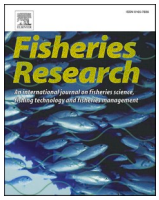
Fisheries Research 270 (2024) 106891

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres



Genomics goes deeper in fisheries science: The case of the blackspot seabream (*Pagellus bogaraveo*) in the northeast Atlantic

Regina L. Cunha [a,b,*], Joana I. Robalo [c], Sara M. Francisco [c], Inês Farias [d], Rita Castilho [a,b,1], Ivone Figueiredo [d,1]

[a] University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal
[b] Centre of Marine Sciences - CCMAR, Campus de Gambelas, 8005-139 Faro, Portugal
[c] MARE – Marine and Environmental Sciences Centre, ARNET-Aquatic Research Network, ISPA Instituto Universitário, Rua Jardim do Tabaco 34, 1149-041 Lisboa, Portugal
[d] Portuguese Institute for Sea and Atmosphere (IPMA), Avenida Doutor Alfredo Magalhães Ramalho 6, 1495-165 Algés, Portugal

ARTICLE INFO

Handled by: Dr. J Viñas

Keywords:
Blackspot seabream
Fish stock assessment
Pagellus bogaraveo
Genotyping-by-sequencing

ABSTRACT

Recent advances in genomics are an essential contributor to the assessment of fish stocks by providing a fine-scale identification of the species' genetic boundaries. The blackspot seabream, *Pagellus bogaraveo*, is a commercial sparid distributed across the northeast (NE) Atlantic and the Mediterranean. Within the NE Atlantic, three *P. bogaraveo* stocks are currently defined: Azores; Atlantic Iberian waters; Celtic Sea and the Bay of Biscay. We used a genotyping-by-sequencing (GBS) approach to better define the spatial scale at which the species occurs in the NE Atlantic. Our findings revealed the existence of an additional genetic cluster in the eastern Atlantic (Gulf of Cádiz) that was not identified in previous studies based on mitochondrial DNA or microsatellite data. The combined effect of ocean circulation patterns, complex bathymetry and the existence of local upwelling may play an important role on the retention of blackspot seabream larvae and adults, providing an explanation for the genetic differentiation between the specimens caught off the Gulf of Cádiz and Peniche (Portugal). Results presented here revealed hidden intra-specific genetic differentiation and can inform a finer-scale sampling to determine the new stock boundaries in the Atlantic Iberian coasts.

1. Introduction

Stock identification is a component of modern fisheries stock assessments (Begg et al., 1999). It depends on models that incorporate a variety of parameters, including, e.g., stock-recruitment relationships (Cadrin et al., 2019) or survival and reproduction rates (Hart, 2001). Traditionally, assessment models applied to fish stocks relied on the assumption that resources are single, homogenous units often delineated by insufficient data (Fujita, 2021). Thus, in cases where management units do not match with groups exhibiting unique demographic dynamics, establishing a clear link between productivity and harvest rates may not be reached (Secor, 2005; Secor, 2014). Misleading results may lead to population declines, particularly in late-reproduction, low fecundity, and high longevity, as in deep-sea fish species (Cheung et al., 2007; Thresher et al., 2007).

Over the last few decades, many studies showed that limited dispersal and low connectivity drive fine-scale genetic structure of marine populations (Benestan, 2019; Schunter et al., 2019), challenging the management paradigm of many exploited species. More recently, novel genomic approaches using a large number of loci from non-model organisms revealed genetic differences within management units, previously thought to be genetically homogeneous (Catarino et al., 2022).

The blackspot seabream (*Pagellus bogaraveo*, Brünnich 1768) is a commercially important benthopelagic sparid distributed across the Eastern Atlantic: from Norway to Cape Blanc in Mauritania, Madeira, the Canary Islands, and the Azores. It is also frequent in the western Mediterranean, including the Strait of Gibraltar, becoming rare eastern of the Strait of Sicily and absent in the Black Sea (Mytilineou et al., 2013; Spedicato et al., 2002; Whitehead, 1986). Additionally, the species was recently recorded in the Levantine Basin (Syrian waters, eastern Mediterranean) (Saad et al., 2020). The species is a protandrous hermaphrodite (most individuals being first functional males and then

* Correspondence to: CCMAR - Campus de Gambelas – University of Algarve, 8005-039 Faro, Portugal.
E-mail address: rcunha@ualg.pt (R.L. Cunha).
[1] Authors contributed equally to this work.

https://doi.org/10.1016/j.fishres.2023.106891
Received 19 May 2023; Received in revised form 30 September 2023; Accepted 19 October 2023
Available online 3 November 2023

developing into females) with a recorded longevity of 20 years (International Council for the Exploitation of the Sea, 2010). *Pagellus bogaraveo* larvae are planktonic, remaining in the water column for 37 days, while juveniles can be found in shallow coastal waters (Teixeira, 2013). Adults are distributed in the continental slope and seamounts, indicating an ontogenetic migration towards deeper waters (down to 600 m deep, Morato et al., 2001). In the NE Atlantic, the spawning season is latitude-dependent and occurs between January and April (Estácio et al., 2001; Sánchez, 1983).

Presently, the International Council for the Exploitation of the Sea (ICES) considers the existence of three *P. bogaraveo* stocks within the NE Atlantic (see Fig. 1 from Jayasinghe et al., 2015, for further details on the subareas): i) Celtic Sea and the Bay of Biscay (ICES subareas 27.6, 7, and 8); ii) Atlantic Iberian waters (ICES subarea 27.9), and iii) the Azores (ICES subarea 27.10) (ICES, 2007) (Fig. 1).

Previous works on *P. bogaraveo* using allozymes, mitochondrial DNA (mtDNA), or microsatellites covered a narrow geographical area of the three putative stocks. Those studies showed no population differentiation between ICES divisions 27.9.a (Galicia) and 27.8.c (Cantabria and Asturias) (Piñera et al., 2007) and marginally significant differences between the Azores (ICES subarea 27.10) and Atlantic Iberian waters (ICES Division 27.9.a) (Bargelloni et al., 2003). A recent study based on mtDNA (*D-loop*) and including specimens from the three putative stocks also supported the differentiation of the Azores stock. Still, no differences were found along Atlantic Iberian waters (Robalo et al., 2020). Given the pelagic larval phase (up to 37 days) of *P. bogaraveo*, the Azorean population is expected to be genetically homogenous. Nonetheless, Stockley et al. (2005) found evidence of mild genetic structuring within the Azores using *D-loop* sequences.

In this study, tissue samples of *P. bogaraveo* were collected from five locations representative of the three putative stocks in the NE Atlantic. We used genotyping-by-sequencing (GBS) and two different approaches (reference and *de novo*) for SNP (single-nucleotide polymorphism) calling to assess fine-scale genetic structure of this important fish resource. High-throughput sequencing may bring forward unsuspected structure in *P. bogaraveo* populations, allowing a better delineation of the current management unit of this commercially valuable species.

2. Methods

2.1. Tissue sampling and DNA extraction

Blackspot seabream fin clips from 91 specimens were collected in five different geographical locations in the NE Atlantic: Île de Sein, Brittany (France – ICES Division 27.7.e); Peniche, Portugal (ICES Division 27.9.a); San Vicente de la Barquera, Cantabria, in the eastern Bay of Biscay (Spain – ICES Division 27.8.c); Gulf of Cádiz (Spain – ICES Division 27.9.a), and Faial, Azores Archipelago (Portugal - ICES Subarea 27.10). Samples from Brittany were collected during the EVHOE bottom trawl survey carried out in the Bay of Biscay (Laffargue et al., 2020). Samples from continental Portuguese waters were collected from commercial landings at Peniche landing port. Fishermen collected samples from commercial fisheries in Cantabria and the Azores. Samples from Cádiz were collected by the IEO (Spanish Institute of Oceanography). The set of samples available is representative of a large portion of the overall species distribution area in the Northeast Atlantic (Fig. 1) and were a subset of the ones used in Robalo et al. (2021). Only the Atlantic populations were considered for the present study. The sampling was opportunistic recurring to captures made by researchers and fishermen. The total number of samples per location are shown in Table 1. Fin clip samples were preserved in 96% ethanol and stored at − 20ºC until DNA extraction.

2.2. Genotyping-by-sequencing (GBS)

All 91 samples were sequenced using the GBS method (Qi et al., 2018), performed by LGC Genomics GmbH (Berlin, Germany). DNA extracted from each individual was digested with the restriction enzyme

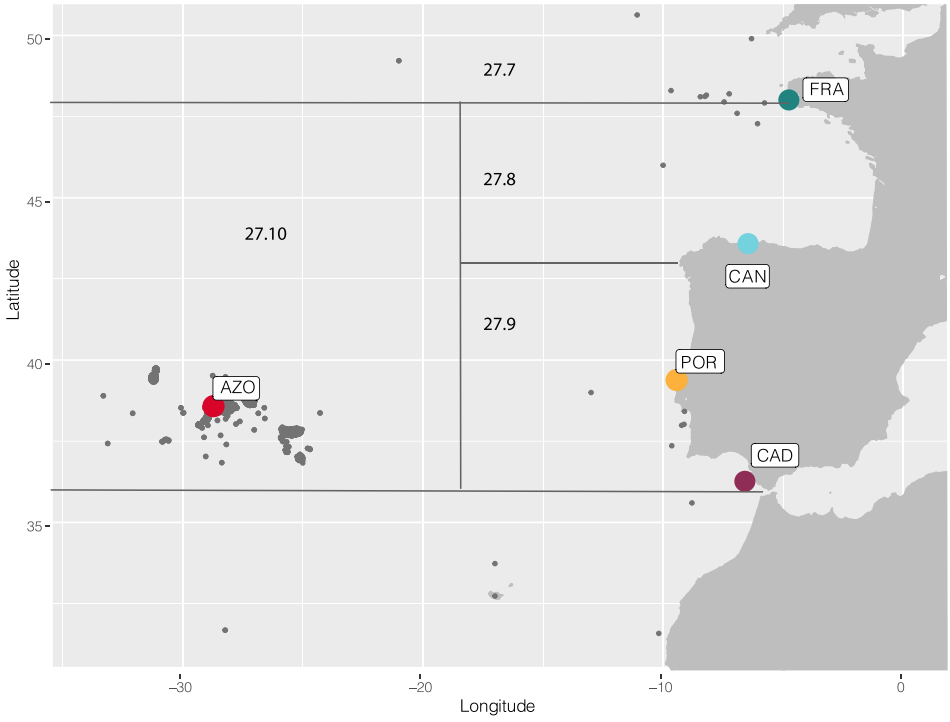


Fig. 1. Sampling sites across the distributional range of *Pagellus bogaraveo* in the northeast Atlantic (yellow: POR-Peniche, Portuguese continental waters; red: AZO-Faial, Azores, Portugal; light blue: CAD – Cádiz, Spain; green: CAN - San Vicente de la Barquera, Cantabria, Spain; purple: FRA - Île-de-Seine, France). Small grey dots are geo-referenced occurrence records from GBIF.org (12 April 2022 GBIF Occurrence Download https://doi.org/10.15468/dl.mytawc). Relevant sub-areas and divisions of FAO fishing areas 27 are shown.

Table 1
Details of *Pagellus bogaraveo* samples used in this study. n: number of individuals sampled, before and after filtering.

Country	Location	Year	Code	Latitude	Longitude	*n*	*n after filter*
Portugal	Faial, Azores	2019	AZO	38.58	-28.72	21	21
Spain	Cádiz	2009–2019	CAD	36.52	-6.28	29	26
France	Île-de-Sein	2019	FRA	48.04	-4.85	6	6
Portugal	Peniche	2019	POR	39.38	-9.44	22	14
Spain	S.Vicente de la Barquera	2020	CAN	43.54	-6.53	13	13

MsII and the barcoded adapters. The common adapter was respectively ligated on the *MsII* cut site of all samples by T4 DNA ligase. Then, 300–700 bp fragments were amplified by PCR using high-fidelity enzymes. Amplification fragments were sequenced on an Illumina NovaSeq 6000 SP FC, 2 × 150 bp (Illumina, Inc., San Diego, CA, United States).

2.3. SNP calling and filtering

FastQC (Andrews, 2010) was used to check the quality of the paired-end reads and the presence of adapters. Two different pipelines, *ipyrad v0.9.81* (Eaton and Overcast, 2020) and *Stacks* version 2.55 (Catchen et al., 2013) with two different types of assembly (reference and *de novo*) for SNP calling, were used to ensure that the results were robust, regardless the methodology. The description of the methods and results based on the *de novo* approach performed with Stacks is presented in Supplementary material_S1.

We used *ipyrad* (https://ipyrad.readthedocs.io/en/master/9-tutorials.html) with a reference-based approach. Since no genome for *P. bogaraveo* is available in the databases, the genome of another Sparidae, *Sparus aurata* Linnaeus, 1758 was used as reference. The following parameters were applied in *ipyrad*: clustering threshold of 0.85; min depth for majority-rule base calling = 6, maximum low quality base calls (Q<20) in a read = 5, max heterozygous sites per locus = 0.5. *ipyrad* does not require a prior trimming step since it includes a built-in trimming step during step 2 of assembly, using the software Cutadapt v4.2 (Martin, 2011). To assemble the data under a set of parameters defined in the parameters file, *ipyrad* goes through seven steps: Step 1 - assigns data to each sample; steps 2–5 - filters low base calls, maps paired-end reads to the reference genome, estimates heterozygosity and consensus allele sequences from clustered reads; step 6 - identifies orthologs across samples; step 7 - filters the orthologs and saves the final data in several possible output formats.

SNP filtering of both assemblies was performed with VCFtools 0.1.17 (Danecek et al., 2011) using a strategy in which genotypes that have more than 5% of missing data were filtered (–max-missing 0.95); a minimum quality score of 30 was required (–minQ 30); loci with a minor allele frequency smaller than 0.05 were removed (–maf 0.05). Then, individuals with more than 20% of missing data were removed using a combination of a script (mawk' $5 > 0.20' out. imiss |cut -f1 > lowDP. indv; https://www.ddocent.com/filtering/) and VCFtools (vcftools –vcf file.vcf –remove lowDP.indv –recode –recode-INFO-all –out file_filt_miss.vcf). VCF files were converted to *genind* and *genlight* R class objects that were then transformed into all required formats.

2.4. Identification of outlier loci

We implemented three different F_{ST} -based methods to identify outlier loci, potentially under selection: OutFLANK (Whitlock and Lotterhos, 2015), PCAdapt (Duforet-Frebourg et al., 2014; Luu et al., 2017) and a Bayesian approach with BayeScan (Foll and Gaggiotti, 2008).

OutFLANK is based on the calculation of the likelihood of F_{ST} values for each locus and estimating the distribution of F_{ST} for neutral markers. It uses a trimmed distribution of F_{ST} values to improve robustness in the presence of outliers. OutFLANK then assigns q-values to individual loci to determine the significance of outliers, considering spatially heterogeneous selection. We used the gl.outflank of the *dartR* R package, with a desired false discovery rate threshold of 0.05 for calculating q-values. OutFLANK generates the null distribution by trimming extreme Fst values from the observed distribution and fitting a modified a x2 probability density distribution on the remaining values. We used default parameters (left and right trim fraction = 0.05; minimum heterozygosity = 0.10), and loci were considered outliers when FDR q-value < 0.05.

PCAadatp stands for "Principal Component Analysis adapted for Outlier Detection with SNP data." This method utilizes principal component analysis (PCA) on SNP data to detect outliers based on their positions in the genetic variation space. It identifies individuals that deviate from the main genetic structure, helping to detect potential outlier loci. We run the function *pcadapt* of the R-package with the same name (Luu et al., 2017). We first performed an analysis with a number of principal components higher than the number of sampling locations (K=6). Then we did a screen plot (a line plot of the eigenvalues of factors or principal components) to choose the value of K. The recommended value of K corresponds to the largest value of K before the plateau of the plot is attained. The statistical test to detect outlier SNPs is the Mahalanobis distance, calculated between the K correlations of the SNP and each axis, along with the mean correlations and scaled by a constant. Assuming the absence of outliers, this distance should follow a chi-square distribution with K degrees of freedom. By default -*p*-values of SNPs with a minor allele frequency smaller than 0.05 are not computed. Loci with Mahalanobis distances that do not follow the distribution of the main bulk of points and have FDR q-value < 0.05 were considered as outliers.

BayeScan uses a Bayesian approach to identify outlier loci based on allele frequency differentiation among populations. It estimates the posterior probability that a given locus is under the effect of selection by comparing two alternative models based on the inclusion or exclusion of selection (Foll and Gaggiotti, 2008). We used Bayescan 2.1 (Foll, 2012) command line with default chain parameters (sample size = 5000; thinning interval = 10; pilot runs = 20; pilot run length = 5000, and additional burn-in = 50,000). The convergence of the Markov Chain Monte Carlo (MCMC) simulation was assessed using the R package, *coda* (Plummer et al., 2006). Loci with false discovery rate (FDR) q-value < 0.05 were considered outliers.

The final outlier dataset is composed of all the outlier loci of the three methods, removing the ones that were common to two or more methods.

2.5. Genetic diversity, structure and differentiation analyses

Population genetic analyses were performed on the all loci and on the outlier datasets that resulted from the *ipyrad* reference-based assembly.

For each sampled location, the mean observed and expected heterozygosities, *Ho* and *He,* respectively, and F_{IS} (inbreeding coefficient) with corresponding confidence intervals were calculated using the function "gl.report.heterozygosity" of the *dartR* R-package (Gruber et al., 2018; Mijangos et al., 2022). The F_{IS} values measure the degree of inbreeding within a population relative to a subpopulation. They range from −1 to +1, where: a $F_{IS} = 0$ suggests that the population and subpopulation have similar genetic diversity, i.e., no inbreeding or outbreeding is occurring; $F_{IS} < 0$ suggests outbreeding (greater

heterozygosity within the population than expected) and $F_{IS} > 0$ suggests inbreeding (greater homozygosity within the population than expected) (Kardos et al., 2016). The Hardy-Weinberg Equilibrium (HWE) test p-values were estimated with the "hw.test" function of the *pegas* (Paradis, 2010) R-package. The "diffCalc") of the *diveRsity* R package was used to estimate the 95% confidence intervals of all estimates based on 10,000 bootstrap iterations.

The *diveRsity* R package (Keenan et al., 2013) was used to evaluate the genetic structure, estimating a non-standardized fixation index F_{ST} (Weir and Cockerham, 1984), a standardized fixation index G'_{ST} (Hedrick, 2005) and differentiation D (Jost, 2008). All these measures are strongly affected by the range of shared alleles between populations, with dominant alleles substantially influencing the behaviour of each metric. Since there is no single definitive measure that fully encapsulates genetic differentiation between populations (for a review see Bird et al., 2011), presenting multiple indicators that are consistent with each other in our datasets demonstrates the reliability of our pairwise findings.

To infer population structure, discriminant analysis of principal components (DAPC) was used using functions implemented in the *ape* (Paradis et al., 2004; Paradis and Schliep, 2019) and *adegenet* (Jombart, 2008) R-packages, respectively. For DAPC, we inferred the number of clusters using the function "snapclust.choose.k", a maximum likelihood-based method (Beugin et al., 2018), also implemented in the *adegenet* R package. The result was combined with Akaike Information Criterion (AIC) (Akaike, 1973) goodness-of-fit statistics, to guide the choice of the optimal number of clusters. Additionally, the "compoplot" function was used to represent the group assignment probability of individuals the several locations.

3. Results

3.1. SNP calling and results from two different assembly strategies

The 91 sequenced individuals generated 144 × 10^6 Illumina paired-end reads. The reference-based assembly performed by *ipyrad* resulted in an average of 2480,271.2 raw reads per individual and after all filtering steps yielded 34,389 SNPs from 80 individuals. After filtering, *Stacks de novo* assembly resulted in 15,936 SNPs from 83 individuals. Results from genetic diversity, structure and differentiation analyses based on *ipyrad* and Stacks assemblies rendered similar results.

3.2. Detection of outlier loci

The three analytical methods (OutFLANK, PCAdapt, and BayeScan) successfully identified putative outlier loci from the all-loci dataset. OutFLANK detected a relatively low number of putative loci under selection (outliers), accounting for just 0.3% (110 loci) of the total. In contrast, PCAdapt and BayeScan identified a substantially higher percentage, capturing 5.5% (1899 loci) and 11.3% (3863 loci), respectively. The final outlier dataset contained 4657 unique loci or 13.6% of the initial data set.

3.3. Genetic differentiation and population structure

Overall, an analysis of the all loci dataset revealed that the observed heterozygosity ($H_O = 0.221$) was marginally lower than the expected heterozygosity ($H_E = 0.243$). A similar trend was noted in the subset of outlier loci, where the observed heterozygosity ($H_O = 0.206$) also fell short of the expected heterozygosity ($H_E = 0.301$), as detailed in Table 2. *P*-values from the HWE test showed no significant deviation from the equilibrium in any location for the all loci dataset. Estimated F_{IS} values for the all loci SNPs dataset were all positive and very low (Table 2 and Supplementary material_Fig. 2_S1) with confidence interval crossing zero in three of the five locations. For the outlier loci dataset, there is a single negative F_{IS} and the remaining positive values are larger than the ones of the all loci dataset (Table 2). Confidence intervals of the F_{IS} values corresponding to the outlier dataset do not cross zero (Supplementary material_Fig. 2_S1).

Allelic richness and nucleotide diversity are fairly similar across all locations; however, both metrics consistently exhibit lower values in the Azores (Table 2). Departures from the Hardy-Weinberg Equilibrium (HWE) were not detected in any of the sampled locations, when all the loci are considered (Table 2). The HWE test was not performed for the outlier dataset because it is composed of loci putatively under selection, which does not meet the assumptions required for this test.

The Fixation (F_{ST} and G'_{ST}) and differentiation (D) indices exhibited varying results for both the all loci and the outlier loci datasets. In the all loci dataset, all F_{ST}, G'_{ST} and D pairwise comparisons yielded significant results, except the following pairs: Portugal/ Cantabria (POR-CAN); Portugal/France (FRA-POR), and Cantabria/France (CAN-FRAN). Any comparison involving the Azores consistently showed significant results for both datasets. Some of the pairwise comparisons that were non-significant when based on the all loci dataset returned significant using the outlier dataset (e.g., CAN-FRA or POR-CAN; see Table 3 for details). On the contrary, the CAD-POR pairwise comparisons returned non-significant in the outlier dataset but significant when based on the all loci dataset.

The Discriminant Analysis of Principal Components (DAPC) conducted on both the complete dataset (Fig. 2 A-B) and outlier loci (Fig. 2 C-D), as identified by the reference-based approach in *ipyrad*, revealed three distinct clusters. Axis 1 separates the cluster that encompasses all samples originating from the Azores and another cluster including all samples from Cádiz, along with a single specimen caught off Peniche, in Portuguese continental waters. The third cluster comprises the

Table 2
Genetic diversity parameters estimated for *Pagellus bogaraveo* from the five locations in the Northeast Atlantic (AZO: Azores; CAD: Gulf of Cádiz; POR: Peniche, Portugal; CAN: Cantabria, Spain; FRA: Île-de-Seine, France) and test for departures from the Hardy-Weinberg Equilibrium (HWE) based on single nucleotide polymorphisms (SNPs). Ho, observed heterozygosity; He, unbiased expected heterozygosity (gene diversity); [C.I.] confidence intervals; π, nucleotide diversity; Ar, allelic richness; F_{IS}, inbreeding coefficient; HWE *p*-values resulting from the Hardy-Weinberg test.

All 34,206 loci	Code	*Ho [C.I.]*	*He [C.I.]*	*π*	*Ar*	*F_{IS} [C.I.]*	*HWE p-value*
	AZO	0.210 [0.125–0.226]	0.216 [0.096–0.212]	0.222	1.507	0.041 [– 0.111–0.088]	0.148
	CAD	0.230 [0.151–0.261]	0.233 [0.148–0.267]	0.238	1.539	0.024 [– 0.011–0.131]	0.164
	FRA	0.222 [– 0.028–0.305]	0.203 [0.023–0.289]	0.232	1.499	0.003 [0.060–0.350]	0.063
	POR	0.220 [0.038–0.237]	0.226 [0.079–0.260]	0.239	1.524	0.046 [0.093–0.269]	0.094
	CAN	0.218 [0.171–0.295]	0.227 [0.147–0.312]	0.238	1.507	0.062 [– 0.059–0.225]	0.129
	TOTAL	0.221 [0.169–0.230]	0.243 [0.202–0.262]	0.245	2.005	0.076 [0.056–0.136]	0.089
Outliers 4265 loci	AZO	0.166 [0.087–0.189]	0.167 [0.047–0.144]	0.171	1.467	0.025 [– 0.240 - – 0.002]	
	CAD	0.226 [0.224–0.317]	0.268 [0.323–0.405]	0.279	1.679	0.116 [0.225–0.330]	
	FRA	0.199 [0.232–0.333]	0.249 [0.283–0.379]	0.209	1.674	0.202 [– 0.222–0.293]	
	POR	0.252 [0.225–0.392]	0.204 [0.151–0.234]	0.279	1.578	-0.078 [– 0.592–0.008]	
	CAN	0.216 [0.128–0.307]	0.266 [0.254–0.398]	0.280	1.692	0.168 [0.108–0.483]	
	TOTAL	0.206 [0.163–0.213]	0.301 [0.343–0.389]	0.304	2.007	0.240 [0.383–0.450]	

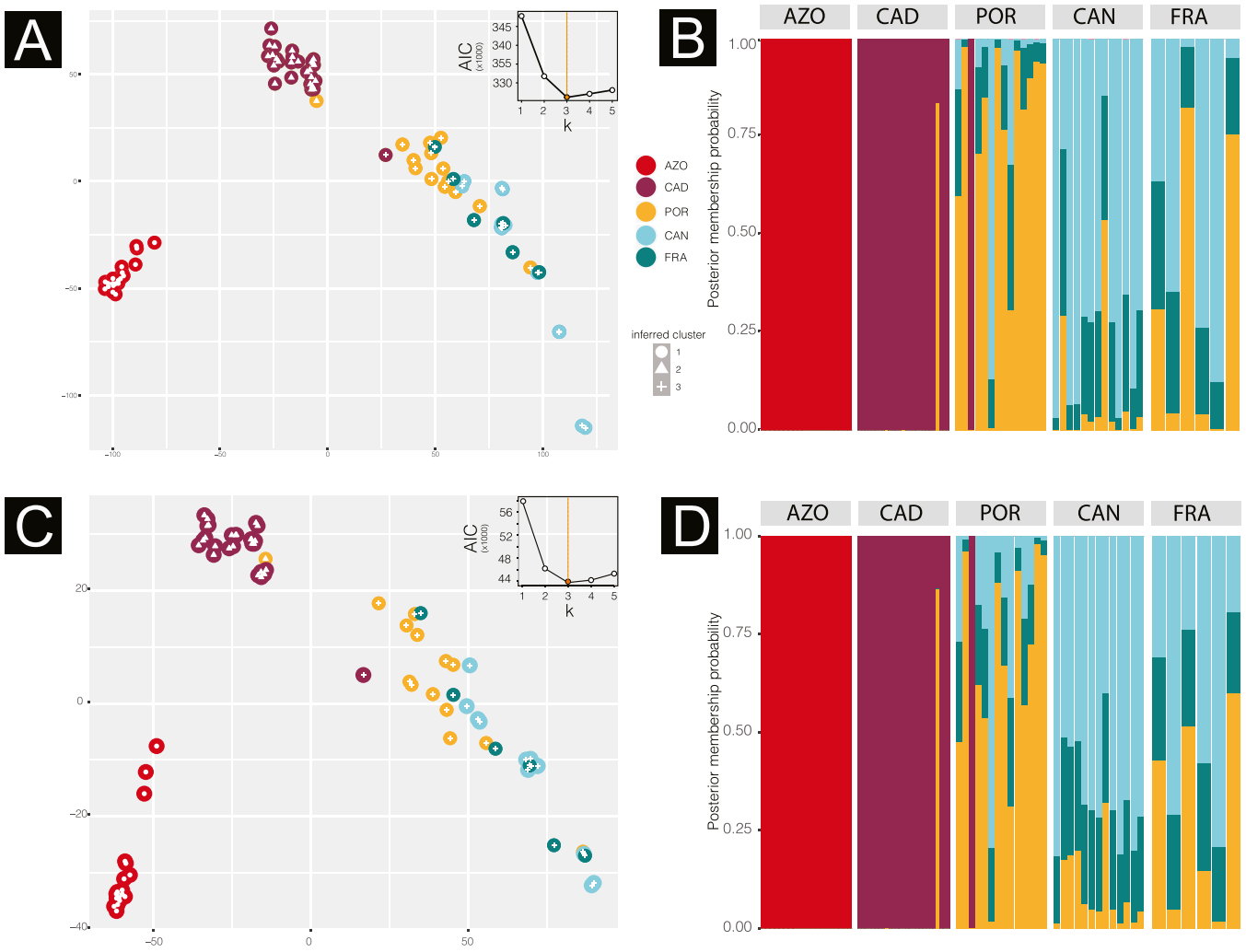


Fig. 2. A. Scatterplot of individuals on the two principal components of the Discriminant Analysis of Principle Components (DAPC) of the all loci dataset obtained from the reference-based approach performed with *ipyrad*. The graph represents the individual groups as symbols and sampling locations as colours. Inset: representation of the Akaike Information Criterion value according to the number of populations K considered. B. Representation of the individual probability of assignment obtained with the *snapclust.choose.k* and *compoplot* functions for the individuals from the all loci dataset. C. Scatterplot of individuals on the two principal components of the DAPC of the outlier loci dataset obtained from the reference-based approach performed with *ipyrad*. The graph represents the individual groups as symbols and sampling locations as colours Inset: representation of the Akaike Information Criterion value according to the number of populations K considered. D. Representation of the individual probability of assignment obtained with the *snapclust.choose.k* and *compoplot* functions for the individuals from the outlier loci dataset. AZO: Azores; CAD: Gulf of Cádiz; POR: Peniche, Portugal; CAN: Cantabria, Spain; FRA: Île-de-Seine, France.

Table 3
Pagellus bogaraveo location pairwise non-standardized fixation index F_{ST} (Weir and Cockerham, 1984), standardized fixation index G'_{ST} (Hedrick, 2005), and a differentiation index *D* (Jost, 2008) for all loci and only outlier loci datasets. Values in bold represent significant values in which the 95% confidence interval does not include zero, and hence the locations are considered statistically genetically-differentiated. AZO: Azores; CAD: Gulf of Cádiz; POR: Peniche, Portugal; CAN: Cantabria, Spain; FRA: Île-de-Seine, France.

	All loci			Outlier loci		
Locations	F_{ST} (Weir and Cockerham, 1984)	G_{ST} (Hedrick, 2005)	D (Jost, 2008)	F_{ST} (Weir and Cockerham, 1984)	G_{ST} (Hedrick, 2005)	D (Jost, 2008)
AZO - CAD	**0.0671**	**0.0562**	**0.0012**	**0.3280**	**0.3169**	**0.0403**
AZO - POR	**0.1319**	**0.1121**	**0.0040**	**0.2982**	**0.2688**	**0.0218**
AZO - CAN	**0.1737**	**0.1511**	**0.0061**	**0.5109**	**0.5274**	**0.1043**
AZO - FRA	**0.1760**	**0.1535**	**0.0052**	**0.2775**	**0.2255**	**0.0094**
CAD - POR	**0.0435**	**0.0363**	**0.0006**	0.0105	0.0152	0.0005
CAD - CAN	**0.0850**	**0.0723**	0.0018	**0.0907**	**0.0867**	0.0118
CAD - FRA	**0.0846**	**0.0731**	**0.0014**	0.0688	**0.0749**	0.0063
POR - CAN	0.0206	0.0174	0.0003	**0.1732**	**0.1661**	**0.0232**
POR - FRA	0.0140	0.0124	0.0001	0.0147	0.0247	0.0005
CAN - FRA	0.0073	0.0082	0.0000	**0.3002**	**0.3004**	**0.0424**

remaining samples from Portuguese continental waters, Cantabria, and the northwest of France, as well as one additional individual from Cádiz. Notably, these findings are corroborated by a DAPC analysis based on the *de novo* reference generated using *Stacks*, as detailed in Supplementary Material S1.

According to the DAPC *compoplot*, individuals from the Azores and Cádiz (with one exception of one individual from Peniche, Portugal that is included in the Cádiz cluster) are allocated to their respective sampling locations with a high degree of confidence (Fig. 2B and D).

4. Discussion

The use of standard genetic approaches (allozymes, microsatellites or mitochondrial DNA) to analyse connectivity and population structure of exploited marine species started decades ago providing important clues for population dynamics (Cuéllar-Pinzón et al., 2016). Recent advances in genomics allowed the integration of demographic histories of marine populations in the clarification of stocks' structure (Benestan, 2019), often providing a fine-scale view of the spatial genetic structure.

In the present study, the management units of *P. bogaraveo* adopted by ICES for the Northeast Atlantic were reassessed by using a wider portion of the genome. *Pagellus bogaraveo* stocks currently considered in the NE Atlantic include the following management units: (1) the Azores; (2) Celtic Sea and the Bay of Biscay, and (3) Atlantic Iberian waters. We identified management unit 1, which is in agreement with previous studies based on mtDNA and microsatellite data, all indicating poor connectivity between the Azores and eastern Atlantic continental margin locations (Stockley et al., 2005). The DAPCs based on the all loci (Fig. 2A) and the outlier (Fig. 2B) SNP datasets, show some visual differentiation between samples from management units 2 (Cantabria and France) and 3 (Atlantic Iberian waters: Peniche, Portugal). However, this separation is not corroborated by the optimal number of genetic clusters generated by the analysis.

The study of Piñera et al. (2007) based on 12 microsatellites of samples of *P. bogaraveo* caught off the Spanish coasts (from the Cantabrian Sea to the Mediterranean, including the Gulf of Cádiz), found no significant differences within the sampled area. Another study (Robalo et al., 2021) based on the mitochondrial control region of a set of NE Atlantic samples, similar to ours, and one location from the Mediterranean (Malaga, Spain) confirmed the genetic cluster within *P. bogaraveo* in the Azores but did not recognize the genetic unit in the Gulf of Cádiz. AMOVA results and corresponding F_{ST} values showed a significant population structure for the entire sampled area attributed by Robalo et al. (2021) to differences between the Azorean and remaining populations. Regardless the dataset used, the significant *P. bogaraveo* location pairwise fixation (F_{ST} and G'_{ST}) and differentiation (D) indices (Table 3) reflect the three genetic units identified in this study: (1) Azores; (2) Iberian Atlantic coasts from the Cantabria to Peniche (Portugal), and (3) Gulf of Cádiz. To analyse the existence of genetic differentiation within the geographic area comprising the Celtic Sea, the Bay of Biscay and the Atlantic Iberian waters, a fine-scale sampling strategy is required to provide a solid scientific basis for the stock delimitation of *P. bogaraveo*. Further, the inclusion of *P. bogaraveo* samples from the Mediterranean would provide valuable information on the connectivity of populations inhabiting adjacent ocean basins.

Our results for all locations using both data sets indicate that observed heterozygosity (H_O) is marginally lower than expected heterozygosity (H_E) (Table 2), which could point to a mild heterozygote deficiency. However, the HWE tests for the all loci dataset showed no significant deviation from the equilibrium, in any location.

The values of the coefficient of inbreeding (F_{IS}) were all positive for the all loci dataset, suggesting evidence for inbreeding (Kardos et al., 2016). However, the corresponding confidence intervals include zero in three of the locations (see Supplementary Material Fig. 2_S1 for further details), implying no strong evidence for the existence of inbreeding (Stacy et al., 2021).

Allelic richness and nucleotide diversity exhibit lower values in the Azores suggesting a reduction of genetic variation of this isolated population most likely due to a founder effect or a genetic bottleneck. This same effect has been widely observed in other commercial fish species that undergone a recent range expansion (Ivanova et al., 2021). Furthermore, samples from the Azores showed the lowest nucleotide diversity value of all analysed populations, also consistent with a more recent expansion.

In terms of the fixation and differentiation indices, the data suggest that differentiation appears to be less pronounced when focusing solely on outlier loci (see in Table 3 the CAD-POR pairwise comparisons that returned non-significant in the outlier dataset but significant when based on all loci). This could imply that neutral loci make a substantial contribution to the genetic differentiation observed in the Cádiz population.

The bathymetry of the Iberian continental slopes and ocean circulation patterns may provide a plausible explanation for the existence of a third Atlantic genetic cluster for *P. bogaraveo*. The presence of abyssal plains (Seine and Horseshoe) and submarine banks (Gorringe and Ampere) adjacent to the Gulf of Cádiz interfere with the spreading of the Mediterranean Outflow Water (MOW), which is a dense water mass that flows from the Mediterranean towards the Atlantic through the Strait of Gibraltar (Iorga and Lozier, 1999). The MOW, modified by surrounding Atlantic water masses, shows an enhanced concentration of nutrients (Van Aken and Becker, 1996). This enriched water mass recirculates eastward towards the Gulf of Cádiz, impelled by the presence of submarine banks (Pascual-Collar et al., 2019) and a cyclonic gyre in this area (Orihuela-García et al., 2023). The surface circulation patterns within the Gulf of Cádiz create a persistent upwelling induced by tidal and prevailing winds (Sala et al., 2018). The combined effect of the cyclonic flow and the eastward movement of nutrient-rich waters towards the Gulf of Cádiz may induce local retention of *P. bogaraveo*, not only of larvae but also of adults that most likely use the area as a feeding ground. Nonetheless, given the geographic proximity between sampling points and present-day oceanic circulation patterns it is conceivable to assume that some individuals might cross regions, which may explain the existence of one sample from Cádiz in the Iberian clade and another from the Portuguese coastal waters (Peniche) in the Cádiz genetic unit (Fig. 2A and C).

The genomic results presented in this study provide a scientific basis to support revising management components adopted for the blackspot seabream in Iberian waters. Further studies are required to set the spatial boundary of the Gulf of Cádiz population, particularly considering additional areas adjacent to this geographic area, specifically the northern Africa and the Mediterranean.

4.1. Conclusions

In this study, we show the relevance of using genomic tools to infer fine-scale genetic structure for sock delimitation. Our results revealed a genetic cluster in the eastern Atlantic that was not previously detected by other molecular markers. The combined effects of local upwelling, induced by tidal and prevailing winds, and specific circulation patterns resultant from a complex bathymetry may play an important role on the retention of blackspot seabream larvae and adults in the Gulf of Cádiz and provide a plausible explanation for the existence of this genetic unit. These results can inform a finer-scale sampling to tune the stock boundaries in the Atlantic Iberian coasts. However, the combination of different methodologies (e.g., morphometrics and life history traits) are required to reach a robust scientific basis for the definition of stock boundaries.

Funding

The authors acknowledge the Portuguese Foundation for Science and Technology (FCT) through the projects: MARE-UIDB/04292/ 2020

granted to MARE (Marine and Environmental Sciences Centre) and CCMAR UIDB/04326/2020 granted to the Centre of Marine Sciences of the University of Algarve; the Portuguese Biological Sampling Program from the EU Data Collection Framework (PNAB/DCF); and PPCENTRO: Projecto da Pequena Pesca na Costa Ocidental Portuguesa, financed by the Operational Program MAR2020 for the European Maritime and Fisheries Fund (EMFF). Regina L. Cunha and Sara Francisco were funded by the transitional norm DL 57/2016.

CRediT authorship contribution statement

Ivone Figueiredo, Inês Farias, Rita Castilho and Joana Robalo conceived the idea and overarched research goals and aims of the manuscript; Sara Francisco collected and/or contributed pre-processed subsets of data; Regina L. Cunha and Rita Castilho analysed the data; Rita Castilho produced the figures; Regina L. Cunha lead the writing of the manuscript and all authors contributed critically to the drafts and gave final approval for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The raw data is available at SRA-NCBI biodata ID 13362372.

Acknowledgements

We are very thankful to the two anonymous reviewers that greatly contributed for improving the earlier versions of this manuscript. The authors acknowledge several researchers and institutions for supplying specimens for analyses: Azores samples were captured by Gisela Dionisio (Naturalist contact@naturalist.pt); Andrea Romero provided Cantabria samples; continental Portugal samples were collected under the PIDDAC Program (Portuguese Government Development Investments and Expenses) and the EU/DG Fisheries' Data Collection Framework (DCF); the Mediterranean samples were provided by Manuel Hidalgo and Miriam Dominguez Rodriguez (IEO - Spanish Institute of Oceanography) and collected during the Alboran survey MEDITS19; Tarifa samples were collected by Carlos Farias (IEO - Spanish Institute of Oceanography); France samples were collected by Pascal Lorance (IFREMER - French Research Institute for Exploitation of the Sea) during the EVOHE bottom trawl survey. We thank the IT Services of the University of Algarve for hosting and maintaining the R2C2 computational cluster facility (http://rcastilho.pt/R2C2/R2C2_cluster.html). We also acknowledge the R community for the generous support given throughout the data analysis of this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fishres.2023.106891.

References

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Csaksi, B.N.P.A.F. (Ed.), 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Andrews, S. FastQC: A quality control tool for high throughput sequence data. in: Accessed 2017 J., ed. Available online at: ⟨http://wwwbioinformaticsbabrahamacuk/projects/fastqc⟩ 2010.

Bargelloni, L., Alarcon, J.A., Alvarez, M.C., Penzo, E., Magoulas, A., Reis, C., Patarnello, T., 2003. Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic-Mediterranean divide. J. Evol. Biol. 16, 1149–1158.

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

Benestan, L., 2019. Population Genomics Applied to Fishery Management and Conservation. In: Oleksiak, M.F., Rajora, O.P. (Eds.), Population Genomics: Marine Organisms. Springer International Publishing, Cham.

Beugin, M.P., Gayet, T., Pontier, D., Devillard, S., Jombart, T., 2018. A fast likelihood solution to the genetic clustering problem. Methods Ecol. Evol. 9, 1006–1016.

Bird, C.E., Holland, B.S., Bowen, B.W., Toonen, R., 2011. Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. Mol. Ecol. 20, 2128–2141.

Cadrin, S.X., Goethel, D.R., Morse, M.R., Fay, G., Kerr, L.A., 2019. So, where do you come from?" The impact of assumed spatial population structure on estimates of recruitment. Fish. Res. 217, 156–168.

Catarino, D.; Jorde, P.E.; Rogers, L.; Albretsen, J.; Jahnke, M.; Sodeland, M.; Mellerud, I.; Andre, C.; Knutsen, H. Finding coarse and fine scale population structure in a coastal species: population demographics meets genomics. Genomics; 2022.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.

Cheung, W.W., Watson, R., Morato, T., Pitcher, T.J., Pauly, D., 2007. Intrinsic vulnerability in the global fish catch. Mar. Ecol. Prog. Ser. 333, 1–12.

Cuéllar-Pinzón, J., Presa, P., Hawkins, S.J., Pita, A., 2016. Genetic markers in marine fisheries: types, tasks and trends. Fish. Res. 173, 194–205.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Duforet-Frebourg, N., Bazin, E., Blum, M.G., 2014. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. Mol. Biol. Evol. 31, 2483–2495.

Eaton, D.A., Overcast, I., 2020. ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformatics 36, 2592–2594.

Estácio, S., Mendonça, A., Krug, H., Menezes, G., Branco, J., Pinho, M., 2001. Aspects of the reproduction of six demersal species captured in the Azores archipelago. Life Mar. Sci. B. 2, 83–94.

Foll, M., 2012. BayeScan v2. 1 user manual. Ecology 20, 1450–1462.

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.

Fujita, R., 2021. The assessment and management of data limited fisheries: future directions. Mar. Policy 133, 104730.

Gruber, B., Unmack, P.J., Berry, O.F., Georges, A., 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol. Ecol. Resour. 18, 691–699.

Hart, D.R., 2001. Individual-based yield-per-recruit analysis, with an application to the Atlantic sea scallop, *Placopecten magellanicus*. Can. J. Fish. Aquat. Sci. 58, 2351–2358.

Hedrick, P.W., 2005. A standardized genetic differentiation measure. Evolution 59, 1633–1638.

ICES. Report of the Working Group on the Biology and Assessment of Deep-sea Fisheries Resources. ICES CM 2007/ACFM: 28; 2007.

International Council for the Exploitation of the Sea, I. Report of the Working Group on the Biology and Assessment of Deep-Sea Fisheries Resources (WGDEEP). ICES; 2010.

Iorga, M.C., Lozier, M.S., 1999. Signatures of the Mediterranean outflow from a North Atlantic climatology: 1. Salinity and density fields. J. Geophys Res. Oceans 104, 25985–26009.

Ivanova, P., Dzhembekova, N., Atanassov, I., Rusanov, K., Raykov, V., Zlateva, I., Yankova, M., Raev, Y., Nikolov, G., 2021. Genetic diversity and morphological characterisation of three turbot (Scophthalmus maximus L., 1758) populations along the Bulgarian Black Sea coast. Nat. Conserv. 43.

Jayasinghe, R.P.K., Amarasinghe, U.S., Newton, A., 2015. Evaluation of marine subareas of Europe using life history parameters and trophic levels of selected fish populations. Mar. Environ. Res 112, 81–90.

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

Jost, L., 2008. G (ST) and its relatives do not measure differentiation. Mol. Ecol. 17, 4015.

Kardos, M., Taylor, H.R., Ellegren, H., Luikart, G., Allendorf, F.W., 2016. Genomics advances the study of inbreeding depression in the wild. Evolut. Appl. 9, 1205–1218.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W., Prodöhl, P.A., 2013. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol. 4, 782–788.

Laffargue, P., Garren, F., Duhamel, E., 2020. Oceanographic cruise. RV Thalass.

Luu, K., Bazin, E., Blum, M.G., 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. 17, 67–77.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10–12.

Mijangos, J.L., Gruber, B., Berry, O., Pacioni, C., Georges, A., 2022. dartR v2: an accessible genetic analysis platform for conservation, ecology and agriculture. Methods Ecol. Evol. 13, 2150–2158.

Morato, T., Solà, E., Grós, M.P., Menezes, G., 2001. Feeding habits of two congener species of seabreams, Pagellus bogaraveo and Pagellus acarne, off the Azores (Northeastern Atlantic) during spring of 1996 and 1997. Bull. Mar. Sci. 69, 16.

Mytilineou, C., Tsagarakis, K., Bekas, P., Anastasopoulou, A., Kavadas, S., Machias, A., Haralabous, J., Smith, C., Petrakis, G., Dokos, J., 2013. Spatial distribution and life-history aspects of blackspot seabream Pagellus bogaraveo (Osteichthyes: Sparidae). J. Fish. Biol. 83, 1551–1575.

Orihuela-García, M.A., Bolado-Penagos, M., Sala, I., Tovar-Sánchez, A., García, C.M., Bruno, M., Echevarría, F., Laiz, I., 2023. Trace metals distribution between the

surface waters of the Gulf of Cadiz and the Alboran Sea. Sci. Total Environ. 858, 159662.
Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26, 419–420.
Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Pascual-Collar, Á., G Sotillo, M., Levier, B., Aznar, R., Lorente, P., Amo-Baladrón, A., Álvarez-Fanjul, E., 2019. Regional circulation patterns of Mediterranean Outflow Water near the Iberian and African continental slopes. Ocean Sci. 15, 565–582.
Piñera, J.A., Blanco, G., Vázquez, E., Sánchez, J.A., 2007. Genetic diversity of blackspot seabream (Pagellus bogaraveo) populations off Spanish Coasts: a preliminary study. Mar. Biol. 151, 2153–2158.
Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R. N. 6, 7–11.
Qi, P., Gimode, D., Saha, D., Schröder, S., Chakraborty, D., Wang, X., Dida, M.M., Malmberg, R.L., Devos, K.M., 2018. UGbS-Flex, a novel bioinformatics pipeline for imputation-free SNP discovery in polyploids without a reference genome: finger millet as a case study. BMC Plant Biol. 18, 117.
Robalo, J.I., Francisco, S.M., Vendrell, C., Lima, C.S., Pereira, A., Brunner, B.P., Dia, M., Gordo, L., Castilho, R., 2020. Against all odds: a tale of marine range expansion with maintenance of extremely high genetic diversity. Sci. Rep. 10 (1), 13.
Robalo, J.I., Farias, I., Francisco, S.M., Avellaneda, K., Castilho, R., Figueiredo, I., 2021. Genetic population structure of the Blackspot seabream (Pagellus bogaraveo): contribution of mtDNA control region to fisheries management. Mitochondrial DNA Part A 32, 115–119.
Saad, A., Masri, M., Sabour, W., 2020. First confirmed record of Sparid Pagellus bogaraveo (Brünnich, 1768) in the Syrian marine waters (Levantine Basin). Mar. Biodivers. Rec. 13 (1).
Sala, I., Navarro, G., Bolado-Penagos, M., Echevarría, F., García, C.M., 2018. High-chlorophyll-area assessment based on remote sensing observations: the case study of Cape Trafalgar. Remote Sens. 10, 165.
Sánchez, F. Biología y pesca del besugo (*Pagellus bogaraveo*) en las subáreas VI, VII y VIII del ICES. ICES CM:11; 1983.
Schunter, C., Pascual, M., Raventos, N., Garriga, J., Garza, J.C., Bartumeus, F., Macpherson, E., 2019. A novel integrative approach elucidates fine-scale dispersal patchiness in marine populations. Sci. Rep. 9, 10796.
Secor, D., 2005. Fish migration and the unit stock: three formative debates. Stock Identification Methods. Elsevier.
Secor, D.H., 2014. Chapter two - the unit stock concept: Bounded fish and fisheries. In: Cadrin, S.X., Kerr, L.A., Mariani, S. (Eds.), Stock identification methods, second edition.,. Academic Press, San Diego.
Spedicato, M.T., Greco, S., Sophronidis, K., Lembo, G., Giordano, D., Argyri, A., 2002. Geographical distribution, abundance and some population characteristics of the species of the genus Pagellus (Osteichthyes: Percirformes) in different areas of the Mediterranean. Sci. Mar. 66, 65–82.
Stacy, R., Palma, J., Correia, M., Wilson, A.B., Andrade, J.P., Castilho, R., 2021. The paradox of retained genetic diversity of Hippocampus guttulatus in the face of demographic decline. Sci. Rep. 11, 10434.
Stockley, B., Menezes, G., Pinho, M.R., Rogers, A.D., 2005. Genetic population structure in the black-spot sea bream (*Pagellus bogaraveo* Brünnich, 1768) from the NE Atlantic. Mar. Biol. 146, 793–804.
Teixeira, J.P.N., 2013. Recruitment dynamics and early life history of the blackspot seabream, *Pagellus bogaraveo* (Perciformes: Sparidae). Universidade dos Açores.
Thresher, R.E., Koslow, J., Morison, A., Smith, D., 2007. Depth-mediated reversal of the effects of climate change on long-term growth rates of exploited marine fish. Proc. Natl. Acad. Sci. USA 104, 7461–7465.
Van Aken, H., Becker, G., 1996. Hydrography and through-flow in the north-eastern North Atlantic Ocean: the NANSEN project. Prog. Oceano 38, 297–346.
Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.
Whitehead, P.J.P. Fishes of the North-eastern Atlantic and the Mediterranean= Poissons de l'Atlantique du Nord-est et de la Méditerranée. (No Title); 1986.
Whitlock, M.C., Lotterhos, K.E., 2015. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. Am. Nat. 186, S24–S36.