Isolation-by-distance and asymmetrical dispersal of an intertidal blenniid across the Atlantic-Mediterranean divide

Rita Castilho Corresp., [1] , **Regina L Cunha** [1] , **Claúdia Faria** [2] , **Eva Velasco** [3] , **Joana I Robalo** [4]

[1] CCMAR-CIMAR Laboratório Associado, Centre for Marine Sciences, Campus de Gambelas, 8005-139 Faro, Portugal

[2] Instituto de Educação, Universidade de Lisboa, Alameda de Universidade, 1649-013, Lisboa, Portugal

[3] Centro Oceanográfico de Gijón, Av. Príncipe de Asturias 70, 33212, Instituto Español de Oceanografía, Gijón, Spain

[4] MARE – Marine and Environmental Sciences Centre, ISPA Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Rua Jardim do Tabaco 34, 1149-041, Lisboa, Portugal

Corresponding Author: Rita Castilho
Email address: rita.castil@gmail.com

Transition zones are of high evolutionary interest because they often harbor unique genetic diversity levels. In the present study, we investigated the phylogeography of the peacock blenny, *Salaria pavo*, a small marine intertidal fish that inhabits rocky habitats of the Mediterranean and the adjacent Atlantic Ocean. We surveyed 170 individuals using mitochondrial and nuclear sequence data from eight locations. Four models of genetic structure were tested: panmixia, isolation-by-distance, secondary contact and phylogeographic break. Results support the isolation-by-distance model combined with asymmetric migration from the Mediterranean to the Atlantic. Additionally, this species displays an imprint of demographic expansion compatible with the last glacial maximum. The hypothesis of a refugium in the Mediterranean cannot be discarded, but the ancestral lineage most probably originated in the Atlantic, where most of the genetic diversity is present.

Isolation-by-distance and asymmetrical dispersal of an intertidal blenniid across the Atlantic-Mediterranean divide

Castilho, R. [a], Cunha, R.L. [a], Faria, C. [c], Velasco, E.[d], Robalo, J.I. [b]

[a] Centre for Marine Sciences, CCMAR-CIMAR Laboratório Associado, Campus de Gambelas, 8005-139 Faro, Portugal.

[b] Unidade de Investigação em Eco-Etologia, ISPA – Instituto Universitário. Rua Jardim do Tabaco 34, 1149-041 Lisboa, Portugal.

[c] Instituto de Educação da Universidade de Lisboa, Alameda de Universidade, 1649-013, Lisboa, Portugal

[d] Instituto Español de Oceanografía, Centro Oceanográfico de Gijón Av. Príncipe de Asturias 70, 33212 Gíjon, Spain.

* Corresponding author at: Centre for Marine Sciences, CCMAR-CIMAR Laboratório Associado, Campus de Gambelas, 8005-139 Faro, Portugal.

Fax: +351 289 800 069.

E-mail address: rcastil@ualg.pt (R. Castilho).

Abstract

Transition zones are of high evolutionary interest because they often harbor unique genetic diversity levels. In the present study, we investigated the phylogeography of the peacock blenny, *Salaria pavo*, a small marine intertidal fish that inhabits rocky habitats of the Mediterranean and the adjacent Atlantic Ocean. We surveyed 170 individuals using mitochondrial and nuclear sequence data from eight locations. Four models of genetic structure were tested: panmixia, isolation-by-distance, secondary contact and phylogeographic break. Results support the isolation-by-distance model combined with asymmetric migration from the Mediterranean to the Atlantic. Additionally, this species displays an imprint of demographic expansion compatible with the last glacial maximum. The hypothesis of a refugium in the Mediterranean cannot be discarded, but the ancestral lineage most probably originated in the Atlantic, where most of the genetic diversity is present.

Keywords: Mediterranean; Atlantic refugium; asymmetric migration; LGM.

INTRODUCTION

Many terrestrial and marine species have often experienced expanding and contracting range shifts over time (Herborg et al. 2007; Reece et al. 2010; Reuschel et al. 2010). These range shifts are generally promoted by geological or climate events that affect temperature and territorial connectivity between locations. The African and Iberian continental margins formed the Gibraltar arch 5.5 MYA (million years ago) producing a land bridge that interrupted the water flow between the Atlantic and Mediterranean adjacent basins. This event, known as the Messinian Salinity Crisis, turned the Mediterranean hypersaline and dried out large expanses of the basin (Duggen et al. 2003; Hsü et al. 1973; Krijgsman 2002). The Mediterranean Sea was then the ground of a drastic contraction-expansion of distributional range processes in marine organisms inhabiting those waters, firstly with the disappearance of the previously existing Tethyan fauna, and secondly by the Mediterranean invasion of Atlantic species through the Strait of Gibraltar when the land bridge receded. Furthermore, the Pleistocene glacial episodes and the consequent fluctuations of the sea level and surface temperature in the Mediterranean and adjacent Atlantic have further shaped the distribution of marine organisms impelling their genetic makeup (Patarnello et al. 2007).

The Mediterranean Sea and its contiguous Northeastern Atlantic Ocean staged several phylogeographic studies on marine fish exploring the relationships between populations inhabiting both regions across a well-defined oceanographic break, the Almeria-Oran Front (AOF). Some species such as *Dicentrarchus labrax*, *Diplodus puntazzo* and *Coryphoblennius galerita*, have shown strong genetic divergence between populations inhabiting both sides of the AOF (Bargelloni et al. 2005; Domingues et al. 2007; Lemaire et al. 2005) while others display evidence of strong genetic flow (e.g., Thalassoma pavo, Chromis chromis and Diplodus sargus

61 Bargelloni et al. 2005; Costagliola et al. 2004; Domingues et al. 2005). It has proven difficult to
62 assign these differences to a single environmental or biological parameter (e.g. Galarza et al.
63 2009b).

64 Species with their areas of distribution centered in the Northeastern Atlantic Ocean and
65 the adjacent Mediterranean Sea provide interesting opportunities to study the evolutionary
66 effects of geographic range shifts. Present in both areas, the peacock blenny, *Salaria pavo,*
67 occurs chiefly around the Western Mediterranean coasts and from the Bay of Biscay south to the
68 Canaries (Zander 1986), being less abundant in the Eastern Mediterranean. This species lives in
69 sheltered rocky habitats and coastal lagoons, in the intertidal, or in the first meters of the
70 subtidal. Contrary to other Blenniidae, the peacock blenny is able to colonize soft substrates
71 (mud and sandy bottoms) and isolated patches of underwater vegetation (Verdiell-Cubedo et al.
72 2006). *Salaria pavo* displays a high tolerance to salinity (from 2 to 65 ‰) and temperature 1 to
73 30 °C (Paris & Quignard 1971; Plaut 1999). Nevertheless, the breeding ecology of the species
74 varies enormously with the availability of breeding grounds (e.g. Almada et al. 1994, and
75 references therein). During the breeding season, males build and defend nests from conspecific
76 males or other intruders and care for the eggs (Gonçalves & Almada 1997). It is known that eggs
77 of *S. pavo* are unable to hatch at temperatures below 15ºC (Westernhagen 1983) and typically the
78 species breed at temperatures above 18ºC, therefore at the LGM, suitable temperatures for the
79 reproduction of *S. pavo* were likely absent in the Bay of Biscay, western Galicia and northern
80 Portugal.

81 The goal of this study was to unveil genetic imprints of the peacock blenny using the
82 mitochondrial D-loop and the first intron of the S7 ribosomal protein nuclear gene sequences.
83 More specifically, we assessed the genetic diversity and population structure of this species over

its sampled distribution range and evaluated connectivity among populations. We evaluated the following biogeographic hypotheses concerning the current distribution of genetic diversity of *S. pavo*: (1) panmixia, whereby there is no discernible geographic or otherwise genetic structure corresponding effectively to a random distribution of haplotypes (Hypothesis 1: Figure 1a); (2) isolation-by-distance (IBD) pattern by which genetic and geographic distances are positively correlated (Wright 1943), and therefore alleles will show a frequency cline pattern between the Atlantic and the Mediterranean (Hypothesis 2: Figure 1b); (3) secondary contact between populations of the two regions, where alleles will transiently show a cline pattern at the contact zone between the two areas (Hypothesis 3: Figure 1c), and (4) genetic phylogeographic break between adjacent regions, wherein a sharp change of allele frequencies is observed between the Atlantic and the Mediterranean (Hypothesis 4: Figure 1d).

MATERIAL AND METHODS

Sampling and generation of molecular data

Samples of *S. pavo* were collected at 8 localities in the Northwestern Mediterranean and Atlantic coast of the Iberian Peninsula (Table 1; Figure 2). No field permits were required as this species is listed as "least concern conservation status" and it was not captured in protected areas. Fish were caught by scuba diving and small fishnets on rocky beaches and fin clips were stored individually in 96% ethanol. Total genomic DNA was extracted from fin or muscle samples preserved in 96% ethanol with the REDExtract-N-Amp kit (Sigma-Aldrich) following the manufacturer's instructions. Voucher specimens are deposited in ISPA (ethanol preserved tissues). We selected two unlinked genes of different genomes to be sequenced: the mitochondrial D-loop and the nuclear S7 ribosomal protein gene (S7, 520 bp, including the first

intron, as described by Chow & Hazama (1998). Nuclear and mitochondrial sequences were obtained from the same individuals. PCR amplification of mitochondrial D-loop and of the S7, were performed with the following pairs of primers: D-loop — LPro1 (5'- ACTCT CACCC CTAGC TCCCA AAG - 3') and HDL1 (5'- CCTGA AGTAG GAACC AGATG CCAG - 3') (Ostellari et al. 1996) and S7— S7RPEX1F (5'- TGG CCT CTT CCT TGG CCG TC - 3') and S7RPEX2R (5'- AAC TCG TCT GGC TTT TCG CC - 3') (Chow & Hazama 1998). PCR amplification reactions were performed in a 20 µl total-reaction volume with 10 µl of REDExtract-N-ampl PCR reaction mix (Sigma–Aldrich), 0.8 µl of each primer (10 µM), 4.4 µl of Sigma-water and 4 µl of template DNA. An initial denaturation at 94ºC for 7 min was followed by 35/30 cycles (denaturation at 94ºC for 30/45s, annealing at 55ºC for 30/45 s, and extension at 72ºC for 1 minute) and a final extension at 72ºC for 7 minutes on a BioRad Mycycler thermal cycler (values D-loop/S7, respectively). The same primers were used for the sequencing reaction, and the PCR products were purified and sequenced in STABVIDA (http://www.stabvida.net/). Sequences for each locus were aligned, edited, and trimmed to a common length using the DNA sequence assembly and analysis software GENEIOUS PRO 7.0 (Biomatters, LTD, Auckland, NZ).

Genetic diversity and population differentiation

The gametic phase of multi-locus genotypes of the nuclear S7 intron was determined using the pseudo-Bayesian approach of Excoffier–Laval–Balding (ELB) algorithm (Excoffier et al. 2003), as implemented in ARLEQUIN 3.5 (Excoffier & Lischer 2010). Gene diversity for both D-loop and S7 fragments was described as haplotype (h) and nucleotide (π) diversities (Nei 1987) were calculated using ARLEQUIN 3.5 (Excoffier & Lischer 2010) for locations with at least five

130 individuals. In order to compare haplotype diversity values, the statistics and asymptotic
131 confidence intervals derived by Salicru et al. (1993) were used for both overall diversity
132 comparison and pairwise comparisons between locations. A median-joining network (Bandelt et
133 al. 1999) was constructed in NETWORK v4.5 (fluxus-engineering.com) to determine the
134 genealogical relationships among haplotypes and to consider their geographical distributions.
135 POWSIM 4.1(Ryman & Palm 2006) was used to assess the power of the data and the suitability
136 of sample sizes to detect significant pairwise fixation at a F_{ST} value of 0.05. Genetic fixation Φ_{ST}
137 (Weir & Cockerham 1984) and differentiation Jost's *D* (Jost 2008) statistics were estimated with
138 diveRsity package 1.9.5 (Keenan et al. 2013) and significance of differentiation was assessed
139 through the calculation of 95% confidence limits using a bias corrected method with 10^4
140 bootstraps.
141 Mobile species subjected to genetic statistical differentiation tests often fail to display
142 minor amounts of population subdivision even if they exist (Palumbi & Warner 2003).
143 Therefore, we used SAShA (Spatial Analysis of Shared Alleles) (Kelly et al. 2010) implemented
144 in the MATLAB environment (Mathworks, Inc.) to test hypothesis 1, i.e., determine the extent to
145 which haplotypes are distributed randomly through space. Non-random distributions of
146 haplotypes can be considered departures from panmixia, and occurrences of the same haplotype
147 in different locations can be considered evidence of gene flow. SAShA generates the observed
148 distribution of geographic distances of each haplotype, as well as a null distribution generated
149 from the same data. SAShA tests for a significant deviation between the arithmetic mean of the
150 observed distance distribution (ODD) and that of the expected distance distribution (EDD). An
151 ODD significantly less than EDD indicates that alleles are under-distributed, and therefore gene

152 flow is restricted. We tested for significance of the difference between ODD and EDD using 10^4
153 permutations.

154 To test hypothesis 2, whether the geographical pattern of genetic differentiation is caused
155 by isolation by distance (IBD) we ran Mantel tests (Mantel 1967) for pairwise matrices between
156 geographical distances (kilometres) of the shortest marine path among locations and genetic
157 differentiation (measured as $\Phi_{ST}/(1-\Phi_{ST})$ and (D/(1- D). Mantel tests (1000 randomizations) were
158 performed using mantel.xla 1.2.4 (Briers 2003).

159

160 **Estimation of gene flow**

161 *S. pavo* adults are not known to undertake active migrations, therefore, instead of referring to
162 migration rates (M), we will refer instead to gene flow (G). G and population size parameter (θ)
163 were inferred using the maximum likelihood (ML) in MIGRATE-N ver. 4.2.6 (Beerli &
164 Felsenstein 1999) among Atlantic and Mediterranean locations in order to determine the degree
165 and direction of migrants across the Atlantic-Mediterranean region. Analyses were first run with
166 a full migration matrix in which gene flow was unrestricted between Atlantic and Mediterranean
167 (asymmetric migration, 4 parameters). To explicitly test other models (including panmixia;
168 immigration into Mediterranean; immigration into Atlantic) we built custom matrices
169 representing gene flow conditions. All G and θ were calculated using F_{ST} estimates and UPGMA
170 as starting points, and taking into account the model of evolution. A Markov Chain Monte Carlo
171 was run for three short chains of 10^4 trees and two long chains of 10^5 trees with a burn-in of 10^3
172 trees and a static heating scheme with start temperatures of 1.00, 1.50, 3.00 and 6.00. Finally,
173 likelihood scores for all migration models were obtained by a thermodynamic integration with
174 Bezier approximation (Gelman & Meng 1998), as implemented in the software. Direct

175 comparison of models was assessed by manually transforming these likelihood scores into Bayes
176 Factors (Kass & Raftery 1995), which was performed using the method described in Beerli and
177 Palczewski (Beerli & Palczewski 2010). MIGRATE-N was run on CCMAR Computational Cluster
178 Facility (http://gyra.ualg.pt/) and on the R2C2 research group cluster facility, provided by the IT
179 department of the University of Algarve.
180
181 **Population demography**
182 Past population demography of *S. pavo* was inferred with the D-loop data using the coalescent
183 Bayesian skyline plot (BSP) model as implemented in BEAST (Ho et al. 2005) employing the
184 Bayesian MCMC coalescent method, a strict clock and the HKY+I+G model of substitution
185 obtained in Modeltest v. 3.7 (Posada & Crandall 1998), using the Akaike information criterion
186 (AIC) (Akaike 1974). Results were visualized in TRACER (Rambaut & Drummond 2007) The
187 Bayesian distribution was generated using results from two independent run of 100 million
188 MCMC steps obtaining effective samples sizes (ESS) of parameter estimates of over 200. We
189 used a mutation rate of 3.6% per million years calculated in previous studies where geological
190 events were available to calibrate the rate of D-loop divergence in marine fish (Donaldson &
191 Wilson 1999) and in the absence of a clock calibration for the D-loop of *S. pavo* we address the
192 rate uncertainty by assuming a higher within-lineage mutation rates of 5% per million years.
193
194 **RESULTS**
195 **D-loop**
196 A total of 131 D-loop sequences (GenBank accession numbers: HQ857214-HQ857383) were
197 obtained. The D-loop data set after alignment consisted of a total of 300 characters comprising

198 52 polymorphic sites (17%) and 10 (3%) parsimony informative sites. Overall, mtDNA diversity
199 was high, with 49 haplotypes recovered. A large proportion of haplotypes (57%) were singletons,
200 i.e., represented by a single individual. Forty haplotypes (82%) were private, i.e. occurred in only
201 one location. On total, 92% of haplotypes had a frequency lower than five individuals. The most
202 frequent haplotype in the Atlantic was shared by 25 individuals, followed by two haplotypes
203 shared by 13 individuals, one present only in the Mediterranean and the other in both
204 Mediterranean and Atlantic (Fig. 2). Regarding the eleven haplotypes shared among locations
205 (Fig. 2), six include individuals from both Atlantic and Mediterranean sampling sites. The
206 presence of many low-frequency closely related haplotypes returns high haplotype diversity
207 (0.952 ± 0.0078) and low nucleotide diversity (3.63% ± 1.84%) of the overall sample, as well as
208 in each locality (Table 1). Haplotype diversity values were not significantly different between
209 locations, according to the test developed by Salicru et al. (1993) ($\chi 2 = 7.15$, $p < 0.05$).
210 The *S. pavo* haplotype network (Fig. 3) has an overall complex pattern of star-like
211 elements, networks that are shallow and dominated by few haplotypes, where rare haplotypes
212 differ from the most common haplotypes by only a few mutations. No evident geographic
213 structure could be depicted from this network, i.e., no discernable association between certain
214 haplotypes and locations can be observed. Although the most frequent haplotype was only
215 present in Portugal and Cadiz, the remaining haplotypes from these localities group together with
216 haplotypes from the Mediterranean. The difference between the overall observed distance
217 distribution (ODD) and the expected distance distribution (EDD) of shared alleles rejected the
218 assumption of panmixia (hypothesis 1) for the D-loop dataset (ODD = 237 km, EDD = 516 km,
219 $p < 0.00001$) (Fig. 4). POWSIM indicated that a F_{ST} of ≥ 0.0248 (time in generations = 150)
220 could be detected with ≥ 95% confidence (95.5% Fisher's exact test, 96.2% chi-square). When

221 F_{ST} was set to zero (simulating no divergence among samples), the proportion of α error of type I
222 (rejecting null hypothesis when true) was lower than 5%.

223 Overall mean pairwise genetic differentiation between the main geographical groups (intra-
224 Mediterranean, between the Mediterranean and Atlantic and intra-Atlantic) showed a tendency
225 for higher Atlantic-Mediterranean values (Fig. 5A). Pairwise location genetic differentiation
226 revealed no association between the levels of differentiation and the three geographical groups
227 considered (Fig. 5B). There is no clear indication of a genetic break (hypothesis 4) between the
228 Mediterranean and Atlantic Ocean as pairwise differentiation values were all within the same
229 range. Isolation by distance model (hypothesis 2) support was equivocal, the null hypothesis of
230 no correlation between geographic and genetic distances was not rejected using Φ_{ST} (r = 0.001; z
231 = 0.503) but was rejected using *D* (r = 0.433; z = 0.008). No haplotype frequency cline
232 (hypothesis 3) could be detected as there were only three haplotypes shared between more than 2
233 locations.

234 Migrate-n was run to determine the level and direction of gene flow across the Almeria-
235 Oran oceanographic boundary. The estimated log Bayes factors based on the Bezier
236 approximation score indicated that the most probable model is the one that contemplates
237 asymmetric migration between the Atlantic and the Mediterranean (Table 2). The number of
238 migrants from the Mediterranean to the Atlantic was ca. three times the number of migrants in
239 the inverse direction.

240 The Bayesian skyline plot indicated that the Western Mediterranean and Atlantic locations
241 of *S. pavo* have experienced a long period of demographic stability in the past, followed by a
242 mild decrease of population size and a quick expansion (Fig. 6). The plot indicates a pronounced
243 ca.100-fold demographic expansion event. The timeframe of this expansion event is totally

dependent on the mutation rate used, and therefore, using 1% dates the event might have started 40,000 years ago, and using 5% dates at 200,000.

First S7 intron

A total of 136 S7 first intron sequences (GenBank accession numbers: JF834709-JF834885) were obtained. The S7 nuclear region data set after alignment consisted of 519 characters, with seven polymorphic sites, among which five with ambiguities. Using the ELB algorithm, we defined 12 closely related alleles, with four abundant and almost ubiquitous alleles, and the remainder represented by only one or two individuals. Overall gene and nucleotide diversities were low, 0.69 ± 0.02 and 0.18 ± 0.14, respectively. The haplotype network (Figure 2) does evidences a lack of geographical structure. The difference between the overall observed distance distribution (ODD) and the expected distance distribution (EDD) of shared alleles does not reject the assumption of panmixia for the S7 dataset (ODD = 522 km, EDD = 509 km, $p < 0.62$).

DISCUSSION

In this study we evaluated four plausible phylogeographic scenarios explaining the genetic differentiation between Mediterranean and Atlantic samples of *Salaria pavo* (Figure 1). The nuclear marker, with only 12 haplotypes displayed comparatively low genetic diversity, probably due to low mutation rates (Harpending 1994). From the haplotype network one can also clearly infer that there is no geographical structure, and the panmixia model is not rejected. We will therefore discuss in more detail the mtDNA results. Pure models of panmixia, secondary contact, and presence of a phylogeographic break do not seem to explain the results obtained, while isolation-by-distance with asymmetric migration between the Atlantic and Mediterranean are more plausible explanations. Before dissecting these results, it is appropriate to address two main

caveats regarding this work. Firstly, we are contrasting our results with well-defined hypotheses that constitute extremes of often less clear biological realities. Isolation-by-distance and clines are not mutually exclusive genetic patterns, as illustrated by ring species (Irwin et al. 2001), in which a series of intermediate subpopulations display a contact zone and are often connected by a cline at the closure of the ring (Bensch et al. 2009). However, our results seem to reject panmixia, secondary contact, and phylogeographic break models, and there is no evidence suggesting that a combination of these would be a better fit. Secondly, the nuclear data display limited variability and no phylogeographic patterns could be identified, showing that S7 was not a good candidate gene for this particular species, although it has been successfully used in other marine fish (Ahti et al. 2016).

Model evaluation

Panmixia (hypothesis 1) was concomitantly rejected by the haplotype network (Figure 3), the spatial analysis of shared alleles (Figure 4) and the Migrate-n results (Table 2). Moreover, the presence of private haplotypes detected in both Mediterranean and Atlantic locations and the fact that some of these were found multiple times on a single location suggests some limitations to gene flow (Hartl & Clark 1997). Results regarding the classical isolation-by-distance regression model (hypothesis 2) were somewhat equivocal: rejection of the model based on Φ_{ST}, and non-rejection based on Jost's *D*. Because Jost's *D* is independent from gene diversity, and it was shown to perform well in evaluating genetic differentiation regardless of haplotype diversity and genetic distance between populations (Bird et al. 2011), we do not entirely reject the isolation-by-distance model. We found no support for hypothesis 3 (secondary contact) as most haplotypes are singletons or are shared between two locations, and no haplotype frequency cline could be

290 detected. There is no evidence for a specific association between haplotype presence and
291 locations, such as detected under a phylogeographic barrier (hypothesis 4).

292 **The Atlantic-Mediterranean continuum and ancestral areas of refugia**

293 Some fish species display a strong genetic discontinuity between each side of the Almeria-Oran
294 oceanographic front, but this pattern is species-dependent. In the same family (e.g. Sparidae
295 Bargelloni et al. 2003; Galarza et al. 2009a) and even within the same genus (e.g. Diplodus: D.
296 puntazzo and D. sargus, Bargelloni et al. 2005), there are species with strong gene flow across
297 the boundary, while others have restricted gene flow. *Salaria pavo* displays no significant
298 differentiation across the Atlanto-Mediterranean boundary and this permeability contrasts with
299 the strong across-boundary differentiation displayed by another intertidal blenniid
300 *Coryphoblennius galerita* (Francisco et al. 2014). The strong thermohaline density gradient
301 nature of the Almeria-Oran oceanographic front, is apparently not sufficient to restrict the
302 mobility of *S. pavo* across the boundary. On the other hand, the paleotemperatures estimated for
303 the summer breeding season during the last glacial maximum were at most 13ºC in Iberian
304 Atlantic and most of west Mediterranean (CLIMAP 1981) are not compatible with the high
305 thermal preferences of *S. pavo.* This species' embryos kept in laboratory arrest their development
306 at temperatures of 15°C or lower (Westernhagen 1983). Considering these conditions, *S. pavo*
307 was at the LGM most likely extirpated from its northern limit, the Bay of Biscay, as well as from
308 North and Central Portugal. These locations represent postglacial colonizations derived from
309 potential refugia located in the Mediterranean or further south in the Atlantic. The most northern
310 location with a representative number of individuals (Sado) displays a significantly lower
311 haplotype diversity values than those found in other Atlantic (Cadiz, Ria Formosa and Olhos

312 d'Água) and Mediterranean locations (Barcelona), which is concordant with a postglacial
313 colonization event.

314 The most probable refugium can be inferred by coalescent theory in which ancestral
315 mitochondrial haplotypes are likely to have given rise to more derived ones because mutation has
316 occurred over a longer period of time (Posada & Crandall 2001). As a consequence, older
317 haplotypes tend to have more connections in a network. Although homoplasy and high mutation
318 rates can bias this pattern, highly connected haplotypes tend to be closely related to ancestral
319 haplotypes (Posada & Crandall 2001). Thus, the presence of highly connected haplotypes in the
320 Atlantic could indicate the Atlantic as the likely major source of *S. pavo* post-glacial
321 recolonization (Fig. 2). However, if gene flow persisted between Atlantic and west
322 Mediterranean during the LGM, both areas may have operated as a vast refugium for the species.
323 The hypothesis of a single refugium located inside the Mediterranean seems the least probable.
324 The asymmetric gene flow rates indicating a large source of migrants from the Mediterranean
325 into the Atlantic provide additional evidence supporting this assumption. However, the
326 asymmetric gene flow detected in the peacock blenny is counterintuitive to expectations based
327 on the prevalent out-of-Mediterranean surface currents (Naranjo et al. 2015). We posit that the
328 unidirectional dispersal direction, also observed in other species (Alberto et al. 2008; Xavier et
329 al. 2011) is disproportionally affected by sporadic storms that alter near-shore counter-currents
330 (Relvas & Barton, 2002) and surface wind patterns rather than yearly or decadal averages of
331 oceanographic conditions.

332

333 **Other factors contributing to the present pattern**

334 Two apparently contradictory results should be noted: (1) strong differentiation between Sado
335 and Algarve (Ria Formosa and Olhos d'Água) located only a few hundred kilometers apart; (2)
336 weak differentiation between Formentera and Sado, over 1000 km apart. *Salaria pavo* differs
337 from other blennids by living preferentially in sheltered rocky habitats, estuaries and lagoons
338 (Zander 1986). Although the pelagic larval duration is of ca. 18 days at a temperature of 21ºC
339 (Westernhagen 1983), it seems likely that larvae of this species can be subject to more efficient
340 retention than those of other blennids of more exposed shores. *Salaria pavo*'s differentiation
341 pattern is consistent with a combination of considerable individual retention with sporadic
342 episodes of range dispersal, which would reconcile the observation of high θ_{ST} values between
343 locations separated by hundreds of kilometers with substantial sharing of haplotypes between
344 other locations.
345 Previously published work hypothesized that the reduced genetic variation detected in *S.*
346 *pavo* could have been the result of a severe bottleneck event (Almada et al. 2009). However, we
347 have found a high number of widely distributed haplotypes, coupled with generally non-
348 significant ΦST values and an expansion signature, findings that do not support the hypothesis of
349 a severe bottleneck. The Bayesian skyline plot (BSP) analysis was used to date shifts in
350 population size of the *S. pavo* lineage A (Fig. 6). Results suggest a recent and rapid 100-fold
351 increase in population size, preceded by a minor decrease that followed an extended period of
352 stability. The lack of a species-specific clock and associated error requires cautious interpretation
353 of age estimates, but the assumed rates of 3.6% and 5% /MY place the expansion unequivocally
354 during the Pleistocene.
355

In summary, we propose that the genetic pattern of *S. pavo* in the Atlanto-Mediterranean region is better explained by a combination of isolation-by-distance and asymmetric migration. The ancestral lineage most probably originated in the Atlantic, where most of the genetic diversity is present. Both dispersal potential and physical factors such as local oceanographic conditions are playing a major role in shaping the genetic structure of this species.

Acknowledgements

We are grateful to Sonia Chenu for her help with the lab work and to Marta Pascual and Ferran Palero, from the University of Barcelona, for their assistance in implementing the Salicrú test of haplotype diversity. This study was funded by the MarinERA project "Marine phylogeographic structuring during climate change: the signature of leading and rear edge of range shifting populations"; by the Eco-Ethology Research Unit' Strategic Plan (PEst-OE/MAR/UI0331/2011), now included in MARE (UID/MAR/04292/2013), and by CCMAR Strategic Plan (PEst-C/MAR/LA0015/2011 and UID/Multi/04326/2013) from Fundação para a Ciência e a Tecnologia - FCT (partially FEDER funded). Collection of specimens complied with the current laws of each country. We dedicate this paper to the loving memory of Prof. Vítor Almada who died during the course of this work.

References

Ahti PA, Coleman RR, DiBattista JD, Berumen ML, Rocha LA, and Bowen BW. 2016. Phylogeography of Indo-Pacific reef fishes: sister wrasses *Coris gaimard* and *C. cuvieri* in the Red Sea, Indian Ocean and Pacific Ocean. *Journal of Biogeography*.

Akaike H. 1974. A new look at the statistical model identifications. *IEEE Transactions on automatic control* 19:716–723.

Alberto F, Massa S, Manent P, Diaz-Almela E, Arnaud-Haond S, Duarte CM, and Serrao EA. 2008. Genetic differentiation and secondary contact zone in the seagrass *Cymodocea nodosa* across the Mediterranean–Atlantic transition region. *Journal of Biogeography* 35:1279-1294.

Almada VC, Gonçalves EJ, Santos AJ, and Baptista MC. 1994. Breeding ecology and nest aggregations in a population of *Salaria pavo* (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology* 45:819–830.

Almada VC, Robalo J, Levy A, Freyhoff J, Bernardi G, and Doadrio I. 2009. Phylogenetic analysis of Peri-Mediterranean blennies of the genus *Salaria*: Molecular insights on the colonization of freshwaters. *Molecular Phylogenetics and Evolution* 52:424–431. 10.1016/j.ympev.2009.03.029

Bandelt H-J, Forster P, and Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16:37–48.

Bargelloni L, Alarcon JA, Alvarez MC, Penzo E, Magoulas A, Palma J, and Patarnello T. 2005. The Atlantic-Mediterranean transition: Discordant genetic patterns in two seabream species, *Diplodus puntazzo* (Cetti) and *Diplodus sargus* (L.). *Molecular Phylogenetics and Evolution* 36:523–535. 10.1016/j.ympev.2005.04.017

Bargelloni L, Alarcon JA, Alvarez MC, Penzo E, Magoulas A, Reis C, and Patarnello T. 2003. Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic-Mediterranean divide. *Journal of Evolutionary Biology* 16:1149–1158. 10.1046/j.1420-9101.2003.00620.x

401 Beerli P, and Felsenstein J. 1999. Maximum likelihood estimation of migration rates and
402 effective population numbers in two populations using a coalescent approach. *Genetics*
403 152:763–773.
404 Beerli P, and Palczewski M. 2010. Unified framework to evaluate panmixia and migration
405 direction among multiple sampling locations. *Genetics* 185:313-326.
406 10.1534/genetics.109.112532
407 Bensch S, Grahn M, Müller N, Gay L, and Åkesson S. 2009. Genetic, morphological, and feather
408 isotope variation of migratory willow warblers show gradual divergence in a ring.
409 *Molecular Ecology* 18:3087-3096.
410 Bird CE, Karl SA, Smouse PE, and Toonen RJ. 2011. Detecting and measuring genetic
411 differentiation. *Phylogeography and population genetics in Crustacea* 19:31-55.
412 Briers R. 2003. Mantel XLA. VBA add-in for Microsoft EXCEL. version 1.2.4 ed.
413 Chow S, and Hazama K. 1998. Universal PCR primers for S7 ribosomal protein gene introns in
414 fish. *Molecular Ecology* 7:1255–1256.
415 CLIMAP. 1981. Seasonal reconstruction of the Earth's surface at the last glacial maximum.
416 Geological Society of America Map Chart Series. Boulder, CO.: Geological Society of
417 America.
418 Costagliola D, Robertson DR, Guidetti P, Stefanni S, Wirtz P, Heiser J, and Bernardi G. 2004.
419 Evolution of coral reef fish *Thalassoma* spp. (Labridae). 2. Evolution of the eastern
420 Atlantic species. *Marine Biology* 144:377–383. 10.1007/s00227-003-1200-y
421 Domingues V, Bucciarelli G, Almada VC, and Bernardi G. 2005. Historical colonization and
422 demography of the Mediterranean damselfish, *Chromis chromis*. *Molecular Ecology*
423 14:4051–4063. 10.1111/j.1365-294X.2005.02723.x

424 Domingues VS, Faria C, Stefanni S, Santos R, Brito A, and Almada VC. 2007. Genetic
425 divergence in the Atlantic-Mediterranean Montagu's blenny, *Coryphoblennius galerita*
426 (Linnaeus 1758) revealed by molecular and morphological characters. *Molecular Ecology*
427 16:3592–3605. 10.1111/j.1365-294X.2007.03405.x
428 Donaldson KA, and Wilson RR. 1999. Amphi-Panamic geminates of Snook (Percoidei:
429 Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA
430 control region of fishes. *Molecular Phylogenetics and Evolution* 13:208–213.
431 10.1006/mpev.1999.0625
432 Duggen S, Hoernle K, van den Bogaard P, Rupke L, and Morgan JP. 2003. Deep roots of the
433 Messinian salinity crisis. *Nature* 422:602–606. 10.1038/nature01553
434 Excoffier L, Laval G, and Balding D. 2003. Gametic phase estimation over large genomic
435 regions using an adaptive window approach. *Human Genomics* 1:7—19.
436 Excoffier L, and Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform
437 population genetics analyses under Linux and Windows. *Molecular Ecology Resources*
438 10:564—567. 10.1111/j.1755-0998.2010.02847.x
439 Francisco SM, Almada VC, Faria C, Velasco EM, and Robalo JI. 2014. Phylogeographic pattern
440 and glacial refugia of a rocky shore species with limited dispersal capability: the case of
441 Montagu's blenny (*Coryphoblennius galerita*, Blenniidae). *Marine Biology* 161:2509-
442 2520.
443 Galarza JA, Carreras-Carbonell J, Macpherson E, Pascual M, Roques S, Turner GF, and Rico C.
444 2009a. The influence of oceanographic fronts and early-life-history traits on connectivity
445 among littoral fish species. *Proceedings of the National Academy of Sciences of the*

United States of America of the United States of America 106:1473–1478. 10.1073/pnas.0806804106

Galarza JA, Turner GF, Macpherson E, and Rico C. 2009b. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). *Canadian Journal of Fisheries and Aquatic Sciences* 66:1478—1490. 10.1139/F09-098

Gelman A, and Meng X. 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical science* 13:163–185.

Gonçalves E, and Almada VC. 1997. Sex differences in resource utilization by the peacock blenny. *Journal of Fish Biology* 51:624-633.

Harpending H. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66:591–600.

Hartl DL, and Clark AG. 1997. *Principles of Population Genetics*. Massachusetts: Sinauer Associates.

Herborg L-M, Weetman D, VanOosterhout C, and Hänfling B. 2007. Genetic population structure and contemporary dispersal patterns of a recent European invader, the Chinese mitten crab, *Eriocheir sinensis*. *Molecular Ecology* 16:231–242. 10.1111/j.1365-294X.2006.03133.x

Ho SW, Phillips M, Cooper A, and Drummond A. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22:1561–1568. 10.1093/molbev/msi145

Hsü K, Ryan W, and Cita M. 1973. Late Miocene desiccation of the Mediterranean. *Nature* 242:240–244. 10.1038/242240a0

469 Irwin DE, Bensch S, and Price TD. 2001. Speciation in a ring. *Nature* 409:333–337.
470 Jost L. 2008. GST and its relatives do not measure differentiation. *Molecular Ecology* 17:4015-
471 4026. 10.1111/j.1365-294X.2008.03887.x
472 Kass RE, and Raftery AE. 1995. Bayes factors. *Journal of the american statistical association*
473 90:773-795. 10.1080/01621459.1995.10476572
474 Keenan K, McGinnity P, Cross TF, Crozier WW, and Prodöhl PA. 2013. diveRsity: An R
475 package for the estimation and exploration of population genetics parameters and their
476 associated errors. *Methods in Ecology and Evolution* 4:782–788. 10.1111/2041-
477 210X.12067
478 Kelly RP, Oliver TA, Sivasundar A, and Palumbi SR. 2010. A method for detecting population
479 genetic structure in diverse, high gene-flow species. *Journal of Heredity* 101:423–436.
480 10.1093/Jhered/Esq022
481 Krijgsman W. 2002. The Mediterranean: Mare Nostrum of Earth Sciences. *Earth Planet Sci Lett*
482 205:1–12. 10.1016/S0025-3227(98)00084-X
483 Lemaire C, Versini JJ, and Bonhomme F. 2005. Maintenance of genetic differentiation across a
484 transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal*
485 *of Evolutionary Biology* 18:70–80. 10.1111/j.1420-9101.2004.00828.x
486 Mantel N. 1967. The detection of disease clustering and a generalized regression approach.
487 *Cancer Reseach* 27:209–220.
488 Naranjo C, Sammartino S, García-Lafuente J, Bellanco MJ, and Taupier-Letage I. 2015.
489 Mediterranean waters along and across the Strait of Gibraltar, characterization and zonal
490 modification. *Deep-Sea Research Part I Oceanographic Research Papers* 105:41-52.

491 Nei M. 1987. Genetic distance and molecular phylogeny. In: Ryman N, and Utter FW, eds.
492 *Population Genetics & Fishery Management*. Seattle: Washington Sea Grant Program,
493 University of Washington, 193–223.
494 Ostellari L, Bargelloni L, Penzo E, Patarnello P, and Patarnello T. 1996. Optimization of single-
495 strand conformation polymorphism and sequence analysis of the mitochondrial control
496 region in *Pagellus bogaraveo* (Sparidae, Teleostei): rationalized tools in fish population
497 biology. *Animal Genetics* 27:423–427. 10.1002/humu.1380020513
498 Palumbi S, and Warner R. 2003. Why Gobies Are Like Hobbits? *science* 299:51.
499 10.1126/science.1080775
500 Paris J, and Quignard J. 1971. La faune ichthyologique des étangs languedociens de Sète a
501 Carnon (Écologie, Éthologie). *Vie et Milieu* 22:301-327.
502 Patarnello T, Volckaert FAMJ, and Castilho R. 2007. Pillars of Hercules: is the Atlantic-
503 Mediterranean transition a phylogeographical break? *Molecular Ecology* 16:4426—4444.
504 10.1111/j.1365-294X.2007.03477.x
505 Plaut I. 1999. Effects of salinity acclimation on oxygen consumption in the freshwater blenny,
506 *Salaria fluviatilis*, and the marine peacock blenny, *S. pavo*. *Marine and Freshwater*
507 *Research* 50:655-659.
508 Posada D, and Crandall KA. 1998. Modeltest: testing the model of DNA substituition.
509 *Bioinformatics* 14:817–818.
510 Posada D, and Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks.
511 *Trends in Ecology & Evolution* 16:37–45. 10.1016/S0169-5347(00)02026-7
512 Rambaut A, and Drummond A. 2007. Tracer v1.5. Available from
513 http://beast.bio.ed.ac.uk/Tracer.

514 Reece J, Bowen B, Joshi K, Goz V, and Larson A. 2010. Phylogeography of two moray eels
515 indicates high dispersal throughout the Indo-Pacific. *Journal of Heredity* 101:391–402.
516 10.1093/jhered/esq036
517 Reuschel S, Cuesta J, and Schubart C. 2010. Marine biogeographic boundaries and human
518 introduction along the European coast revealed by phylogeography of the prawn
519 *Palaemon elegans*. *Molecular Phylogenetics and Evolution*.
520 10.1016/j.ympev.2010.03.021
521 Ryman N, and Palm S. 2006. POWSIM: a computer program for assessing statistical power
522 when testing for genetic differentiation. *Molecular Ecology Notes* 6:600-602.
523 Salicru M, Menendez ML, Morales D, and Pardo L. 1993. Asymptotic distribution of (h, φ)-
524 entropies. *Communications in Statistics - Theory and Methods*:2015–2031.
525 Verdiell-Cubedo D, Oliva-Paterna F, and Torralva M. 2006. Condition of *Salaria pavo* in the
526 Mar Menor coastal lagoon (SE Iberian Peninsula): potential influence of environmental
527 variables on juveniles. *Journal of Applied Ichthyology* 22:407-413.
528 Weir BS, and Cockerham CC. 1984. Estimating F-statistics for the analysis of population
529 structure. *Evolution* 38:1358–1370. 10.2307/2408641
530 Westernhagen Hv. 1983. Observations on the reproductive and larval biology of *Blennius pavo*
531 (Pisces: Teleostei). *Helgoland Marine Research* 36:323–335.
532 Wright S. 1943. Isolation by distance. *Genetics* 28:114–138.
533 Xavier R, Zenboudji S, Lima FP, HARRIS DJ, Santos AM, and Branco M. 2011.
534 Phylogeography of the marine isopod*q* (Rezig, 1989) in North African Atlantic and
535 western Mediterranean coasts reveals complex differentiation patterns and a new species.
536 *Biological Journal of the Linnean Society* 104:419-431.

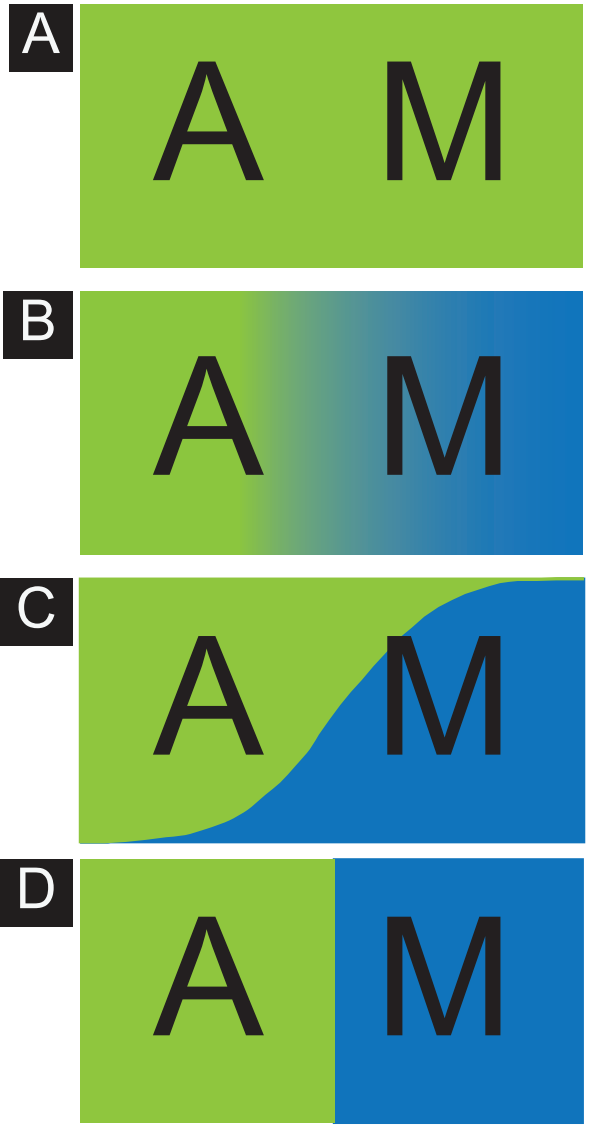
537 Zander CD. 1986. Blenniidae. In: Whitehead PJP, Bauchot M-L, Hureau J-C, Nielsen J, and
538 Tortonese E, eds. *Fishes of the North-eastern Atlantic and the Mediterranean*. Paris:
539 UNESCO, 1096–1112.

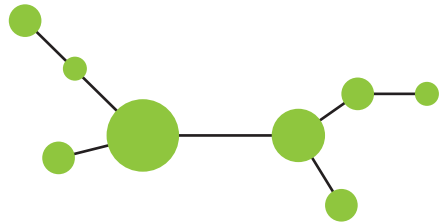
540

Figure 1(on next page)

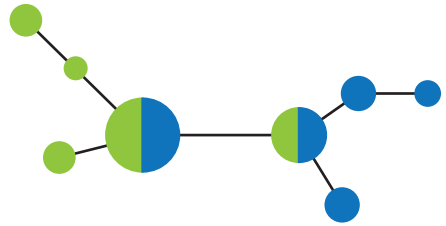
Schematics of four models.

Schematics of four models of haplotype frequency distribution and haplotype networks that are expected to result from the scenarios involving panmixia (A), isolation-by-distance (B), secondary contact (C) and phylogeographic barrier (D).

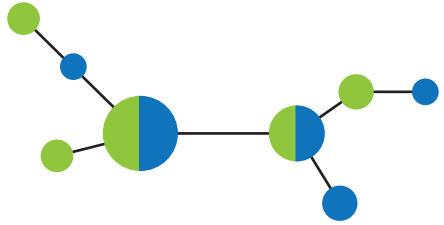




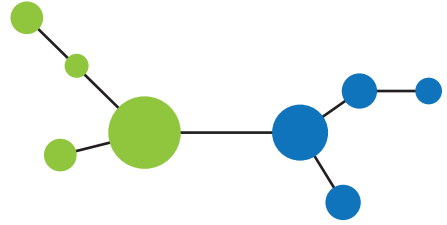
Hypothesis 1: Panmixia.
A panmictic population is one in which every individual has an equal chance of mating with another individual. There is no discernable population structure.



Hypothesis 2: Isolation by distance.
Under models of isolation by distance, many neutral alleles will show cline patterns, especially along geographic axes with the least gene flow.



Hypothesis 3: Secondary contact between populations of the two regions.
With secondary contact, neutral alleles will transiently show a cline pattern at the contact zone between the two populations.
The clines along the secondary contact zone will form even if the allele frequency difference between the two populations is modest.



Hypothesis 4: Genetic phylogeographic break between adjacent regions.
A sharp geographic boundary between two clades usually assumed to be a result of geographic barriers to dispersal, cryptic species boundaries, or recent contacts between historically allopatric populations.

Figure 2(on next page)

Distribution of D-loop haplotypes of *Salaria pavo* on each location.

Distribution of D-loop haplotypes of *Salaria pavo* on each location. Two-letter codes refer to the name of locations in Table 1. Numbers in parenthesis represent the sampling size. Large pies display the proportion of individuals that have unique haplotypes (dark blue) and the proportion of individuals that share haplotypes (light blue). The inset with large pie refers to the totality of samples pooled together. The inset with small circles refers to haplotypes numbers (see Annex 1). Small pies represent the frequency of haplotypes that are shared among individuals. Colours allow comparing the presence of common haplotypes that are present in locations. The biogeographical break of the Almeria-Oran front is represented in red (AOF).

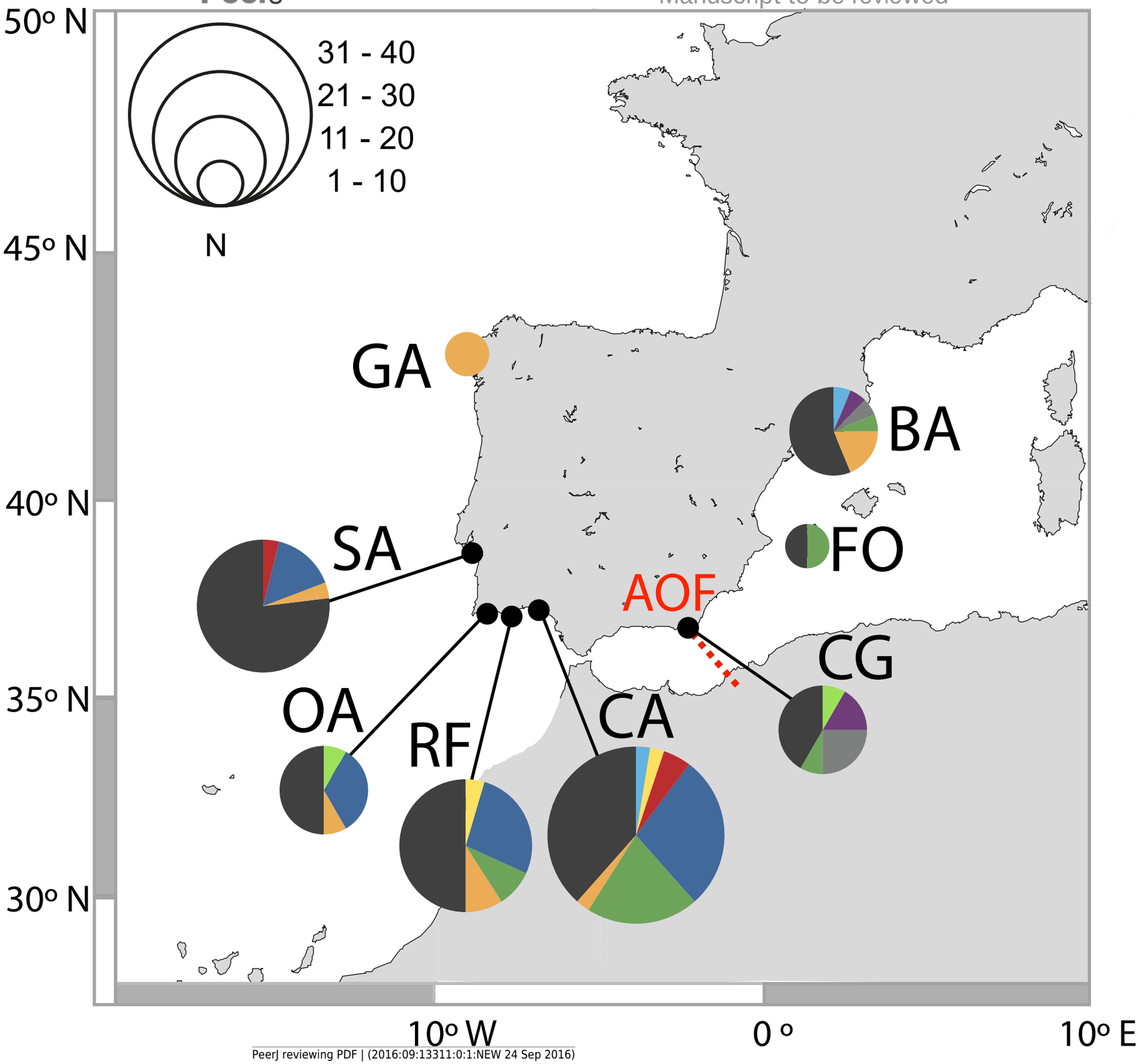

50º N
45º N
40º N
35º N
30º N
10º W
0 º
10º E
31 - 40
21 - 30
11 - 20
1 - 10
N
GA
BA
FO
SA
AOF
CG
OA
RF
CA

Figure 3(on next page)

Median-joining post-processed haplotype network for *Salaria pavo* .

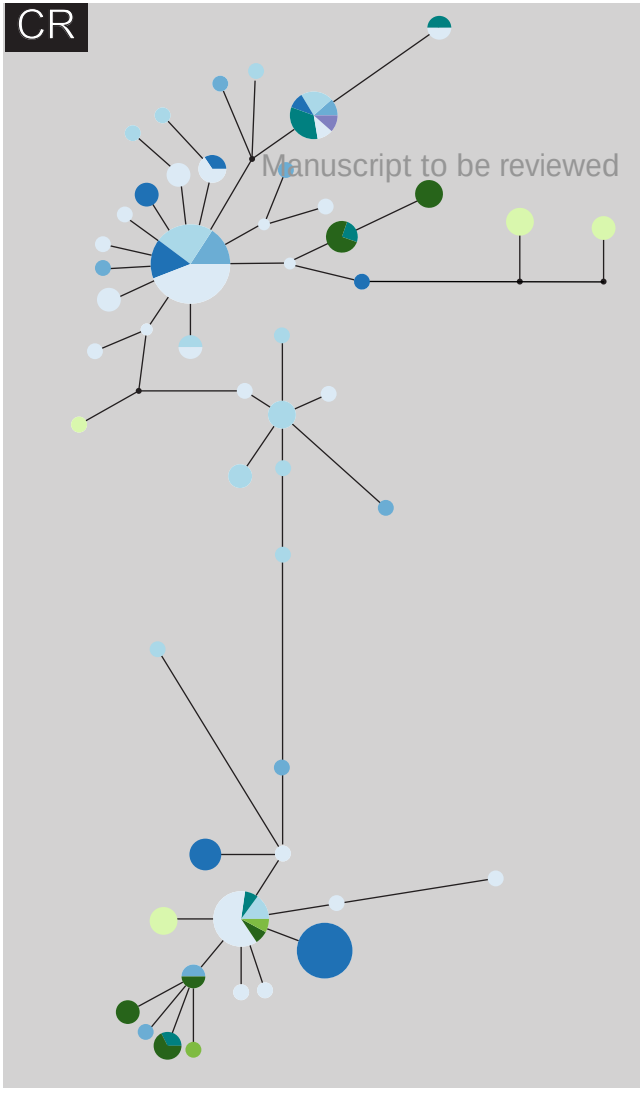
Median-joining post-processed haplotype network for *Salaria pavo*. The area of the circles is proportional to the frequency of individuals in the sample. Lines are proportional to mutations. Black dots represent median-vectors, or putative haplotypes not sampled or extinct. Colours represent collection location (see key).

Mediterranean

- Barcelona
- Formentera
- Cabo Gata

Atlantic

- Cadiz
- Ria Formosa
- Olhos d'Água
- Sado
- Galicia



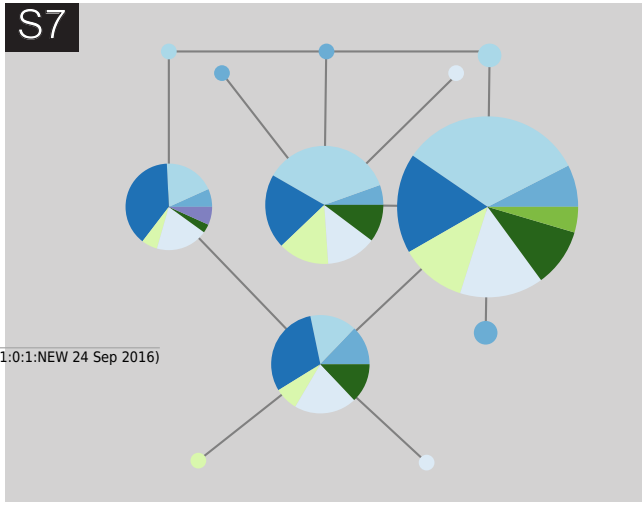
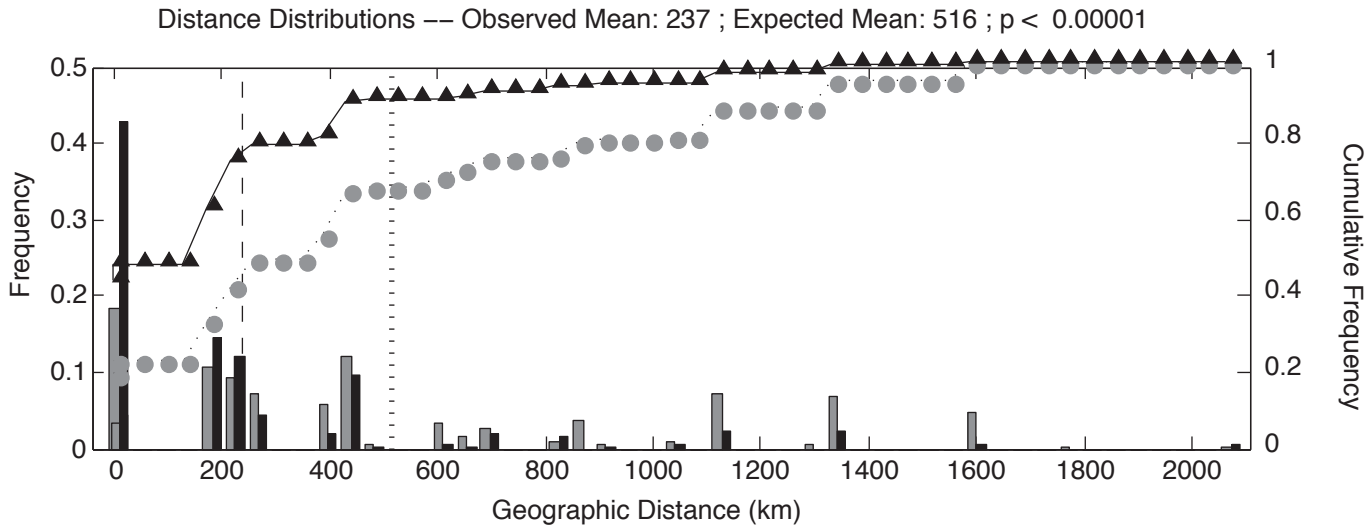


Figure 4(on next page)

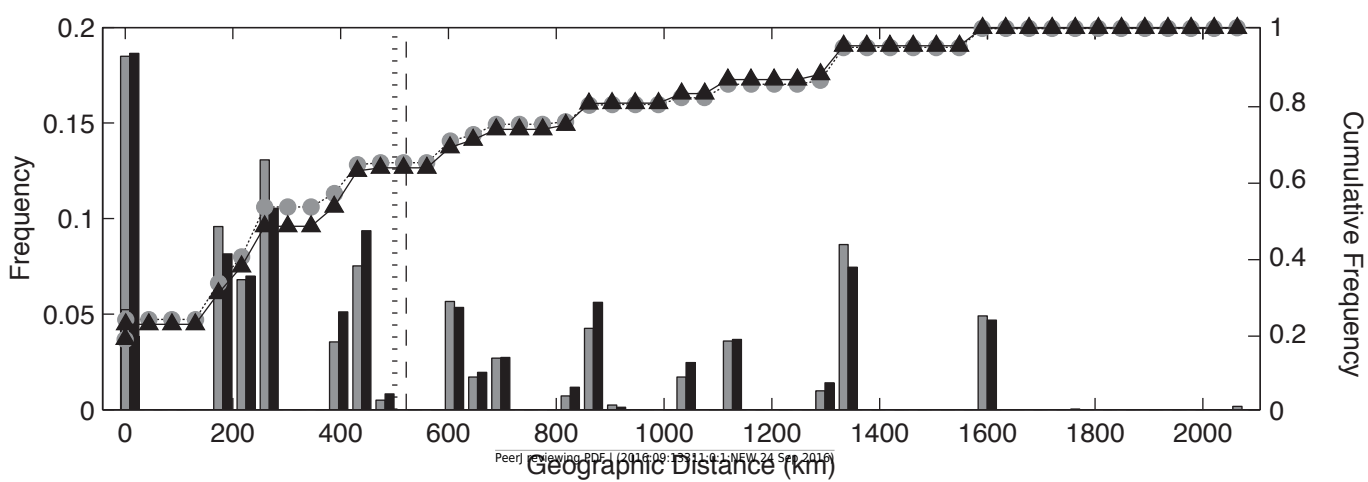
Spatial analysis of shared CR mtDNA distribution of *Salaria pavo* .

Spatial analysis of shared CR mtDNA distribution of *Salaria pavo*. Histograms represent the frequency of alleles between locations distance classes. Expected means and significance value were calculated with 1,000 randomized permutations of the data set. Vertical lines represent the mean of frequencies. Triangles and circles are the cumulative frequency of alleles at increasing distance. *p*- value is the probability that the observed mean is greater than the expected.

CR


Distance Distributions –– Observed Mean: 237 ; Expected Mean: 516 ; p < 0.00001
Frequency
Cumulative Frequency
Geographic Distance (km)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.8
1
200
400
600
800
1000
1200
1400
1600
1800
2000

S7


Distance Distributions –– Observed Mean: 521 ; Expected Mean: 508; p = 0.6161
Frequency
Cumulative Frequency
Geographic Distance (km)
0
0.05
0.1
0.15
0.2
0.4
0.6
0.8
1
200
400
600
800
1000
1200
1400
1600
1800
2000

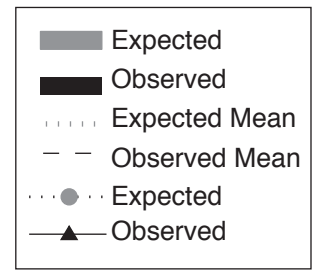

Expected
Observed
Expected Mean
Observed Mean
Expected
Observed

Figure 5(on next page)

Mitochondrial differentiation *D*Jost and *G*st statistics between location pairs

Mitochondrial differentiation *D*Jost and *G*st statistics between location pairs with N > 10. Location codes as in Fig.2, BA = Barcelona, CG = Cabo de Gata, CA = Cadiz, OL = Olhos de Água, RF = Ria Formosa, SA = Sado. Significance of differentiation indicated with an asterisk was assessed through the calculation of 95% confidence limits using a bias corrected bootstrapping method. Line at 0.0248 indicates ≥ 95% confidence (95.5% Fisher's exact test, 96.2% chi-square) threshold detection of F_{ST}.

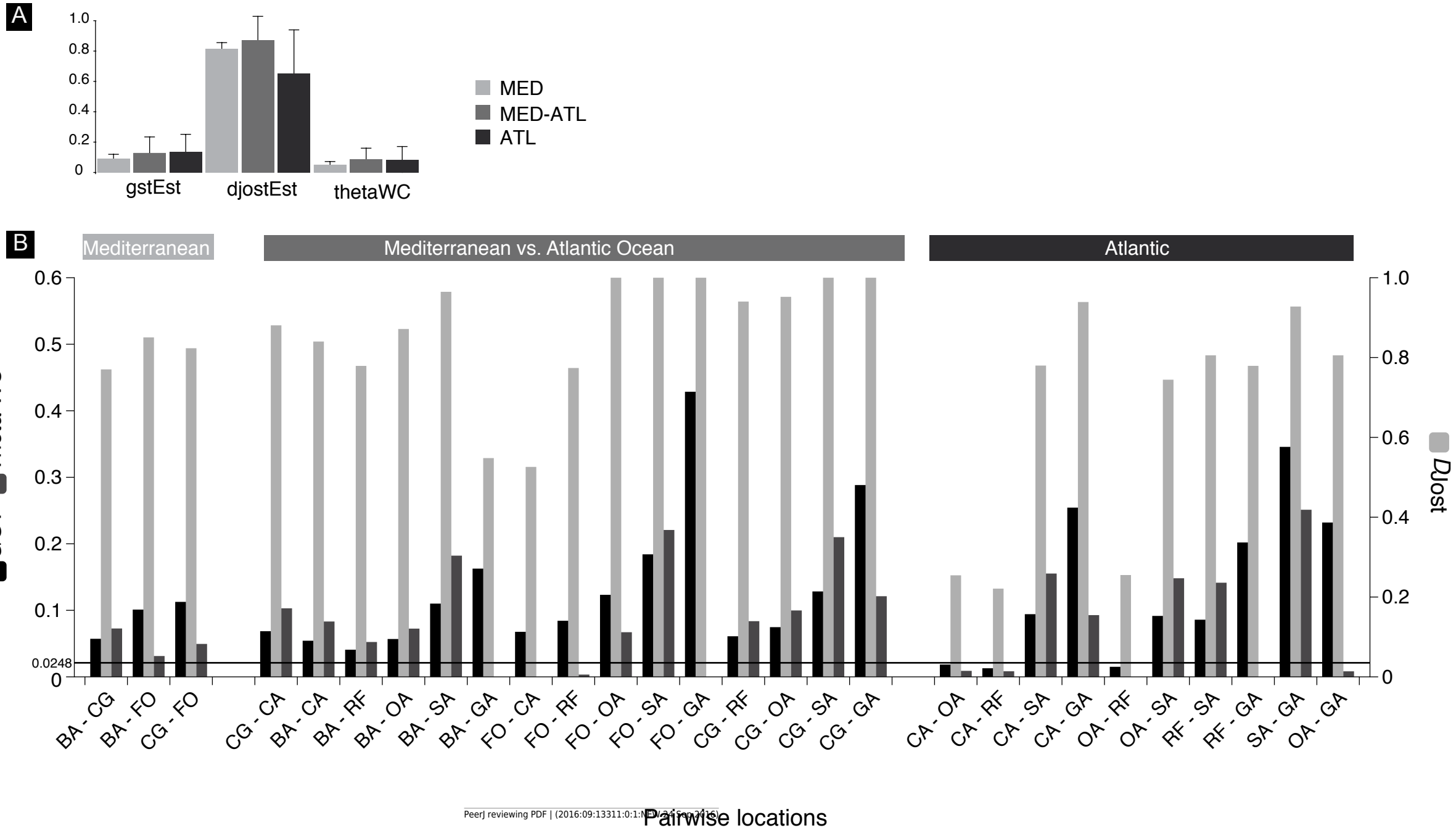

A
1.0
0.8
0.6
0.4
0.2
0
gstEst
djostEst
thetaWC
MED
MED-ATL
ATL
B
Mediterranean
Mediterranean vs. Atlantic Ocean
Atlantic
GST
Theta WC
DJost
0.6
0.5
0.4
0.3
0.2
0.1
0.0248
0
1.0
0.8
0.6
0.4
0.2
0
BA - CG
BA - FO
CG - FO
CG - CA
BA - CA
BA - RF
BA - OA
BA - SA
BA - GA
FO - CA
FO - RF
FO - OA
FO - SA
FO - GA
CG - RF
CG - OA
CG - SA
CG - GA
CA - OA
CA - RF
CA - SA
CA - GA
OA - RF
OA - SA
RF - SA
RF - GA
SA - GA
OA - GA
Pairwise locations

Figure 6(on next page)

Bayesian skyline reconstructions showing the historical demographic trends for *Salaria pavo* for D-loop sequences.

Bayesian skyline reconstructions showing the historical demographic trends for *Salaria pavo* for D-loop sequences. Time, in thousands of years, is shown on the x-axis. Along the y-axis is the expressed population size estimated in units of Neτ (Ne: effective population size, τ: mutation rate per haplotype per generation). The central dark horizontal line in the plot is the median value for effective population size; the light lines are the upper and lower 95% HPD for those estimates. The grey rectangle corresponds to the period of the last glacial maximum (LGM) for a 3.5% molecular clock (A), 5% (B) and 10% (C).

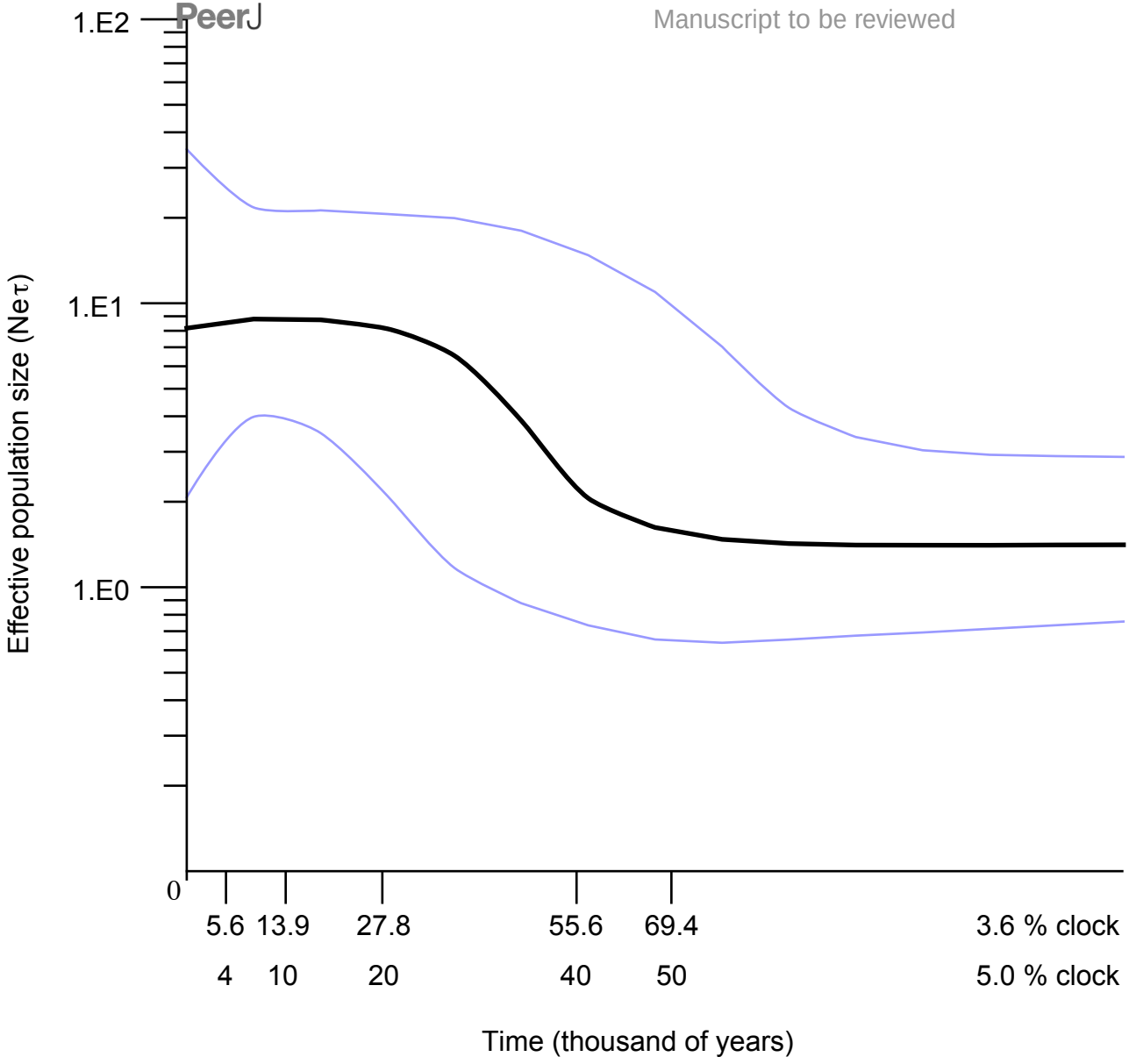

1.E2
1.E1
1.E0
0
Effective population size (Neτ)
5.6
13.9
27.8
55.6
69.4
3.6 % clock
4
10
20
40
50
5.0 % clock
Time (thousand of years)

Table 1(on next page)

Sampling locations of *Salaria pavo*.

Sample locations, sample abbreviation code, sample sizes and summary statistics for a sequence fragment of the mtDNA D-loop and the first intron of S7 nuclear gene of *Salaria pavo.*

Table 1 Sample locations, sample abbreviation code, sample sizes and summary statistics for a sequence fragment of the mtDNA D-loop and the first intron of S7 nuclear gene of *Salaria pavo.*

			Mitochondrial D-loop						**First intron of S7 gene**				
Region	Locations	Code	N	NH	NP	Haplotype diversity ± s.d.	Nucleotide diversity ± s.d. (%)	PS	N	NH	Gene diversity ± s.d.	Nucleotide diversity ± s.d. (%)	Observed Heterozygosity
	Barcelona	BA	16	9	3	0.92 ± 0.04	6.65 ± 3.48	36	15	5	0.65 ± 0.07	0.17 ± 0.00	0.53
M	Formentera	FO	2	2	1				3	1			
	Cabogata	CG	12	6	2	0.86 ± 0.06	1.63 ± 0.96	9	13	4	0.64 ± 0.07	0.15 ± 0.00	0.46
	Cadiz	CA	39	19	13	0.88 ± 0.04	1.62 ± 0.90	25	22	6	0.72 ± 0.05	0.19 ± 0.00	0.41
	Ria Formosa	RF	23	13	8	0.94 ± 0.03	1.15 ± 0.68	17	40	6	0.62 ± 0.04	0.16 ± 0.00	0.45
A	Olhos de Água	OA	12	7	6	0.91 ± 0.08	1.50 ± 0.90	16	12	5	0.78 ± 0.07	0.23 ± 0.00	0.42
	Sado	SA	26	7	4	0.72 ± 0.08	1.56 ± 0.88	12	30	4	0.73 ± 0.03	0.19 ± 0.00	0.40
	Galicia	GA	1	1	1				1	1			

N, number of individuals per location; NH, haplotype richness; NP, number of private haplotypes; PS, number of polymorphic sites

Table 2(on next page)

Gene flow amongst Atlantic and Western Mediterranean locations of peacock blenny *Salaria pavo.*

Gene flow amongst Atlantic and Western Mediterranean locations of peacock blenny *Salaria pavo* estimated in Migrate-n for mitochondrial DNA data. Differences between each alternative model and model with highest rank are in column dBézier. Exponentiated model differences (column EXP dBézier) are used to estimate model probability dividing EXP dBézier by the sum of all EXP dBézier column values.

1 Table 2. Comparison of four biogeographic models for *Salaria pavo*.

Marker	Models		Bézier	Log Bayes factor	Probability
mtDNA	Model 1	Panmixia	-1266.8	19.1	0.000
	Model 2	ATL ↔MED	-1247.7	0.0	0.940
	Model 3	ATL → MED	-1250.5	2.8	0.060
	Model 4	MED → ATL	-1262.0	14.3	0.000

2

3