENEPOP 4.0 (Rousset, 2008). To assess genetic variation between samples, observed (Ho) and expected (He) heterozygosity, and F_{IS} estimate (Weir and Cockerham, 1984) were calculated using GENETIX 4.0.5 (Belkhir, 2004) and GENEPOP 4.0 (Rousset, 2008), respectively. Global and by locus Hardy-Weinberg (HW) tests were conducted with GENEPOP 4.0 using the exact test. Outlier loci, potentially under selection, were identified using independent approaches namely Bayescan version 2.1 (Foll and Gaggiotti, 2008) and *pcadapt* R package (Luu et al., 2017). Bayescan relies on differences in allele frequencies between subpopulations to identify candidate loci under selection. Subpopulation specific F_{ST} coefficients are divided into two components by logistic regression. The population-specific component β is shared by all loci, and the locus-specific component α is shared by all populations. When the locus-specific component α is significantly different from zero means that the component is necessary to explain the diversity pattern and points out a departure from neutrality at the given locus. Specifically, positive values of alpha indicate diversifying selection whereas negative values suggest balancing/purifying selection. For each locus, a posterior probability is computed and SNPs with prior odds (PO) greater than a threshold are considered outliers since PO express how likely the model is with selection compared to neutral model. Here, we ran the program with 50,000 burn-in period, 100,000 number of iterations, 20 pilot runs and thinning interval size set to 50. We set prior odds (PO) to 10 since the data set comprised less than 1,000 SNPs and we set the false discovery rate (FDR) to 5% to identify putative outlier loci. The method implemented in *pcadapt* R package relies on PCA to detect population structure, and it allows the identification of genetic markers potentially involved in biological adaptation considering their relationship with the population structure. First, a PCA is used to ascertain population structure and identify the number of principal components K to consider. Then, each SNP is regressed by the K principal components. The results are reported as a vector of z -scores between each SNP, and the first K components obtained with the linear regression. Based on this vector, the outliers are then identified using a multi-dimensional approach. The Mahalanobis distance is used as a statistic to measure the distance between each point and its mean. It calculates the distance between the covariance matrix of the z -scores and the vector of the z -score means. Finally, Mahalanobis distances are transformed in p-values based on the correlations between SNPs and the K principal components to performed multiple hypothesis testing. The q-value procedure is recommended to choose a threshold for p-values FDR approach (Benjamini and Hochberg, 1995). The *pcadapt* function with a large number of PCs (K=20) was applied to the dataset to identify the optimal number of principal components K by using scree plot (Cattell's rule) or scoreplot (population structure). As a result, *pcadapt* function was run with the optimal K (K=4) by using Mahalanobis method and a threshold for MAF of 0.05 selecting $\alpha = 0.1$ for the threshold of putative outlier SNPs. We then evaluated the overlap in loci to create datasets of putative neutral and outlier SNPs that could suggest different patterns of genetic divergence. We used Venn diagram to visualize SNPs that were common to each method. To create the outliers dataset, we kept the union of all SNPs identified as outliers

by each method. Then, we removed those SNPs from the whole dataset to obtain the neutral dataset. We performed the following genetic analysis on the whole, neutral and outlier SNPs datasets using the resulting datasets of SNPs to compare potential alternative patterns of divergence. We estimated global and pairwise values of F_{ST} index (Weir and Cockerham, 1984) and relative significance using the exact test implemented in GENEPOP 4.0 to assess genetic differentiation among populations. To investigate population structure, we ran STRUCTURE v.2.3.4 (Pritchard et al., 2000) using the datasets of whole, neutral and outlier SNPs as input. The software was run assuming admixture, correlated allele frequency and testing the scenarios without prior information on their geographical locations. We ran the analysis for K from 1 to 10 with 500,000 iterations and burn-in period at 50,000. STRUCTURE HARVESTER online (Earl and VonHoldt, 2012) was used to evaluate the best K was performed according to Pritchard's criterion (Pritchard et al., 2000). CLUMPP 1.1 (Jakobsson and Rosenberg, 2007) was utilised to align and merge the multiple runs for each K and DISTRUCT 1.1 (Rosenberg, 2004) to visualise the posterior membership probability of each individual to clusters from 1 to 10. Using the groups identified by STRUCTURE analysis, a locus-by-locus analysis of molecular variance (AMOVA, n= 1000 permutations) was carried out using ARLEQUIN 3.5. (Excoffier and Lischer, 2010). To validate the results, we performed a principal coordinate analysis (PCoA) using GENALEX 6.5 (Peakall and Smouse, 2012). PCoA is a multivariate technique that allows to investigate relative genetic distances among population samples and to visualise the major patterns within multiple populations samples. We computed the method using the matrix of pairwise codominant genetic distance among all pairs of population samples, with the default options of Triangular Distance Matrix and Covariance-Standardized. TESS3 algorithm, combining genetic and geographic data to compute spatial ancestry coefficients, was also used to estimate and visualize population structure (Caye et al., 2016, 2018). The method uses individual spatial coordinates to better discriminate among putative populations using a graph based non-negative matrix factorization (Caye et al., 2016). The data were converted to tess3 matrix format using *tess2tess3* function implemented in *tess3r* R package (Caye et al., 2018). The algorithm was run by using *tess3* function to estimate spatial population structure for K values from 1 to 10 with 20 repetitions for each K. To identify the K that better described the data cross validation score was inspected, and ancestry coefficients were analysed only for values of K from 2 to 4. The corresponding Qmatrix of each K was visualized by using *plotQ* function implemented in *pophelper* R package. The values of the Qmatrix were interpolated on the geographic map of the study area using *ggtess3Q* function, showing the distribution of gene pools over the seascape.

We represented the spatial structure between geographical sites with distance-based Moran's eigenvector map (dbMEMs) variables obtained through a distance matrix. In order to determine the dbMEMs, we first calculated the spatial distance matrix from geographical coordinates (longitude and latitude of centroids) containing the distances between the locations using the *gcd.hf* function. Then, we computed the distance-based Moran Eigenvector's Maps (dbMEMs) using the *pcnm* function in *vegan* R package (Oksanen et al.,

2022). In this way, we were able to use spatial variables (dbMEMs) as independent vectors, representing the spatial structure associated with the neighbourhood network across different scales, as explanatory variables in the following analysis (Borcard and Legendre, 2002). Environmental variables were downloaded from E.U. Copernicus Marine Service Information according to their availability over the analysis period from 2014 to 2018 (CMEMS, `http://marine.copernicus.eu/`). From the available products we selected three environmental variables, as reported in Supplementary materials, based on their influence on the distribution of European hake and their variation between sampling locations (Milano et al., 2014; Coll et al., 2016). To detect multicollinearity, we applied the function *vifstep* in *usdm* R package (Naimi, 2017; Naimi et al., 2014) in which highly correlated variables with Variance Inflation Factor (VIF) values greater than 10 were excluded. In our seascape genomic analysis we considered as environmental factors the resulting set of two out of three variables (salinity (sal, psu) and chlorophyll concentrations (chl, milligram m^{-3}) and as spatial variables the resulting set of three out of four dbMEMs after multicollinearity analysis using *vifstep* function. We evaluated the effect of spatial distribution and/or environmental factors on genetic structure at seven macro-areas using a redundancy analysis (RDA), a constrained ordination method used in multivariate analysis (Legendre and Legendre, 2012). We first performed three distinct RDAs using both whole, neutral and outlier SNPs as response variable and we combined the spatial factors (dbMEMs) and the environmental factors as explanatory variables. Then, we performed a partial RDA both on whole, neutral and outlier SNPs to assess only the influence of environmental factors excluding the effect of the spatial factors on the genetic differentiation and vice versa. All RDAs were performed using *rda* function in *vegan* R package (Oksanen et al., 2022). Following Selmoni et al. (2020) we extracted the first principal component that reaches the 80% of cumulative variance to use them as response variables to compute RDA. The principal components were produced performing a PCA on the standardized matrix using the *prcomp* function, after transforming the matrix of allele frequencies with the Hellinger approach using the *decostand* function. We applied the ANOVA and marginal ANOVAs with 1,000 permutations to assess the significance of the global model and of each variable. We calculated the adjusted coefficient of determination (R^2_{adj}) using the *RSquareAdj* function in *vegan* R package to determine the variation explained by the model (Oksanen et al., 2022). Then, we used the *ordistep* function with 1,000 permutations to perform both forward and backward selection of explanatory variables that best explained the variability of the response variable (optimal model). Also in this case, we evaluated the significance of the model and each variable as explained above. Finally, we used the *varpart* function to partition the variation in genetic data with respect to environmental and spatial variables. We assessed the proportion of variability explained by each set of variables through the adjusted coefficient of determination (R^2_{adj}) and Venn's diagrams.

4.3 Results

We found significant linkage disequilibrium among 15 pairs of loci considering results over all population samples ($p < 0.05$ after FDR correction). Over all loci, there was a significant heterozygosity deficit (after FDR correction) among all population samples. Therefore, we removed all loci that in at least one sample were involved in linkage disequilibrium and presented departure from Hardy-Weinberg equilibrium. The whole resulting dataset employed in all downstream analyses was constituted by 261 SNP loci. F_{IS} values ranged between -0.02 for MDQ and 0.04 for NDR and GHZ (Table 4.18). Over all loci, significant heterozygosity deficits were found in 3 population samples (NDR, GHZ and TOR) out of 15 (after FDR correction) (Table 4.18). Observed (*Ho*) and expected (*He*) heterozygosity varied between 0.30 for ANB, GTU, TOR and CAS and 0.33 for AGA, ETP, MLG, MDQ and GHZ and between 0.30 for GTU and 0.34 for ETP and GHZ, respectively, with a mean value of 0.32 in both cases (Table 4.18).

Sample	Ho	He	F_{IS}
AGA	0.33	0.33	0.01
MHD	0.32	0.32	0.01
HUE	0.32	0.33	0.02
CDZ	0.32	0.32	0.02
ETP	0.33	0.34	0.02
MLG	0.33	0.33	0.02
RQT	0.32	0.32	-0.01
MDQ	0.33	0.33	-0.02
NDR	0.32	0.33	0.04*
GHZ	0.33	0.34	0.04*
ANB	0.3	0.31	0.01
TBK	0.32	0.32	-0.01
GTU	0.3	0.3	0.02
TOR	0.3	0.31	0.03*
CAS	0.3	0.31	0.02

Table 4.18: Genetic diversity statistics for each population samples of *Merluccius merluccius* including mean observed (*Ho*) and expected (*He*) heterozygosity, and F_{IS} estimation with significant values (*) ($p < 0.05$ after FDR correction). Code of population samples are given in Table 4.17

Methods for outlier detection identified different numbers of loci as outliers: *pcadapt* identified 55 outlier SNPs and *Bayescan* identified 59. The panel of SNPs identified by both methods resulted in an outlier dataset of 31 SNP loci. The neutral dataset included the 178 markers which were not detected as outlier, considering both balancing and diversifying outliers, by any of the two methods. Pairwise F_{ST} values ranged from -0.0036 (MLG vs. GHZ) to 0.1196 (CDZ vs. GTU), from -0.0034 (MLG vs. GHZ) to 0.0188 (CDZ vs. GTU) and from -0.0061 (MLG vs. GHZ) to 0.3729 (CDZ vs. GTU), with an overall significant F_{ST} of 0.0377, 0.0047 and 0.1360 for whole, neutral and outlier datasets respectively ($p < 0.001$). Pairwise comparisons of genetic differentiation displayed a high degree of structure in the whole and outlier datasets with 80 and 91 out of 105 significant tests respectively (Suppl. Table 4.20). Whereas the level of structure was moderate for neutral dataset with 21 out of 105 significant tests (Suppl. Table 4.20). Atlantic population samples were found genetically differentiated from Mediterranean population samples for all datasets. However, only for whole and outlier dataset, Atlantic population samples exhibited significant levels ($\alpha = 0.05$) of genetic divergence from all Mediterranean population samples. Furthermore, outlier dataset unveiled significant levels ($\alpha = 0.05$) of genetic differentiation within Atlantic population samples between AGA and HUE, AGA and CDZ, and MHD and CDZ. The levels of genetic differentiation, based on significant pairwise F_{ST} comparisons, between population samples inside the Mediterranean Sea varied considerably among the different datasets. Detailed results are reported in Table 4.20 in the Supplementary materials. The clustering pattern of population samples resulting from STRUCTURE analysis was concordant with the structure suggested by the pairwise F_{ST} (Figure 4.44). Based on Pritchard's criterion, two clusters were identified for the whole and neutral datasets (Figure 4.48A, 4.48B). The first cluster grouped the Atlantic, North Alboran (except RQT), South Alboran population samples and GHZ, while the second cluster grouped RQT, ANB, North Tunisia and North Spain population samples (Figure 4.44A, 4.44B). For the outlier dataset three clusters were identified (Suppl. Figure 4.48C) with the cluster one that grouped the same population samples as in the two other datasets, while the second previous cluster is divided into two different clusters: one grouping RQT and North Spain population samples and another grouping ANB and North Tunisia population samples (Figure 4.44C). PCoA analysis also confirmed STRUCTURE results (Figure 4.45). The first principal coordinate showed a separation between Atlantic, North Alboran (except RQT), South Alboran and GHZ population samples from RQT, ANB, North Tunisia and North Spain population samples with 83.4, 72.3 and 83% of variance for whole, neutral and outlier datasets respectively. North Spain and RQT population samples were separated from ANB and North Tunisia population samples along the second principal coordinate accounting for 9.9, and 12.8% of the variance for the whole and outlier datasets respectively (Figure 4.45A, 4.45C). Based on cross-validation criterion for TESS3, three main ancestral groups (Figure 4.46A and Suppl. Figure 4.49A, 4.50A) were identified for whole dataset corresponding to the following geographical clines: (1) Atlantic, North Alboran (except RQT) and South Alboran samples and GHZ; (2) ANB and North Tunisia samples; (3) RQT and North Spain samples. Two main ancestral groups (Figure 4.46B and Suppl. Figure 4.49B,

4.50B) were identified for neutral dataset corresponding to: (1) Atlantic, South Alboran samples and GHZ; (2) ANB, North Alboran, North Tunisia and North Spain samples. For outlier dataset, a subtle substructure were detected with four ancestral groups identified that correspond to the following geographical area: (1) Atlantic samples (except AGA); (2) AGA and North and South Alboran samples; (3) ANB and North Tunisia samples; (4) North Spain samples (Figure 4.46C and Suppl. Figure 4.49C, 4.50C).

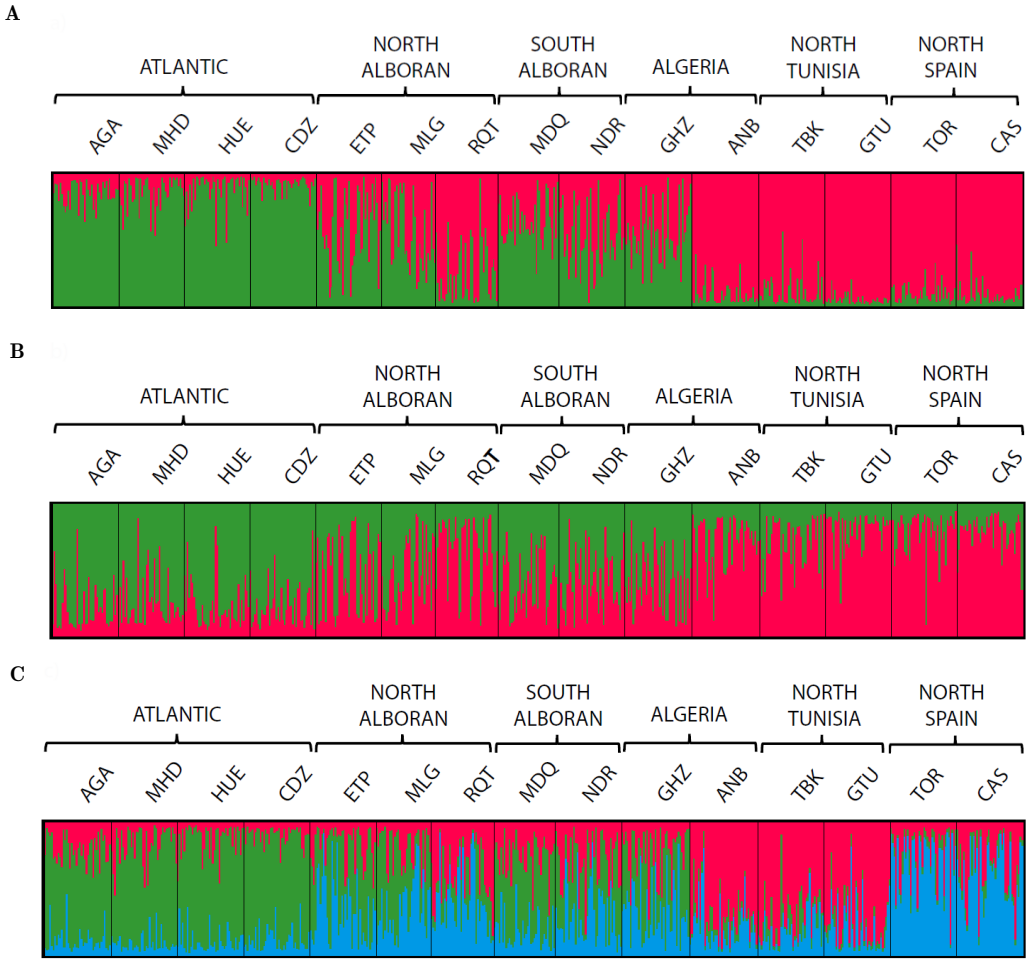

Figure 4.44: Graphical representation of the Bayesian clustering analysis of STRUCTURE for whole (A), neutral (B) and outlier (C) SNPs datasets. Code of population samples are given in Table 4.17.

The global RDA model performed on whole, neutral and outlier SNPs identified significant patterns of variation ($p < 0.05$). The percentage of genetic variation explained was 0.73% for the whole dataset of loci, 0.40% for the neutral ones and 5.40% for the outlier ones, with the first two axes accounting respectively for 0.46%, 0.24% and 4.59% of the explained variation. The forward and backward procedure using *ordistep* function selected spatial variables (dbMEM2 and dbMEM3) for the whole and outlier datasets of SNPs and no spatial variables for neutral dataset of loci, while salinity and chlorophyll were identified for all RDA on

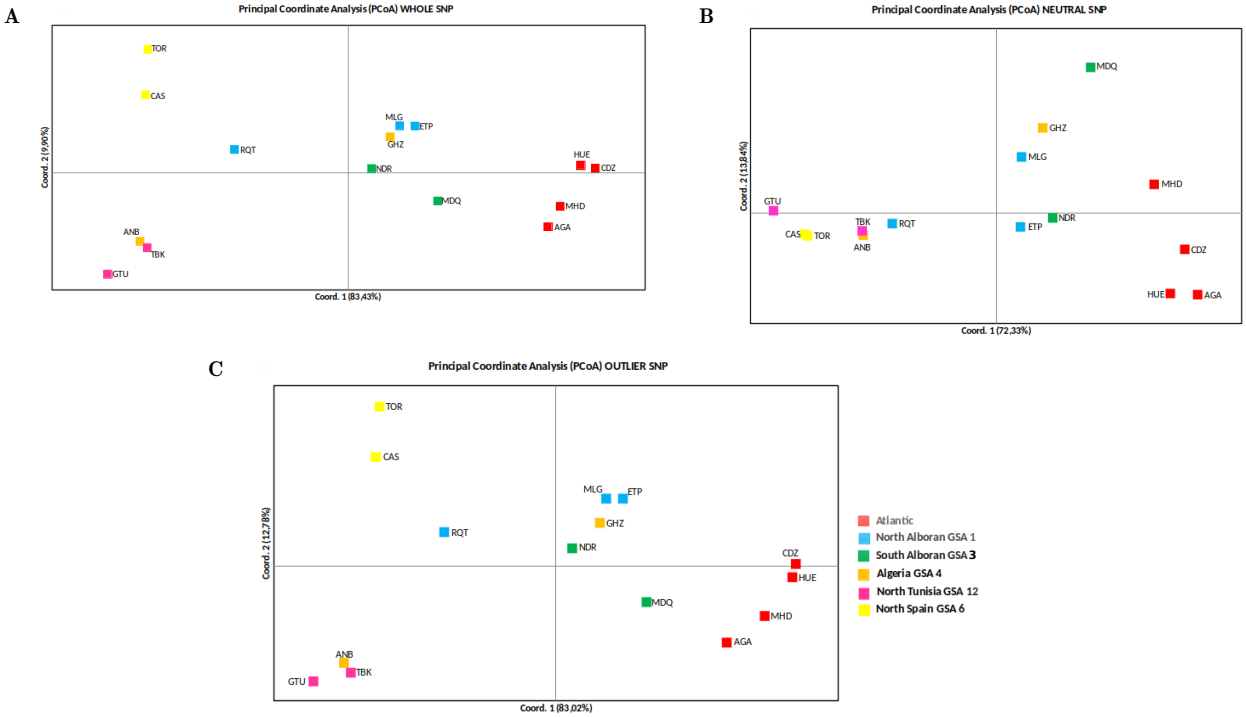

Figure 4.45: Principal coordinate analysis (PCoA) of whole (A), neutral (B) and outlier (C) SNPs datasets. Code of population samples are given in Table 4.17

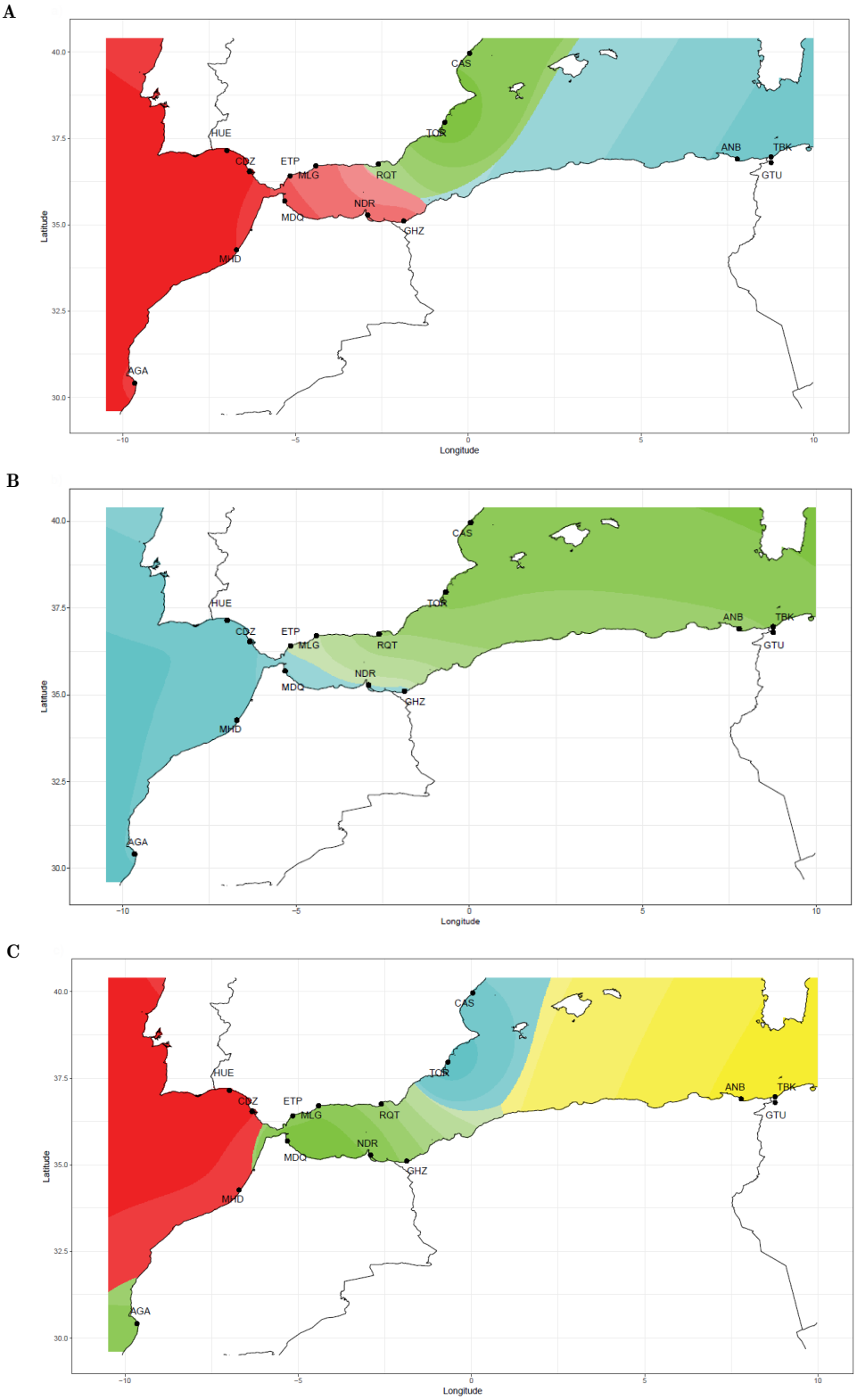

Figure 4.46: Interpolation of the values of the ancestry coefficient on the geographical map of the study area for whole (A), neutral (B) and outlier (C) SNPs datasets. Code of population samples are given in Table 4.17

whole, neutral and outlier SNPs (Figure 4.47, Table 4.19). From marginal ANOVAs, spatial and environmental variables selected were all significant predictors of whole, neutral and outlier genetic variation ($p < 0.01$). The results of variation partitioning based on neutral SNPs indicated that only environmental variables explained the individual fraction of variance, specifically chlorophyll (0.3%) explained approximately the same individual fraction of variance as the salinity (0.1%) with the highest amount of explained variance represented by shared variance (0.6%) (Suppl. Figure 4.51B). Moreover, the environmental variables explained more of the individual fraction of variance in the outlier SNPs (17.1%), whereas spatial variables explained 3%. Shared variance (3%) resulted in the same amount of explained variance as spatial variables (Suppl. Figure 4.51C). As for outlier SNPs, the environmental variables explained more of the individual fraction of variance in the whole SNPs (5.2%), whereas spatial variables explained 0.7% (Suppl. Figure 4.51A). Shared variance (0.8%) resulted in the same amount of explained variance as spatial variables. Partial RDA performed with spatial variables as condition confirmed that most of the genetic variation for all datasets, particularly for outlier dataset, was mainly explained by environmental variables. Especially, in the partial RDA performed on whole dataset, the global model considering as condition the spatial variables explained 0.47% of the genetic variation, while using the environmental variables as condition the model explained 0.27% of the genetic variation. When we assessed the effect of environmental variables on whole SNPs while controlling for the effect of spatial variables, sal and chl were significant variables. On the contrary, controlling the effect of environmental variables, spatial variables (dbMEM2, dbMEM3) were significant. (Suppl. Figure 4.52 and Table 4.19). In the partial RDA performed on neutral dataset, the global model considering as condition the spatial variables explained 0.39% of the genetic variation, while using the environmental variables as condition the model did not explain genetic variation. When we assessed the effect of environmental variables on neutral SNPs while controlling for the effect of spatial variables, sal and chl were significant variables. (Suppl. Figure 4.53 and Table 4.19). In the partial RDA performed on outlier dataset, the global model considering as condition the spatial variables explained 3.13% of the genetic variation, while using the environmental variables as condition the model explained 1.54% of the genetic variation. When we assessed the effect of environmental variables on outlier SNPs while controlling for the effect of spatial variables, sal and chl were significant variables. On the contrary controlling the effect of environmental variables, spatial variables (dbMEM2, dbMEM3) were significant (Suppl. Figure 4.54 and Table 4.19).

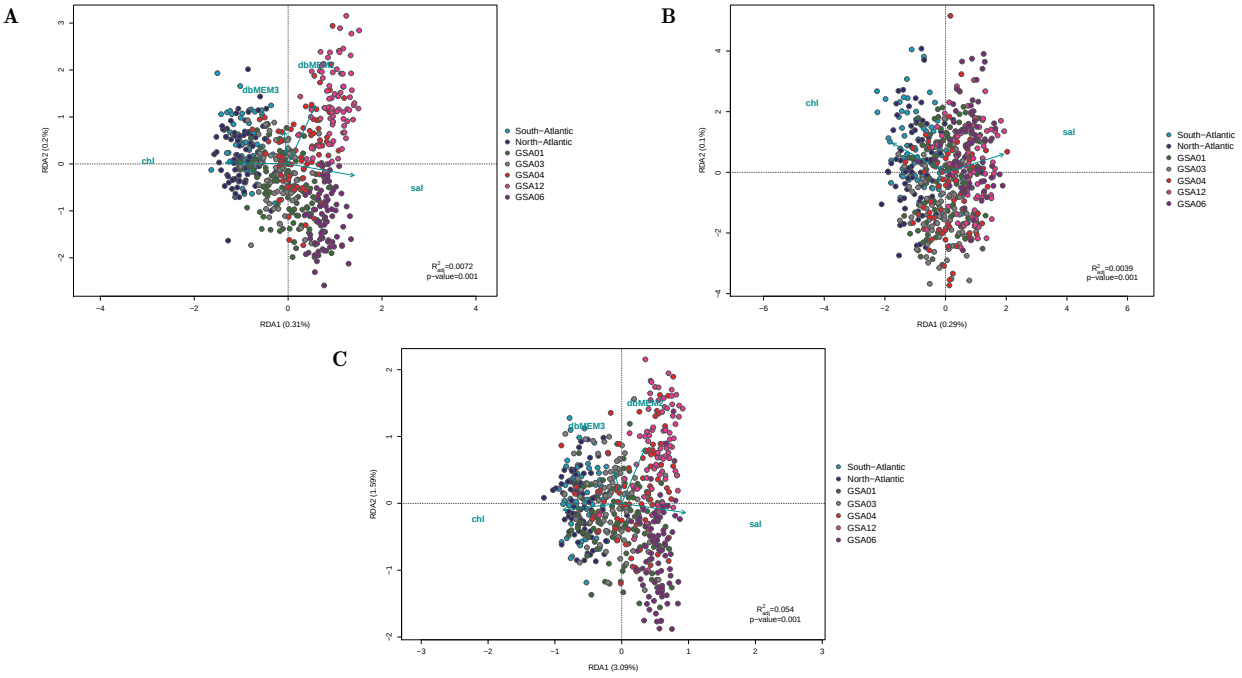

Figure 4.47: Redundancy analysis (RDA) performed on whole (A), neutral (B) and outlier (C) SNPs datasets following the forward and backward selection of variables using *ordistep* function. Significant variables found as significant predictors are shown with arrows. Code of population samples are given in Table 4.17.

		Selected variables (ordistep function)			
SNPs	Analysis	Environmental	Spatial	p	R^2_{adj}
Whole	RDA			0.0010	0.0072
		sal***			
		chl**			
			dbMEM2***		
			dbMEM3**		
	partial RDA			0.0010	0.0047
		sal***			
		chl***			
	partial RDA			0.0010	0.0027
			dbMEM2***		
			dbMEM3***		
Neutral	RDA			0.0010	0.0039
		sal***			
		chl*			
	partial RDA			0.0010	0.0039
		sal***			
		chl*			
Outlier	RDA			0.0010	0.0540
		sal***			
		chl***			
			dbMEM2***		
			dbMEM3**		
	partial RDA			0.0010	0.0375
		sal***			
		chl***			
	partial RDA			0.0010	0.0227
			dbMEM2***		
			dbMEM3***		
			dbMEM4**		

Significant explanatory variables are indicated with the following symbols:

* $p < 0.05$

** $p < 0.01$

*** $p = 0.001$

Table 4.19: Results from RDA and partial RDA for each response variables divided in whole, neutral and outlier SNPs datasets.

4.4 Discussion

The identification of existing misalignment of current management units with the biological units can be attained with the understanding of population structure of marine species (Kerr et al., 2017). In the present study we applied a molecular approach using SNP loci as genetic markers to examine population structure of European hake in the Alboran Sea and adjacent waters. The current work confirmed the presence of considerable levels of differentiation in European hake populations in the Alboran Sea and among adjacent waters. The level of genetic variability, based on observed and expected heterozygosity, was found in the same order of magnitude as in previous studies that used SNP markers (Milano et al., 2014; Westgaard et al., 2017; Leone et al., 2019).

The application of different approaches for population differentiation provided inferences related to both demographic processes indicated by variation at the neutral SNP loci and adaptive processes identified by variation at outlier SNP loci. This strategy enabled to detect genetic variation over shorter distances compared to other studies. Indeed, the use of different datasets of SNPs loci unveiled a strong differentiation among geographical samples in the European hake within the Alboran Sea and adjacent waters. The presence of a west-to-east genetic cline and the separation between Atlantic and Mediterranean population samples with Alboran Sea resulting as a transition area was mainly identified by using the whole genetic dataset as well as the selected dataset of only neutral markers. A further detection of a fine scale segregation patterns was pointed out by the analysis conducted on outlier SNPs loci that demonstrated a higher resolution in revealing finer spatial scale patterns compared to neutral markers. In details, the analysis based on outlier loci resulted in discriminating the European hake samples from the North Spain from the East Algeria-Tunisian samples. The results obtained from previous studies on European hake clustered Atlantic population samples of Morocco (Roldán et al., 1998), Portugal (Lundy et al., 1999; Castillo et al., 2004) or Spain (Tanner et al., 2014; Castillo et al., 2005) closer to Mediterranean population samples than to Northeast Atlantic population samples from Galician shelf area to up Northern Atlantic areas. This reported pattern of differentiation is in line with our results. Despite some genetic differentiation between Atlantic and Mediterranean population samples, our findings suggested that Alboran Sea population samples were genetically closer to Atlantic population samples than to the remaining Mediterranean population samples.

Several studies argued that the Gibraltar Strait acts as a geographical barrier to gene flow, separating the populations structure of European hake across Atlantic Ocean and Mediterranean Sea, with limited connectivity between population samples (Lundy et al., 1999; Castillo et al., 2004, 2005; Pita et al., 2010). Nevertheless, none of these studies conducted simultaneous analyses with population samples on either side of the Gibraltar Strait, both in northern-southern and eastern-western directions, and neither in the close vicinity of the Gibraltar Strait.

Our results, with a more exhaustive sampling design across the Gibraltar Strait, supported the existence

of ongoing gene flow from the Atlantic Ocean to the Mediterranean Sea with a west-to-east genetic gradient from the Atlantic component into the Mediterranean Sea. As previously suggested by Milano et al. (2014), the Alboran Sea resulted as a transition zone between Atlantic and Mediterranean water bodies with a limited differentiation between the northern and the southern area only detected by neutral marker. However, whether this gene flow is the result of passive dispersal of eggs and larvae, or the active migration of juvenile or adult must be determined. Considering the Mediterranean Sea area, a limited gene flow was detected between Alboran Sea and Eastern Mediterranean Sea waters because of the genetic divergence observed both between Alboran Sea and North Tunisia, and North of Spain population samples. The Almeria Oran front (Tintore et al., 1988) has been widely reported as a genetic break for several marine species (Patarnello et al., 2007), including European hake (Cimmaruta et al., 2005; Pita et al., 2014). The role of Almeria-Oran front in acting as an ecological barrier or as a connectivity moderator is not so clear due to differentiation patterns observed between neighbouring GSAs. In the case of Algeria population samples, Annaba showed similarities with Tunisia population samples while Ghazaouet showed similarity with South Alboran population samples. The same occurred with Roquetas population sample of North Alboran Sea located in a transitional area between North Alboran Sea and North Spain that was genetically closer to North Spain population samples than to North Alboran ones. Indeed, these patterns of differentiation within and among GSAs are probably due to the proximity of these population samples to the GSAs boundaries. The exact location of the genetic pattern of segregation compared to the Almeria-Oran front should be further investigated with an extensive sampling design covering the whole GSAs taking also into consideration the fisheries ground for each GSA.

Furthermore, in this study, a first attempt to investigate certain factors that can explain the observed pattern of genetic variation was made. Indeed, Milano et al. (2014) found significant correlations between allele frequencies of several SNP outlier loci and seawater surface temperature and salinity and argued that European hake populations might be adapted to local conditions. A previous study based on allozyme loci showed a strong correlation between genetic variation and salinity and temperature values, suggesting that these two environmental factors may play a role for selective processes and maintaining the genetic structure of European hake (Cimmaruta et al., 2005). The seascape analysis performed to investigate the influence of environmental variables on population structure of European hake underlined those environmental factors such as salinity and chlorophyll better explained the genetic variation as opposed to spatial factors suggesting their role in shaping the genetic population structure. However, a larger proportion of the genetic variation was explained by clusters in the outlier loci dataset, confirming the higher power of resolution of these markers than neutral ones to detect fine scale population structure.

Whatever the factors underlying the evidence of genetic structure across European hake populations across its distributional range, it should be noted the power of outlier loci in revealing segregation patterns over shorter distances. The complex spatial structure and genetic connectivity among and within GSAs of Alboran Sea and Western Mediterranean Sea should be taken into account in the definition of management

units for the effective integration of biologically based information into management. The limit between GSAs should be further investigated. Therefore, the assessment and management of European hake within the western Mediterranean Sea should be carried out jointly among countries as already suggested (Benchoucha et al., 2012). Finally, the ongoing gene flow between the Atlantic and Mediterranean populations of the European hake suggests that a cohesive stock management of this species among the ICES, CECAF and GFCM should be considered.

4.5 Supplementary materials

4.5.1 Molecular data

Sampling, DNA extraction and SNP genotyping procedure reported here was performed in the framework of TRANSBORAN project. A muscle sample was taken from around 40 European hake individuals per location and stored in non-denaturated ethanol 96% (Figure 4.43 and Table 4.17). DNA was extracted from 593 tissue samples using the pureLink DNA Invitrogen Kit, following the manufacturer's protocol. A custom Genotyping By Sequencing (GBS) assay was developed by Thermo Fisher Scientific, from a set of 917 SNP of European hake which were selected from (Milano et al., 2014; Leone et al., 2019), using the following criteria and ranking: (1) outlier loci, (2) significant P value and common to Atlantic and Mediterranean populations, (3) significant P value for Atlantic populations and (4) higher FST values. This set of markers was passed through a Thermo Fisher Scientific's quality control process. The quality check was performed using European Hake, *Merluccius merluccius*, reference genome (GenBank assembly accession: GCA_900312545.1). 591 SNP markers which passed the quality step were then submitted to AgriSeqTM primer design phase. The primer designs were in-silico checked for specificity and sensitivity of the intended target/marker regions using hake reference genome. A total of 574 SNPs were designed and were contained within 554 amplicons/target regions. Next, targeted sequencing was performed using the developed custom GBS hake panel. The DNA samples were prepared for sequencing using the AgriSeqTM HTS Library Kit (Applied Biosystems). DNA samples were quantified using the Quant-iT DNA Assay kit (Thermo Fisher Scientific) on the Fluoroskan™ Microplate Fluorometer. DNA concentrations were normalized to 3.3 ng/µl for a total of 10 ng DNA per 10 µl reaction. Normalized DNA was combined with the Ion AgriSeq primer panel and AgriSeq amplification master mix. For amplification of genomic targets, the following thermocycling programs were used: 99°C for 2 min, then 16 cycles of 99°C for 15s and 60°C for 4 min. Amplicons were then prepared for ligation with pre-ligation enzyme digestion at 50°C for 10 min, 55°C for 10 min, and 60°C for 20 min. IonCode TM Barcode Adapters 1-1248 kit were ligated to the digested products with barcoding enzyme and buffer. Labelled amplicons were then pooled, cleaned up, amplified and normalized. Following library preparation, libraries were loaded onto an Ion 540 TM sequencing Chip Kit via the Ion 540TM Kit-Chef and

Ion Chef. Sequencing was then performed on the Ion S5 system (Thermo Fisher, Inc. Waltham, MA). A total of 73M of reads per sequencing run were generated. The reads were then de-multiplexed to individual samples using barcode sequences. For each sample, the sequenced reads from the targeted regions were then mapped to the hake reference genome using TMAP- Torrent Mapping Alignment Program followed by genotyping using TVC-Torrent Variant Caller. The genotypes were reported in different formats TOP, TOP/BOT and actual alleles using AgriSum Toolkit.

4.5.2 Environmental variables

Monthly mean values of physical and biochemical environmental variables were downloaded from E.U. Copernicus Marine Service Information (`https://marine.copernicus.eu/`). Salinity (sal, psu) and temperature (temp, °C) data were extracted from the Global Ocean Physics Reanalysis (`https://doi.org/10.48670/moi-00021`) from 2014 to 2018 on a grid with 1/12° x 1/12° horizontal resolution (approximatively 8 km) and 50 vertical levels of thickness increasing with deph (0-5500 m). Chlorophyll concentrations (chl, milligram m-3) were extracted from the Global Ocean Biogeochemistry Hindcast (`https://doi.org/10.48670/moi-00019`) from 2014 to 2018 on a grid with 1/4° x 1/4° horizontal resolution and 75 vertical levels of thickness increasing with depth (0-5500 m). Average values were extracted over the temporal windows of 5 years (2014-2018) based on macro-areas extension keeping only surface values.

4.5.3 Results

a

WHOLE SNP LOCI		ATLANTIC				NORTH ALBORAN			SOUTH ALBORAN		ALGERIA		NORTH TUNISIA		NORTH SPAIN	
		AGA	MHD	HUE	CDZ	ETP	MLG	RQT	MDQ	NDR	GHZ	ANB	TBK	GTU	TOR	CAS
ATLANTIC	AGA	-														
	MHD	0.0018	-													
	HUE	0.0036	0.0033	-												
	CDZ	0.0078	0.0045	0.0007	-											
NORTH ALBORAN	ETP	**0.0150**	**0.0153**	**0.0147**	**0.0185**	-										
	MLG	**0.0166**	**0.0163**	**0.0167**	**0.0198**	-0.0019	-									
	RQT	**0.0497**	**0.0545**	**0.0602**	**0.0660**	**0.0193**	**0.0143**	-								
SOUTH ALBORAN	MDQ	**0.0112**	**0.0110**	**0.0173**	**0.0172**	0.0051	0.0045	**0.0244**	-							
	NDR	**0.0169**	**0.0188**	**0.0224**	**0.0260**	0.0017	0.0002	**0.0106**	0.0036	-						
ALGERIA	GHZ	**0.0179**	**0.0170**	**0.0180**	**0.0227**	-0.0005	-0.0036	**0.0146**	0.0051	-0.0017	-					
	ANB	**0.0801**	**0.0855**	**0.0950**	**0.1038**	**0.0457**	**0.0410**	**0.0112**	**0.0499**	**0.0314**	**0.0387**	-				
NORTH TUNISIA	TBK	**0.0778**	**0.0840**	**0.0911**	**0.0969**	**0.0419**	**0.0391**	**0.0116**	**0.0451**	**0.0283**	**0.0354**	0.0004	-			
	GTU	**0.0965**	**0.1034**	**0.1118**	**0.1196**	**0.0583**	**0.0534**	**0.0208**	**0.0613**	**0.0422**	**0.0494**	0.0053	0.0026	-		
NORTH SPAIN	TOR	**0.0903**	**0.0933**	**0.0943**	**0.1030**	**0.0387**	**0.0338**	**0.0103**	**0.0567**	**0.0330**	**0.0340**	**0.0207**	0.0214	**0.0291**	-	
	CAS	**0.0856**	**0.0874**	**0.0919**	**0.0992**	**0.0356**	**0.0332**	0.0064	**0.0505**	**0.0294**	**0.0309**	0.0114	0.0124	0.0198	0.0002	-

b

NEUTRAL SNP LOCI		ATLANTIC				NORTH ALBORAN			SOUTH ALBORAN		ALGERIA		NORTH TUNISIA		NORTH SPAIN	
		AGA	MHD	HUE	CDZ	ETP	MLG	RQT	MDQ	NDR	GHZ	ANB	TBK	GTU	TOR	CAS
ATLANTIC	AGA	-														
	MHD	0.0016	-													
	HUE	-0.0033	0.0010	-												
	CDZ	0.0003	-0.0003	0.0002	-											
NORTH ALBORAN	ETP	0.0015	0.0012	0.0011	0.0020	-										
	MLG	0.0043	0.0021	0.0031	0.0034	-0.0005	-									
	RQT	**0.0099**	0.0089	**0.0095**	**0.0092**	0.0015	0.0014	-								
SOUTH ALBORAN	MDQ	0.0065	0.0024	0.0062	0.0039	0.0009	0.0003	0.0075	-							
	NDR	0.0022	0.0009	0.0011	0.0010	-0.0020	-0.0007	0.0031	0.0027	-						
ALGERIA	GHZ	0.0026	0.0014	0.0024	0.0049	-0.0015	-0.0034	0.0030	0.0010	-0.0017	-					
	ANB	**0.0104**	0.0077	**0.0080**	**0.0122**	0.0003	0.0020	0.0004	0.0098	0.0031	0.0033	-				
NORTH TUNISIA	TBK	**0.0122**	0.0088	0.0091	0.0096	-0.0006	0.0008	0.0002	0.0085	0.0025	0.0036	-0.0013	-			
	GTU	**0.0185**	**0.0180**	**0.0164**	**0.0188**	0.0056	0.0063	0.0023	0.0127	0.0075	0.0067	0.0026	-0.0011	-		
NORTH SPAIN	TOR	**0.0152**	**0.0119**	**0.0126**	**0.0139**	0.0027	0.0035	-0.0003	**0.0103**	0.0048	0.0062	-0.0026	-0.0018	-0.0019	-	
	CAS	**0.0158**	**0.0115**	**0.0135**	**0.0143**	0.0017	0.0057	0.0012	**0.0110**	0.0061	0.0060	-0.0024	-0.0008	-0.0007	-0.0019	-

c

OUTLIER SNP LOCI		ATLANTIC				NORTH ALBORAN			SOUTH ALBORAN		ALGERIA		NORTH TUNISIA		NORTH SPAIN	
		AGA	MHD	HUE	CDZ	ETP	MLG	RQT	MDQ	NDR	GHZ	ANB	TBK	GTU	TOR	CAS
ATLANTIC	AGA	-														
	MHD	0.0070	-													
	HUE	**0.0207**	0.0122	-												
	CDZ	**0.0359**	**0.0166**	-0.0007	-											
NORTH ALBORAN	ETP	**0.0525**	**0.0602**	**0.0626**	**0.0646**	-										
	MLG	**0.0525**	**0.0640**	**0.0731**	**0.0750**	-0.0047	-									
	RQT	**0.1592**	**0.1917**	**0.2152**	**0.2251**	**0.0707**	**0.0559**	-								
SOUTH ALBORAN	MDQ	**0.0237**	**0.0336**	**0.0473**	**0.0493**	**0.0208**	**0.0147**	**0.0845**	-							
	NDR	**0.0590**	**0.0806**	**0.0924**	**0.0965**	0.0073	0.0008	**0.0372**	**0.0128**	-						
ALGERIA	GHZ	**0.0548**	**0.0667**	**0.0747**	**0.0753**	-0.0008	-0.0061	**0.0503**	0.0100	-0.0047	-					
	ANB	**0.2533**	**0.2957**	**0.3279**	**0.3359**	**0.1755**	**0.1592**	**0.0566**	**0.1746**	**0.1176**	**0.1444**	-				
NORTH TUNISIA	TBK	**0.2415**	**0.2852**	**0.3130**	**0.3210**	**0.1663**	**0.1561**	**0.0587**	**0.1651**	**0.1119**	**0.1351**	0.0055	-			
	GTU	**0.2887**	**0.3342**	**0.3634**	**0.3729**	**0.2091**	**0.1955**	**0.0809**	**0.2088**	**0.1526**	**0.1773**	0.0003	0.0017	-		
NORTH SPAIN	TOR	**0.2701**	**0.2992**	**0.3142**	**0.3212**	**0.1282**	**0.1131**	**0.0441**	**0.1861**	**0.1091**	**0.1127**	**0.1067**	**0.1155**	**0.1296**	-	
	CAS	**0.2494**	**0.2798**	**0.2994**	**0.3064**	**0.1221**	**0.1053**	**0.0229**	**0.1657**	**0.0908**	**0.0996**	**0.0720**	**0.0739**	**0.0912**	0.0028	-

Table 4.20: Pairwise F_{ST} values among population samples of *Merluccius merluccius* computed on whole (A), neutral (B), outlier (C) datasets respectively. Significant values are represented in bold ($p = 0.05$ after FDR correction). Code of population samples are given in Table 4.17.

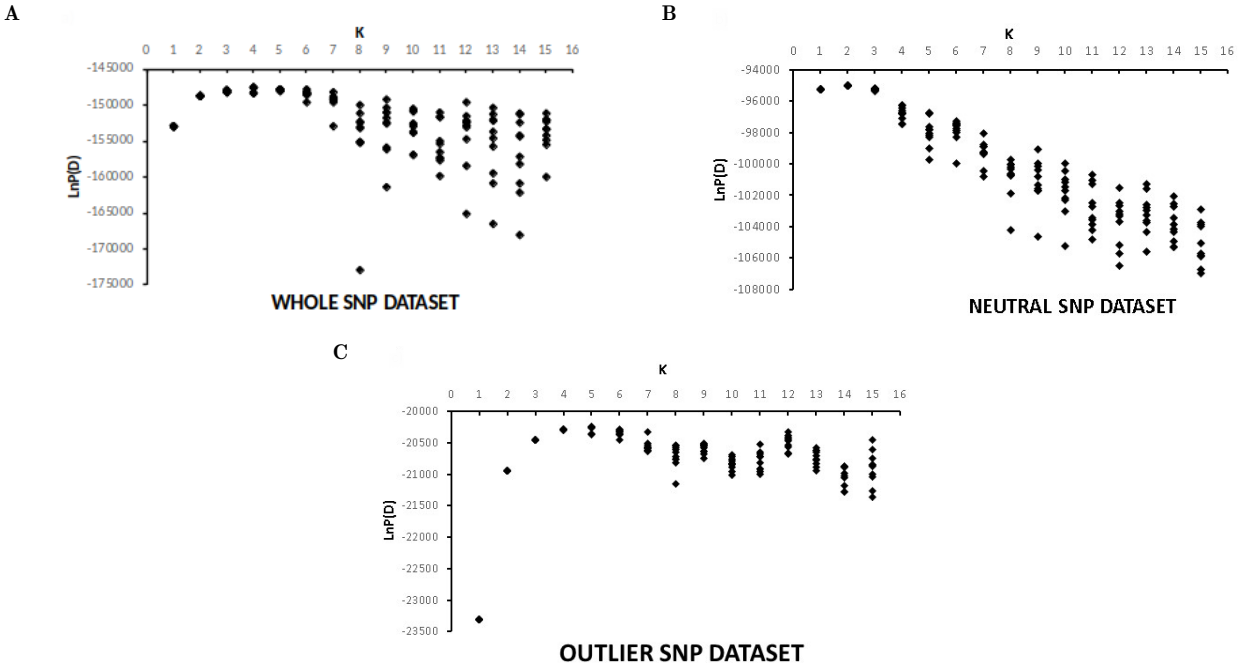

Figure 4.48: Plot of LnP(D) as a function of the number of clusters (K) across the 10 runs for whole (A), neutral (B), outlier (C) SNPs datasets.

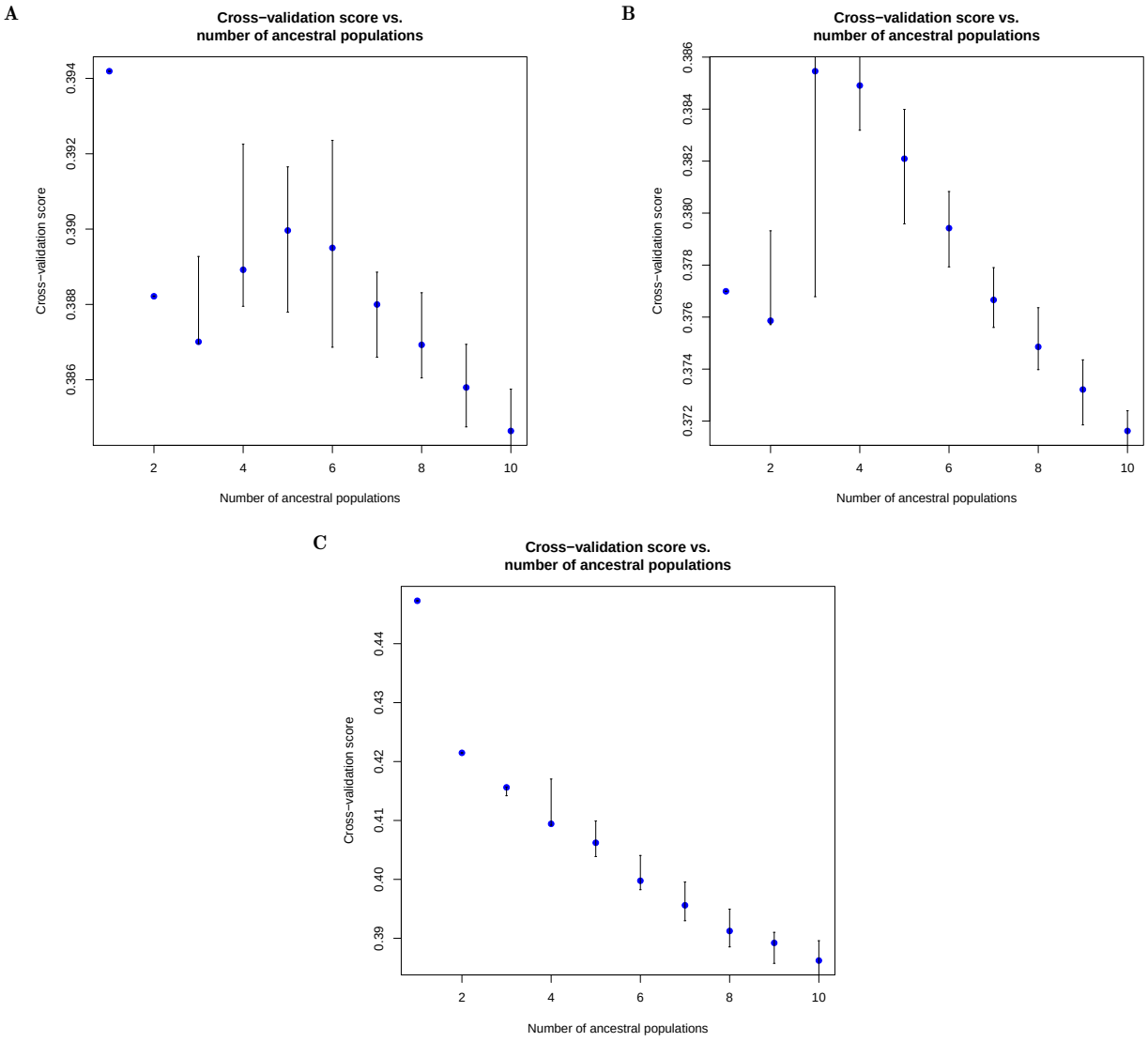

Figure 4.49: Cross-validation score profile resulted from tess3 function showing putative ancestral population of whole (A), neutral (B) and outlier (C) SNPs datasets. The optimal number of clusters (K) was decided following the cross-validation criterion for tess3 in which the best K values corresponds either to a plateau or an increase in the curve.

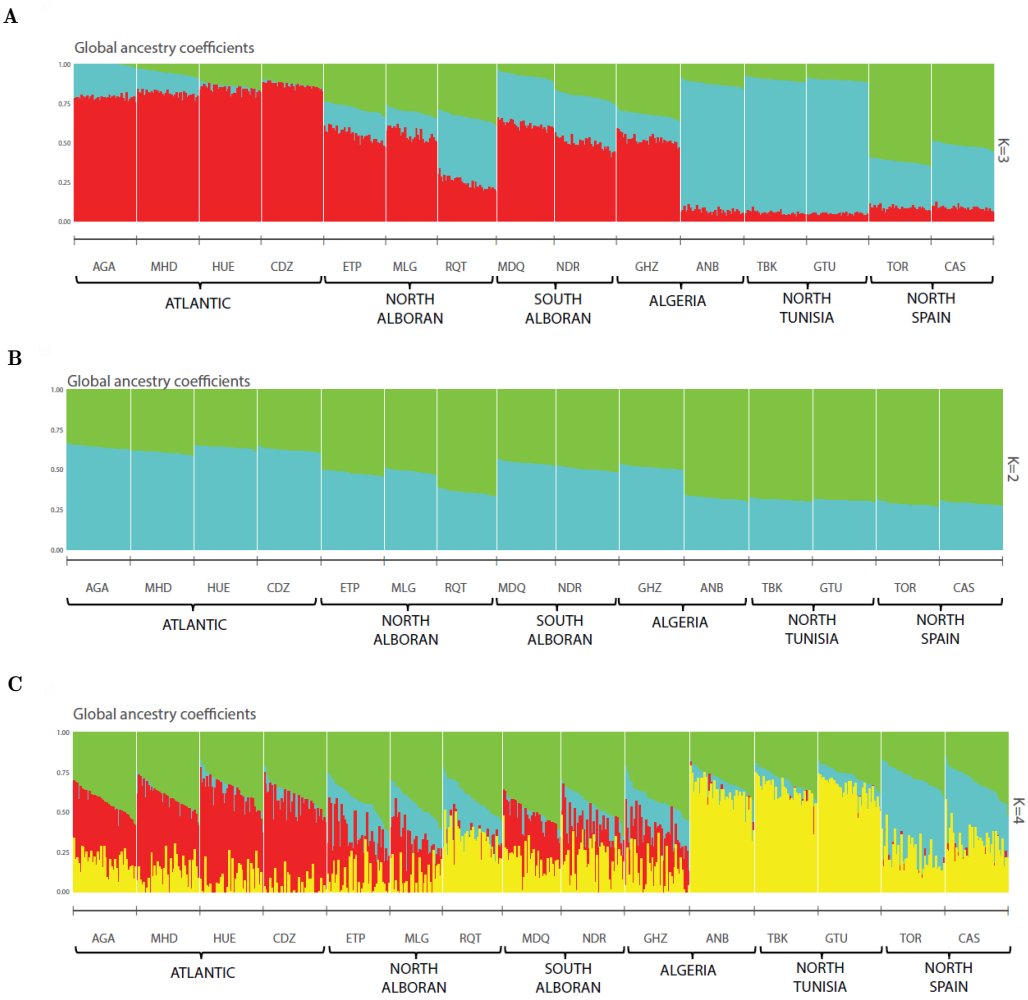

Figure 4.50: Barplot representation of ancestry coefficient matrix of whole (A), neutral (B), outlier (C) SNPs datasets. Code of population samples are given in Table 4.17

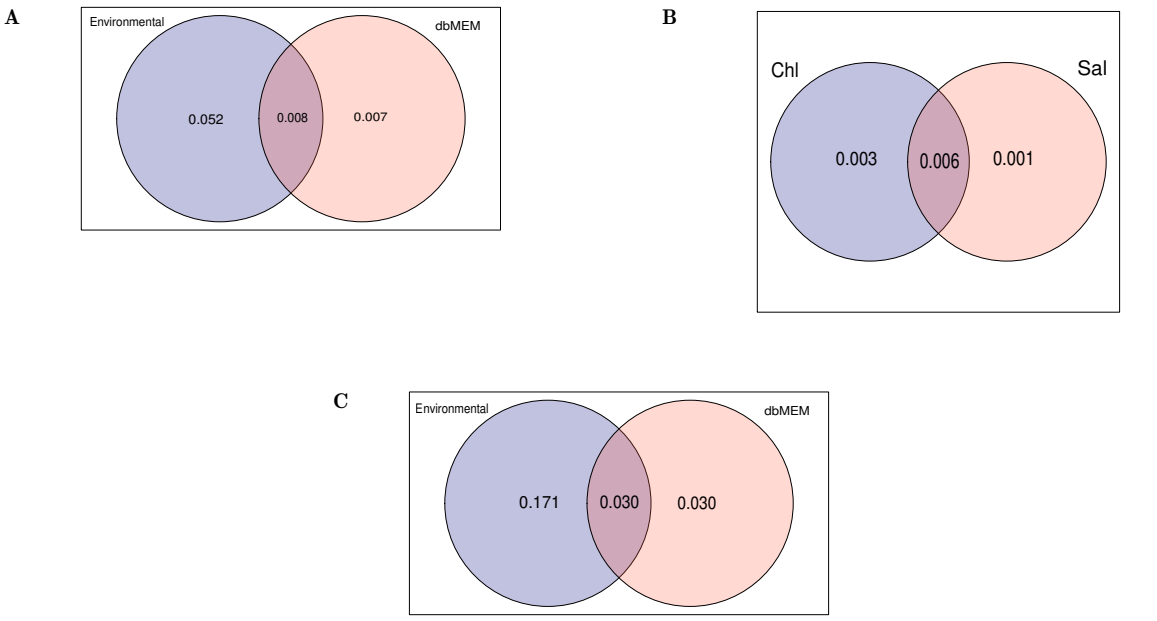

Figure 4.51: Results from variation partitioning analysis on whole (A), neutral (B) and outlier (C) SNPs datasets expressed by the adjusted coefficient of determination (R^2_{adj}). Venn's diagram shows the independent and shared variance explained by environmental and spatial variables.

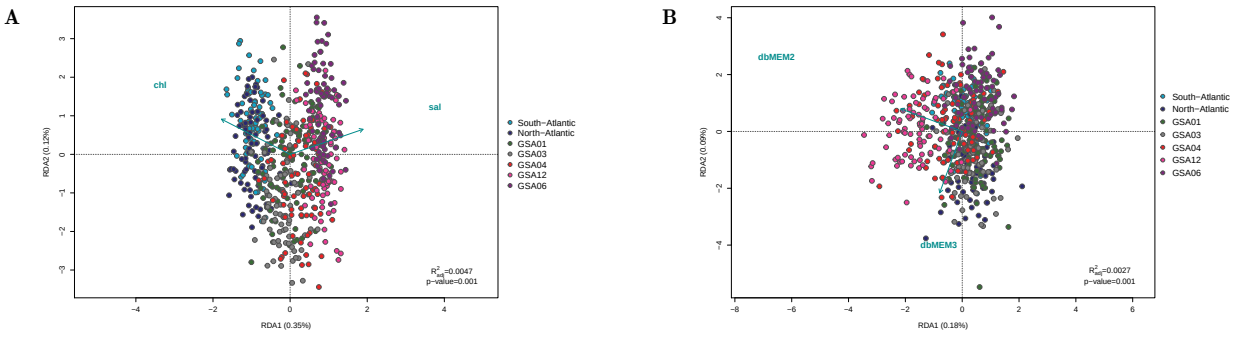

Figure 4.52: Partial redundancy analysis (RDA) conditioned on spatial (A) and environmental (B) variables performed on whole SNPs. Environmental and spatial variables are shown with arrows.

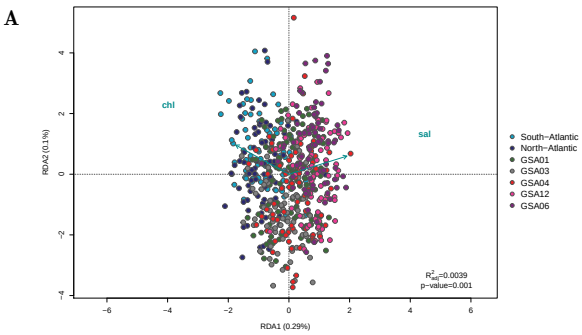

Figure 4.53: Partial redundancy analysis (RDA) conditioned on spatial variables performed on neutral SNPs. Environmental and spatial variables are shown with arrows.

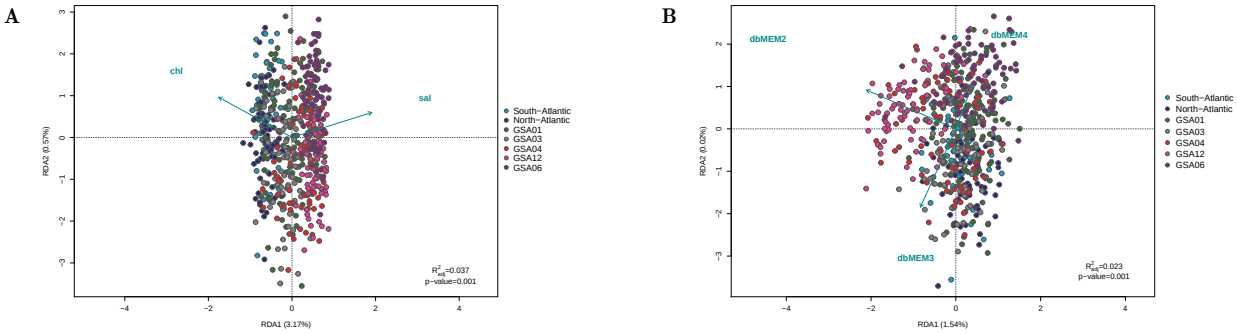

Figure 4.54: Partial redundancy analysis (RDA) conditioned on spatial (A) and environmental (B) variables performed on outlier SNPs. Environmental and spatial variables are shown with arrows.

Chapter 5
General discussion

Understanding stock structure is an essential requirement before any stock assessment can be provided for management plans. Stock identification methods are directly relevant for the appropriate definition of stock units. The description of stock structure using a holistic approach contributed to the general understanding of the biological features and processes acting in marine species population.

Within this project, population structure of common sole (*Solea solea*) in the Mediterranean Sea was investigated as the first case study using multiple tracers with a range of spatio-temporal scales. Multidisciplinary approaches demonstrated evidence of complex spatial structure of the common sole population in the Mediterranean Sea suggesting the presence of local adaptation acting at smaller scale within a large scale differentiation pattern due to neutral processes. The results of multiple tracers were combined in a holistic integrated approach. All tracers were complementary and consistent in contributing to a scientific and valid indication of the population structure of common sole species in the Mediterranean areas assessed. The results obtained from the otolith analysis almost substantially corresponded with the ones provided by the genetic analyses helping in integrating the information from a biological perspective. In fact, both multidisciplinary and holistic approaches developed here contributed to fill the gaps in the understanding of the mechanisms acting at different spatio-temporal scales, as well as the processes and factors shaping population structure of this flatfish stock. In particular, genomics approaches pointed out two main processes structuring common sole populations: a long term and longitudinal genetic drift differentiation at wide geographical scale, and a local adaptation driven by distance segregation, environment features and/or demographic processes spatially structured at fine geographical scale. These mechanisms were further investigated and confirmed by holistic approach.

Even using a reduced number of population samples especially for certain tracers, the results of this project attested that implementing a holistic perspective of the population structure of marine species targeted by commercial fisheries is necessary in order to properly provide scientific advice and take scientifically oriented decisions on the most likely stock structure for assessment and management purposes (ICES, 2022). Indeed, once a comprehensive review of stock structure and a synthetic perspective through holistic approach was

provided, alternative assessment and management options should be proposed to evaluate their limitations. For instance, in the decision-making steps a Management Strategy Evaluation (MSE) should be adopted to measure the consequences of alternative management options under simulated scenarios that incorporates biological population structure and bioeconomic models. Simulation models can consider different mixed-stock structures and dispersion models to evaluate the degree of connectivity (Kerr et al., 2017; Cadrin, 2020).

In this sense, the integration of an increasing amount of information on the dynamics and boundaries of the stocks is essential for stock units delineation within an adaptive spatial management framework, as demonstrated by MED_UNITs project. Especially, the strategy to explore the role of environmental variables at spatial scales in delineating the population structure pointed out from the combination of molecular and otolith data was successfully in identifying stock units. As a results, the new stocks configuration was tested under multiple management scenarios trying to develop a framework for a proper management focused on reducing bias in assessment and management schemes in order to protect the spatial structure of the marine resources (Spedicato et al., 2022).

The work conducted in my study represented a first approach to the configuration of stock units of common sole in the Mediterranean Sea. At this stage, proper stock units of common sole in the whole Mediterranean Sea area should not be proposed at biologically valuable regional scales due to the complexity of the population structure pointed out in this work. Although there is limitation in data availability, the preliminary knowledge for future application of spatial stock assessment models was determined by this project. Hence, future studies could use these results to better design the sampling collection for specific applied questions. The implementation of further investigation is necessary to provide reliable and robust scientific advice.

As second case study, population structure of European hake (*Merluccius merluccius*) in the Alboran Sea and adjacent waters was assessed using only one tracer. Genetic analysis on data from genome wide SNPs markers suggested higher levels of structure with a westward–eastward differentiation gradient among populations derived by a long-term genetic divergence, as well as a potential subtle differentiation at fine geographical scale using putatively adaptive SNPs. Multiple tracers and holistic approaches for this species were applied inside TRANSBORAN project. A more complete view of population structure, stock units boundaries and mechanisms involved in the spatial distribution of European hake in Alboran Sea, and adjacent waters will be shortly provided by the final project report. The population structure of European hake inside the whole Mediterranean Sea was investigated within the MED_UNITs project, a continuous longitudinal gradient was suggested according to single and combined phenotypic otolith data and consistent with the occurrence of both several stocks and complex spatial structure (Morales-Nin et al., 2022; Spedicato et al., 2022). This information is consistent with the genetic evidence provided by MED_UNITs results suggesting three potential stocks at large basin scale: western Mediterranean, central Mediterranean and

eastern Mediterranean. The results achieved by MED_UNITs project drew the attention to the ability of high quality SNPs loci in unveiling pattern of differentiation (Spedicato et al., 2022).

Regarding the potential of innovative genomics tools, analysis conducted on both species stressed out the power of outlier SNPs, as diagnostic tool, in revealing subtle population structure in marine organisms that express different life traits behavior and/or various pattern of segregation and connectivity depending on the marine realm they belong to. Indeed, in the Mediterranean Sea case study of common sole outlier detection revealed local selection under highly structured differentiation, while the same panel of molecular markers in the Atlantic Ocean populations of common sole revealed local selection under high levels of gene flow (Diopere et al., 2018).

In both case studies, a seascape genomics analyses was applied trying to unravel the relative contribution of space and environmental variables on neutral and outliers genetic structure understanding gene flow and local adaptation processes, respectively. A recent study conducted on two fish species with contrasting dispersal abilities demonstrated the ability of this approach in revealing which variables contributed to shape species population boundaries (Boulanger et al., 2022).

Signatures of local selection driven by spatial and environmental variables were identified by outlier SNP markers in both case studies with contrasting outcomes. Results of this study confirmed that the association of these loci to environmental variables can provide information on the factors shaping the populations structures. Future directions in seascape genomics approach should lead to the combination of population genomics using outlier SNPs, associated with a range of environmental variables, and both habitat and climate models to forecast population structure reaction to environmental fluctuations (Jahnke and Jonsson, 2022). For instance, the distribution of several important marine species may be affected by increasing sea surfaces temperatures (Stanley et al., 2018).

Among other challenges, the expected change in stocks distribution due to climate change may need to be taken into account by conservation and management strategies (Pinsky et al., 2018). In fact, climate change altering environmental variables can influence evolutionary processes by forcing populations to migrate or evolve life-history traits to adapt to climate change. Because of either migratory or adaptation behavior may modify up-to-date marine population structures, it is necessary to evaluate the impact of climate change on the performance of stock identification methods in future study (Boulanger et al., 2022; Goethel et al., 2021). In addition to natural and artificial tracers generally used in stock identification methods, simulations considering different scenarios induced by climate change could assist spatial population dynamics models in advancing on potentially unexpected patterns (Goethel et al., 2021), or implementing other scenarios under the management strategy evaluation framework (Jacobsen et al., 2022).

As conclusion, these three years of PhD project allowed me to develop different methods and methodologies in the framework of stock identification. The developed approaches could be applied, according to data availability, to new target species, taking also into consideration new techniques, for instance, connected to

climate change. There is the need to adopt a holistic approach with multiple techniques to improve information on stock structure for resource management to develop stock assessment applications that includes an evaluation of the most appropriate spatial populations structure of marine species. In fact, multidisciplinary and holistic approach allows to enhance the chance to define stock units considering the ecological, oceanographic, and evolutionary processes contributing to the population structure of marine species.

The results obtained in my PhD project and in all the projects I have had the opportunity to participate in, seem to confirm the potential power of stock identification methods in delineating the stock boundaries, in providing evidence of local adaptation and in advancing knowledge for fisheries assessment and management. In particular, the results of the analysis performed on the target species *Solea solea* and *Merluccius merluccius* with the purpose of the identification of actual spatial stock units will be useful in stock assessment and management plans that include spatially explicit local specificities, as well as other scientific studies.

References

Abaunza, P., Murta, A., Campbell, N., Cimmaruta, R., Comesaña, A., Dahle, G., Santamaría, M.G., Gordo, L., Iversen, S., MacKenzie, K., et al., 2008. Stock identity of horse mackerel (*Trachurus trachurus*) in the northeast atlantic and mediterranean sea: Integrating the results from different stock identification approaches. Fisheries Research 89, 196–209.

Archer, F.I., Adams, P.E., Schneiders, B.B., 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources 17, 5–11.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al., 2000. Gene ontology: tool for the unification of biology. Nature genetics 25, 25–29.

AsyMMETR, O., 2013. Differences in elemental chemistry and co stable isotope composition between left and right otoliths of a flatfish, the common sole *Solea solea*. Vie et milieu-life and enVironment 63, 169–179.

Avigliano, E., 2022. Optimizing the methodological design in fish stock delineation from otolith chemistry: review of spatio-temporal analysis scales. Reviews in Fisheries Science & Aquaculture 30, 330–345.

Bacha, M., Jemaa, S., Hamitouche, A., Rabhi, K., Amara, R., 2014. Population structure of the european anchovy, engraulis encrasicolus, in the sw mediterranean sea, and the atlantic ocean: evidence from otolith shape analysis. ICES Journal of Marine Science 71, 2429–2435.

Begg, G.A., Friedland, K.D., Pearce, J.B., 1999. Stock identification and its role in stock assessment and fisheries management: an overview. Fisheries research 43, 1–8.

Bekkevold, D., Höjesjö, J., Nielsen, E.E., Aldvén, D., Als, T.D., Sodeland, M., Kent, M.P., Lien, S., Hansen, M.M., 2020. Northern european *Salmo trutta* (l.) populations are genetically divergent across geographical regions and environmental gradients. Evolutionary applications 13, 400–416.

Belkhir, K., 2004. Genetix 4.05, logiciel sous windows tm pour la génétique des populations. http://www. genetix. univ-montp2. fr/genetix/genetix. htm .

Benchoucha, S., Pérez-Gil, J., Ainouche, N., Jarbui, O., Baro, J., Elouamari, N., Ben-Merien, S., Hamida, C., García-Jiménez, M., Bernardon, M., et al., 2012. Advance in preparing a joint assessment of european hake, *Merluccius merluccius*, stock for gsas 01, 02, 03 and 04 of the gfcm (algeria, morocco and spain). fao-copemed ii occasional paper n^{o}14 .

Benestan, L., Oleksiak, M., Rajora, O., 2019. Population genomics: marine organisms .

Benestan, L., Quinn, B.K., Maaroufi, H., Laporte, M., Clark, F.K., Greenwood, S.J., Rochette, R., Bernatchez, L., 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in american lobster (*Homarus americanus*). Molecular Ecology 25, 5073–5092.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57, 289–300.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Annals of statistics , 1165–1188.

Benzekri, H., Armesto, P., Cousin, X., Rovira, M., Crespo, D., Merlo, M.A., Mazurais, D., Bautista, R., Guerrero-Fernández, D., Fernandez-Pozo, N., et al., 2014. De novo assembly, characterization and functional annotation of senegalese sole (*Solea senegalensis*) and common sole (*Solea solea*) transcriptomes: integration in a database and design of a microarray. Bmc Genomics 15, 1–18.

Berg, P.R., Jorde, P.E., Glover, K.A., Dahle, G., Taggart, J.B., Korsbrekke, K., Dingsør, G.E., Skjæraasen, J.E., Wright, P.J., Cadrin, S.X., et al., 2021. Genetic structuring in atlantic haddock contrasts with current management regimes. ICES Journal of Marine Science 78, 1–13.

Berger, A.M., Deroba, J.J., Bosley, K.M., Goethel, D.R., Langseth, B.J., Schueller, A.M., Hanselman, D.H., 2021. Incoherent dimensionality in fisheries management: consequences of misaligned stock assessment and population boundaries. ICES Journal of Marine Science 78, 155–171.

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D.T., Barth, J.M., Beacham, T.D., Maes, G.E., Martinsohn, J.T., Miller, K.M., Naish, K.A., et al., 2017. Harnessing the power of genomics to secure the future of seafood. Trends in Ecology & Evolution 32, 665–680.

Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., Bernatchez, L., Mallet, M.A., 2019. Seascape genomics of eastern oyster (*Crassostrea virginica*) along the atlantic coast of canada. Evolutionary Applications 12, 587–609.

Berner, D., 2019. Allele frequency difference afd–an intuitive alternative to fst for quantifying genetic population differentiation. Genes 10, 308.

Berryman, A.A., 2002. Population: a central concept for ecology? Oikos 97, 439–442.

Bonhomme, V., Picq, S., Claude, J., Gaucherel, C., 2014. Momocs: outline analysis using r. Journal of Statistical Software 56, 24–p.

Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological modelling 153, 51–68.

Boulanger, E., Benestan, L., Guerin, P.E., Dalongeville, A., Mouillot, D., Manel, S., 2022. Climate differently influences the genomic patterns of two sympatric marine fish species. Journal of Animal Ecology 91, 1180–1195.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using diamond. Nature methods 12, 59–60.

Burke, N., Brophy, D., King, P.A., 2008. Otolith shape analysis: its application for discriminating between stocks of irish sea and celtic sea herring (*Clupea harengus*) in the irish sea. ICES Journal of Marine Science 65, 1670–1675.

Cadrin, S.X., 2020. Defining spatial structure for fishery stock assessment. Fisheries Research 221, 105397.

Cadrin, S.X., Karr, L.A., Mariani, S., 2014a. Stock identification methods: an overview. Stock identification methods , 1–5.

Cadrin, S.X., Kerr, L.A., Mariani, S., 2013. Stock identification methods: applications in fishery science .

Cadrin, S.X., Kerr, L.A., Mariani, S., 2014b. Interdisciplinary evaluation of spatial population structure for definition of fishery management units, in: Stock identification methods. Elsevier, pp. 535–552.

Campana, S.E., 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. Marine Ecology Progress Series 188, 263–297.

Campana, S.E., Thorrold, S.R., 2001. Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? Canadian Journal of Fisheries and Aquatic Sciences 58, 30–38.

Carroll, S.P., Hendry, A.P., Reznick, D.N., Fox, C.W., 2007. Evolution on ecological time-scales.

Casey, J., Jardim, E., Martinsohn, J.T., 2016. The role of genetics in fisheries management under the eu common fisheries policy. Journal of fish biology 89, 2755–2767.

Casey, J., Pereiro, J., 1995. European hake (*M. merluccius*) in the north-east atlantic, in: Hake. Springer, pp. 125–147.

Castillo, A.G., Alvarez, P., Garcia-Vazquez, E., 2005. Population structure of *Merluccius merluccius* along the iberian peninsula coast. ICES Journal of Marine Science 62, 1699–1704.

Castillo, A.G., Martinez, J.L., Garcia-Vazquez, E., 2004. Fine spatial structure of atlantic hake (*Merluccius merluccius*) stocks revealed by variation at microsatellite loci. Marine Biotechnology 6, 299–306.

Catanese, G., Montes, I., Iriondo, M., Estonba, A., Iudicone, D., Procaccini, G., 2016. High resolution snps selection in *Engraulis encrasicolus* through taqman openarray. Fisheries Research 177, 31–38.

Caye, K., Deist, T.M., Martins, H., Michel, O., François, O., 2016. Tess3: fast inference of spatial population structure and genome scans for selection. Molecular Ecology Resources 16, 540–548.

Caye, K., Jay, F., Michel, O., François, O., 2018. Fast inference of individual admixture coefficients using geographic data. The Annals of Applied Statistics 12, 586–608.

Caye, K., Jumentier, B., Lepeule, J., François, O., 2019. Lfmm 2: Fast and accurate inference of gene-environment associations in genome-wide studies. Molecular biology and evolution 36, 852–860.

Chang, M.Y., Geffen, A.J., 2013. Taxonomic and geographic influences on fish otolith microchemistry. Fish and Fisheries 14, 458–492.

Chang, M.Y., Geffen, A.J., Kosler, J., Dundas, S.H., Maes, G.E., 2012. The effect of ablation pattern on la-icpms analysis of otolith element composition in hake, *Merluccius merluccius*. Environmental biology of fishes 95, 509–520.

Cheung, W.W., Lam, V.W., Sarmiento, J.L., Kearney, K., Watson, R., Pauly, D., 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish and fisheries 10, 235–251.

Ciannelli, L., Fisher, J.A., Skern-Mauritzen, M., Hunsicker, M.E., Hidalgo, M., Frank, K.T., Bailey, K.M., 2013. Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: a review. Marine Ecology Progress Series 480, 227–243.

Cimmaruta, R., Bondanelli, P., Nascetti, G., 2005. Genetic structure and environmental heterogeneity in the european hake (*Merluccius merluccius*). Molecular ecology 14, 2577–2591.

Cline, A.J., Hamilton, S.L., Logan, C.A., 2020. Effects of multiple climate change stressors on gene expression in blue rockfish (sebastes mystinus). Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 239, 110580.

Coll, M., Steenbeek, J., Sole, J., Palomera, I., Christensen, V., 2016. Modelling the cumulative spatial–temporal effects of environmental drivers and fishing in a nw mediterranean marine ecosystem. Ecological Modelling 331, 100–114.

Constenla, M., Soler-Membrives, A., Besada, V., Carrassón, M., 2021. Impact assessment of a large river on the sediments and fish from its continental shelf: using *Solea solea* as sentinel in the ebro river mouth (nw mediterranean, spain). Environmental Science and Pollution Research , 1–16.

Coop, G., Witonsky, D., Di Rienzo, A., Pritchard, J.K., 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185, 1411–1423.

Cope, J.M., Punt, A.E., 2009. Drawing the lines: resolving fishery management units with simple fisheries data. Canadian Journal of Fisheries and Aquatic Sciences 66, 1256–1273.

Cowen, R.K., Sponaugle, S., 2009. Larval dispersal and marine population connectivity .

Cuveliers, E., Geffen, A., Guelinckx, J., Raeymaekers, J., Skadal, J., Volckaert, F., Maes, G., 2010. Microchemical variation in juvenile solea solea otoliths as a powerful tool for studying connectivity in the north sea. Marine Ecology Progress Series 401, 211–220.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12, 499–510.

Delerue-Ricard, S., Darnaude, A.M., Raeymaekers, J.A., Dundas, S.H., Skadal, J., Volckaert, F.A., Geffen, A.J., 2019. Extensive larval dispersal and restricted movement of juveniles on the nursery grounds of sole in the southern north sea. Journal of Sea Research 155, 101822. URL: `https://www.sciencedirect.com/science/article/abs/pii/S1385110118303071?via%3Dihub`.

Diopere, E., Vandamme, S.G., Hablützel, P.I., Cariani, A., Van Houdt, J., Rijnsdorp, A., Tinti, F., Consortium, F., Volckaert, F.A., Maes, G.E., 2018. Seascape genetics of a flatfish reveals local selection under high levels of gene flow. ICES Journal of Marine Science 75, 675–689.

Do Prado, F.D., Vera, M., Hermida, M., Bouza, C., Pardo, B.G., Vilas, R., Blanco, A., Fernández, C., Maroso, F., Maes, G.E., et al., 2018. Parallel evolution and adaptation to environmental factors in a marine flatfish: Implications for fisheries and aquaculture management of the turbot (*Scophthalmus maximus*). Evolutionary Applications 11, 1322–1341.

Du Pontavice, H., Randon, M., Lehuta, S., Vermard, Y., Savina-Rolland, M., 2018. Investigating spatial heterogeneity of von bertalanffy growth parameters to inform the stock structuration of common sole, *Solea solea*, in the eastern english channel. Fisheries Research 207, 28–36.

Duruz, S., Sevane, N., Selmoni, O., Vajana, E., Leempoel, K., Stucki, S., Orozco-terWengel, P., Rochat, E., Dunner, S., Consortium, N., et al., 2019. Rapid identification and interpretation of gene–environment associations using the new r. sambada landscape genomics pipeline. Molecular ecology resources 19, 1355–1365.

Earl, D.A., VonHoldt, B.M., 2012. Structure harvester: a website and program for visualizing structure output and implementing the evanno method. Conservation genetics resources 4, 359–361.

El-Geziry, T., Bryden, I., 2010. The circulation pattern in the mediterranean sea: issues for modeller consideration. Journal of Operational Oceanography 3, 39–46.

Excoffier, L., Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. Molecular ecology resources 10, 564–567.

FAO, 1999. Report of the first session of the scientific advisory committee:rome, italy, 23–26 march 1999. FAO Fisheries Report No. 601., Rome, 1999 .

FAO, 2019. Working group on stock assessment of demersal species (wgsad) benchmark session for the assessment of european hake in gsas 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 22, 23 and 26. FAO headquarters URL: https://www.fao.org/gfcm/technical-meetings/detail/en/c/1313336/.

FAO, 2020a. Fao fisheries and aquaculture division - fishstatj - software for fishery and aquaculture statistical time series. FAO Fisheries and Aquaculture Division [online]. Rome. [Cited 3 November 2021]. URL: https://www.fao.org/fishery/statistics/software/fishstatj/en.

FAO, 2020b. The state of mediterranean and black sea fisheries 2020. General Fisheries Commission for the Mediterranean. Rome. URL: https://doi.org/10.4060/cb2429en.

FAO, 2020c. The state of world fisheries and aquaculture 2020. Sustainability in action. .

FAO, 2021. Gfcm 2030 strategy for sustainable fisheries and aquaculture in the mediterranean and the black sea. rome URL: https://doi.org/10.4060/cb7562en.

FAO, 2022. Mediterranean and black sea (major fishing area 37). fisheries and aquaculture division [online]. [Cited Thursday, June 30th 2022] URL: https://www.fao.org/fishery/en/area/37/en.

FAO-GFCM, 2021. Report of the working group on stock assessment of demersal species (wgsad) – benchmark session for the assessment of common sole in gsa 17. Scientific Advisory Committee on Fisheries (SAC) .

Ferhani, K., Bekrattou, D., Mouffok, S., 2021. Body morphometric and otolith shape analyses of anchovy (*Engraulis encrasicolus* (linnaeus, 1758)) in the algerian basin. Acta Adriatica: International Journal of Marine Sciences 62, 159–170.

Ferreira, I., Santos, D., Moreira, C., Feijó, D., Rocha, A., Correia, A.T., 2019. Population structure of *Chelidonichthys lucerna* in portugal mainland using otolith shape and elemental signatures. Marine Biology Research 15, 500–512.

Fick, S.E., Hijmans, R.J., 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology 37, 4302–4315.

Fiorentino, F., Massutí, E., Tinti, F., Somarakis, S., Garofalo, G., Russo, T., Facchini, M., Carbonara, P., Kapiris, K., Tugores, P., et al., 2014. Stock units: Identification of distinct biological units (stock units) for different fish and shellfish species and among different gfcm-gsa. STOCKMED Deliverable 3, 310.

Flanagan, S.P., Jones, A.G., 2017. Constraints on the f st–heterozygosity outlier approach. Journal of Heredity 108, 561–573.

Foll, M., Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics 180, 977–993.

Forester, B.R., Lasky, J.R., Wagner, H.H., Urban, D.L., 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Molecular Ecology 27, 2215–2233.

Francis, R.M., 2017. pophelper: an r package and web app to analyse and visualize population structure. Molecular ecology resources 17, 27–32.

Frichot, E., François, O., 2015. Lea: An r package for landscape and ecological association studies. Methods in Ecology and Evolution 6, 925–929.

Frichot, E., Schoville, S.D., Bouchard, G., François, O., 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Molecular biology and evolution 30, 1687–1699.

Gačić, M., Civitarese, G., Eusebi Borzelli, G., Kovačević, V., Poulain, P.M., Theocharis, A., Menna, M., Catucci, A., Zarokanellos, N., 2011. On the relationship between the decadal oscillations of the northern ionian sea and the salinity distributions in the eastern mediterranean. Journal of Geophysical Research: Oceans 116.

Gagnaire, P.A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., Bierne, N., 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evolutionary applications 8, 769–786.

Geffen, A., Jarvis, K., Thorpe, J., Leah, R., Nash, R., 2003. Spatial differences in the trace element concentrations of irish sea plaice pleuronectes platessa and whiting *Merlangius merlangus* otoliths. Journal of Sea Research 50, 247–256.

Geffen, A.J., Morales-Nin, B., Gillanders, B.M., 2016. Fish otoliths as indicators in ecosystem based management: results of the 5th international otolith symposium (ios2014). Marine and Freshwater Research 67, i–iv.

Goethel, D.R., Berger, A.M., 2017. Accounting for spatial complexities in the calculation of biological reference points: effects of misdiagnosing population structure for stock status indicators. Canadian Journal of Fisheries and Aquatic Sciences 74, 1878–1894.

Goethel, D.R., Bosley, K.M., Langseth, B.J., Deroba, J.J., Berger, A.M., Hanselman, D.H., Schueller, A.M., 2021. Where do you think you're going? accounting for ontogenetic and climate-induced movement in spatially stratified integrated population assessment models. Fish and Fisheries 22, 141–160.

Grati, F., Scarcella, G., Polidori, P., Domenichetti, F., Bolognini, L., Gramolini, R., Vasapollo, C., Giovanardi, O., Raicevich, S., Celić, I., et al., 2013. Multi-annual investigation of the spatial distributions of juvenile and adult sole (*Solea solea* l.) in the adriatic sea (northern mediterranean). Journal of Sea Research 84, 122–132.

Guarnieo, I., Franzellitti, S., Ungaro, N., Tommasini, S., Piccinetti, C., Tinti, F., 2002. Control region haplotype variation in the central mediterranean common sole indicates geographical isolation and population structuring in italian stocks. Journal of Fish Biology 60, 1459–1474.

Günther, T., Coop, G., 2013. Robust identification of local adaptation from allele frequencies. Genetics 195, 205–220.

Harrell Jr, F.E., Harrell Jr, M.F.E., 2019. Package 'hmisc'. CRAN2018 2019, 235–6.

Hart, A.J., Ginzburg, S., Xu, M., Fisher, C.R., Rahmatpour, N., Mitton, J.B., Paul, R., Wegrzyn, J.L., 2020. Entap: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Molecular ecology resources 20, 591–604.

Hauser, L., Carvalho, G.R., 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9, 333–362.

Hawkins, S., Bohn, K., Sims, D., Ribeiro, P., Faria, J., Presa, P., Pita, A., Martins, G., Neto, A., Burrows, M., et al., 2016. Fisheries stocks from an ecological perspective: Disentangling ecological connectivity from genetic interchange. Fisheries Research 179, 333–341.

Healey, A.J., Farthing, M.W., Nunoo, F.K., Potts, W.M., Sauer, W.H., Skujina, I., King, N., de Becquevort, S., Shaw, P.W., McKeown, N.J., 2020. Genetic analysis provides insights into species distribution and population structure in east atlantic horse mackerel (*Trachurus trachurus* and *T. capensis*). Journal of fish biology 96, 795–805.

Hemmer-Hansen, J., Therkildsen, N.O., Pujolar, J.M., 2014. Population genomics of marine fishes: next-generation prospects and challenges. The Biological Bulletin 227, 117–132.

Hendricks, S., Anderson, E.C., Antao, T., Bernatchez, L., Forester, B.R., Garner, B., Hand, B.K., Hohenlohe, P.A., Kardos, M., Koop, B., et al., 2018. Recent advances in conservation and population genomics data analysis. Evolutionary Applications 11, 1197–1211.

Higgins, R.M., Danilowicz, B.S., Balbuena, J.A., Daníelsdóttir, A.K., Geffen, A.J., Meijer, W.G., Modin, J., Montero, F.E., Pampoulie, C., Perdiguero-Alonso, D., et al., 2010. Multi-disciplinary fingerprints reveal the harvest location of cod gadus morhua in the northeast atlantic. Marine Ecology Progress Series 404, 197–206.

Hsieh, C.h., Yamauchi, A., Nakazawa, T., Wang, W.F., 2010. Fishing effects on age and spatial structures undermine population stability of fishes. Aquatic Sciences 72, 165–178.

Husson, F., Josse, J., Le, S., Mazet, J., 2017. Factominer: Multivariate exploratory data analysis and data mining URL: `https://CRAN.R-project.org/package=FactoMineR`.

Hüssy, K., Albertsen, C.M., Hemmer-Hansen, J., Vinther, M., Serre, S.H., Thomsen, T.B., Eero, M., 2022. Where do you come from, where do you go: early life stage drift and migrations of cod inferred from otolith microchemistry and genetic population assignment. Canadian Journal of Fisheries and Aquatic Sciences 79, 300–313.

Hussy, K., Mosegaard, H., Albertsen, C.M., Nielsen, E.E., Hemmer-Hansen, J., Eero, M., 2016. Evaluation of otolith shape as a tool for stock discrimination in marine fishes using baltic sea cod as a case study. Fisheries Research 174, 210–218.

ICES, 2021. Stock identification methods working group (simwg). ICES Scientific Reports 3, 37 pp. URL: `https://doi.org/10.17895/ices.pub.8290`.

ICES, 2022. Stock identification methods working group (simwg). ICES Scientific Reports 4, 66 pp. URL: `http://doi.org/10.17895/ices.pub.20937001`.

Iwata, H., Ukai, Y., 2002. SHAPE: A Computer Program Package for Quantitative Evaluation of Biological Shapes Based on Elliptic Fourier Descriptors. Journal of Heredity 93, 384–385. URL: `https://doi.org/10.1093/jhered/93.5.384`.

Izzo, C., Ward, T.M., Ivey, A.R., Suthers, I.M., Stewart, J., Sexton, S.C., Gillanders, B.M., 2017. Integrated approach to determining stock structure: implications for fisheries management of sardine, *Sardinops sagax*, in australian waters. Reviews in Fish Biology and Fisheries 27, 267–284.

Jacobsen, N.S., Marshall, K.N., Berger, A.M., Grandin, C., Taylor, I.G., 2022. Climate-mediated stock redistribution causes increased risk and challenges for fisheries management. ICES Journal of Marine Science 79, 1120–1132.

Jahnke, M., Jonsson, P.R., 2022. Biophysical models of dispersal contribute to seascape genetic analyses. Philosophical Transactions of the Royal Society B 377, 20210024.

Jakobsson, M., Rosenberg, N.A., 2007. Clumpp: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806.

Jeffery, N.W., Stanley, R.R., Wringe, B.F., Guijarro-Sabaniel, J., Bourret, V., Bernatchez, L., Bentzen, P., Beiko, R.G., Gilbey, J., Clément, M., et al., 2017. Range-wide parallel climate-associated genomic clines in atlantic salmon. Royal Society open science 4, 171394.

Jeffreys, H., 1961. Theory of probability, ed. 3 oxford university press.

Jemaa, S., Bacha, M., Khalaf, G., Dessailly, D., Rabhi, K., Amara, R., 2015. What can otolith shape analysis tell us about population structure of the european sardine, *Sardina pilchardus*, from atlantic and mediterranean waters? Journal of Sea Research 96, 11–17.

Jombart, T., 2008. adegenet: a r package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jombart, T., 2017. A tutorial for the spatial analysis of principal components (spca) using adegenet 2.1.0. Imperial College London URL: `https://github.com/thibautjombart/adegenet/blob/master/tutorials/tutorial-spca.pdf`.

Jombart, T., Devillard, S., Dufour, A.B., Pontier, D., 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101, 92–103.

Kassambara, A., 2017. Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra. volume 2. Sthda.

Kassambara, A., Mundt, F., 2017. Factoextra: Extract and visualize the results of multivariate data analyses URL: `http://www.sthda.com/english/rpkgs/factoextra`.

Keating, J.P., Brophy, D., Officer, R.A., Mullins, E., 2014. Otolith shape analysis of blue whiting suggests a complex stock structure at their spawning grounds in the northeast atlantic. Fisheries Research 157, 1–6.

Kerr, L.A., Hintzen, N.T., Cadrin, S.X., Clausen, L.W., Dickey-Collas, M., Goethel, D.R., Hatfield, E.M., Kritzer, J.P., Nash, R.D., 2017. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. ICES Journal of Marine Science 74, 1708–1722.

Kindt, R., Kindt, M.R., 2015. Package 'biodiversityr'. R Proj .

Kitada, S., Nakajima, K., Hamasaki, K., 2017. Population panmixia and demographic expansion of a highly piscivorous marine fish *Scomberomorus niphonius*. Journal of fish biology 91, 1435–1448.

Korta, M., García, D., Santurtún, M., Goikoetxea, N., Andonegi, E., Murua, H., Álvarez, P., Cerviño, S., Castro, J., Murillas, A., 2015. European hake (*Merluccius merluccius*) in the northeast atlantic ocean. Hakes: biology and exploitation , 1–37.

Kotoulas, G., Bonhomme, F., Borsa, P., 1995. Genetic structure of the common sole solea vulgaris at different geographic scales. Marine Biology 122, 361–375.

Kvamme, C., Bekkevold, D., Nash, R.D.M., Jansson, E., Sørvik, A.G.E., Taggart, J.B., Skaala, Ø., Dahle, G., Glover, K., Quintela, M., 2020. Genetic analysis redraws the management boundaries for the european sprat .

Lecomte, J.B., Le Pape, O., Baillif, H., Nevoux, M., Vermard, Y., Savina, M., Veron, M., Lehuta, S., Hunter, E., Rivot, E., 2020. State-space modeling of multidecadal mark–recapture data reveals low adult dispersal in a nursery-dependent fish metapopulation. Canadian Journal of Fisheries and Aquatic Sciences 77, 342–354.

Legendre, P., Legendre, L., 2012. Numerical ecology. Elsevier.

Lehnert, S.J., DiBacco, C., Van Wyngaarden, M., Jeffery, N.W., Lowen, J.B., Sylvester, E.V., Wringe, B.F., Stanley, R.R., Hamilton, L.C., Bradbury, I.R., 2019. Fine-scale temperature-associated genetic structure between inshore and offshore populations of sea scallop (*Placopecten magellanicus*). Heredity 122, 69–80.

Leone, A., Álvarez, P., García, D., Saborido-Rey, F., Rodriguez-Ezpeleta, N., 2019. Genome-wide snp based population structure in european hake reveals the need for harmonizing biological and management units. ICES Journal of Marine Science 76, 2260–2266.

Levene, H., et al., 1960. Contributions to probability and statistics. Essays in honor of Harold Hotelling 278, 292.

Liggins, L., Treml, E.A., Riginos, C., 2019. Seascape genomics: contextualizing adaptive and neutral genomic variation in the ocean environment, in: Population genomics: Marine organisms. Springer, pp. 171–218.

Longmore, C., Fogarty, K., Neat, F., Brophy, D., Trueman, C., Milton, A., Mariani, S., 2010. A comparison of otolith microchemistry and otolith shape analysis for the study of spatial variation in a deep-sea teleost, *Coryphaenoides rupestris*. Environmental biology of fishes 89, 591–605.

Lundy, C.J., Moran, P., Rico, C., Milner, R.S., Hewitt, G.M., 1999. Macrogeographical population differentiation in oceanic environments: a case study of european hake (*Merluccius merluccius*), a commercially important fish. Molecular Ecology 8, 1889–1898.

Luu, K., Bazin, E., Blum, M.G., 2017. pcadapt: an r package to perform genome scans for selection based on principal component analysis. Molecular ecology resources 17, 67–77.

Marval-Rodríguez, A., Renán, X., Galindo-Cortes, G., Acuña-Ramírez, S., Jiménez-Badillo, M.d.L., Rodulfo, H., Montero-Muñoz, J.L., Brulé, T., De Donato, M., 2022. Assessing the speciation of *Lutjanus campechanus* and *Lutjanus purpureus* through otolith shape and genetic analyses. Fishes 7, 85.

McLean, D.L., Ferreira, L.C., Benthuysen, J.A., Miller, K.J., Schläppy, M.L., Ajemian, M.J., Berry, O., Birchenough, S.N., Bond, T., Boschetti, F., et al., 2022. Influence of offshore oil and gas structures on seascape ecological connectivity. Global Change Biology 28, 3515–3536.

Meirmans, P.G., 2015. Seven common mistakes in population genetics and how to avoid them. Molecular ecology 24, 3223–3231.

Mercier, L., Darnaude, A.M., Bruguier, O., Vasconcelos, R.P., Cabral, H.N., Costa, M.J., Lara, M., Jones, D.L., Mouillot, D., 2011. Selecting statistical models and variable combinations for optimal classification using otolith microchemistry. Ecological Applications 21, 1352–1364.

Mérigot, B., Letourneur, Y., Lecomte-Finiger, R., 2007. Characterization of local populations of the common sole *Solea solea* (pisces, soleidae) in the nw mediterranean through otolith morphometrics and shape analysis. Marine Biology 151, 997–1008.

Midway, S.R., Cadrin, S.X., Scharf, F.S., 2014. Southern flounder (*Paralichthys lethostigma*) stock structure inferred from otolith shape analysis. Fishery Bulletin 112.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G.R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A.J., et al., 2014. Outlier snp markers reveal fine-scale genetic structuring across european hake populations (*Merluccius merluccius*). Molecular ecology 23, 118–135.

Morales-Nin, B., Pérez-Mayol, S., MacKenzie, K., Catalán, I.A., Palmer, M., Kersaudy, T., Mahé, K., 2022. European hake (*Merluccius merluccius*) stock structure in the mediterranean as assessed by otolith shape and microchemistry. Fisheries Research 254, 106419.

Morales-Nin, B., Pérez-Mayol, S., Palmer, M., Geffen, A.J., 2014. Coping with connectivity between populations of *Merluccius merluccius*: An elusive topic. Journal of Marine Systems 138, 211–219.

Morat, F., Letourneur, Y., Dierking, J., Pécheyran, C., Bareille, G., Blamart, D., Harmelin-Vivien, M., 2014. The great melting pot. common sole population connectivity assessed by otolith and water fingerprints. PLoS One 9, e86585.

Moreira, C., Froufe, E., Sial, A., Caeiro, A., Vaz-Pires, P., Correia, A., 2018. Population structure of the blue jack mackerel (*Trachurus picturatus*) in the ne atlantic inferred from otolith microchemistry. Fisheries Research 197, 113–122.

Moreira, C., Froufe, E., Vaz-Pires, P., Correia, A., 2019. Otolith shape analysis as a tool to infer the population structure of the blue jack mackerel, *Trachurus picturatus*, in the ne atlantic. Fisheries Research 209, 40–48.

Moreira, C., Froufe, E., Vaz-Pires, P., Triay-Portella, R., Méndez, A., Castro, J.P., Correia, A.T., 2022. Unravelling the spatial-temporal population structure of *Trachurus picturatus* across the north-east atlantic using otolith fingerprinting. Estuarine, Coastal and Shelf Science 272, 107860.

Naimi, B., 2017. usdm: uncertainty analysis for species distribution models. r package version 1.1–18. R Documentation https://cran.r-project.org/web/packages/usdm/usdm.pdf .

Naimi, B., Hamm, N.A., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography 37, 191–203.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., Hohenlohe, P.A., 2013. Genotyping-by-sequencing in ecological and conservation genomics. Molecular ecology 22, 2841.

Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., et al., 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nature communications 3, 1–7.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J., Weedon, J., 2022. vegan: Community ecology package URL: `https://github.com/vegandevs/vegan`. r package version 2.6-5.

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M., Portnoy, D.S., 2018. These aren't the loci you'e looking for: Principles of effective snp filtering for molecular ecologists. Molecular ecology 27, 3193–3206.

Ovenden, J.R., Berry, O., Welch, D.J., Buckworth, R.C., Dichmont, C.M., 2015. Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. Fish and Fisheries 16, 125–159.

Paris, J.R., Sherman, K.D., Bell, E., Boulenger, C., Delord, C., El-Mahdi, M., Fairfield, E.A., Griffiths, A.M., Gutmann Roberts, C., Hedger, R.D., et al., 2018. Understanding and managing fish populations: keeping the toolbox fit for purpose. Journal of Fish Biology 92, 727–751.

Patarnello, T., Volckaert, F.A., Castilho, R., 2007. Pillars of hercules: is the atlantic–mediterranean transition a phylogeographical break? Molecular ecology 16, 4426–4444.

Peakall, R., Smouse, P., 2012. Genalex 6.5: genetic analysis in excel. population genetic software for teaching and research–an update. bioinformatics 28: 2537e2539.

Perry, R.I., Cury, P., Brander, K., Jennings, S., Möllmann, C., Planque, B., 2010. Sensitivity of marine systems to climate and fishing: concepts, issues and management responses. Journal of Marine Systems 79, 427–435.

Pinsky, M.L., Palumbi, S.R., 2014. Meta-analysis reveals lower genetic diversity in overfished populations. Molecular ecology 23, 29–39.

Pinsky, M.L., Reygondeau, G., Caddell, R., Palacios-Abrantes, J., Spijkers, J., Cheung, W.W., 2018. Preparing ocean governance for species on the move. Science 360, 1189–1191.

Pita, A., Casey, J., Hawkins, S.J., Villarreal, M.R., Gutiérrez, M.J., Cabral, H., Carocci, F., Abaunza, P., Pascual, S., Presa, P., 2016. Conceptual and practical advances in fish stock delineation.

Pita, A., Pérez, M., Balado, M., Presa, P., 2014. Out of the celtic cradle: The genetic signature of european hake connectivity in south-western europe. Journal of sea research 93, 90–100.

Pita, A., Perez, M., Cervino, S., Presa, P., 2011. What can gene flow and recruitment dynamics tell us about connectivity between european hake stocks in the eastern north atlantic? Continental Shelf Research 31, 376–387.

Pita, A., Presa, P., Perez, M., 2010. Gene flow, multilocus assignment and genetic structuring of the european hake. Thalassas 26, 129–133.

Primo, A.L., Azeiteiro, U.M., Marques, S.C., Martinho, F., Baptista, J., Pardal, M.A., 2013. Colonization and nursery habitat use patterns of larval and juvenile flatfish species in a small temperate estuary. Journal of Sea Research 76, 126–134.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P., 2000. Association mapping in structured populations. The American Journal of Human Genetics 67, 170–181.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.r-project.org/`.

Randon, M., Le Pape, O., Ernande, B., Mahé, K., Volckaert, F.A., Petit, E.J., Lassalle, G., Le Berre, T., Réveillac, E., 2020. Complementarity and discriminatory power of genotype and otolith shape in describing the fine-scale population structure of an exploited fish, the common sole of the eastern english channel. Plos one 15, e0241429.

Randon, M., Réveillac, E., Le Pape, O., 2021. A holistic investigation of tracers at population and individual scales reveals population structure for the common sole of the eastern english channel. Estuarine, Coastal and Shelf Science 249, 107096.

Randon, M., Réveillac, E., Rivot, E., Du Pontavice, H., Le Pape, O., 2018. Could we consider a single stock when spatial sub-units present lasting patterns in growth and asynchrony in cohort densities? a flatfish case study. Journal of sea research 142, 91–100.

Rashidabadi, F., Abdoli, A., Tajbakhsh, F., Nejat, F., Avigliano, E., 2020. Unravelling the stock structure of the persian brown trout by otolith and scale shape. Journal of Fish Biology 96, 307–315.

Recasens, L., Lombarte, A., Morales-Nin, B., Tores, G., 1998. Spatiotemporal variation in the population structure of the european hake in the nw mediterranean. Journal of fish biology 53, 387–401.

Reiss, H., Hoarau, G., Dickey-Collas, M., Wolff, W.J., 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries 10, 361–395.

Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M., Holderegger, R., 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology 24, 4348–4370.

Roberts, D.G., Ayre, D.J., 2010. Panmictic population structure in the migratory marine sparid *Acanthopagrus australis* despite its close association with estuaries. Marine Ecology Progress Series 412, 223–230.

Rodríguez-Ezpeleta, N., Bradbury, I.R., Mendibil, I., Álvarez, P., Cotano, U., Irigoien, X., 2016. Population structure of atlantic mackerel inferred from rad-seq-derived snp markers: Effects of sequence clustering parameters and hierarchical snp selection. Molecular Ecology Resources 16, 991–1001.

Rogers, S., 1992. Environmental factors affecting the distribution of sole (*Solea solea* (l.)) within a nursery area. Netherlands Journal of Sea Research 29, 153–161.

Roldán, M.I., García-Marín, J.L., Utter, F.M., Pla, C., 1998. Population genetic structure of european hake, *Merluccius merluccius*. Heredity 81, 327–334.

Rolland, J.L., Bonhomme, F., Lagardère, F., Hassan, M., Guinand, B., 2007. Population structure of the common sole (*Solea solea*) in the northeastern atlantic and the mediterranean sea: revisiting the divide with epic markers. Marine Biology 151, 327–341.

Rosenberg, N.A., 2004. Distruct: a program for the graphical display of population structure. Molecular ecology notes 4, 137–138.

Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for windows and linux. Molecular ecology resources 8, 103–106.

Sabatini, L., Bullo, M., Cariani, A., Celić, I., Ferrari, A., Guarniero, I., Leoni, S., Marčeta, B., Marcone, A., Polidori, P., et al., 2018. Good practices for common sole assessment in the adriatic sea: Genetic and morphological differentiation of *Solea solea* (linnaeus, 1758) from *S. aegyptiaca* () and stock identification. Journal of Sea Research 137, 57–64.

Saltalamacchia, F., Berg, F., Casini, M., Davies, J.C., Bartolino, V., 2022. Population structure of european sprat (*Sprattus sprattus*) in the greater north sea ecoregion revealed by otolith shape analysis. Fisheries Research 245, 106131.

Segovia, N.I., González-Wevar, C.A., Haye, P.A., 2020. Signatures of local adaptation in the spatial genetic structure of the ascidian *Pyura chilensis* along the southeast pacific coast. Scientific reports 10, 1–14.

Selmoni, O., Lecellier, G., Vigliola, L., Berteaux-Lecellier, V., Joost, S., 2020. Coral cover surveys corroborate predictions on reef adaptive potential to thermal stress. bioRxiv .

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611. URL: `https://doi.org/10.1093/biomet/52.3-4.591`, doi:`10.1093/biomet/52.3-4.591`.

Simoncelli, S., Fratianni, C., Pinardi, N., Grandi, A., Drudi, M., Oddo, P., Dobricic, S., 2019. Mediterranean sea physical reanalysis (cmems med-physics)[data set]. Copernicus Monitoring Environment Marine Service (CMEMS) URL: `https://doi.org/10.25423/MEDSEA_REANALYSIS_PHYS_006_004`.

Singh, M., Kashyap, A., Ansari, J., Serajuddin, M., 2022. Spatial variations in the shape and chemistry of sagittal otoliths in *Channa punctatus* (channidae) populations of ganga basin, india. Inland Water Biology , 1–13.

Skliris, N., 2014. Past, present and future patterns of the thermohaline circulation and characteristic water masses of the mediterranean sea, in: The Mediterranean Sea. Springer, pp. 29–48.

Smith, P., Jamieson, A., Birley, A., 1990. Electrophoretic studies and the stock concept in marine teleosts. ICES Journal of Marine Science 47, 231–245.

Spedicato, M.T., Cannas, R., Mahe, K., Morales, B., Tsigenopoulos, C., Zane, L., Kavadas, S., Maina, Irida, S.G., Sartor Paolo, Bandelj, V., Russo, T., Fiorentino, F., 2022. Study on advancing fisheries assessment and management advice in the mediterranean by aligning biological and management units of priority species med_units [November 2021. 1330 pp. 681]. URL: `http://doi.org/10.2926/909535`.

Stanley, R., Jeffery, N., 2017. Cartdist: Re-projection tool for complex marine systems.

Stanley, R.R., DiBacco, C., Lowen, B., Beiko, R.G., Jeffery, N.W., Van Wyngaarden, M., Bentzen, P., Brickman, D., Benestan, L., Bernatchez, L., et al., 2018. A climate-associated multispecies cryptic cline in the northwest atlantic. Science Advances 4, eaaq0929.

Stucki, S., Orozco-terWengel, P., Forester, B.R., Duruz, S., Colli, L., Masembe, C., Negrini, R., Landguth, E., Jones, M.R., Consortium, N., et al., 2017. High performance computation of landscape genomic models including local indicators of spatial association. Molecular Ecology Resources 17, 1072–1089.

Tanner, S.E., Pérez, M., Presa, P., Thorrold, S.R., Cabral, H.N., 2014. Integrating microsatellite dna markers and otolith geochemistry to assess population structure of european hake (*Merluccius merluccius*). Estuarine, Coastal and Shelf Science 142, 68–75.

Tanner, S.E., Reis-Santos, P., Cabral, H.N., 2016. Otolith chemistry in stock delineation: a brief overview, current challenges and future prospects. Fisheries Research 173, 206–213.

Teruzzi, A., Bolzon, G., Cossarini, G., Lazzari, P., Salon, S., Crise, A., Solidoro, C., 2019. Mediterranean sea biogeochemical reanalysis (cmems med-biogeochemistry)[data set]. Copernicus Monitoring Environment Marine Service (CMEMS) URL: `https://doi.org/10.25423/MEDSEA_REANALYSIS_BIO_006_008`.

Thioulouse, J., Dray, S., Dufour, A.B., Siberchicot, A., Jombart, T., Pavoine, S., 2018. Multivariate Analysis of Ecological Data with ade4. Springer. doi:`10.1007/978-1-4939-8850-1`.

Tintore, J., La Violette, P., Blade, I., Cruzado, A., 1988. A study of an intense density front in the eastern alboran sea: the almeria–oran front. Journal of Physical Oceanography 18, 1384–1397.

Tuset, V., Lozano, I., González, J., Pertusa, J., García-Díaz, M., 2003. Shape indices to identify regional differences in otolith morphology of comber, *Serranus cabrilla* (l., 1758). Journal of Applied Ichthyology 19, 88–93.

Valenzuela-Quiñonez, F., 2016. How fisheries management can benefit from genomics? Briefings in Functional Genomics 15, 352–357.

Vasconcelos, R.P., Reis-Santos, P., Tanner, S., Fonseca, V., Latkoczy, C., Günther, D., Costa, M.J., Cabral, H., 2007. Discriminating estuarine nurseries for five fish species through otolith elemental fingerprints. Marine Ecology Progress Series 350, 117–126.

Vaz, A.C., Scarcella, G., Pardal, M.A., Martinho, F., 2019. Water temperature gradients drive early life-history patterns of the common sole (*Solea solea* l.) in the northeast atlantic and mediterranean. Aquatic Ecology 53, 281–294.

de Villemereuil, P., Frichot, É., Bazin, É., François, O., Gaggiotti, O.E., 2014. Genome scan methods against more complex models: when and how much should we trust them? Molecular ecology 23, 2006–2019.

Waldman, J.R., 1999. The importance of comparative studies in stock analysis. Fisheries Research 43, 237–246.

Waples, R.S., Punt, A.E., Cope, J.M., 2008. Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9, 423–449.

Weir, B.S., Cockerham, C.C., 1984. Estimating f-statistics for the analysis of population structure. evolution , 1358–1370.

Welch, D.J., Newman, S.J., Buckworth, R.C., Ovenden, J.R., Broderick, D., Lester, R.J., Gribble, N.A., Ballagh, A.C., Charters, R.A., Stapley, J., et al., 2015. Integrating different approaches in the definition of biological stocks: A northern australian multi-jurisdictional fisheries example using grey mackerel, *Scomberomorus semifasciatus*. Marine Policy 55, 73–80.

Westgaard, J.I., Staby, A., Aanestad Godiksen, J., Geffen, A.J., Svensson, A., Charrier, G., Svedäng, H., André, C., 2017. Large and fine scale population structure in european hake (*Merluccius merluccius*) in the northeast atlantic. ICES Journal of Marine Science 74, 1300–1310.

Willette, D., Allendorf, F., Barber, P., Barshis, D., Carpenter, K., Crandall, E.D., Cresko, W., Fernandez-Silva, I., Matz, M., Meyer, E., et al., 2014. So, you want to use next-generation sequencing in marine systems? insight from the pan-pacific advanced studies institute. Bulletin of Marine Science 90, 79–122.

Wujdi, A., Kim, H., Oh, C.W., 2022. Population structure of indian mackerel (*Rastrelliger kanagurta*) in java and bali island, indonesia inferred from otolith shape. Sains Malaysiana 51, 39–50.

Zdobnov, E.M., Apweiler, R., 2001. Interproscan–an integration platform for the signature-recognition methods in interpro. Bioinformatics 17, 847–848.