

# Genetic variation of the spiny spider crab *Maja brachydactyla* in the northeastern Atlantic

Graciela Sotelo<sup>1,\*</sup>, Paloma Morán<sup>1</sup>, Luis Fernández<sup>2</sup>, David Posada<sup>1</sup>

<sup>1</sup>Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Campus Lagoas-Marcosende, Universidad de Vigo, 36310 Vigo, Spain

<sup>2</sup>Departamento de Biología Animal, Biología Vegetal y Ecología, Campus A Zapateira, Universidad de A Coruña, 15071 A Coruña, Spain

**ABSTRACT:** The spiny spider crab *Maja brachydactyla* is a decapod crustacean that inhabits the northeastern Atlantic, from the British Isles to Senegal, and is the target of intense fishing activity. In the absence of previous genetic studies, we describe the genetic variability and structure of this species along Atlantic European coasts, paying special attention to the northwestern Iberian Peninsula (Galicia), where this species may be overexploited. Analysis of 2 mitochondrial fragments and 9 microsatellites indicates that these populations maintain appreciable levels of genetic diversity, most of which are found within localities. This scenario could be explained by quite large effective population sizes and noticeable levels of gene flow. *M. brachydactyla* in the northeastern Atlantic do not show any signs of a genetic bottleneck, and effectively conform to a diverse metapopulation with limited structure.

**KEY WORDS:** Genetic diversity · Gene flow · Effective population size · MtDNA · Microsatellites · Crab population genetics · Phylogeography

—Resale or republication not permitted without written consent of the publisher—

## INTRODUCTION

The spiny spider crab *Maja brachydactyla* Balss, 1922 is a decapod crustacean found in the northeastern Atlantic, from the British Isles to Senegal (Sotelo et al. 2008). This crab has a complex life cycle that has been widely studied for the northwestern region of the Iberian Peninsula (Galicia) (Corgos 2004). Here, planktonic larvae are released in coastal areas and develop through 3 stages in 2 to 3 wk. After settlement, juveniles remain in shallow waters (<15 m) over rocky-sandy bottoms for 2 to 3 yr until they undergo their terminal moult and reach sexual maturity, usually during summer. Adults begin autumn migrations to deeper waters (10 to 30 m) in order to reproduce, and, in the following spring, females return to juvenile habitats for larval release. Females are able to store sperm and to fertilize several broods without additional copulations. In Galicia, the mean number of broods per year is 3, but this number decreases towards the north of the

distribution depending on water temperature. This leads to variable and overlapping cohorts, especially taking into account that an adult can live 2 to 3 yr.

European populations of *Maja brachydactyla* are subject to high exploitation pressure, and are one of the crustaceans with the highest commercial importance along these coasts. Annual landings have been growing during recent years after a decline in the 1980s, as the FAO statistics indicate. In 2005 alone, 5000 tonnes were caught. The main fishery regions in Europe are France and the United Kingdom. On the Iberian Peninsula, the most important area is Galicia, which is the focus of our study, and where the exploitation rate exceeds 90% (Freire et al. 2002).

Despite their relevance, little is known about the dynamics and structure of these crab populations across their whole range, especially from a genetic perspective. This information is fundamental to assess the status of the species and the need for conservation strategies. While the harmful effects of overexploita-

\*Email: gsotelo@uvigo.es

tion have been documented in several marine fishes (see Hutchings 2000), less attention has been paid to marine invertebrates, which are often characterized by large population sizes, high potential for dispersal and wide geographic distribution (see Zane et al. 2000). These factors, in conjunction with the absence of obvious physical barriers in marine environments, support the prediction of genetic homogeneity among populations, with isolation by distance at a macrogeographic scale (Mathews 2007). This theoretical pattern has been observed in several instances (e.g. Triantafyllidis et al. 2005), but rejected in others (e.g. Mathews 2007). In fact, the presence or absence of genetic structure in marine invertebrates seems to be the result of a combination of ecological, behavioural and physical mechanisms that should be explored in each particular case (Thorpe et al. 2000).

In the present study we describe the genetic variability and structure of *Maja brachydactyla* populations in the northeastern Atlantic, both at microgeographic (Galicia) and macrogeographic scales (Europe). For this purpose, we assessed the genetic variation in 2 mitochondrial DNA fragments and at 9 microsatellite loci. Detailed knowledge of the amount and distribution of this variation could be very important to preserve this relevant fishery resource.

## MATERIALS AND METHODS

**Sampling.** We analyzed 299 individuals from 15 localities in the northeastern Atlantic region, with a particular focus on the northwestern Iberian coast (Galicia) (Fig. 1, Table 1). Samples were collected between 2003 and 2005 by nets or SCUBA diving, and a pair of walking legs from each specimen was preserved in 100% ethanol for molecular analysis. Muscle tissue was used to extract genomic DNA following a Chelex resin protocol (Estoup et al. 1996).

**Mitochondrial DNA sequencing.** We amplified 2 fragments of mitochondrial DNA (mtDNA) by polymerase chain reaction (PCR) using universal primers: 560 bp of the 16S rRNA gene with 16Sar and 16Sbr (Palumbi et al. 1991) and 710 bp of the cytochrome oxidase I (COI) gene with LCO1490 and HCO2198 (Folmer et al. 1994). Reactions were carried out in a final volume of 20  $\mu$ l, containing 1  $\mu$ l of DNA extraction, 2  $\mu$ l of 10 $\times$  PCR buffer (160 mM  $(\text{NH}_4)_2\text{SO}_4$ , 670 mM Tris-HCl pH 8.8, 0.1% Tween 20), 1  $\mu$ l of 50 mM  $\text{MgCl}_2$ , 1  $\mu$ l of each primer (20  $\mu$ M), 1  $\mu$ l of 10 mM dNTP Mix (Applied Biosystems), 0.2  $\mu$ l BIOTAQ polymerase (5 U  $\mu$ l<sup>-1</sup>, Bioline) and 13  $\mu$ l of sterile bidistilled water. PCR profiles were as follows: 5 min at 95°C, 35 cycles of 20 s at 95°C, 20 s at 55°C, 20 s at 72°C and 7 min at 72°C for 16S and 5 min at 95°C, 30 cycles of 20 s at 95°C, 20 s

at 42°C, 30 s at 72°C and 7 min at 72°C for COI. For verification, PCR products were run in 2% agarose gels stained with ethidium bromide. Prior to sequencing, single strand conformation polymorphism (SSCP) analysis was used to screen for different haplotypes, as each haplotype results in a different mobility pattern when migrating through a gel in non-denaturing conditions (Sunnucks et al. 2000). We used vertical gels (23 cm long  $\times$  21 cm wide  $\times$  1 mm thick) containing 8% acrylamide (50:1 acrylamide/bis-acrylamide). Then, 5  $\mu$ l of unpurified PCR product were mixed with 5  $\mu$ l of denaturing loading buffer, heated at 95°C for 5 min and then chilled on ice until loaded on the gels. Gels ran at 8°C and 30 W for 4 and 8 h (16S and COI genes, respec-

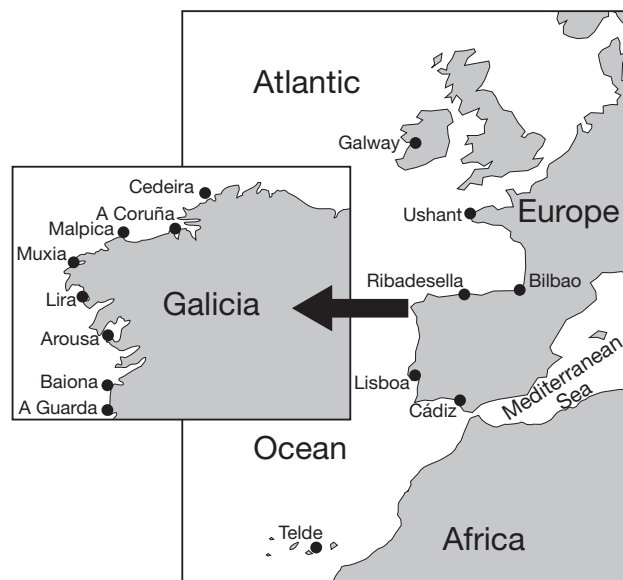


Fig. 1. Map of sampled localities

Table 1. Collecting sites. Sample size (n), year(s) of collection and geographical coordinates are indicated

Locality	n	Year	Coordinates
Galway (Ireland)	25	2004	53° 43' N, 9° 02' W
Ushant (France)	25	2004	48° 27' N, 5° 05' W
Bilbao (Spain)	4	2004	43° 15' N, 2° 58' W
Ribadesella (Spain)	20	2005	43° 28' N, 5° 04' W
Cedeira (Spain)	10	2003	43° 15' N, 2° 58' W
A Coruña (Spain)	26	1999/2003	43° 22' N, 8° 23' W
Malpica (Spain)	20	2005	43° 19' N, 8° 48' W
Muxía (Spain)	10	2004	43° 06' N, 9° 13' W
Lira (Spain)	36	2003/2004	42° 48' N, 9° 09' W
Arousa (Spain)	37	1999/2004	42° 30' N, 8° 52' W
Baiona (Spain)	17	2004	42° 07' N, 8° 51' W
A Guarda (Spain)	18	2004/2005	41° 54' N, 8° 53' W
Lisboa (Portugal)	22	2005	38° 43' N, 9° 08' W
Cádiz (Spain)	25	2005	36° 32' N, 6° 17' W
Telde (Canary Islands)	4	2006	28° 00' N, 15° 25' W

tively), and the resulting bands were visualized by silver-staining. In order to assess and confirm the nucleotide differences between the different haplotypes detected by SSCP, all samples with different mobility patterns, as well as random samples of the most frequent patterns, were sequenced with the forward PCR primers indicated above. Previously, PCR products were purified with the GFX PCR DNA and gel band purification kit (Amersham Biosciences) following manufacturer's instructions. Sequencing reactions were performed with dRhodamine terminator chemistry (Applied Biosystems), precipitated with ethanol and run in an automated sequencer (ABI PRISM 310, Applied Biosystems). Electropherograms were visualized with the program DNA Sequencing Analysis 3.7 (Applied Biosystems).

**Microsatellite genotyping.** We also genotyped the same individuals for 9 microsatellite loci (Sotelo et al. 2007): mMb9, mMb12, mMb15FAM, mMb15NED, mMb23, mMb33, mMb307, mMb339 and mMb403.

**Mitochondrial data analysis.** Sequences were aligned with ClustalW (Thompson et al. 1994) and revised by eye. Best-fit nucleotide substitution models were selected under the Akaike information criterion (AIC) using Modeltest V3.6 (Posada & Crandall 1998) and PAUP\* 4b10 (Swofford 2003). For each locality, several variability measures were calculated with DnaSP V4.10 (Rozas et al. 2003): number of haplotypes ( $k$ ), number of segregating sites ( $S$ ), haplotype diversity ( $h$ ) and nucleotide diversity ( $\pi$ ).

To evaluate population growth, we used Fu's  $F_S$  (Fu 1997) and  $R_2$  (Ramos-Onsins & Rozas 2002) statistics, also available in DnaSP. The 95% confidence intervals for these were generated by coalescent simulations conditioned on theta ( $\theta$ ) (1000 replicates). In addition, we analyzed the fit of the mismatch distribution to those expected under demographic and spatial expansion with Arlequin V3.01 (Excoffier et al. 2005). The sum of squared deviations (SSD) between the observed and the expected distributions was used as a test statistic, and the significance was assessed after 1000 bootstraps. Inferred expansions were dated following the expression  $T = \tau/2\mu k$ , where  $\tau$  is the estimated number of generations since the expansion,  $\mu$  is the mutation rate per site per generation and  $k$  is the sequence length. We assumed a mutation rate of 1% per lineage per million years (average estimate reported in Ketmaier et al. 2003) and a mean generation time of 2 yr (see Freire et al. 2002).

We used several approaches to estimate effective population sizes ( $N_e$ ). First, we used the expression  $N_e = \theta/\mu$ , where  $\theta$  was estimated from nucleotide diversity (Tajima 1996), while  $\mu$  was estimated as above. Also, we used LAMARC V2.1.2b (Kuhner 2006) to simultaneously estimate effective population size and

population growth for each locality. Here we performed a likelihood analysis with default starting values. Each run consisted of 10 short chains with 1000 recorded parameter sets, a sampling interval of 20 and a burn-in of 2000; and 2 long chains with 50 000 recorded parameter sets, a sampling interval of 20 and a burn-in of 5000. We also used heating, with 3 relative temperatures: 1, 1.1 and 2.

To explore population differentiation, several pairwise indexes were calculated with Arlequin:  $F_{ST}$ , Slatkin's linearized  $F_{ST}$ 's; TrN (Tamura & Nei 1993) distances and  $D_A$  (Nei & Li 1979) distances. The latter two take into account the nucleotide differences between haplotypes. In every case, p-values were obtained after 1000 bootstraps and corrected for multiple tests with the modified false discovery rate (FDR) procedure of Benjamini & Yekutieli (2001) as described in Narum (2006). To detect any association between the genetic and the geographical distances between populations, Mantel tests were performed with 10 000 permutations using PopTools (available at [www.cse.csiro.au/poptools/](http://www.cse.csiro.au/poptools/)). Geographical distances were calculated with ArcGIS (ESRI) as the shortest path following the coastline.

Subsequently, genetic and geographic data were combined/related to assess population structure. We used 2 approaches that do not require the localities being grouped *a priori*. First, we used SAMOVA V.1.0 (Dupanloup et al. 2002), an annealing procedure to define clusters of adjacent populations that maximize the proportion of genetic variance due to differences between groups ( $F_{CT}$ ). We tested different numbers of groups, and each run consisted of 1000 annealing replicates. Second, we used Barrier V.2.2 (Manni et al. 2004), a program that implements Monmonier's maximum difference algorithm, to identify the areas where genetic change is highest. In a first step, all sampling localities are distributed according to geographic distances and connected by a network of triangles. Then, each connection is associated with the corresponding genetic distance and the edge with the highest distance is the beginning of the first barrier. This barrier is extended through adjacent edges until it reaches the limits of the triangulation or closes on itself. Many barriers, as many as indicated by the user, can be computed one after another in decreasing order of importance.

Finally, to infer the evolutionary processes that could have shaped the observed geographical distribution of the recorded haplotypes, a nested clade phylogeographic analysis (NCPA; Templeton et al. 1995) was implemented. First, a haplotype network was constructed using the program TCS V1.21 (Clement et al. 2000). Before building the nested design, ambiguities in the network were resolved following different criteria (see Pfenninger & Posada 2002). To calculate the

NCPA statistics and their significance (10 000 replicates), the program Geodis V2.0 (Posada et al. 2000) was used. In the NCPA, clade distances ( $D_c$ ) measure the geographical spread of a clade; nested clade distances ( $D_n$ ) measure the geographical distance of a clade from the geographical centre of the next higher nested clade; and interior-tip distances ( $I-T_c$  and  $I-T_n$ ) measure differences in  $D_c$  and  $D_n$  between interior and tip clades. The revised version of the inference key (Templeton 2004, available at <http://darwin.uvigo.es/software/geodis.html>) was used to deduce whether the observed distance patterns are best attributed to restricted gene flow or to historical events, such as range expansion, past fragmentation, or long-distance colonization.

**Microsatellite data analysis.** Several summary statistics were calculated in Genepop V3.4 (Raymond & Rousset 1995): allele frequencies, number of alleles per locus ( $A$ ), observed ( $H_o$ ) and expected heterozygosities ( $H_e$ ) by locus and by population as well as departures from Hardy-Weinberg expectations. The latter were inferred using an exact probability test, where the p-value was estimated by a Markov chain method with 10 000 dememorizations, 100 batches and 1000 iterations per batch.  $F_{stat}$  V.2.9.3 (Goudet 2001) was used to calculate allelic richness ( $A_r$ ), and Genetix 4.03 (Belkhir et al. 2004) was used to compute pairwise  $F_{ST}$  between populations and their significance (after 1000 permutations and FDR adjustment). We also estimated the standardized measure  $G'_{ST}$  (Hedrick 2005) using RecodeData V.0.1 (Meirmans 2006). This statistic indicates the level/degree of differentiation proportional to the maximum possible given the observed within-population variation, although its statistical significance is not easy to calculate. In order to take into account the size differences between alleles, the  $F_{ST}$ -analogue  $R_{ST}$  was calculated with Arlequin (significance was again evaluated through 1000 permutations and FDR correction). In addition, a matrix of pairwise chord distances between populations was obtained with Genetix. This distance was chosen because it has been suggested as being one of the most suitable for microsatellites (Takezaki & Nei 1996). As with mitochondrial data, association between these differentiation indexes and geographical distances was analyzed with Mantel tests (10 000 permutations using PopTools) and population structure was further assessed with SAMOVA and Barrier.

Genotype biases were evaluated with MicroChecker (Van Oosterhout et al. 2004), and null alleles were consistently found at 2 loci across all populations: mMb33 and mMb307. As null alleles may lead to erroneous estimates of population differentiation, their frequency was estimated with FreeNA (Chapuis & Estoup 2007), which also recalculates  $F_{ST}$  between populations

by the ENA ('excluding null alleles') method, although without supplying p-values. In addition, it provides a genotype dataset corrected for null alleles using the INA ('including null alleles') method. The latter was then used to re-estimate the chord distances. Both corrected measures were also used to perform Mantel tests and Barrier analysis. SAMOVA was not applied to this dataset because this correction is not suitable for estimation of  $F$ -statistics and analogues (Chapuis & Estoup 2007), on which this algorithm relies.

Assuming equilibrium, effective population sizes were estimated from the expected heterozygosities per sample, over all loci. Given the low frequency of multi-step changes estimated previously for these loci (Sotelo et al. 2007), we assumed a stepwise mutation model. Under this model,  $N_e = \{[1/(1 - H_e)]^2 - 1\}/8\mu$  (Estoup & Cornuet 1999). The mutation rate assumed was  $10^{-3}$  gamete<sup>-1</sup> generation<sup>-1</sup>, which can be considered a consensus estimate (see Ellegren 2004, Yue et al. 2007). Effective population sizes were also estimated with ONeSAMP (Tallmon et al. 2008), a program that uses approximate Bayesian computation. The upper and lower bounds on the prior distribution for  $N_e$  were 2 and 5000, respectively. After many attempts, the LAMARC runs for microsatellites continuously reported 'maximization failures' and 'overflow errors', which prevented us from obtaining meaningful results.

## RESULTS

### Mitochondrial data

The final alignments for the 16S and COI fragments included 518 and 548 bp, respectively. For both fragments, the SSCP patterns and the corresponding sequences were always in agreement. The 16S fragment showed no polymorphic sites, and no further analyses were performed. For the COI fragment, 41 haplotypes, defined by 38 variable positions, were found. In this fragment, 11 sites were parsimony-informative, and 33 out of 40 changes were synonymous. The best-fit model of nucleotide substitution was HKY+I (Hasegawa et al. 1985), but with a low Akaike weight (15%), so for the subsequent analyses we used the averaged parameter estimates over all models (Posada & Buckley 2004) for base frequencies ( $A = 0.2730$ ,  $C = 0.1911$ ,  $G = 0.1739$ ,  $T = 0.3619$ ), relative substitution rates ( $AC = 0.4562$ ,  $AG = 12.7578$ ,  $AT = 0.9435$ ,  $CG = 0.0001$ ,  $CT = 7.8745$ ) and proportion of invariable sites ( $p\text{-inv} = 0.8041$ ).

Genetic diversity was high and similar across populations (Table 2), with the exception of Bilbao and Telde, that were less diverse, but for which only 4 individuals were available. The demographic analyses

Table 2. Genetic variability across populations. For COI (cytochrome oxidase I), the estimates indicated are: number of segregating sites ( $S$ ), number of haplotypes ( $k$ ), haplotype diversity ( $h$ ) followed by the 95% confidence interval [CI ( $h$ )], and nucleotide diversity ( $\pi$ ) followed by the 95% confidence interval [CI ( $\pi$ )]. The last 2 columns correspond to the observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities over the 9 microsatellite loci

Locality	COI						Microsatellites			
	$S$	$k$	$h$	CI ( $h$ )		$\pi$	CI ( $\pi$ )		$H_o$	$H_e$
Galway	10	6	0.7300	0.6550	0.8050	0.0033	0.0013	0.0053	0.6526	0.7098
Ushant	7	6	0.7267	0.6527	0.8007	0.0031	0.0012	0.0050	0.6462	0.7205
Bilbao	3	2	0.5000	0.2350	0.7650	0.0027	0.0007	0.0047	0.6389	0.7593
Ribadesella	13	10	0.8316	0.7566	0.9066	0.0039	0.0016	0.0062	0.6450	0.7521
Cedeira	8	6	0.7778	0.6408	0.9148	0.0035	0.0013	0.0057	0.6977	0.7204
A Coruña	14	12	0.9139	0.8819	0.9459	0.0050	0.0023	0.0077	0.6535	0.7739
Malpica	9	9	0.8947	0.8537	0.9357	0.0044	0.0019	0.0069	0.6816	0.7930
Muxía	4	4	0.7333	0.6133	0.8533	0.0030	0.0011	0.0049	0.7176	0.7797
Lira	9	10	0.8222	0.7792	0.8652	0.0035	0.0015	0.0055	0.6090	0.7676
Arousa	17	15	0.8829	0.8459	0.9199	0.0038	0.0016	0.0060	0.6342	0.7588
Baiona	9	5	0.7721	0.7021	0.8421	0.0044	0.0019	0.0069	0.6966	0.7454
A Guarda	8	6	0.7320	0.6360	0.8280	0.0028	0.0010	0.0046	0.6026	0.7475
Lisboa	16	11	0.9264	0.8974	0.9554	0.0052	0.0024	0.0080	0.6379	0.7588
Cádiz	9	8	0.8567	0.8107	0.9027	0.0035	0.0014	0.0056	0.6888	0.7796
Telde	2	2	0.6667	0.4627	0.8707	0.0024	0.0005	0.0043	0.6571	0.8286

suggested fluctuating sizes at least for some populations (Table 3). Fu's  $F_s$  and  $R_2$  statistics rejected constant population size at Ribadesella, Cedeira, A Coruña and Arousa, and, with the exception of the smallest samples, most  $F_s$  values were negative. According to the mismatch distributions, the hypothesis of a sudden/spatial expansion could not be rejected in most populations, with a mean expansion estimate ( $\tau$ ) ranging from 2 to 4. The onset of the expansion was dated between 100 000 and 188 000 yr ago, but the confidence intervals were quite wide. In order to assess a possible lack of power for the mismatch analysis, we performed a very small simulation experiment.

Using the program Recodon (Arenas & Posada 2007), we simulated 10 datasets with 20 individuals each, sampled from a population with constant size. The remaining simulation parameters were specified to produce samples with haplotype and nucleotide diversities similar to those observed in the present study. In all 10 datasets, the mismatch analysis did not reject the expansion model. On the other hand, in order to increase statistical power, we pooled the localities according to differentiation analyses, and all together. In the former case,  $R_2$  fitted a constant model for all localities,  $F_s$  detected size changes for the Cantabrian and Galician groups and all the mismatch distributions

Table 3. Demographic estimates from COI. The last 3 columns indicate the estimates for the mismatch distributions. SSD: sum of squared deviations between observed and expected distributions under a sudden or spatial expansion;  $\tau$ : expansion parameter;  $T$ : time from the expansion, in kiloyears ago; n.a.: least square procedure to fit model mismatch distribution and observed distribution did not converge. Significant values are indicated in bold

Locality	Fu's $F_s$	$R_2$	SSD		$\tau$		$T$	
			Sudden	Spatial	Sudden	Spatial	Sudden	Spatial
Galway	-0.1860	0.0979	0.0630	0.0489	2.629	2.381	119.94	108.62
Ushant	-0.3950	0.1177	0.0358	0.0257	2.458	2.223	112.14	101.41
Bilbao	1.7160	0.4330	0.2563	0.1596	4.121	3.489	188.00	159.17
Ribadesella	<b>-4.2260</b>	<b>0.0770</b>	0.0044	0.0029	2.442	2.294	111.41	104.65
Cedeira	-1.9530	<b>0.1094</b>	0.0219	0.0150	2.557	2.373	116.65	108.26
A Coruña	<b>-4.2780</b>	0.0914	0.0152	0.0145	3.234	3.183	147.54	145.21
Malpica	-2.5870	0.1264	0.0122	0.0116	3.102	3.059	141.51	139.55
Muxía	0.1370	0.1930	0.0172	0.0126	2.637	2.262	120.30	103.19
Lira	-2.8330	0.1041	0.0135	0.0100	2.405	2.345	109.72	106.98
Arousa	<b>-8.2980</b>	<b>0.0554</b>	n.a.	0.0023	n.a.	2.265	n.a.	103.33
Baiona	0.9410	0.1394	0.0296	0.0171	3.817	3.231	174.13	147.40
A Guarda	-1.1670	0.0984	0.0286	0.0224	2.202	2.037	100.46	92.93
Lisboa	-3.6900	0.0816	n.a.	0.0151	n.a.	3.340	n.a.	152.37
Cádiz	-1.7930	0.1074	0.0141	0.0136	2.292	2.314	104.56	105.57
Telde	1.5300	0.3333	0.2651	<b>0.2424</b>	2.289	2.322	104.43	105.93

failed to reject the expansion model. When all samples were pooled, all tests were in agreement with an expansion.

The estimates of  $N_e$  were very different depending on the method used. While those based on nucleotide diversity were similar across localities and oscillated around 200 000 (Table 4), those reported by LAMARC were highly variable (Table A1 in Appendix 1 of the supplementary material available at [http://www.int-res.com/articles/suppl/m362p211\\_app.pdf](http://www.int-res.com/articles/suppl/m362p211_app.pdf)), ranging from <200 000 to >30 000 000. Moreover, for the LAMARC estimates, the likelihood scarcely improved along the Markov chain Monte Carlo (MCMC) despite running a reasonable number (2 200 000) of generations.

Two haplotypes, 3 and 4, occurred at almost all localities and, together with haplotypes 2 and 11, represented more than half of the individuals (Table 5, Fig. A1 in Appendix 1). The remaining haplotypes were mostly restricted to single populations. There were several haplotypes at each locality, usually quite similar, with 2 nucleotide differences on average. Consequently, population differentiation was, in general, low, despite the distance measure used, and there was no correlation with geographical distances (no Mantel test was significant). Several pairwise  $F_{ST}$  values were 0 (18 out of 105) and only those involving Bilbao, Telde and Cádiz, as well as some values from Malpica and Baiona, were >0.1 (Table 6). After the FDR correction, 27 comparisons were significant, mostly related to Cádiz. The TrN-corrected values were even lower (with the exception of Bilbao), and the number of significant tests decreased (15). Nevertheless, Cádiz maintained its distinctiveness, as also corroborated by  $D_A$  estimates (Table A2 in Appendix 1).

Table 4. Effective population sizes estimates. See Table A1 in Appendix 1 for estimates obtained from LAMARC and ONE-SAMP (unreliable; see 'Discussion'). COI based on  $\theta$ ; microsatellites based on  $H_e$

Locality	COI	Microsatellites
Galway	165500	1359
Ushant	154000	1475
Bilbao	137500	2032
Ribadesella	198000	1909
Cedeira	175000	1474
A Coruña	252500	2320
Malpica	220000	2793
Muxía	148500	2450
Lira	177000	2190
Arousa	192000	2024
Baiona	222500	1803
A Guarda	142000	1835
Lisboa	261500	2024
Cádiz	176500	2449
Telde	122000	4128

A different number of SAMOVA clusters (2 to 8) were assessed. In all these partitions, >90% of the variation was allocated within populations. Bilbao and Telde always formed their own separate groups, as did Cádiz. The remaining localities clustered together depending on the number of partitions. The results of Barrier with the different distances were in agreement with those from SAMOVA. The most important discontinuities separated Bilbao, Telde and Cádiz. Within the remaining localities, other barriers appeared, in different order, forming 4 geographical groups: Galway, Ushant, Ribadesella and Cedeira (1); A Coruña, Malpica, Muxía, Lira and Baiona (2); Arousa (3); and A Guarda and Lisboa (4).

In the haplotype network, the 2 most frequent haplotypes (3 and 4) were separated by 2 mutational steps and surrounded by the remaining haplotypes, which, in general, differed from them by a single substitution (Fig. 2). Reticulations in this network were likely the result of homoplasies produced by parallel and reverse changes, rather than by recombination (although see Tsaousis et al. 2005). The nesting design resulted in 3 nesting levels. Haplotypes 3 and 4 were the centre of the 2 main clades at the 1-step level. The NCPA led to a pattern of restricted gene flow with isolation by distance for 3 clades: 1-1, 1-11 and 2-1 (we used 0.01 as the critical p-value).

### Microsatellite data

All 9 microsatellite loci were highly variable both with respect to size range and number of alleles (Table 7). The least variable were mMb23 and mMb9, with 7 and 8 alleles, respectively, while the most diverse was mMb15NED, with 35 alleles. This is presumably the reason why  $H_o$  and  $H_e$  were quite different across loci:  $H_e$  ranged from 0.57 (mMb9) to 0.94 (mMb15NED). Despite these differences, all loci were in Hardy-Weinberg equilibrium (HWE), except for mMb33 and mMb307, which showed a clear heterozygote deficit (Table 7). These deviations were most probably due to the presence of null alleles, as suggested by Micro-Checker. The estimated mean frequency of null alleles for mMb33 and mMb307 was high (51 and 21%, respectively), while negligible (<5%; see Chapuis & Estoup 2007) for the other loci.

All populations displayed similar levels of microsatellite variation (Table 2). Except Bilbao, Muxía and Telde, all populations showed departures from HWE, but again this was attributable to the 2 loci indicated above. In fact, when removing them from the dataset, all populations adjusted to HWE. The allelic richness was almost identical across populations and differed depending on the locus: the mean value was 2.61

Table 5. Absolute frequency of COI haplotypes across populations

Haplotype	Galway	Ushant	Bilbao	Ribadesella	Cedeira	A Coruña	Malpica	Muxía	Lira	Arousa	Baiona	A Guarda	Lisboa	Cádiz	Telde	Total
1	1											2				3
2	2	3		1		1	1	1	2	1		2	3	2		19
3	5	3		3	1	6	4	2	12	4	7	2	4	3		56
4	12	12	1	8	5	3	1	5	9	11	3	9	3	2		84
5	3	5		2					2	1		1	2			16
6	2	1			1					2			1			7
7		1				1										2
8			3	1		3	2	2	1	1						13
9				1												1
10				1												1
11				1		3	5		4	5	4	1	1	8	2	34
12				1												1
13				1		3							1	2		7
14					1											1
15					1											1
16					1				2	1						4
17						2				1			1			4
18						1										1
19						1										1
20						1										1
21						1										1
22							2									2
23							2									2
24							2			1						3
25							1									1
26									1							1
27									2							2
28								1								1
29										3						3
30										3						3
31										1						1
32										1						1
33										1						1
34											2					2
35											1					1
36												3				3
37													1		2	3
38													3			3
39														4		4
40														3		3
41														1		1
Total	25	25	4	20	10	26	20	10	36	37	17	18	22	25	4	299

for mMb23 and 5.25 for mMb15NED (Table 7). The estimates of  $N_e$  according to  $H_e$  were around 2000 (Table 4), while those obtained with ONeSAMP were extremely low (in general, <100; Table A1 in Appendix 1) and showed a clear correlation with sample sizes ( $R^2 = 0.76$ ,  $p = 0.00001$ ).

Apparently, the geographic distribution of the allele frequencies did not follow any clear pattern (Fig. 3).  $F_{ST}$  and  $G'_{ST}$  values were in general low, being 0 for 24 out of 105 comparisons (Table 8). The highest values corresponded to Galway and Telde (northern and southern limits of the sampling area). After FDR correction, only 6 comparisons remained significant, most

of them involving Galway. When null alleles were excluded, in general,  $F_{ST}$  values decreased, with the exception of Galway (Table A3 in Appendix 1). Differentiation was still lower when measured with  $R_{ST}$  (Table A3 in Appendix 1).

The Mantel tests showed that  $F_{ST}$ ,  $G'_{ST}$  and  $R_{ST}$  were significantly correlated with geographical distances. When we tested 2 SAMOVA groups, Galway was separated from the other localities, and, as the number of clusters increased, Cádiz defined another group, as did Telde. The northernmost and the southernmost populations always appeared apart from the remaining localities, but the composition of the other groups was quite variable depending on the partition evaluated. In all cases, >97% of total variance was located within populations. The Barrier analysis, with  $G'_{ST}$  and  $F_{ST}$  corrected for null alleles, identified a barrier surrounding 5 adjacent populations on the Galician coast: A Coruña, Malpica, Muxía, Lira and Arousa. Moreover, Galway and Ushant clustered together and Cádiz formed its own group. Chord distances were similar between populations, with the exception of Telde and Bilbao, and also Cedeira and Muxía, which seemed more distant (Table A4 in Appendix 1). This led to the allocation of boundaries separating these localities when imported into Barrier. In this case, there was no correlation with geographical distances.

## DISCUSSION

We studied the levels of genetic variation and population structure of *Maja brachydactyla* on the north-eastern Atlantic coast, with a particular focus on the northwestern region of the Iberian Peninsula (Galicia), but also including samples collected near the latitudinal limits of the species (Ireland and Canary Islands). Two main results were obtained: first, these populations are quite diverse, and, second, there is a weak geographical structure.

The levels of genetic diversity observed from mtDNA ( $\pi = 0.003$  to  $0.005$ ) are in the range reported

Table 6. Pairwise differentiation indexes between populations based on COI data.  $F_{ST}$  values are below and linearized  $F_{ST}$  are above the diagonal. Significant  $F_{ST}$  values after the FDR (false discovery rate) procedure are indicated in bold. Statistical significance could not be estimated for the linearized  $F_{ST}$

	Galway	Ushant	Bilbao	Ribadesella	Cedeira	A Coruña	Malpica	Muxía	Lira	Arousa	Batona	A Guarda	Lisboa	Cádiz	Telde
Galway															
Ushant	0														
Bilbao	<b>0.2541</b>	0.2565													
Ribadesella	0	0	0.1709												
Cedeira	0	0	0.2316	0											
A Coruña	<b>0.0817</b>	<b>0.0990</b>	0.1321	0.0296	0.0730										
Malpica	<b>0.1305</b>	<b>0.1453</b>	0.1713	0.0719	<b>0.1189</b>	0.0015									
Muxía	0	0	0.1169	0	0	0.0441	0.0982								
Lira	0.0315	0.0568	<b>0.2155</b>	0.0095	0.0407	0.0098	0.0340	0.0197							
Arousa	0.0220	0.0327	<b>0.1663</b>	0	0	0.0226	0.0372	0.0092	0.0193						
Batona	0.0993	<b>0.1357</b>	<b>0.2863</b>	0.0627	0.1101	0.0150	0.0188	0.0910	0	0.0471					
A Guarda	0	0	0.2523	0	0	<b>0.0884</b>	<b>0.1275</b>	0	0.0527	0.0208	<b>0.1185</b>				
Lisboa	0.0480	0.0579	<b>0.1919</b>	0.0206	0.0565	0.0033	0.0298	0.0508	0.0163	0.0213	0.0446	0.0653			
Cádiz	<b>0.1481</b>	<b>0.1557</b>	<b>0.2434</b>	<b>0.0882</b>	<b>0.1332</b>	0.0312	0.0141	<b>0.1357</b>	<b>0.0678</b>	<b>0.0520</b>	0.0529	<b>0.1341</b>	0.0500		
Telde	<b>0.2896</b>	<b>0.2919</b>	0.4167	0.2050	<b>0.2638</b>	0.1235	0.0754	0.2913	0.1837	0.1330	0.1679	<b>0.2693</b>	0.1273	0.0645	

for other marine invertebrates in this area: 0.001 to 0.007 for *Meganyctiphanes norvegica* (Papetti et al. 2005) and 0.001 to 0.005 for *Sepia officinalis* (Pérez-Losada et al. 2007). The same observation was made for microsatellites ( $H_e = 0.7$  to 0.8, while, for example, the mean value found for *Loligo forbesi* was 0.8; Shaw et al. 1999). Such diversity could, in principle, not be expected given the intensive fishing activity focussed on this crustacean. For instance, it has been estimated that 90% of Galician crabs are caught every season (Freire et al. 2002). Nevertheless, an important counteracting factor could be the high fertility of the spiny spider crab, which would lead to large effective population sizes. The females spawn several times a year, and the number of eggs per clutch could be >500 000 in the largest females (L. Fernández unpubl. data), even though another characteristic of crabs is their high mortality in the first benthic instars (see Lee et al. 2006).

The estimated effective population sizes differed by orders of magnitude depending on the method and/or markers used. Values were much higher for mtDNA than for microsatellites. In the case of LAMARC, these estimates are unreliable: the likelihood of the data did not change along the MCMC after many attempts, suggesting that the likelihood surface is too flat for meaningful estimation. On the other hand, the estimates obtained from ONeSAMP seem incredibly small, and the fact that they were highly correlated with sample size suggests that they cannot be precise. Also, the different nature of the estimators might contribute to explain some of these differences. Estimates based on the coalescent (LAMARC, ONeSAMP) represent a summary of the historical effective sample size, while estimates based on  $\theta$  and  $H_e$  assume that the population is at equilibrium. Given the diversity observed, the effective population sizes should not be too small, but reliable estimates need to be calculated with more markers and higher sample sizes.

While the neutrality tests ( $F_u$ 's  $F_S$ ,  $R_2$ ) failed to reject a constant population size for most locations (but not all), the mismatch analyses failed to reject an expansion model in all samples. One explanation for this is the lack of statistical power to reject the null hypothesis. According to computer simulations, the power of  $F_S$  and  $R_2$  seems to be reasonable (Ramos-Onsins & Rozas 2002). Statistical power has not been calculated for the SSD statistic of the mismatch distribution, although for the MAE statistic it seems to be low (Ramos-Onsins & Rozas 2002). Our simulation experiment, despite its small size, suggests that the SSD test might not be very powerful. An alternative explanation for these results is that all the populations have fluctuated in time, but it was only detected in some cases with the  $F_S$  and  $R_2$  tests. Several pieces of evi-

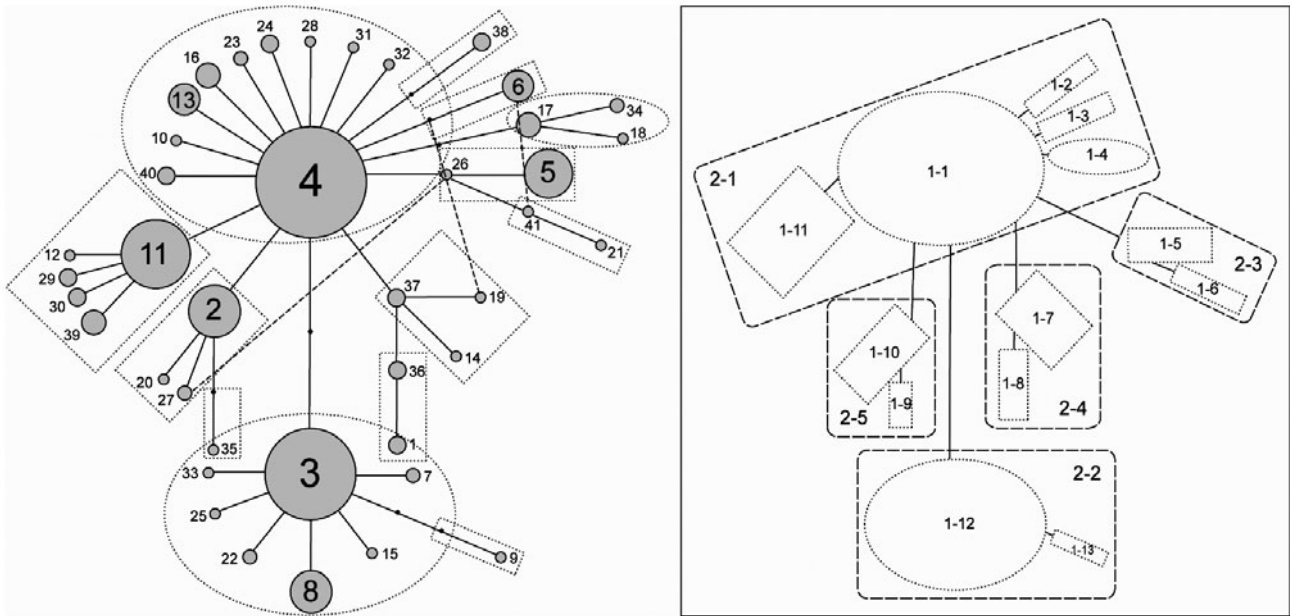


Fig. 2. Haplotype network and nested design for COI. Circles represent haplotypes with sizes proportional to their frequency. Solid lines connecting haplotypes represent mutational steps, and black dots are missing or unsampled haplotypes. Dashed lines connecting haplotypes represent reticulations resolved before constructing the cladogram. Dashed line boxes over the network enclose 1-step level clades. In the diagram on the right, they are outlined and nested into 2-step level clades that complete the design

dence point in this direction. Most  $F_S$  values were negative, with the exception of the smallest samples, and, when all localities were pooled, all tests were in agreement with an expansion. The  $\tau$  estimates were very similar for all samples, which would be difficult to explain if some populations fluctuated and others did not. Finally, the observed high levels of haplotype diversity coupled with low levels of nucleotide diversity also fit a recent expansion scenario (Rogers 1995). Such expansion would have taken place around 100 000 to 150 000 yr ago, during the Late Pleistocene. Again, more samples and markers would be advisable to confirm this expansion.

Table 7. Summary statistics for microsatellite loci. Size range, total number of alleles ( $A$ ), mean allelic richness ( $A_r$ ), observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities and  $F$ -statistics. Departures from Hardy-Weinberg expectations (significant  $F_{IS}$  values) are indicated in bold

Locus	Size (bp)	$A$	$A_r$	$H_o$	$H_e$	$F_{IS}$	$F_{ST}$	$F_{IT}$
m15FAM	130–190	13	3.397	0.73	0.75	0.0220	0.0011	0.0230
m12	159–204	12	3.305	0.62	0.71	0.1319	0.0074	0.1383
m33	179–221	18	3.955	0.25	0.78	<b>0.6856</b>	0.0263	0.6939
m23	224–244	7	2.489	0.57	0.58	0.0247	0.0071	0.0315
m15NED	233–303	35	5.273	0.94	0.94	-0.0010	0.0042	0.0032
m9	337–355	8	2.546	0.51	0.57	0.1030	0.0130	0.1147
m307	160–182	12	3.678	0.42	0.75	<b>0.4386</b>	0.0224	0.4512
m339	191–235	23	4.610	0.88	0.88	-0.0053	0.0031	-0.0021
m403	230–290	19	4.327	0.83	0.85	0.0326	0.0068	0.0392

Both mtDNA and microsatellites indicate that, although most of the variation is distributed within populations, there is some differentiation among them. Pairwise  $F_{ST}$  and analogues were, in general, small, but a few were significant. Populations in the latitudinal limits of the sampling area were detected as the most divergent: Cádiz (the southernmost locality after Telde) according to mtDNA and Galway (the northernmost locality) according to microsatellites. The main boundaries defined by the SAMOVA and Barrier analyses seem to fit this scenario, identifying also some north to south structuring across the remaining populations. Restricted gene flow with isolation by distance

was inferred by the Mantel test for microsatellites and by the NCPA for mtDNA. The NCPA was able to infer this situation because the most frequent haplotypes have colonized almost the whole range, and, at the same time, several alleles have only recently emerged at each locality. Although it has recently been suggested that the NCPA might spuriously claim restricted gene flow under panmixia (Panchal & Beaumont 2007), it seems clear that despite being quite homogeneous, the Atlantic *Maja brachydactyla* populations do not conform exactly to a single random mating unit. Rather, the NCPA

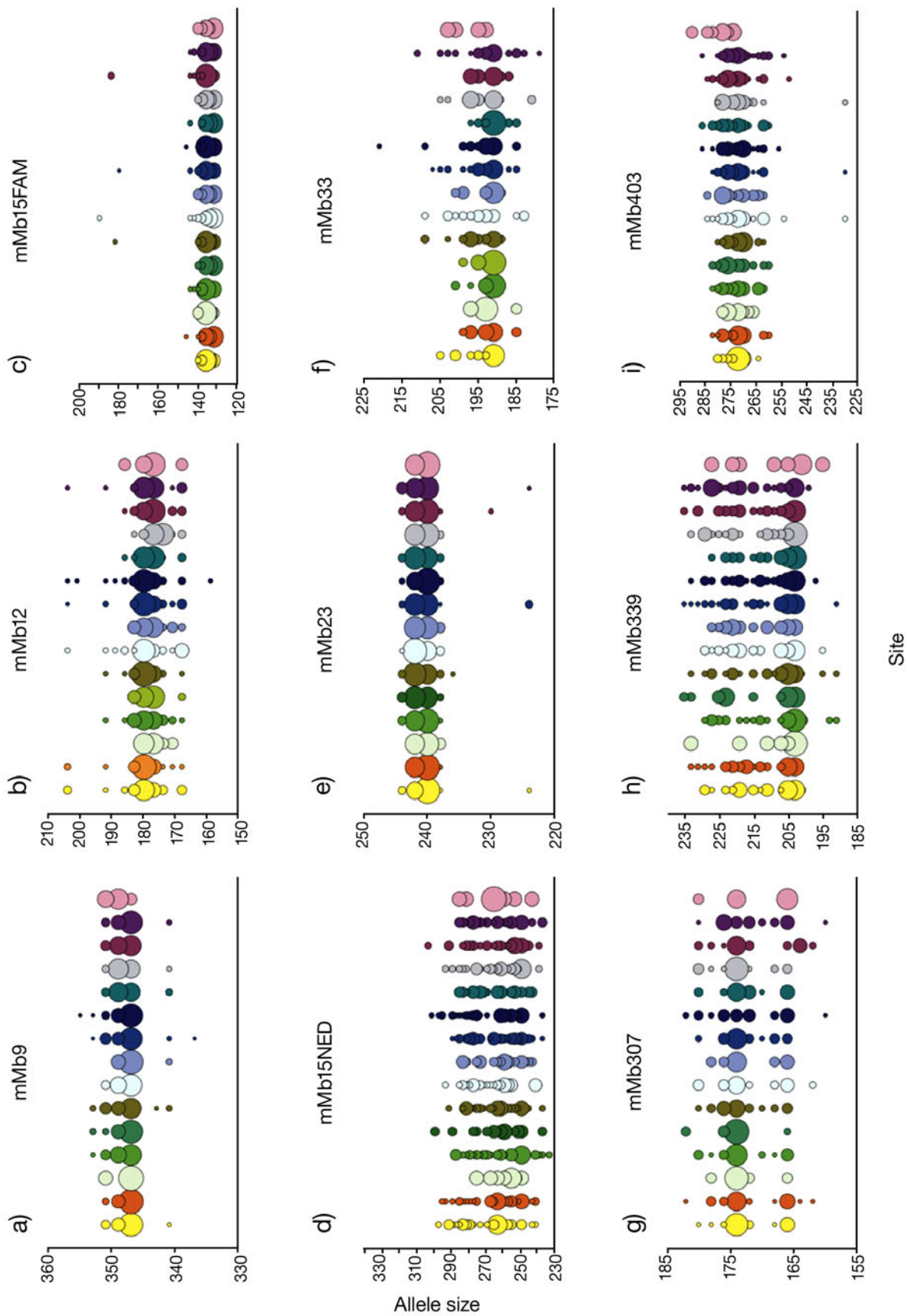


Fig. 3. Allele frequency distribution across populations for 9 microsatellite loci. Each row corresponds to 1 allele, the size of which is indicated on the y-axis. Circle size is proportional to the frequency. Each column (and circle colour) corresponds to 1 sampled locality. They are arranged from left to right in a north to south direction (as indicated in Table 1)

Table 8. Pairwise differentiation indexes between populations based on microsatellite data.  $F_{ST}$  values are below and  $G'_{ST}$  are above the diagonal. Significant  $F_{ST}$  values after the FDR procedure are indicated in bold

	Galway	Ushant	Bilbao	Ribadesella	Cedeira	A Coruña	Malpica	Muxía	Lira	Arousa	Baiona	A Guarda	Lisboa	Cádiz	Telde
Galway															
Ushant	0.0066														
Bilbao	0.0165	0													
Ribadesella	0.0035	0.0069	0.0026												
Cedeira	0.0139	0.0169	0.0303	0											
A Coruña	0.0165	0.0052	0	0.0124	0.0147										
Malpica	<b>0.0314</b>	0.0087	0	0.0214	0.0192	0.0010									
Muxía	0.0052	0	0	0	0	0	0								
Lira	0.0093	0.0049	0	0.0017	0	0	0.0006	0							
Arousa	<b>0.0219</b>	0	0	0.0051	0.0244	0.0049	0.0102	0	0.0052						
Baiona	<b>0.0232</b>	0.0214	0.0403	0	0	0.0155	0.0156	0	0.0073	0.0178					
A Guarda	0.0233	0.0173	0.0105	0.0045	0.0040	0.0126	0.0158	0.0011	0.0127	<b>0.0250</b>	0.0070				
Lisboa	0.0164	0.0224	0	0.0054	0.0022	0.0078	0.0190	0	0.0065	0.0107	0.0019	0			
Cádiz	<b>0.0256</b>	0.0129	0.0071	0.0147	0.0213	0.0089	0.0057	0	0.0060	0.0002	0.0114	<b>0.0280</b>	0.0095		
Telde	0.0648	0.0478	0.0170	0.0317	0.0549	0.0377	0.0149	0.0125	0.0258	0.0280	0.0324	0.0143	0.0088	0.0220	

seems powerful enough to distinguish localized restrictions to gene flow—expected for this coastal species at this geographical scale—when focusing on particular clades.

In this species, gene flow can be achieved by adults and larvae. Spider crab adults perform reproductive migrations, and meeting of individuals from neighbour localities along the migration corridors should connect coastal populations (see Freire et al. 2002, Corgos 2004). Besides, larval exchange has traditionally been proposed as a main factor in homogenizing marine populations. Larval movements have been surveyed in different decapod species and in different habitats, showing different behaviours. Sometimes zoeal stages are already retained inshore, while in other cases they are transported offshore and then the megalopae are return to intertidal areas (Zeng & Naylor 1996, Lee et al. 2006). Nevertheless, Teske et al. (2007) have suggested that the amount of genetic structure within marine biogeographic regions strongly depends on the presence of free-swimming larvae, but not on their particular characteristics, such as duration, retention, or active dispersal. It seems logical to think that spider crab larvae will often be retained inshore, but that, in some cases, they will drift by ocean currents. In this sense, the distribution of the genetic variation could be explained to some extent by the currents in this area. The northern localities, mainly Galway, are under the influence of the North Atlantic Current (flowing to the north) and also under the Biscay Bay Current, characterized by a loop circulation pattern, which would be responsible for the mixing along the French and Cantabrian coasts (from Ushant to Cedeira). In contrast, the western Iberian coast is dominated by the Portugal Current, which shows a north- or south-ward direction depending on the season and which could be an important factor in promoting gene flow in this region. Between the above currents lies the central area of the Galician coast (from A Coruña to Baiona), which is characterized by a deeply indented shape. The southern limit of the Portugal Current is defined by the Azores Current, which moves from north to south and into the Strait of Gibraltar including Cádiz. Finally, the Canary Current flows from north to south along the African coast and through the Canary Islands.

Genetic distinctiveness of populations in these southern regions has been documented for other species within the NE Atlantic, for example, for the euphausiid *Meganyctiphanes norvegica* in Cádiz Bay (Zane et al. 2000, Papetti et al. 2005) and for the cuttlefish *Sepia officinalis* in the Canary Islands (Pérez-Losada et al. 2007). But an absence of structure throughout the whole region has also been described for other species, like the European lobster *Homarus gammarus* (Triantafyllidis et al. 2005) or the Norway lobster *Nephrops norvegicus* (Stamatis et al. 2006).

Our data present some limitations. First, the sample sizes are unequal across localities and small for some of them, especially for Bilbao and Telde, with only 4 individuals. Unquestionably, this could distort the differentiation estimates on which the SAMOVA and Barrier analyses are based. Second, there are important gaps between Lisboa and Cádiz and between Cádiz and the Canary Islands that could lead to ambiguous results in the allocation of barriers. In this sense, it could be impossible to discriminate if there is a real genetic discontinuity between the populations separated by a boundary or if there is actually a cline that is not detected due to insufficient sampling (Manni et al. 2004). Moreover, the accuracy of both methods (SAMOVA and Barrier) decreases in the presence of high gene flow between groups (Dupanloup et al. 2002).

Another factor that could not be completely ruled out is the influence of artificial movements of individuals due to commercial farming near the coast, in the vicinity of the natural habitat of this species. Crabs from different geographical origins are kept together in order to supply local markets, mostly during closed fishing season; furthermore, during water renewals, foreign larvae might be released into the environment. In Galicia, these types of activities began in the 1980s, and, nowadays, the number of farms is considerable (around 140 to 200). However, and due to the lack of available data, it is difficult to estimate their exact contribution to the genetic structure of European populations of *Maja brachydactyla*—an area that deserves further study.

In conclusion, *Maja brachydactyla* populations show high diversity and some genetic structure along the northeastern Atlantic. We have found no genetic evidence for overexploitation, but, before any suggestion can be made for fisheries management, more detailed studies should be carried out.

**Acknowledgements.** We thank all the people who kindly provided samples for this study: J. Cuesta, H. Fernández-Palacios, D. Latrouite, A. Morán, I. Mosqueira, A. Santos, O. Tully and P. Verísimo. In addition, the fishermen's associations from A Guarda, Baiona and Malpica were very helpful. Thanks to P. Tarroso for helping with coastline distances. We are also grateful to P. Alvarinho and N. Santamaría for laboratory assistance. This work was partially funded by a grant from the Spanish Ministerio de Ciencia y Tecnología (VEM203-2047). G.S. was supported by a PhD fellowship from the Xunta de Galicia.

#### LITERATURE CITED

- Arenas M, Posada D (2007) Recodon: coalescent simulation of coding DNA sequences with recombination, migration and demography. *BMC Bioinformatics* 8:458
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24: 621–631
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9: 1657–1659
- Corgos A (2004) Estrategia vital, estructura espacial y dinámica poblacional de la centolla, *Maja squinado* (Decapoda: Majidae). PhD thesis, University of A Coruña
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 11:2571–2581
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5:435–445
- Estoup A, Cornuet JM (1999) Microsatellite evolution: inferences from population data. In: Goldstein DB, Schlötterer C (eds) *Microsatellites, evolution and applications*. Oxford University Press, New York, p 49–65
- Estoup A, Largiadere CR, Perrot E, Chourrout D (1996) Rapid one-tube extraction for a reliable PCR detection of fish polymorphic markers and transgenes. *Mol Mar Biol Biotechnol* 5:295–298
- Excoffier L, Laval G, Schneider S (2005) Arlequin Ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3:294–299
- Freire J, Bernardez C, Corgos A, Fernandez L, Gonzalez-Gurriaran E, Sampedro MP, Verísimo P (2002) Management strategies for sustainable invertebrate fisheries in coastal ecosystems of Galicia (NW Spain). *Aquat Ecol* 36:41–50
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Goudet J (2001)  $F_{stat}$ , a program to estimate and test gene diversities and fixation indices (Version 2.9.3). Available from [www2.unil.ch/popgen/softwares/fstat.htm](http://www2.unil.ch/popgen/softwares/fstat.htm)
- Hasegawa M, Kishino K, Yano T (1985) Dating the human–ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59:1633–1638
- Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature* 406:882–885
- Ketmaier V, Argano R, Caccone A (2003) Phylogeography and molecular rates of subterranean aquatic stenasellid isopods with a peri-Tyrrhenian distribution. *Mol Ecol* 12: 547–555
- Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–770
- Lee JT, Coleman RA, Jones MB (2006) Population dynamics and growth of juveniles of the velvet swimming crab *Necora puber* (Decapoda: Portunidae). *Mar Biol* 148:609–619
- Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Hum Biol* 76:173–190
- Mathews L (2007) Evidence for restricted gene flow over

- small spatial scales in a marine snapping shrimp *Alpheus angulosus*. *Mar Biol* 152:645–655
- Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60:2399–2402
- Narum S (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conserv Genet* 7:783–787
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273
- Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1991) The simple fool's guide to PCR, Version 2.0. Department of Zoology, University of Hawaii, Honolulu
- Panchal M, Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution* 61:1466–1480
- Papetti C, Zane L, Bortolotto E, Bucklin A, Patarnello T (2005) Genetic differentiation and local temporal stability of population structure in the euphausiid *Meganycitiphanes norvegica*. *Mar Ecol Prog Ser* 289:225–235
- Pérez-Losada M, Nolte MJ, Crandall KA, Shaw PW (2007) Testing hypotheses of population structuring in the Northeast Atlantic Ocean and Mediterranean Sea using the common cuttlefish *Sepia officinalis*. *Mol Ecol* 16:2667–2679
- Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylomatophora): fragmentation, corridor migration, and secondary contact. *Evolution* 56:1776–1788
- Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst Biol* 53:793–808
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol Ecol* 9:487–488
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092–2100
- Raymond M, Rousset F (1995) Genepop (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249
- Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* 49:608–616
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Shaw PW, Pierce GJ, Boyle PR (1999) Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Mol Ecol* 8:407–417
- Sotelo G, Moran P, Posada D (2007) Identification and characterization of microsatellite loci in the spiny spider crab *Maja brachydactyla*. *Conserv Genet* 8:245–247
- Sotelo G, Moran P, Posada D (2008) Genetic identification of the northeastern Atlantic spiny spider crab as *Maja brachydactyla* Balss, 1922. *J Crustac Biol* 28:76–81
- Stamatis C, Triantafyllidis A, Moutou KA, Mamuris Z (2006) Allozymic variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. *ICES J Mar Sci* 63:875–882
- Sunnucks P, Wilson ACC, Beheregaray LB, Zenger K, French J, Taylor AC (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol Ecol* 9:1699–1710
- Swofford DL (2003) PAUP\* phylogenetic analysis using parsimony (\*and other methods), Version 4. Sinauer Associates, Sunderland, MA
- Tajima F (1996) The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* 143:1457–1465
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399
- Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) ONE-SAMP: a program to estimate effective population size using approximate Bayesian computation. *Mol Ecol Notes* 8:299–301
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526
- Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Mol Ecol* 13:789–809
- Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140:767–782
- Teske P, Papadopoulos I, Zardi G, McQuaid C, Edkins M, Griffiths C, Barker N (2007) Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: planktonic, abbreviated and direct development. *Mar Biol* 152:697–711
- Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: 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
- Thorpe JP, Sole-Cava AM, Watts PC (2000) Exploited marine invertebrates: genetics and fisheries. *Hydrobiologia* 420:165–184
- Triantafyllidis A, Apostolidis AP, Katsares V, Kelly E and others (2005) Mitochondrial DNA variation in the European lobster (*Homarus gammarus*) throughout the range. *Mar Biol* 146:223–235
- Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Mol Biol Evol* 22:925–933
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538
- Yue GH, David L, Orban L (2007) Mutation rate and pattern of microsatellites in common carp (*Cyprinus carpio* L.). *Genetica* 129:329–331
- Zane L, Ostellari L, Maccatrozzo L, Bargelloni L, Cuzin-Roudy J, Buchholz F, Patarnello T (2000) Genetic differentiation in a pelagic crustacean (*Meganycitiphanes norvegica*: Euphausiacea) from the North East Atlantic and the Mediterranean Sea. *Mar Biol* 136:191–199
- Zeng CS, Naylor E (1996) Occurrence in coastal waters and endogenous tidal swimming rhythms of late megalopae of the shore crab *Carcinus maenas*: implications for onshore recruitment. *Mar Ecol Prog Ser* 136:69–79

# Genetic variation of the spiny spider crab *Maja brachydactyla* in the northeastern Atlantic

Graciela Sotelo<sup>1,\*</sup>, Paloma Morán<sup>1</sup>, Luis Fernández<sup>2</sup>, David Posada<sup>1</sup>

<sup>1</sup>Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Campus Lagoas-Marcosende, Universidad de Vigo, 36310 Vigo, Spain

<sup>2</sup>Departamento de Biología Animal, Biología Vegetal y Ecología, Campus A Zapateira, Universidad de A Coruña, 15071 A Coruña, Spain

*Marine Ecology Progress Series 362:211–223 (2008)*

\*Email: gsotelo@uvigo.es

## Appendix 1

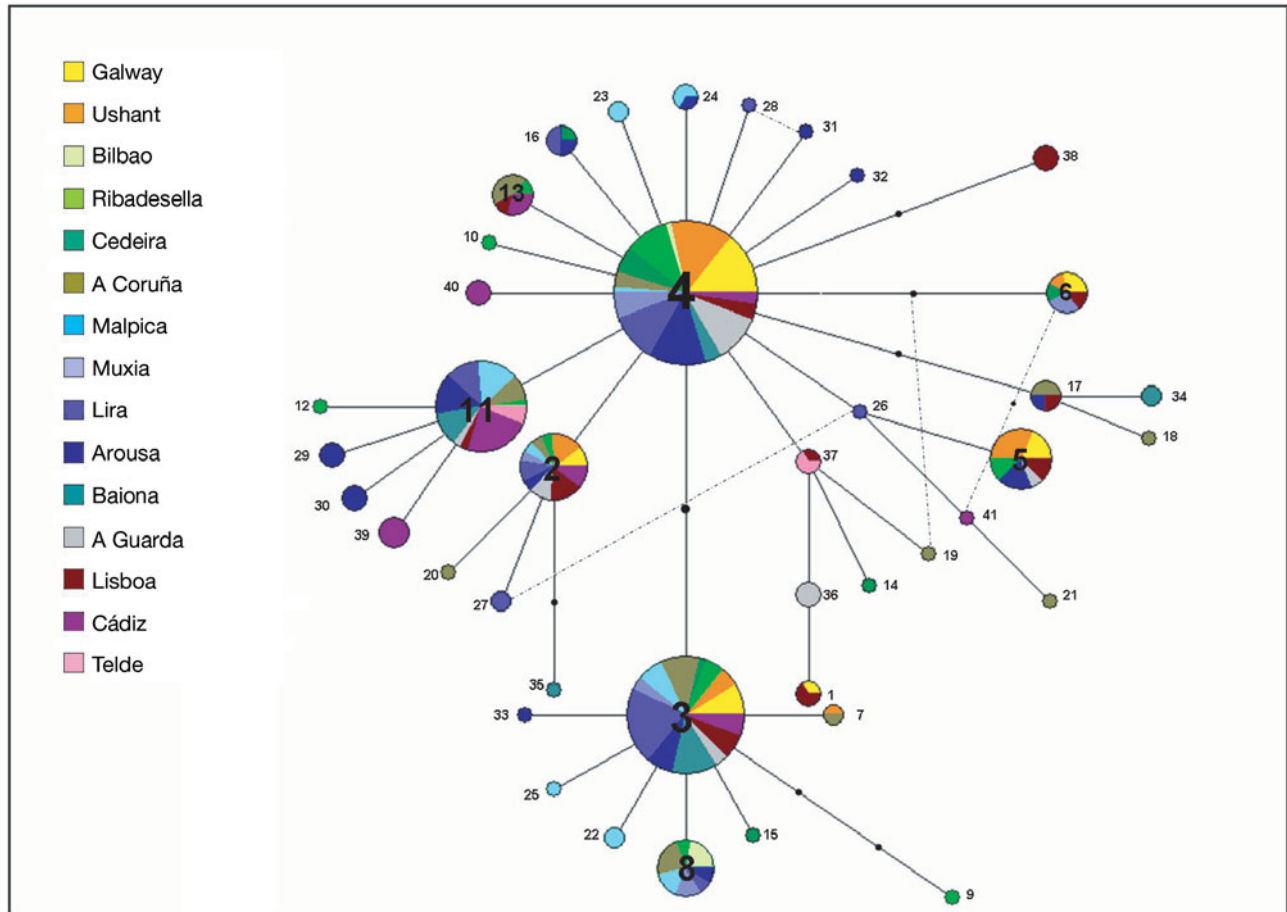


Fig. A1. Haplotype network for COI. Circles represent haplotypes with sizes proportional to their frequency and colours representing the localities where they were sampled. Solid lines connecting haplotypes represent mutational steps, and black dots are missing or unsampled haplotypes. Dashed lines connecting haplotypes represent reticulations resolved before constructing the cladogram

Table A1. Effective population sizes estimates (considered unreliable) from LAMARC and ONeSAMP. n is the sample size

Locality	n	COI LAMARC	Microsatellites ONeSAMP
Galway	25	413 000	127
Ushant	25	322 500	55
Bilbao	4	n.a	6
Ribadesella	20	3 827 000	49
Cedeira	10	33 507 500	19
A Coruña	26	1 223 500	42
Malpica	20	514 500	85
Muxía	10	182 500	16
Lira	36	509 500	353
Arousa	37	3 725 500	126
Baiona	17	303 000	33
A Guarda	18	1 902 500	67
Lisboa	22	1 589 000	106
Cádiz	25	411 500	139
Telde	4	n.a.	5

n.a.: not available

Table A2. Pairwise distances between populations based on COI data. TrN values are below the diagonal and  $D_A$ , above the diagonal. Significant values after FDR procedure are indicated in bold

	Galway	Ushant	Bilbao	Ribadesella	Cedeira	A Coruña	Malpica	Muxía	Lira	Arousa	Baiona	A Guarda	Lisboa	Cádiz	Telde
Galway															
Ushant	0														
Bilbao	<b>0.3526</b>	<b>0.4037</b>													
Ribadesella	0	0	0.9967												
Cedeira	0	0.0112	0.3354	0											
A Coruña	0.0419	0.0750	0.0780	0.0065	0										
Malpica	0.0808	<b>0.1215</b>	0.1361	0.0249	0.0524	0									
Muxía	0.0247	0.0714	0.1491	0	0.0078	0	0								
Lira	0.0146	0.0360	0.2442	0	0.0149	0.0107	0.0101	0							
Arousa	0.0349	0.0556	<b>0.3297</b>	0.0108	0.0106	0.0537	0.0593	0.0597	0.0490						
Baiona	0.0547	0.0977	0.1392	0.0088	0.0236	0	0	0	0.0009	0.0350					
A Guarda	0	0.0215	<b>0.4427</b>	0.0119	0	0.0737	0.1157	0.0872	0.0620	0.0374	0.0969	0			
Lisboa	0	0.0051	<b>0.2444</b>	0	0	0.0326	0.0689	0.0155	0.0312	0.0364	0.0441	0			
Cádiz	<b>0.1175</b>	0.1337	<b>0.4226</b>	0.0730	0.1013	<b>0.1128</b>	<b>0.1121</b>	<b>0.1594</b>	<b>0.1165</b>	0.0205	0.0955	0.1029	<b>0.0886</b>		
Telde	0.1622	0.2026	0.5654	0.1163	0.1161	0.1284	0.1592	0.2798	0.1897	0.0422	0.1387	0.0891	0.0403	0.0340	

Table A3. Pairwise differentiation indexes between populations based on microsatellite data.  $F_{ST}$  values from the corrected dataset applying the ENA method are below the diagonal and  $R_{ST}$  from the original dataset, above the diagonal. There are no significance estimates for the first index and none of the  $R_{ST}$  values are significant after FDR procedure

	Galway	Ushant	Bilbao	Ribadesella	Cedeira	A Coruña	Malpica	Muxía	Lira	Arousa	Baiona	A Guarda	Lisboa	Cádiz	Telde
Galway		0.0420	0.0788	0.0589	0.0358	0.0135	0.0092	0.0603	0.0253	0.0001	0.0261	0.0431	0.0396	0.0423	0.0387
Ushant	0.0075		0	0	0	0	0	0	0	0.0011	0.0079	0	0.0052	0.0013	0.0365
Bilbao	0.0026	0.0016		0	0	0	0	0	0	0	0	0	0	0.0082	0.0103
Ribadesella	0.0085	0.0071	0		0	0	0	0	0	0.0053	0	0	0	0.0444	0.0078
Cedeira	0.0188	0.0079	0	0		0	0	0	0	0	0	0	0	0	0
A Coruña	0.0172	0.0090	0	0.0033	0.0072		0	0.0056	0	0	0	0	0	0.0336	0.0097
Malpica	0.0291	0.0097	0.0050	0.0115	0.0078	0.0021		0	0	0	0	0	0	0	0.0278
Muxía	0.0176	0	0	0.0008	0	0.0001	0		0	0.0082	0.0130	0	0	0	0.0314
Lira	0.0167	0.0069	0	0	0	0	0	0		0	0	0	0	0.0217	0.0049
Arousa	0.0260	0.0066	0.0024	0.0025	0.0174	0.0090	0.0140	0	0.0055	0	0	0	0.0005	0.0225	0.0075
Baiona	0.0298	0.0220	0.0099	0.0018	0.0006	0.0094	0.0087	0	0.0081	0.0223	0	0	0.0212	0.0449	0
A Guarda	0.0265	0.0161	0.0051	0	0.0007	0.0117	0.0104	0	0.0116	0.0163	0	0	0	0.0012	0.0247
Lisboa	0.0240	0.0261	0	0	0.0032	0.0118	0.0202	0	0.0083	0.0100	0.0040	0.0007	0	0	0
Cádiz	0.0284	0.0128	0.0091	0.0082	0.0114	0.0095	0.0060	0	0.0043	0.0038	0.0106	0.0160	0.0066	0	0.0613
Telde	0.0525	0.0404	0.0114	0.0157	0.0251	0.0373	0.0207	0	0.0247	0.0336	0.0086	0.0091	0.0071	0.0196	0

Table A4. Pairwise chord distances between populations based on microsatellite data. Distance values from the original dataset are below the diagonal and values from the corrected dataset applying the INA method, above the diagonal

	Galway	Ushant	Bilbao	Ribadesella	Cedeira	A Coruña	Malpica	Muxía	Lira	Arousa	Baiona	A Guarda	Lisboa	Cádiz	Telde
Galway		0.0280	0.0660	0.0310	0.0470	0.0250	0.0350	0.0410	0.0330	0.0340	0.0340	0.0350	0.0350	0.0340	0.0760
Ushant	0.0300		0.0520	0.0270	0.0340	0.0270	0.0300	0.0340	0.0260	0.0220	0.0350	0.0290	0.0320	0.0300	0.0790
Bilbao	0.0670	0.0460		0.0640	0.0790	0.0590	0.0660	0.0690	0.0680	0.0610	0.0610	0.0640	0.0630	0.0720	0.1000
Ribadesella	0.0290	0.0320	0.0660		0.0390	0.0270	0.0320	0.0370	0.0280	0.0250	0.0280	0.0310	0.0270	0.0290	0.0770
Cedeira	0.0490	0.0420	0.0830	0.0460		0.0390	0.0450	0.0370	0.0340	0.0400	0.0450	0.0380	0.0380	0.0410	0.0860
A Coruña	0.0270	0.0280	0.0560	0.0310	0.0450		0.0270	0.0320	0.0200	0.0240	0.0320	0.0230	0.0300	0.0290	0.0810
Malpica	0.0370	0.0320	0.0640	0.0370	0.0490	0.0300		0.0390	0.0230	0.0250	0.0360	0.0280	0.0330	0.0260	0.0710
Muxía	0.0400	0.0360	0.0680	0.0380	0.0450	0.0350	0.0410		0.0390	0.0390	0.0430	0.0360	0.0370	0.0390	0.0760
Lira	0.0270	0.0270	0.0640	0.0280	0.0350	0.0220	0.0260	0.0390		0.0240	0.0310	0.0290	0.0290	0.0250	0.0790
Arousa	0.0320	0.0220	0.0540	0.0290	0.0440	0.0250	0.0290	0.0400	0.0250		0.0340	0.0270	0.0320	0.0270	0.0790
Baiona	0.0320	0.0370	0.0690	0.0310	0.0460	0.0330	0.0360	0.0440	0.0270	0.0310		0.0390	0.0260	0.0300	0.0700
A Guarda	0.0330	0.0330	0.0620	0.0390	0.0430	0.0280	0.0340	0.0420	0.0320	0.0340	0.0380		0.0290	0.0350	0.0740
Lisboa	0.0330	0.0410	0.0670	0.0380	0.0440	0.0320	0.0360	0.0430	0.0310	0.0370	0.0300	0.0280		0.0310	0.0710
Cádiz	0.0320	0.0360	0.0700	0.0350	0.0490	0.0330	0.0290	0.0450	0.0250	0.0310	0.0300	0.0420	0.0380		0.0760
Telde	0.0840	0.0900	0.1090	0.0850	0.0970	0.0870	0.0740	0.0850	0.0780	0.0840	0.0830	0.0790	0.0830	0.0810	0