



OPEN

# Highly regional population structure of *Spondyliosoma cantharus* depicted by nuclear and mitochondrial DNA data

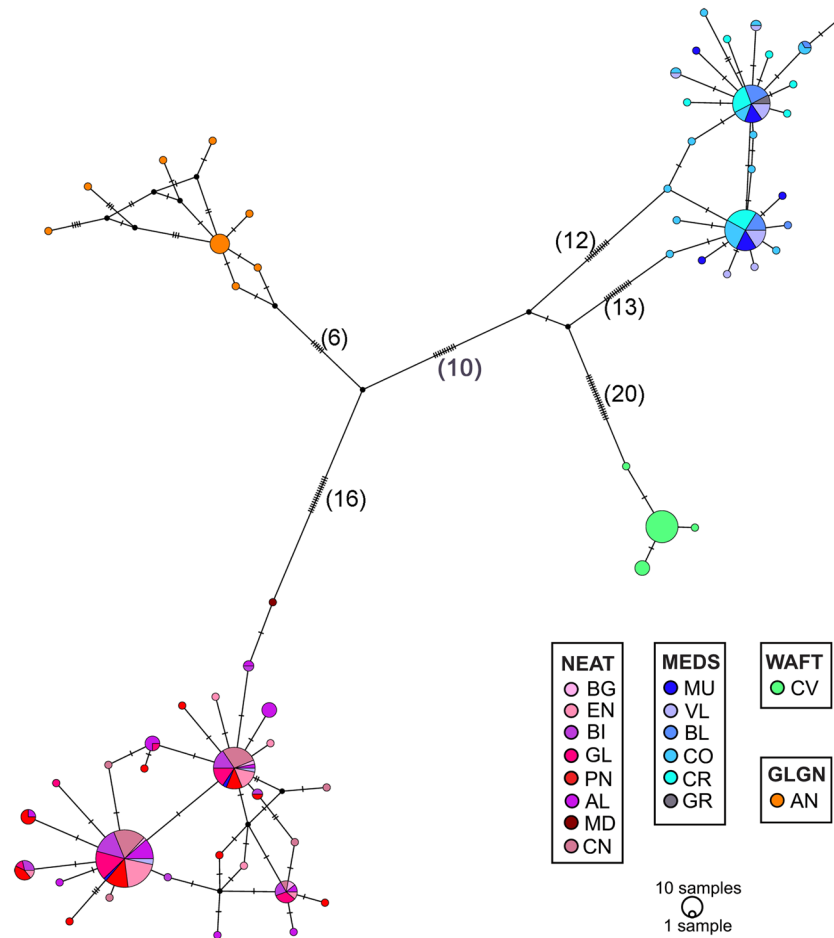
Ana Neves <sup>1,2\*</sup>, Ana Rita Vieira<sup>1,2</sup>, Vera Sequeira<sup>1,2</sup>, Rafaela Barros Paiva<sup>2</sup>, Leonel Serrano Gordo<sup>1,2</sup> & Octávio S. Paulo <sup>1,3</sup>

Resolution of population structure represents an effective way to define biological stocks and inform efficient fisheries management. In the present study, the phylogeography of the protogynous sparid *Spondyliosoma cantharus*, in the East Atlantic and Mediterranean Sea, was investigated with nuclear (S7) and mitochondrial (cytochrome b) DNA markers. Significant divergence of four regional genetic groups was observed: North Eastern Atlantic, Mediterranean Sea, Western African Transition (Cape Verde) and Gulf of Guinea (Angola). The two southern populations (Cape Verde and Angola) each comprised reciprocally monophyletic mtDNA lineages, revealed low levels of diversity in Cape Verde and high diversity for Angola despite being represented by only 14 individuals. A complete divergence between North Atlantic and Mediterranean populations was depicted by the mitochondrial marker, but a highly shared nuclear haplotype revealed an incomplete lineage sorting between these regions. Bayesian skyline plots and associated statistics revealed different dynamics among the four regions. Cape Verde showed no expansion and the expansion time estimated for Angola was much older than for the other regions. Mediterranean region seems to have experienced an early population growth but has remained with a stable population size for the last 30000 years while the North Atlantic population has been steadily growing. The lack of genetic structuring within these regions should not be taken as evidence of demographic panmixia in light of potential resolution thresholds and previous evidence of intra-regional phenotypic heterogeneity.

An essential part of sustainable management of fisheries is the identification of biological fish stocks, as it is the unit at which population assessments should be undertaken and management measures are applied<sup>1,2</sup>. Determining the spatial patterns of demographic independence and mixing among post-juvenile fish populations is critical for sustainability of fisheries<sup>3</sup>. Molecular genetic methods have been used for a long time to infer stock structure in fishes, as they proved to be efficient tools at identifying demographically independent units for fisheries management<sup>4</sup>. Overfishing such units can lead to a collapse of the fishery since a recovery of population sizes from migration is unlikely to occur and with the loss of a genetic stock, a species also loses the individuals that are adapted to a particular habitat through evolution, i.e. local adaptation<sup>5</sup>.

Sparids are highly important fisheries resources, with numerous species being commercially exploited. One of these species is the black seabream, *Spondyliosoma cantharus*, that occurs over a wide range of the eastern Atlantic, from Scandinavia to Namibia, around Madeira, Cape Verde and the Canary Islands. It is also common in the Mediterranean Sea and the western Black Sea<sup>6</sup>. Landings have reached almost 10 000 t per year in the last decade, with a major input from the north and central Eastern Atlantic, showing an increasing trend since 1980<sup>7</sup>. Life history characteristics of *S. cantharus* make it particularly vulnerable to local over-exploitation since it is a protogynous hermaphrodite, slow-growing, long-lived<sup>8</sup>, showing habitat specificity during the spawning season, with spawning aggregations and male nest guarding behaviours<sup>9,10</sup>.

<sup>1</sup>Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016, Lisboa, Portugal. <sup>2</sup>MARE – Marine and Environmental Sciences Centre, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016, Lisboa, Portugal. <sup>3</sup>cE3c - Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal. \*email: [amneves@fc.ul.pt](mailto:amneves@fc.ul.pt)



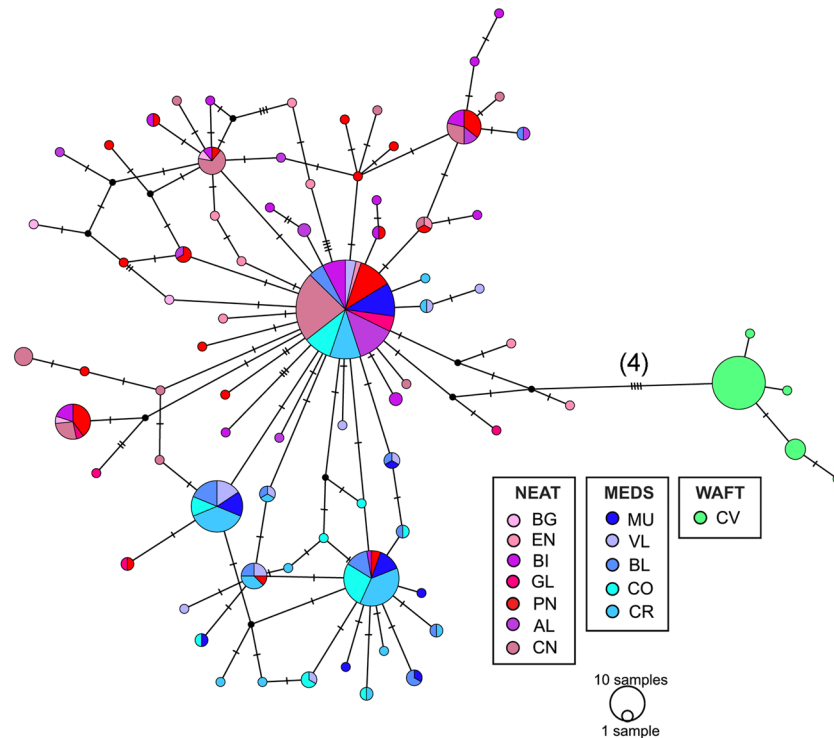
**Figure 1.** Parsimony network for *Spondyliosoma cantharus* calculated with mtDNA cytochrome *b* sequences (663 bp). The area of the circles is proportional to each haplotype frequency. Dashes represent mutations and the number of mutations between group regions are given between brackets. Colours refer to the population in which haplotypes were found. Acronyms for populations as in sampling section.

Despite the wide geographical distribution of *S. cantharus*, little information on stock structure is available for the species. Only two studies on body morphology<sup>11</sup> and otolith shape and isotopes ratios<sup>12</sup> have analysed the population structure of the species, revealing clear distinct morphotypes for the Canary Islands and Angola and evident phenotypic differences among several of the European areas analysed. The use of genetic approaches along with phenotypic ones is advised in order to create an interdisciplinary perspective and increase the probability of correct stock identification<sup>13</sup>. In this study the phylogeography and population structure of *S. cantharus* was explored across the East Atlantic and Mediterranean Sea using both mitochondrial (cytochrome *b*) and nuclear (S7) DNA, in order to identify putative stocks and compare the results to those found for the phenotypic traits.

## Results

A total of 263 *cytb* sequences were obtained, aligned, and trimmed to 663 bp, with 63 distinct haplotypes identified (GenBank accession nos. MH545766 – MH545828) showing 99 polymorphic sites yielding 81 transitions and 26 transversions. For S7, 366 sequences (corresponding to 183 individuals) with 483 bp were analysed and 85 haplotypes were identified (GenBank accession nos. MH545829 – MH545911), with 90 polymorphic sites representing 44 transitions, 34 transversions and 16 sites showed insertion/deletion variation in a total of 7 indel events with an average indel length event of 2.857.

**Population structure.** The haplotype network for the *cytb* of *S. cantharus* revealed four distinct groups (Fig. 1). The first group included mainly sequences from the North Eastern Atlantic (NEAT) areas (BG, EN, BI, GL, PN, AL and CN), with one haplotype shared among 61 individuals out of 263 and including three sampled in the Mediterranean Sea (MEDS), one from MU and two from VL. A second group, represented by MEDS sequences, showed two main haplotypes shared by 31 and 26 individuals from all the MEDS sampled areas. The two other groups were composed exclusively by private haplotypes, 4 for Cape Verde (West African Transition region, WAFT) and 8 for Angola (Gulf of Guinea region, GLGN). A similar pattern was present for the S7 haplotype network (Fig. 2), although not so distinct, since one haplotype occurred 110 times and was shared by all



**Figure 2.** Parsimony network for *Spondyliosoma cantharus* calculated with nDNA first intron S7 sequences (483 bp). The area of the circles is proportional to each haplotype frequency. Dashes represent mutations and the number of mutations between Cape Verde and the remaining populations is given between brackets. Colours refer to the population in which haplotypes were found. Acronyms for populations as in sampling section.

	cytb			S7			
	NEAT	MEDS	WAFT	GLGN	NEAT	MEDS	WAFT
NEAT		<b>32.08</b>	<b>46.81</b>	<b>27.16</b>		<b>0.045</b>	<b>0.323</b>
MEDS	<b>0.914</b>		<b>35.31</b>	<b>30.31</b>	<b>0.0505</b>		<b>0.320</b>
WAFT	<b>0.970</b>	<b>0.895</b>		<b>41.98</b>	<b>0.293</b>	<b>0.288</b>	
GLGN	<b>0.940</b>	<b>0.861</b>	<b>0.969</b>				

**Table 1.** Gene flow among regions for *Spondyliosoma cantharus* represented by  $F_{ST}$  (below diagonal) and  $\Phi_{ST}$  (above diagonal). Significant values of  $p < 0.05$  assessed by permutation test with  $1 \times 10^5$  replicates are shown in bold. NEAT – North Eastern Atlantic; MEDS – Mediterranean Sea; WAFT – West African Transition; GLGN – Gulf of Guinea.

sampled areas from the NEAT and MEDS. However, only three more haplotypes were shared between NEAT and MEDS areas (Fig. 2, Supplementary Table S1) showing a distinct separation of the haplotypes occurring in NEAT and MEDS individuals. The Cape Verde samples showed, again, only private haplotypes. None of the DNA samples from AN could be amplified for the S7 region and therefore AN was excluded from S7 results.

A significant population structuring for both *cytb* ( $F_{ST} = 0.928$ ,  $p < 0.000$ ) and the S7 ( $F_{ST} = 0.618$ ,  $p < 0.000$ ) was defined. Pairwise  $F_{ST}$  and  $\Phi_{ST}$  values for the populations within each region (Supplementary Table S2), showed no structure among the populations with all non-significant values. Conversely, a significant genetic structure, with high  $F_{ST}$  and  $\Phi_{ST}$  values between all regions, was estimated for both markers (Table 1). AMOVA indicated  $> 92\%$  (*cytb*) and  $\sim 60\%$  (S7) of variance to be due to differences among regions (Table 2).

Mantel test showed significant correlations between genetic and geographical distances for *cytb* ( $r = 0.50$ ,  $p < 0.001$ ) but not for S7 ( $r = 0.51$ ,  $p = 0.083$ ). However, when the mantel test was carried out for populations within each region, isolation by distance was not found for either *cytb* ( $r = 0.06$ ,  $p = 0.370$  and  $r = 0.14$ ,  $p = 0.298$  for NEAT and MEDS, respectively) or S7 ( $r = 0.56$ ,  $p = 0.098$  and  $r = -0.23$ ,  $p = 0.623$ , NEAT and MEDS, respectively).

**Diversity analysis.** A large proportion of singleton haplotypes were obtained for *cytb* (73%) and S7 (62%). The WAFT region showed the lower values for all the diversity indices in both DNA markers (Fig. 3). The GLGN region showed high values for all indices estimated with *cytb* marker, despite the low number of individuals sampled in this region (Fig. 3a). For *cytb* the MEDS region showed the higher diversity indices. However, values

	source of variation	df	variance components	percentage of variation	Fixation index
<b>cytb</b> 4 groups: CV vs AN vs NEAT vs MEDS	Among groups	3	<b>16.20</b>	92.44	F <sub>ST</sub> 0.928
	Among populations within groups	10	<b>0.06</b>	0.36	
	Within populations	249	<b>1.26</b>	7.19	
<b>S7</b> 3 groups: NEAT vs MEDS vs CV	Among groups	2	<b>1.61</b>	61.51	F <sub>ST</sub> 0.615
	Among populations within groups	10	<b>0.19</b>	0.7.30	
	Within populations	353	<b>0.81</b>	31.19	

**Table 2.** Hierarchical analyses of molecular variance (AMOVA) for *cytb* and S7 sequences of *Spondyliosoma cantharus*. Probability values were obtained after a permutation test with  $1 \times 10^5$  replicates and significance  $p < 0.05$  level are indicated in bold. AN – Angola; CV – Cape Verde; NEAT – North Eastern Atlantic; MEDS – Mediterranean Sea.

for this region were inflated due to only five individuals of MU and VL populations that had shared haplotypes with the NEAT region (Supplementary Table S1). For S7 marker MEDS region showed slightly higher haplotype diversity, while NEAT showed higher values for the rest of the indices (Fig. 3b).

**Demographic analysis.** Neutrality results suggest demographic expansion for NEAT and MEDS regions with significant negative values estimated (Table 3). For WAFT region the estimated values were negative but not significant (Table 3). For GLGN region only  $R_2$  test yield significant value for population growth, but given the small sample size of this region, the  $R_2$  test is expected to give the more reliable results<sup>14</sup>.

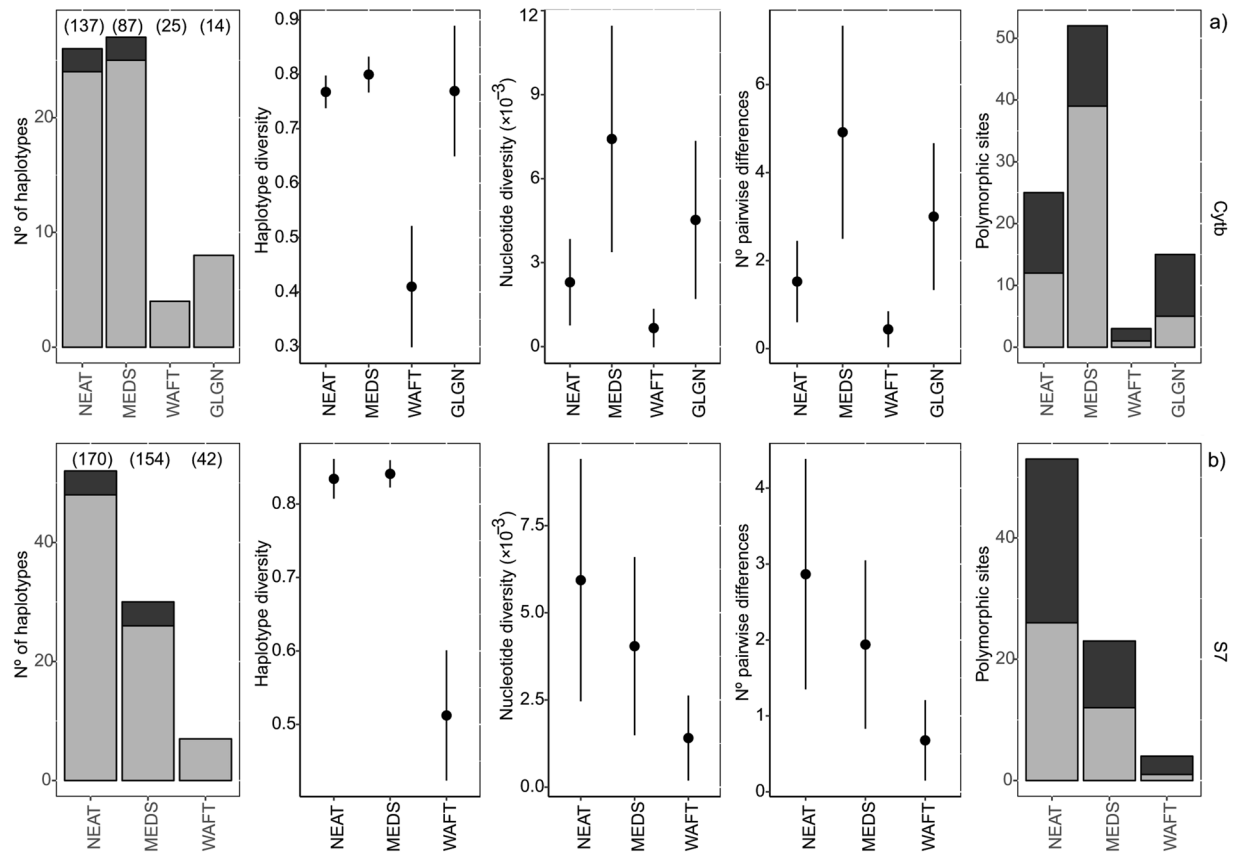
The mismatch distributions calculated for the different regions did not differ statistically from those expected for populations experiencing a demographic expansion and a spatial expansion, with exception of MEDS (Table 3, Supplementary Fig. S1). The average time estimated for demographic and spatial expansion showed similar values, reflecting an expansion period during the Pleistocene for GLGN region, while the other regions showed an expansion during the last glacial period. However, these values should be considered with caution due to the wide confidence intervals found. The tMRCA estimates were similar to the expansion estimates from mismatch analyses (Table 3). The Bayesian skyline plot showed no signatures of expansion for the WAFT region, with an effective population size much smaller than the other regions, while NEAT region has been steadily growing (Fig. 4). MEDS region seems to have experienced a population growth but has remained with a stable effective population size for the last 30 000 years (Fig. 4). The GLGN region showed an earlier expansion time and a slight increasing population trend with a peak about 100 000 years ago (Fig. 4).

## Discussion

In the present study the genetic structure of *S. cantharus* was analysed with nuclear and mitochondrial markers in 14 sampling areas along the geographical distribution of the species. The results revealed high regional population structure, with four clear groups in the analysed area: NEAT, MEDS, WAFT (represented by Cape Verde) and GLGN (represented by Angola). All the analyses clearly disclosed genetic structure among these regions, with each of the southern populations (GLGN and WAFT) comprising reciprocally monophyletic endemic clades and evidence of limited migration between NEAT and MEDS.

The Atlantic Ocean and the Mediterranean Sea are connected by the Strait of Gibraltar since the end of the Messinian Salinity Crisis (some 5.33 million years ago)<sup>15</sup>. The Strait of Gibraltar with 12.9 km wide and 286 m deep can represent a phylogeographical break, but no obvious relationship has been established between species dispersal ability or life history, and the observed patterns of genetic isolation between Atlantic and Mediterranean populations<sup>15</sup>. In the Sparidae family, divergent patterns have been observed for species with similar biology. A clear differentiation from Atlantic and Mediterranean populations have been found for species such as *Lithognathus mormyrus*, *Dentex dentex*<sup>16</sup>, *Diplodus puntazzo*<sup>17</sup> but not for others like *Pagrus pagrus*<sup>15,18</sup>, *Pagellus bogaraveo*<sup>16</sup>, *P. erythrinus*<sup>19</sup>, *Diplodus sargus*<sup>17,20</sup>. For *S. cantharus*<sup>16</sup>, found different patterns when using mitochondrial DNA, where clear differentiation between Atlantic and Mediterranean populations was detected, while no differentiation was found for allozymes data. In the present study the *cytb* marker evidenced a clear separation from NEAT and MEDS regions, with only two shared haplotypes that occurred in all NEAT populations and the two nearest MEDS populations, MU and VL. This points for a secondary contact with unidirectional gene flow from the Atlantic to the Mediterranean. The nuclear allele sharing could reflect a combination of nuclear introgression and incomplete lineage sorting. In fact, even if mutation rates were similar, mitochondrial DNA is expected to have lineage sorting four times faster than the nuclear loci, since the former have a single copy per individual and is inherited uniparentally and thus has a smaller genetically effective population size<sup>21</sup>. For protogynous fishes such as *S. cantharus*, this sorting difference may be somewhat smaller since all breeders contribute to the mitochondrial gene pool of the species. The mtDNA genomic compartment in this species can be  $\frac{1}{2}$  and not  $\frac{1}{4}$  as for most species, because the same individual will mature as female and change sex along its life and consequently, “all” individuals will function as females first, transmitting a copy of mtDNA into the population.

*S. cantharus* populations within each region showed no population structure, with very low pairwise  $F_{ST}$  and  $\Phi_{ST}$  values for both DNA markers, and lack of IBD found among samples collected within NEAT and MEDS regions. Within each region shared mitochondrial and nuclear haplotypes were present in all populations and nucleotide diversity showed no defined geographical trend with similar values for most populations. The high number of singletons found is also common among marine fishes, where the usual haplotype distribution for



**Figure 3.** Standard diversity measures by regions for *Spondyliosoma cantharus* (a) *cytb* and (b) *S7*. For each region, the number of haplotypes, haplotype diversity, nucleotide diversity, the mean number of pairwise differences and the number of polymorphic sites are shown. Number of haplotypes and polymorphic sites are divided in private/informative (grey) and shared/non informative (black). Vertical bars denote standard deviations. Number of individuals analysed are given in parenthesis above each region bar of the number of haplotypes graphic. NEAT – North Eastern Atlantic; MEDS – Mediterranean Sea; WAFT – West African Transition; GLGN – Gulf of Guinea.

large fish populations is a small number in medium to high frequencies, with most occurring in low frequencies or in single copy<sup>22</sup>.

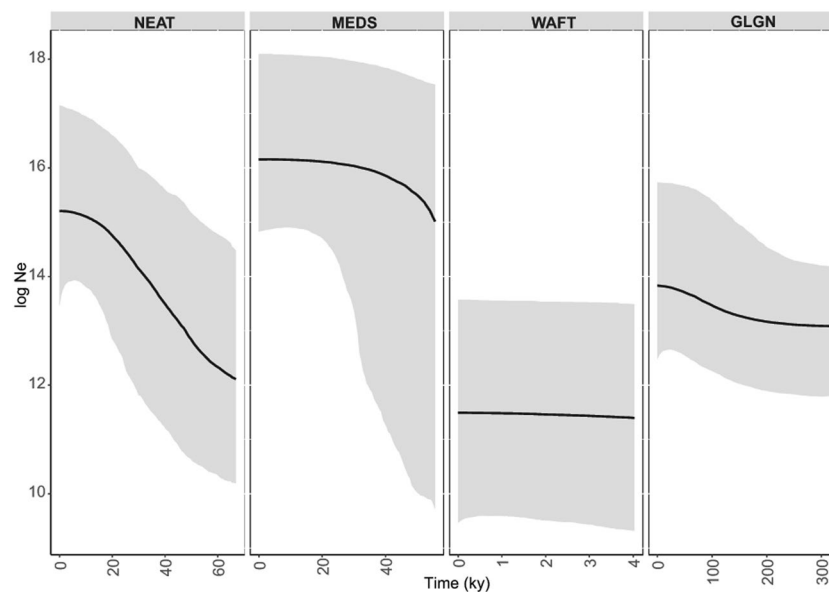
Despite efforts to gather samples from multiple areas within the Gulf of Guinea only samples from one population, Angola, could be collected. *Cytb* results showed haplotypic diversity values similar to the NEAT and MEDS regions and higher nucleotide diversity despite the low number of individuals analysed. Expansion time for GLGN region dates from late Pleistocene, and these populations are likely to have maintained large number of individuals during glacial cycles since this area can function as tropical refugia where the mild conditions<sup>23</sup> allowed to maintain a high number of breeders, justifying the high diversity found in the small number of individuals analysed. No data for the *S7* could be obtained, since this DNA marker could not be amplified for the samples in this population. Nuclear DNA experience lower mutation rates than mitochondrial DNA<sup>24</sup> and thus slower sorting among population is expected to occur, as discussed above for NEAT and MEDS regions. Unfortunately, the absence of nuclear DNA information for the GLGN region preclude to explore the occurrence of shared haplotypes with the northern regions, as described for other species such as *Pomatomus saltatrix*<sup>25</sup>.

The diversity indices found for both DNA markers in WAFT population were much lower than those estimated for the other areas. The much recent time of expansion estimated for this region can justify these low values since the population might have still had no time to highly increase diversity, persisting a single strong founder recolonization effect<sup>26</sup>. This region is strongly influenced by large-scale oceanic circulation and dominant currents which are likely to promote a physical barrier to larval drifting across the archipelago<sup>27</sup> serving to restrict gene flow. The fact that fishing has been increasing in this region since 2005<sup>28</sup> can also contribute to the decline of the genetic pool available in the archipelago, which can eventually lead to genetic drift in small populations<sup>29</sup> reducing the potential of the population to adapt to rapid environmental changes.

The population expansion time in the NEAT and MEDS regions, as estimated by BSP, dates from the Last Glacial Period (LGP). These values must be looked with caution, since no molecular clock calibration is available for the species and the substitution rates inferred at population levels are much higher than those for species level<sup>30–32</sup> holding a more recent time since expansion than that estimated from BSP and mismatch distribution. Nevertheless, it is likely that during the climatic oscillations that occurred over the LGP, the species had refugia in two different areas, one in the Mediterranean and one in the southern North Eastern Atlantic and that

	NEAT	MEDS	WAFT	GLGN
<b>neutrality tests</b>				
Tajima's D	<b>-1.912</b>	<b>-1.694</b>	-1.108	-1.484
Fs	<b>-22.014</b>	<b>-8.188</b>	-1.653	-1.936
R <sub>2</sub>	<b>0.0299</b>	0.0448	0.1018	<b>0.0826</b>
<b>mismatch distribution</b>				
demographic expansion				
τ	1.516 (0.707–2.879)		0.5 (0–1.189)	9.2 (0.133–80)
t (ky)	143 (64–454)		71 (0–170)	1318 (19–11463)
SSD	0.0013	<b>0.0011</b>	0.0074	0.0199
Hri	0.0413	<b>0.0888</b>	0.1675	0.0494
spatial expansion				
τ	0.980 (0.543–2.763)		0.526 (0.099–1.534)	6.797 (0.187–97.582)
t (ky)	140 (78–396)		75 (14–220)	974 (27–13982)
SSD	0.0013	<b>0.1073</b>	0.0073	0.0181
Hri	0.0413	<b>0.0888</b>	0.1675	0.0494
<b>BEAST Bayesian Skyline</b>				
t <sub>MRC</sub> A (ky)	225 (67–425)	160 (56–309)	81 (4–197)	713 (325–1147)

**Table 3.** Demographic parameters of *Spondyliosoma cantharus* based on *cytb* and S7. Neutrality tests: Fs (Fu's), D (Tajima's) and R2 test. Mismatch distributions: τ (age of expansion in units of mutational time) t (time in thousand years), SSD (sum of square deviation) and Hri (Harpending's Raggedness index). Bayesian Skyline: t<sub>MRC</sub>A (time to most recent common ancestor in thousand years). Significant values of probability assessed by a permutation test with  $1 \times 10^5$  replicates are shown in bold and 95% CI are given in parenthesis. NEAT – North Eastern Atlantic; MEDS – Mediterranean Sea; WAFT – West African Transition; GLGN – Gulf of Guinea.



**Figure 4.** Bayesian skyline plot showing changes in effective population size through time (thousand years before present, ky) for *Spondyliosoma cantharus* for each region. The black line represents the median posterior estimate of the effective population size. The grey area shows the 95% highest posterior density limits. NEAT – North Eastern Atlantic; MEDS – Mediterranean Sea; WAFT – West African Transition; GLGN – Gulf of Guinea.

these populations have been evolving separately since then. Refugia in North Eastern Atlantic is probable to have occurred near south Iberian Peninsula. Although no latitudinal cline in allelic richness is present, populations from Portuguese coast showed slightly higher values.

The relatively long pelagic larval phase duration (PLD) reported for *S. cantharus*<sup>33</sup> could also influence the low degree of genetic differentiation found along the European coasts, as the correlation between the PLD and  $F_{ST}$  estimates has been found to be very strong in recent modelling approaches<sup>34</sup>. However, phenotypic information available on the species points for segregation among NEAT populations, with clear morphological features of body and otoliths shape varying across the sampled areas<sup>11,12</sup>. The lack of within region genetic structure, with the low values of  $F_{ST}$  and  $\Phi_{ST}$ , can be due to just enough gene flow existing to homogenize neutral markers despite the limited exchange of individuals on average among sites. In fact, despite the proved value of neutral genetic markers in elucidating patterns of genetic connectivity among wild populations, they might not be appropriate or sensitive enough to detect shallow genetic structure or to provide information about adaptive genetic variation<sup>35</sup>. The values of divergence for quantitative trait loci between populations are expected to be much higher than those obtained with neutral markers<sup>35</sup>. Markers with high mutational rates (STRs) or a genome wide analysis (SNPs) or even candidate genes for environmental responses/morphological traits could provide a deeper insight of species stock structure (see<sup>36</sup> and references therein). Several authors have already detected, for other species, no population structure with molecular markers but differentiation with phenotypical traits (eg<sup>2,37,38</sup>), pointing to the importance from a fisheries management point of view that several methodologies should be gathered to establishing an effective knowledge of the species population structure.

## Conclusion

The obtained results reveal a large scale structure for *S. cantharus* populations. The Cape Verde and Angola regions revealed two distinct clades and the Northeast Atlantic and the Mediterranean populations showed a clear separation but with incomplete lineage sorting from a common ancestor revealed by the nuclear marker. Similar diversity was estimated for all populations analysed with exception of Cape Verde. The shallow diversity, low  $N_e$ , and seemingly isolated nature may compromise the resilience of the Cape Verde population to increasing fishing pressure and environmental changes and this should be taken in account and considered for future studies in the area.

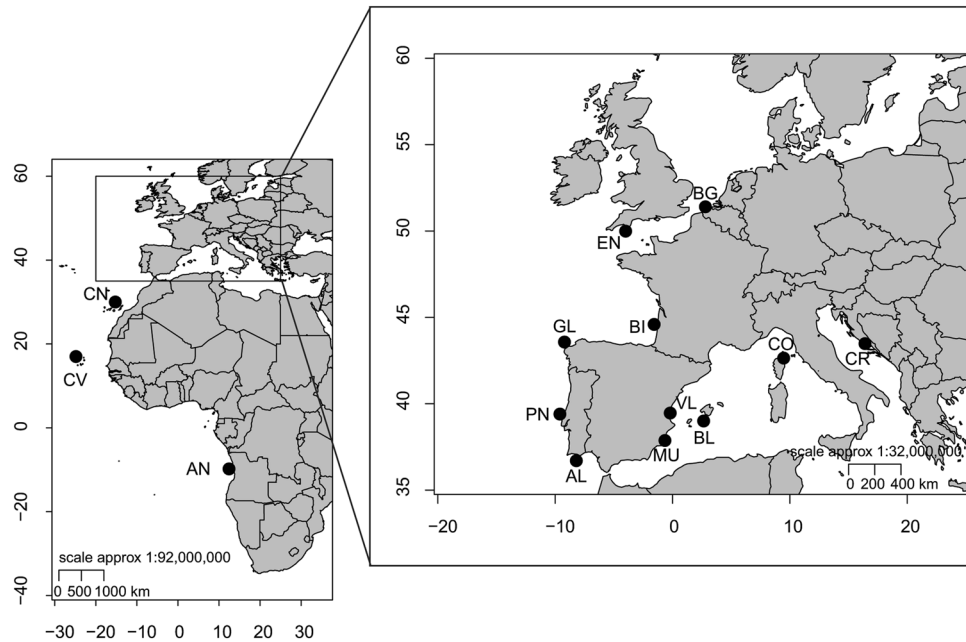
Despite the evidence from the available phenotypic traits data of significant differences within NEAT populations, genetic data showed no differentiation among populations for both NEAT and MEDS regions. This indicates that gene flow exists among locations although with some restriction as suggested by some differentiation in derived, more recent mutations that account for morphological differentiation. The markers used in this study are good for analysing event at an evolutionary timescale but less powerful for detecting events occurring at an ecological timescale. The establishment of unequivocal stock structure of the species will benefit with the use of markers with high mutation rates. The knowledge on stock structure of exploited species is essential in a world where the rising demand for fish protein and consequent increase in fishing pressure combined with global warming boosts the challenges faced by these species.

## Material and Methods

**Sampling.** A total of 263 black seabream samples were collected from 9 locations along the Eastern Atlantic and 5 locations in the Mediterranean Sea (Fig. 5, Supplementary Table S1). Samples were obtained from scientific cruises, carried by national fisheries institutes from Belgium, England, France, Spain and Angola, and purchased at the market in Portugal (Peniche and Algarve), Canary Island, Cape Verde, Corsica and Croatia. Individuals from English Channel (EN), Bay of Biscay (BI), Galicia (GL), Algarve (AL), Canary Island (CN) and Angola (AN) were frozen and sent to Lisbon. After arrival a piece of fin was removed and preserved in 96% ethanol. Samples from all the other locations, Belgium (BG), Peniche (PN), Cape Verde (CV), Murcia (MU), Valencia (VL), Balearic Islands (BL), Corsica (CO) and Croatia (CR), were taken from fresh fish and stored in 96% ethanol.

**DNA extraction, amplification, sequencing and alignment.** Genomic DNA was extracted from fin samples with the REExtract-N-Amp kit (Sigma-Aldrich) following the manufacturer's instructions. Mitochondrial DNA cytochrome *b* (*cytb*) and nuclear DNA first intron of the *S7* ribosomal protein gene (*S7*) were amplified by polymerase chain reaction (PCR). For *cytb*, amplification of the region of interest was obtained with pair of primers purposely designed by the authors, based on 9 aligned *cytb* complete sequences of *Spondyliosoma cantharus* obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank>) (accession numbers: DQ198007, EU036508/9, EF439602, EF439237, EF427607/8 and EU224083/4): BRBcytBF (forward) 5'-GCTGACTCATCCGAAATCTT-3' and BRBcytBR (reverse) 5'-ATGTAGGGTCTTCAACTGG-3'. *S7* amplifications were performed with the following pair of primers: S7RPEX1F (forward) 5'-TGGCCTCTTCCTTGGCCGTC-3' and S7RPEX2R (reverse) 5'-AACTCGTCTGGCTTTTCGCC-3'<sup>39</sup>. PCR amplification reactions were performed in a 20  $\mu$ l total-reaction volume with 10  $\mu$ l of REExtract-N-ampl PCR reaction mix (Sigma-Aldrich), 0.8  $\mu$ l of each primer (10 mM), 4.4  $\mu$ l of bidistilled water and 4  $\mu$ l of template DNA. An initial denaturation at 94 °C for 5/3 min was followed by 30 cycles (denaturation at 94 °C for 45 s, annealing at 56 °C for 30/60 s, and extension at 72 °C for 1 minute) and a final extension at 72 °C for 10 minutes on a BioRad Mycycler thermal cyler (values *cytb*/*S7*, respectively). PCR products were purified with the SureClean kit (Bioline) following the manufacturer's protocol, and the same primers were used for the sequencing reaction provided by Macrogen (<http://www.macrogen.com>, 716 samples) and STABVIDA (<http://www.stabvida.net/>, 108 samples).

Sequences were aligned using Clustal W<sup>40,41</sup> in BIOEDIT<sup>42</sup>. *S7* haplotypes of length-variant heterozygotes were determined using Mixed Sequence Reader<sup>43</sup>, and manual adjustments were made whenever necessary. Allelic states of *S7* sequences were estimated using the Bayesian programme PHASE 2.1<sup>44-46</sup>. Five runs with different seeds for the random number generator and 500 iterations as burn-in, 500 main iterations and a thinning interval = 1 were performed to check for consistency across results. All runs returned consistent allele identities.



**Figure 5.** Sampling locations of *Spondyliosoma cantharus* in the Eastern Atlantic Ocean and the Mediterranean Sea. The position of samples with black dots is only indicative. Acronyms for locations as in sampling section.

**Population structure.** The relationship between haplotypes was analysed by haplotype networks built with the software PopART<sup>47</sup> using a TCS network<sup>48</sup>. For *cytb* network some additional GenBank sequences from Madeira (MD) and Greece (GR) were added to extend the sampling area (GenBank accession nos. EF439237, EU036508/9).

Hierarchical analysis of molecular variance (AMOVA)<sup>49</sup> was used to estimate genetic structure among and between regions. Estimates of genetic divergence among regions and among populations within region were calculated with the fixation index  $F_{ST}$  and  $\Phi_{ST}$ <sup>49</sup> as implemented in ARLEQUIN V3.5<sup>50</sup>, with p-values being corrected for multi comparisons by Benjamini-Hochberg (BH) false discovery rate method<sup>51</sup>.

The correlation between geographical distance, measured along the coastline, and  $F_{ST}$  was computed with the Mantel test<sup>52,53</sup>, also in ARLEQUIN with  $1 \times 10^5$  permutations.

**Diversity analysis.** Diversity indices (number of haplotypes, % of private haplotypes, haplotype and nucleotide diversities, mean pairwise differences and number of polymorphic loci)<sup>54,55</sup> were estimated for all populations using ARLEQUIN V3.5<sup>50</sup>. Graphical presentation of diversity indices was obtained using ggplot2 package<sup>56</sup> for R<sup>57</sup>.

**Demographic analysis.** Demographic history for the region structure depicted for *cytb* by the haplotype networks and AMOVA analyses was investigated in ARLEQUIN V3.5 with neutrality tests, Tajima's  $D$ <sup>58</sup> and Fu's  $F_S$ <sup>59</sup>, the  $R_2$ <sup>14</sup> was also estimated using package pegas<sup>60</sup> for R<sup>57</sup>. Mismatch distribution analysis was also performed for each region and compared to the distribution expected in populations affected by sudden expansion ( $1 \times 10^5$  replicates), under the assumption of selective neutrality, in which a unimodal distribution is expected<sup>61</sup>. The sum of squared deviation (SSD) and raggedness index (Hri) were used to detect departure between observed and expected distributions. If evidence of expansion was found (p values > 0.05) the  $\tau$  parameter of demographic and spatial expansions was used to estimate the time since the expansion (t, in years), using the equation  $t = (\tau \times n) / (2 \times \mu \times k)$ , where  $\mu$  is the nucleotide mutation rate,  $k$  is the sequenced number of nucleotides and  $n$  is the generation time (equal to age for sexual maturity)<sup>62</sup>. Since there is no information on *cytb* mutation rate for *S. cantharus*, the values used in the present study were 2% per nucleotide per Myr based on<sup>16</sup>. The generation time used was 3.8 years, according to<sup>63</sup>.

Past population demography of *S. cantharus* was also inferred using Bayesian skyline plot (BSP)<sup>64</sup>, employing the Bayesian MCMC coalescent method and a strict clock as implemented in BEAST 1.8.4<sup>65</sup>. Best-fit model of nucleotide substitution for *cytb* was estimated with ModelFinder in IQ-TREE<sup>66</sup>. The TN + G4 was the model selected and an average mutation rate of 2% per nucleotide per Myr was used. The Bayesian distribution was generated with  $1 \times 10^8$  Markov Chain Monte Carlo (MCMC) steps, and the convergence of parameters was visually checked with effective samples sizes (ESS) of estimates over 200 with TRACER 1.7.1<sup>67</sup>. The time to most recent common ancestor (tMRCA) and the median and corresponding credibility intervals of the BSP were assessed for each population.



## Data availability

The sequences generated under this study are available from the GenBank database and accession numbers are given in Results section.

Received: 1 July 2019; Accepted: 20 February 2020;

Published online: 04 March 2020

## References

- Begg, G. A., Friedland, K. D. & Pearce, J. B. Stock identification and its role in stock assessment and fisheries management: an overview. *Fish. Res.* **43**, 1–8 (1999).
- Abaunza, P. *et al.* Stock identity of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean Sea: integrating the results from different stock identification approaches. *Fish. Res.* **89**, 196–209 (2008).
- Welch, D. *et al.* Integrating different approaches in the definition of biological stocks: a northern Australian multi-jurisdictional fisheries example using grey mackerel. *Scomberomorus semifasciatus*. *Mar. Policy* **55**, 73–80 (2015).
- Ward, R. D. Genetics in fisheries management. *Hydrobiologia* **420**, 191–201 (2000).
- Gopalakrishnan, A., Jayasankar, J., Shah, P. & Shalin, S. Genetic stock characterization of fish using molecular markers in *Course manual summer school on advanced methods for fish stock assessment and fisheries management*. Lecture Note Series No. 2/2017 308–316 (CMFRI, Kochi, 2017).
- Bauchot, M. L. & Hureau, J. C. Sparidae in: *Fishes of the North-Eastern Atlantic and the Mediterranean* (eds. Whitehead, P. J. P., Bauchot, M. L., Hureau, J. C., Nielsen, J. & Tortonese, E.) 883–907 (UNESCO, 1986).
- FAO Fisheries and Aquaculture Department. <http://www.fao.org/fishery/statistics/collections/en> (2010–2018).
- Neves, A., Vieira, A. R., Sequeira, V., Paiva, R. B. & Gordo, L. S. Modelling the growth of a protogynous sparid species, *Spondyliosoma cantharus* (Teleostei, Sparidae). *Hydrobiologia* **797**, 265–275 (2017).
- Russell, B., Pollard, D. & Carpenter, K. E. *Spondyliosoma cantharus*. The IUCN Red List of Threatened Species 2014: e.T170258A1303321. 10.2305/IUCN.UK.2014-3.RLTS.T170258A1303321.en (2014).
- Pinder, A. C., Velterop, R., Cooke, S. J. & Britton, J. R. Consequences of catch-and-release angling for black bream *Spondyliosoma cantharus*, during the parental care period: implications for management. *ICES J. Mar. Sci.* **74**, 254–62 (2017).
- Neves, A., Vieira, A. R., Sequeira, V., Paiva, R. B. & Gordo, L. S. Phenotypic changes in the body of black seabream, *Spondyliosoma cantharus* (Teleostei: Sparidae), along the eastern Atlantic. *Estuar. Coast. Shelf. Sci.* **214**, 31–37 (2018).
- Neves, A. *et al.* Otolith shape and isotopic ratio analyses as a tool to study *Spondyliosoma cantharus* population structure. *Mar. Environ. Res.* **143**, 93–100 (2019).
- Cadrin, S. X. & Secor, D. H. Accounting for spatial population structure in stock assessment: past, present, and future in: *The future of fisheries science in North America* (eds. Beamish, R. J. & Rothschild, B. J.) 405–426 *Fish. Fish. Ser.* 31 (Springer, 2009).
- Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100 (2002).
- Patarnello, T., Volckaert, F. & Castilho, R. Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break? *Mol. Ecol.* **16**, 4426–4444 (2007).
- Bargelloni, L. *et al.* Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic–Mediterranean divide. *J. Evol. Biol.* **16**, 1149–1158 (2003).
- Bargelloni, L. *et al.* The Atlantic–Mediterranean transition: Discordant genetic patterns in two seabream species, *Diplodus puntazzo* (Cetti) and *Diplodus sargus* (L.). *Mol. Phylogenet. Evol.* **36**, 523–535 (2005).
- Ball, A. O., Beal, M. G., Chapman, R. W. & Sedberry, G. R. Population structure of red porgy, *Pagrus pagrus*, in the Atlantic Ocean. *Mar. Biol.* **150**, 1321–1332 (2007).
- Angiulli, E., Sola, L., Ardizzone, G., Fassatoui, C. & Rossi, A. R. Phylogeography of the common pandora *Pagellus erythrinus* in the central Mediterranean Sea: sympatric mitochondrial lineages and genetic homogeneity. *Mar. Biol. Res.* **12**, 4–15 (2016).
- González-Wangüemert, M., Froufe, E., Pérez-Ruzafa, A. & Alexandrino, P. Phylogeographical history of the white seabream *Diplodus sargus* (Sparidae): Implications for insularity. *Mar. Biol. Res.* **7**, 250–260 (2011).
- Palumbi, S. R., Cipriano, F. & Hare, M. P. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* **55**, 859–868 (2001).
- Luzier, C. W. & Wilson, R. R. Analysis of mtDNA haplotypes of kelp bass tests for sibling-dominated recruitment near marine protected areas of the California Channel Islands. *Mar. Ecol. Prog. Ser.* **277**, 221–230 (2004).
- Schefuß, E., Sinninghe Damsté, J. S. & Jansen, J. H. F. Forcing of tropical Atlantic sea surface temperatures during the mid-Pleistocene transition. *Paleoceanography* **19**, PA4029 (2004).
- Ballard, J. W. O. & Whitlock, M. C. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744 (2004).
- Reid, K. *et al.* Secondary contact and asymmetrical gene flow in a cosmopolitan marine fish across the Benguela upwelling zone. *Heredity* **117**, 307–315 (2016).
- Bernatchez, L., Dodson, J. J. & Boivin, S. Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. *J. Fish Biol.* **35**, 233–244 (1989).
- Cunha, R. L., Lopes, E. P., Reis, D. M. & Castilho, R. Genetic structure of *Brachidontes puniceus* populations in Cape Verde archipelago shows signature of expansion during the last glacial maximum. *J. Mollus. Stud.* **77**, 175–181 (2011).
- FAO Fishery and aquaculture country profiles. Cabo Verde (2008) country profile fact sheets in *FAO Fisheries and Aquaculture Department*, <http://www.fao.org/fishery/facp/CPV/en> (2005–2018).
- Smith, P. J. Genetic diversity of marine fisheries resources: possible impacts of fishing. *FAO Fisheries Technical Paper* No. 344 (Rome, FAO, 1994).
- Ho, S. Y., Phillips, M. J., Cooper, A. & Drummond, A. J. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* **22**, 1561–1568 (2005).
- Ho, S. Y. & Shapiro, B. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* **11**, 423–434 (2011).
- Subramanian, S. & Lambert, D. M. Time dependency of molecular evolutionary rates? Yes and no. *Genome Biol. Evol.* **3**, 1324–1328 (2011).
- Raventós, N. & Macpherson, E. Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Mar. Biol.* **138**, 1115–1120 (2001).
- Faurby, S. & Barber, P. H. Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Mol. Ecol.* **21**, 3419–3432 (2012).
- Conover, D. O., Clarke, L. M., Munch, S. B. & Wagner, G. N. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *J. Fish Biol.* **69**, 21–47 (2006).
- Kochzius, M. Trends in fishery genetics in: *The future of fisheries science in North America* (eds. Beamish, R. J. & Rothschild, B. J.) *Fish. Fish. Ser.* **31**, 453–493 (Springer, Dordrecht, 2009).
- Marin, K., Coon, A., Carson, R., Debes, P. V. & Fraser, D. J. Striking phenotypic variation yet low genetic differentiation in sympatric lake trout (*Salvelinus namaycush*). *PLoS ONE* **11**, e0162325 (2016).

38. Vieira, A. R. *et al.* Genetic and morphological variation of the forkbeard, *Phycis phycis* (Pisces, Phycidae): evidence of panmixia and recent population expansion along its distribution area. *PLoS ONE* **11**, e0167045 (2016).
39. Chow, S., Hazama, K. & Universal, P. C. R. primers for S7 ribosomal protein gene introns in fish. *Mol. Ecol.* **7**, 1255–1256 (1998).
40. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic. Acids. Res.* **22**, 4673–4680 (1994).
41. Larkin, M. A. *et al.* Clustal W and Clustal X Version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
42. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic. Acids. Symp. Ser.* **41**, 95–98 (1999).
43. Chang, C.-T. *et al.* Mixed Sequence Reader: A program for analyzing dna sequences with heterozygous base calling. *Sci. World J.* **2012**, 1–10 (2012).
44. Stephens, M., Smith, N. J. & Donnelly, P. A. new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
45. Stephens, M. & Donnelly, P. A. comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
46. Stephens, M. & Scheet, P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* **76**, 449–462 (2005).
47. Leigh, J. W. & Bryant, D. PopART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
48. Clement, M., Snell, Q., Walke, P., Posada, D. & Crandall, K. TCS: estimating gene genealogies. *Proc. 16th Int. Parallel Distrib. Process Symp.* **2**, 184 (2002).
49. Excoffier, L., Smouse, P. E. & Quattr, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
50. Excoffier, L. & Lischer, H. E. L. An integrated software package for population genetics data analysis. *Mol. Ecol. Resour.* **10**, 564–67 (2010).
51. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B. Met.* **57**, 289–300 (1995).
52. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
53. Smouse, P. E., Long, J. C. & Sokal, R. R. Multiple regression and correlation extensions of the Mantel Test of matrix correspondence. *Syst. Zool.* **35**, 627–632 (1986).
54. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
55. Nei, M. *Molecular evolutionary genetics* (Columbia University Press, 1987).
56. Wickham, H. *ggplot2: Elegant graphics for data analysis* (Springer-Verlag, 2009).
57. R Core Team. R: a language and environment for statistical computing. *r foundation for statistical computing*. <http://www.R-project.org/> (2015).
58. Tajima, F. The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601 (1989).
59. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–25 (1997).
60. Paradis, E. *pegas: an R package for population genetics with an integrated-modular approach*. *Bioinformatics* **26**, 419–420 (2010).
61. Rogers, A. R. & Harpending, H. C. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
62. Yang, C., Lian, T., Wang, Q.-X., Huang, Y. & Xiao, H. Preliminary study of genetic diversity and population structure of the Relict Gull *Larus relictus* (Charadriiformes Laridae) using mitochondrial and nuclear genes. *Mitochondr. DNA* **27**, 4246–4249 (2016).
63. Neves, A., Vieira, A. R., Sequeira, V., Paiva, R. B. & Gordo, L. S. Insight on reproductive strategy in Portuguese waters of a commercial protogynous species, the black seabream *Spondyliosoma cantharus* (Sparidae). *Fish. Res.* **206**, 85–95 (2018).
64. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192, <https://doi.org/10.1093/molbev/msi103> (2005).
65. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
66. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
67. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

## Acknowledgements

Authors are very grateful to the people who helped in the sample collection from the several studied areas, particularly to Dr. M. Angeles Blanco (IEO, Spain), Dr. Eric Durieux (CNRS, France), Dr. Jim Ellis (CEFAS, United Kingdom), Dr. Virgílio Estêvão (INIP, Angola), Dr. Angel Fernandez (IEO, Spain), Dr. Rui Freitas (FECM/UniCV, Cape Verde), Dr. José Antonio González (ULPGC, Spain), Dr. Carme Alomar Mascaró (IEO, Spain), Dr. Sanja Matic-Skoko (IZOR, Croacia), Dr. Kelle Moreau (ILVO, Belgium), Dr. Michele Salaun (Ifremer, France), Dr. Carmen van-Dunem Santos (UAN, Angola) and Dr. Antonio Solla (IEO, Spain). We are also grateful to Dr. Frederico Almada, Dr. Sara M. Francisco and Dr. Sofia Seabra for their help in sample analyses. This study was partially supported by Fundação para a Ciência e a Tecnologia (FCT), through the strategic project UID/MAR/04292/2019 granted to MARE, the grant attributed to Ana Neves (SFRH/BD/92769/2013) and the research contracts attributed to Ana Rita Vieira (CEECIND/01528/2017) and Vera Sequeira (CEECIND/02705/2017).

## Author contributions

A.N., L.S.G. and O.S.P. developed the concept of the study. A.N., V.S., A.R.V., R.B.P. and L.S.G. coordinated the samples collecting. A.N. A.R.V. and R.B.P. designed and carried out molecular analyses. A.N. analysed the data with supervision of L.S.G. and O.S.P. A.N. wrote the paper and V.S., A.R.V., R.B.P., L.S.G. and O.S.P. revised and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-61050-x>.

**Correspondence** and requests for materials should be addressed to A.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020