Original Article

Lineage splitting, secondary contacts and genetic admixture of a widely distributed marine invertebrate

Pérez-Portela R [1, 2] *, Rius M [3,4], Villamor A[5]

[1] Center for Advanced Studies of Blanes (CEAB - CSIC), Blanes (Girona), Catalunya, Spain

[2] Rosenstiel School of Marine & Atmospheric Science (RSMAS), University of Miami, Miami, USA

[3] Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, European Way, SO14 3ZH, United Kingdom

[4] Department of Zoology, University of Johannesburg, Auckland Park, 2006, Johannesburg, South Africa.

[5] Laboratorio di Ecologia Sperimentale, Università degli Studi di Bologna, Ravenna, Italy

* Corresponding author:

Rosenstiel School of Marine & Atmospheric Science (RSMAS), University of Miami

4600 Rickenbacker Cswy

Miami, FL 33149

perezportela@gmail.com

nº of words: 6931

Estimate number of pages for tables and figures: 6

Short running head: “Lineage splitting and admixture in a marine invertebrate”

ABSTRACT

Aim: We explore the genetic structure of the widespread starfish, *Marthasterias* spp, in order to: a) identify historical causes of genetic divergence, b) test the effect of past climatic events

on populations' demography, and c) explore main barriers to gene flow.

Location: North and south-east Atlantic Ocean and Mediterranean Sea.

Methods: We amplified and sequenced three mitochondrial genes and one nuclear intron, and genotyped five nuclear microsatellite loci from 337 specimens. We reconstructed the phylogeny and phylogeography of *Marthasterias* spp using the obtained DNA sequences, and used the microsatellite loci to explore major genetic discontinuities along the European coast.

Results: Our results suggested the existence of two allopatric species, one in the Northern Hemisphere (*M. glacialis*) and another in the Southern Hemisphere (*Marthasterias* sp.). This allopatric split could be attributed to recent changes in oceanographic circulation of the eastern Atlantic Ocean that may have acted as barrier to gene flow. Mitochondrial divergence between European lineages could be attributed to potential vicariance during Pleistocene glacial periods, but was not supported by nuclear markers, which may be indicative of recent genetic admixture. Secondary contact after glacial periods and gene flow across the Gibraltar Strait may explain this pattern. Genetic structure of *M. glacialis* based on nuclear markers did not show much divergence among geographical areas although most populations were significantly differentiated.

Main conclusions: The phylogeography of the widely distributed genus *Marthasterias* has most likely been shaped by recent changes in climate and oceanographic patterns. Major changes of ocean current patterns initially resulted in splitting between Northern and Southern Hemisphere lineages. Subsequently, glacial periods most likely enhanced vicariance of European lineages, followed by a postglacial expansion facilitating secondary contacts and genetic admixture.

Keywords: DNA, Echinodermata, gene flow, marine barriers, microsatellites, phylogeography, speciation, sea stars

INTRODUCTION

Molecular tools enable researchers to conduct phylogenetic analyses of geographically contextualized genetic data. This has dramatically enhanced our understanding of the geographic ordination of genetic lineages and helped inferring evolutionary history of many taxa (Avise, 2000). Phylogeography is a discipline that integrates historical, phylogenetic and geographical data to test biogeographic hypothesis, explore speciation processes, solve taxonomic conflicts and trace out strategies for conservation of biodiversity (Avise, 2000; Beheregaray, 2008). One of the main challenges of phylogeography is to conduct comprehensive analyses of widespread species, mainly because information is often limited or non-existing in many parts of the species range (Beheregaray, 2008; Hickerson *et al.,* 2010). Additionally, the complexity of phylogeographic patterns requires coupling appropriated sampling with the study of multilocus data (Beheregaray, 2008; Debiasse *et al.,* 2014). Nevertheless, most phylogeographic studies of widespread species include a restricted geographical coverage and/or use few genetic markers (e.g. Wangensteen *et al.* 2012; Borrero-Pérez *al*., 2011). Thus, there is a need for comprehensive studies that couple large datasets of multiple genetic markers with a comprehensive geographic sampling to provide a broad view of the interplay among evolutionary histories, past biogeographic discontinuities, and present gene flow patterns.

In the marine environment, gametes and larvae are the main dispersal drivers of benthic organisms. Several studies have demonstrated that there is not a direct relationship between planktonic duration and levels of genetic divergence among populations (Siegel *et al.*, 2008; Weersing & Toonen 2009; Hunter & Halanych 2010, among others). Together with the larval dispersal potential, water currents, substrate availability and recruitment success determine the genetic structure observed in benthic species (Siegel *et al.,* 2008).

Oceanographic circulation profoundly affects species distributions and the genetic structure of populations (Siegel *et al.,* 2008). Oceanic current regimes at several spatial scales are capable to act as physical barriers defining biogeographic breaks such as the Almeria-Oran Front (AOF) that splits the Atlantic and Mediterranean biogeographic provinces (Patarnello *et al.,* 2007; among other examples). In addition, past climatic events, mainly the Quaternary climatic fluctuations have marked the biogeography of marine species due to repeated contractions and expansions of the ice sheets (e.g. Patarnello *et al*., 2007; Maggs *et al.,* 2008). For example, Pleistocene glaciations caused sea level drops of up to 150 meters, driving dramatic range shifts along the European coastline, and modifying circulation patterns across most marine corridors of the European shelf (Patarnello *et al.,* 2007; Maggs *et al.,* 2008). As a consequence, most of the current genetic patterns of marine coastal organisms have been affected by past vicariance events, due to periodical closing of corridors that prevented larval interchange and migration (Wilson & Veraguth 2010).

Recent molecular studies of widely distributed echinoderms along the eastern Atlantic and the Mediterranean Sea show the existence of contrasting phylogeographic patterns, and different effects of the Atlanto-Mediterranean transition on gene flow, even among species with similar life history traits (*e.g*. Zulliger *et al*. 2009; Borrero-Pérez *al*. 2011; Wangensteen *et al.* 2012; Penant *et al*. 2013, Garcia-Cisneros *et al*. 2016, among others). For instance, the most common sea urchins in this area, *Paracentrotus lividus* and *Arbacia lixula*, show gene flow restriction across the AOF. Whereas *A. lixula* displays no genetic differentiation over distances of thousands of kilometres within basins (Wangensteen *et al.* 2012), *P. lividus* presents genetic differentiation of Mediterranean populations following a pattern of isolation by distance despite having a long-lived planktotrophic larva (Penant *et al.* 2013). In contrast, the AOF seems to have little influence on genetic connectivity of many other species. The starfish *Astropecten auranciacus* also displayed significant genetic differentiation between

Atlantic and Mediterranean populations, although in this species that divergence was attributed to a pattern of isolation by geographical distance more than the disruptive effect of the AOF (Zulliger *et al*., 2009). In the widespread holothurian, *Holothuria mammata*, a weak differentiation was found between Macaronesian, South Portuguese and Western Mediterranean populations, with a significant genetic break in the Aegean Sea (Borrero-Pérez *al*., 2011). Another example is the brittlestar *Ophiothrix* sp., which shows genetic homogeneity across the same region. This could be explained by its large population size, connectivity between distant populations and a recent demographic expansion (Pérez-Portela *al*., 2013; Taboada and Pérez-Portela 2016).

Interestingly, complex patterns of cryptic speciation have been also observed across the Atlantic-Mediterranean area in several echinoderm taxa (e.g. Boissin *et al*., 2011; Pérez-Portela *et al*., 2013; Weber *et al.* 2015; Taboada & Pérez-Portela 2016, among others). For instance, up to six lineages attributable to different cryptic species of *Ophioderma* were discovered, some of them in sympatry. A complex mixture of processes including vicariance, demographic expansions and adaptive divergence may have driven the diversification of this species complex during the Pleistocene (Boissin *et al*., 2011). Another echinoderm species that may contain cryptic species is the broadcast spawner starfish *Marthasterias glacialis* (Linnaeus, 1758). A previous genetic study of *M. glacialis* found two sympatric mitochondrial lineages across south-western Europe (Pérez-Portela *et al.*, 2010). However, this study was based on only one mitochondrial gene, so could not attribute these divergent mitochondrial lineages to isolated evolutionary units. *Marthasterias glacialis* has a planktotrophic larva that remains in the water column for several months before settlement and is widely distributed across contrasting climatic and biogeographic areas, from Norway to the Macaronesia Islands, the Mediterranean Sea, and South Africa (Mortensen, 1927; Barker & Nichols, 1983), although records of its presence in the rest of Africa’s Atlantic coast are lacking. Taxonomic

discrepancies between descriptions of *M. glacialis* in the Northern and Southern Hemispheres may actually hide the existence of a species complex (Mortensen, 1933; Clark, 1974). Its discontinuous distribution together with its large morphological variability have generated a large taxonomic debate, and several forms and even species, have been separated and synonymised several times (Mortensen, 1927, 1933; Clark, 1974). Nevertheless, to date *M. glacialis* is the only formally described and accepted species of the genus *Marthasterias* (World Register of Marine Species, http://www.marinespecies.org/aphia.php?p=taxdetails&id=123224).

In this work we study *M. glacialis* to clarify the phylogenetic relationships among marine regions, understand the historical causes of genetic divergence and to explore the existence of cryptic evolutionary units. We then explore barriers to gene flow and past events (e.g. climatic fluctuations) to identify the relative effect of both long- and short-term processes on genetic patterns across the entire species range. Finally, we study both phylogeographic and gene flow patterns across major biogeographic disjunctions in the North Atlantic and Mediterranean area to identify potential drivers of genetic breaks.

MATERIALS AND METHODS

Sample collection and genetic markers used

A total of 337 specimens of *Marthasterias glacialis* were collected from 21 sites in the North Eastern Atlantic, Western and Central Mediterranean, Cape Verde and South Africa (see Table 1, Fig. 1). Despite intensive fieldwork across the Eastern Mediterranean (Crete and Rhodes in Greece; Sicily in Italy), the Canary Islands (El Hierro and Tenerife) and Cape Verde (Sal), only a few samples could be included in this study due to the low density of individuals observed in these sites. Total DNA was extracted following the protocols

153 described in Pérez-Portela *et al.* (2010).

154 Fragments of three mitochondrial genes, the Cytochrome c oxidase subunit I, the 16S
155 rRNA, and a fragment including the putative control region (hereafter COI, 16S and CR), and
156 the nuclear first internal transcribed spacer (hereafter ITS1) were amplified and sequenced for
157 a subset of samples (see Tables 1 and 2). Details about fragment amplification, sequencing
158 process, analyses of saturation and selection on the DNA fragments are included as
159 Supporting Information (Appendix S1). The new DNA sequences obtained in this study have
160 been deposited in Genbank (Accession numbers "*pending*"). Within- and among- lineage
161 divergences were explored using samples from several geographical regions.

162 In order to detect discrepancies between mito-nuclear markers and better understand
163 the phylogeography of the species, we used sequences of several mitochondrial and nuclear
164 fragments together with microsatellite loci. Five nuclear microsatellite loci were isolated (see
165 details of microsatellite isolation, genotyping methods and variation details in Appendix S1)
166 and used for genotyping individuals from 12 European sites for population genetic analyses
167 (see Table 1).

168

169 **Phylogeny and phylogeography**

170 In order to determine the evolutionary relationships, origin and distribution of
171 different lineages in *Marthasterias*, several mitochondrial fragments with different mutation
172 rates, and one nuclear intron were used to obtain a robust phylogenetic reconstruction and
173 more support than in previous molecular analyses of the species.

174 Number of haplotypes, haplotype diversity (*Hd*) and nucleotide diversity (π) were
175 calculated per population and locus. The complete datasets of COI, 16S, CR and ITS1 were
176 separately used to construct unrooted networks with the software Network 4.5.1.0
177 (http://www.fluxus-engineering.com/sharenet.htm), which assumes the median-joining

network in the absence of recombination.

Phylogenetic trees were initially reconstructed by Bayesian inference analyses (BI) for every mitochondrial marker (COI, 16S and CR) and the nuclear ITS1 separately. Mitochondrial markers were then concatenated and phylogenetic trees performed under the BI and maximum likelihood (ML) criterion, implementing the evolution models estimated on every partition. Lineage divergence between clades was calculated by a Kimura 2- parameters model (K2P), for comparison with other studies on other species, in Mega 5.2.2 (http://www.megasoftware.net/). BI reconstruction and divergence times were estimated with the software Beast 2.1.2 (http://beast.bio.ed.ac.uk/) applying COI divergence rates between 2.2% and 2.8% Ma (see details about the phylogenetic analyses, divergence times and mutation rates in Appendix S1).

Genetic structure in *M. glacialis* along the European coast

We used the mitochondrial gene COI and five nuclear microsatellite loci to obtain a detailed picture of the genetic structure and barriers to gene flow in *M. glacialis* from twelve European populations (Table 1). We obtained genetic differentiation measures across four main European marine ecoregions (Celtic Sea, South European Atlantic Shelf, Azores-Canaries-Madeira, and Western Mediterranean) (Spalding *et al.*, 2007), which are known to be divided by major oceanographic breaks, among populations within ecoregions, and the potential clustering of mitochondrial lineages within the Mediterranean Sea (see Result section for a full explanation). Only localities including more than 13 specimens were considered, with the exception of Madeira, which was included because its geographical relevance. The population of Scilly Island (GBR) was considered for analyses with COI but it had to be excluded for the microsatellite analyses because of the difficulties to amplify nuclear fragments.

The number of haplotypes (*h*), private haplotypes (*Np*), haplotype diversity (*Hd*), and nucleotide diversity (π) were calculated for the COI. Haplotype richness (*Hk*) was also calculated after rarefaction (to the minimum sample size) with the software Contrib 1.02 (Petit *et al.*, 1998) per population. From the microsatellite loci, descriptors as allelic richness (*Ra*), private alleles (*Np*), observed and expected heterozygosity (H_o and H_e), fixation index (F_{IS}), and fit to Hardy Weinberg equilibrium (HWE), were computed in Fstat 2.9.3.2 (http://www2.unil.ch/popgen/softwares/fstat.htm).

A Bayesian clustering analysis, performed with the software Structure 2.3 (pritchardlab.stanford.edu/structure.html) was used to infer the optimal number of major homogeneous genetic units (*K*). An initial run with the whole dataset was performed with a *K* from 1 to 14. A second run with a *K* from 1 to 6 was computed including only the Mediterranean populations, Costa Brava and Ischia, in order to check for potential clustering of individuals belonging to divergent mitochondrial lineages. Ten independent replicates with one million Markov chain Monte Carlo (MCMC) analyses were performed for each run, and 100,000 burn-in period under the "admixture model" implemented. The best value of *K* was selected comparing the rate of change in the likelihood of *K*, using the ad hoc statistic ΔK in Structure Harvester (Evanno *et al.*, 2005).

Population structure and population grouping within marine ecoregions was investigated with analyses of molecular variance (AMOVA) of COI and microsatellite loci, using pairwise distances (Φ_{ST}) for COI, and F_{ST} for microsatellites, and 16,000 permutations in Arlequin 3.0 (Excoffier *et al.*, 2005). Populations were grouped according to pre-defined marine ecoregions, considering genetic subdivisions caused by 'biogeographical breaks'. For further analyses of population genetic structure within ecoregions, genetic distances (Φ_{ST} or F_{ST}) were also assessed with 16,000 permutations. For the microsatellite loci, F_{ST} were calculated grouping individuals of the same mitochondrial lineages found in sympatry in the

Mediterranean, independently of the original locality within this basin. This strategy was used to explicitly test whether the mitochondrial lineages found (see Results section) correspond to isolated genetic units. The frequency of null alleles for microsatellite loci was estimated using the expectation maximization algorithm with the software FreeNA (http://www1.montpellier.inra.fr/CBGP/software/FreeNA/). In most cases, the corrections only affected the third decimal place in the F_{ST} values and they did not affect their significance. Consequently, the effect of presence of null alleles was not considered in further analyses. The fit to a pattern of isolation by geographical distance across the north-east Atlantic and Mediterranean area, including the two mitochondrial lineages, was tested in Arlequin using the Mantel test procedure with 16,000 permutations. Φ_{ST} matrices of COI and F_{ST} of microsatellites were represented in a multidimensional scaling graph performed with Systat 9. A discriminant analysis of principal components (DAPC) was also applied for the microsatellite dataset to assess genetic structure using populations as groups with the "adegenet" package (Jombart *et al.* 2010) in R 3.2.1 (R Core Team, 2014).

The demographic history of *Marthasterias* in the Northern Hemisphere was inferred from the COI. We *a priori* tested demographic processes and genetic differences across marine ecoregions, as the smaller homogeneous biogeographic units, and lineages using coalescent-based Bayesian skyline plots (BSP) in BEAST (see details in Appendix S1).

To explore historical migration across the Gibraltar Strait analyses of gene flow were performed between the South European Atlantic Shelf and the Western Mediterranean regions using Lamarc 2.1.9 (http://evolution.genetics.washington.edu/lamarc/index.html) with COI. This Bayesian approach based on coalescence is computational highly demanding, consequently we clustered all populations from the South European Atlantic Shelf in one group and the Western Mediterranean in another. They could not be properly applied for microsatellites due to the difficulties to reach coalescence. Our search strategy in Lamarc

253 2.1.9 is detailed in Appendix S1.

254

255 **RESULTS**

256

257 **Phylogeny and phylogeography**

258 A total of 261 new DNA sequences were generated in this study (see Table 1), and
259 283 specimens were genotyped using microsatellites. A subset from the total collection was
260 selected, and 55 sequences of the 16S, 55 of the CR, and 39 of ITS1 were obtained from
261 specimens of different marine ecoregions. Genetic diversity descriptors obtained for every
262 fragment are summarized in Table 2.

263 Mitochondrial haplotype networks and phylogenetic trees always showed three
264 divergent lineages, although the relationship between lineages was different among markers
265 (Appendices S2 and S3). Lineage I included individuals from the Celtic Sea, South European
266 Atlantic Shelf, Azores-Canaries-Madeira and Western Mediterranean ecoregions; Lineage II
267 contained only Western Mediterranean individuals; and Lineage III formed by the Cape
268 Verde and the Agulhas Bank (South Africa) samples. The three lineages did not share any
269 haplotype (see Fig. 2 and Appendix S3) for any marker. Two lineages appeared in sympatry
270 in the Mediterranean (see Fig. 2 and Appendix S2) as previously described (Pérez-Portela *et*
271 *al.*, 2010) (Table 2 and Fig. 2). Within the mitochondrial Lineage I four haplotypes (H_1,
272 H_5, H_8, and H_22) appeared in all North Atlantic and Mediterranean sites, whereas other
273 four frequent haplotypes (H_10, H_13, H_19, H_24) were absent in Azores and Madeira.

274 The nuclear ITS1 network recovered two monophyletic clades, Clade A and Clade B.
275 Clade B grouped all alleles found in South Africa, and matched with the mitochondrial
276 Lineage III (Figs. 2 and Appendix S2). Clade A included sequences from individuals of the
277 mitochondrial lineages I and II, with no evidences of further splitting. The concatenated

phylogenetic trees of mitochondrial markers based on both BI and ML were congruent between them and corroborated this divergence and geographical distribution (Appendix S2).

Divergence between genetic lineages varied when different markers were compared (Table 3). Results showed an earlier split between Lineage III and an ancestor of Lineage I and Lineage II according to the mitochondrial concatenated tree (Appendix S2), but this relationship was not fully consistent with results using markers separately (Appendix S3). This is likely due to different resolution and coalescence of the different sequences. The splitting between the North and South Atlantic monophyletic groups was estimated 1.78 and 1.414 Ma (posterior probability = 1.0), and divergence between Lineage I and II between 1.239 and 0.957 Ma (posterior probability = 0.96).

Genetic structure in *M. glacialis* along the European coast

Diversity and fixation indexes of the 13 populations of *M. glacialis* included in the population genetic analyses and for the different loci are presented in Table 4 and Appendix S1. Deviation from the HWE is likely due to high levels of inbreeding as observed in other marine benthic invertebrates. According to FreeNA null alleles did not have large effect on the genetic differentiation observed. Mixing of different cohort ages could not largely cause deviation from the HWE since all specimens collected within sites were in the same size range, and substructuring within populations was not detected in the Bayesian analysis explained below.

The Bayesian analysis of populations based on microsatellite genotypes, attributed the best value of *K* to three genetic clusters. The assignment of individuals to each of the three clusters showed a general pattern of panmixia (Fig. 3) with no evidences of large genetic breaks. Only the populations of Sagres, São Vicente and Madeira displayed a large proportion of individuals assigned to a specific cluster (Fig. 3a and Fig.1b). The second run did not detect

any clustering of the Mediterranean specimens related to the mitochondrial lineages (Fig. 3b).

Analysis of molecular variance showed significant differences among ecoregions for the COI (Φ_{CT} = 0.132, P = 0.00) but not for microsatellites (F_{CT} = -0.009, P = 0.661). Significant differences among populations within ecoregions were found for both types of markers (Φ_{SC} = 0.016 and 0.0063, respectively for COI and microsatellites, $P < 0.001$) (Table 5). Most F_{ST} pairwise comparisons between populations were significant (see Table 6). The graphical representations of the pairwise Φ_{ST} and F_{ST} matrices are shown in Fig. 4a. On the MDS of pairwise Φ_{ST} of COI we detected a clustering of populations, nearly related to different ecoregions (CB+ISC: Western Mediterranean, GBR+UK: Celtic Sea, AZ+MAD+ALG: Macaronesian Islands and Algarve, and all the other populations from the South European Atlantic Shelf). In the MDS of microsatellites Sagres, São Viçente, Azores and Madeira appeared genetically distant from other populations (Fig. 4a). Genetic differentiation caused by geographical distance was not detected (Mantel test for COI: r = -0.481, P = 0.938, and microsatellite loci: r = -0.214, P = 0.847). The DAPC analysis (Fig. 4b) was in agreement with the MDS based on the F_{ST} matrix of microsatellite loci (Fig. 4a), except for the Azores population that was not clearly separated from the other Atlantic and Mediterranean populations.

Additionally, no significant differentiation was detected between Mediterranean individuals of the mitochondrial Lineage I and II (F_{ST} = 0.008; P = 0.291), suggesting free interbreeding and absence of reproductive barriers between the two mitochondrial lineages found in the Mediterranean Sea.

The South European Atlantic Shelf and the Western Mediterranean (Lineage I) ecoregions showed signs of demographic expansion as demonstrated by the skyline plots based on the COI (Fig. 5). Celtic Sea and Lineage II from the Western Mediterranean did not present any signal of expansion, although confidence intervals were wide likely due to smaller

sample size of these geographical areas compared to the South European Atlantic Shelf. In the Azores-Canaries-Madeira ecoregion there was not a clear signal of recent expansion but it suggested a smooth increase in population size.

Historical migration rates estimated from Lamarc demonstrated higher level of gene flow from the Atlantic basin (South European Atlantic shelf ecoregion) to the Mediterranean basins (Western Mediterranean ecoregion) than in the reverse way (Fig. 6).

DISCUSSION

Our study showed the existence of three divergent mitochondrial lineages across the widespread distribution of the genus *Marthasterias*. These lineages most likely originated from historical vicariance correspond to two allopatric species with anti-tropical distribution that are currently found in separate hemispheres. The genetic structure of the Northern Hemisphere populations was mainly determined by the geographic isolation of the Atlantic islands, and gene flow barriers such as the strong upwelling in South Portugal. In addition, Pleistocene glaciations may have had a large effect on the distribution of genetic diversity of this species, with secondary contacts facilitating genetic admixture between previously isolated populations.

The Southern Hemisphere samples corresponded to a distant lineage from South Africa, supported by both mitochondrial and nuclear DNA. The matching among the mitochondrial Lineage III, the nuclear monophyletic Clade B and the disjunct geographical distribution suggest the existence of a different species in South Africa, which is consistent with previous morphological data (Clark, 1974). Therefore, the South African form should be considered as a different evolutionary unit of *Marthasterias* from the Northeast Atlantic. Phylogenetic trees suggested an earlier divergence of the South African lineage (between 1.78

and 1.414 Ma; posterior probability of 1.0), followed by the separation of the Northeast Atlantic and Mediterranean lineages (between 1.239 and 0.957 Ma; posterior probability of 0.96). It can be hypothesized that Quaternary climatic fluctuations over the last 2.4 Ma, which are known to have promoted divergence events and extinction processes of many coastal species (Maggs *et al*. 2008; Waltari & Hickerson, 2013), likely triggered the split of *Marthasterias* lineages between the Northern and Southern Hemispheres. The intensification of the Benguela upwelling during the transition between the late Pliocene and early Pleistocene, about 1.8 Ma (Marlow *et al*., 2000), together with the establishment of the current oceanographic circulation along the Eastern Atlantic, promoted genetic discontinuity between both hemispheres (Cunningham & Collins, 1998). Although *Marthasterias* has feeding larva capable of drifting in the plankton for more than 120 days (Barker & Nichols, 1983), the present day Atlantic oceanographic circulation is characterised by clockwise and anticlockwise gyres that prevent mixing water masses between both hemispheres, acting as a physical barrier to larval interchange. Divergence patterns over these biogeographical breaks have also been observed in copepods (Blanco-Bercial *et al.*, 2011) and the mussel *Mytilus*, although strong signals of mitochondrial introgression was observed in this last genus (Gérard *et al*., 2008). The specimen of *Marthasterias* from Cape Verde was genetically distant to all the other haplotypes, although our phylogenetic reconstruction clustered it more related to the South African lineage than to the European ones. Cape Verde is located on the limit between the North and South Atlantic waters and is a recognised biogeographical breakpoint (Quinteiro *et al.*, 2007). Future research along the western coast of the Africa is needed to detect potential areas of hybridisation between *Marthasterias* lineages and the existence of other independent evolutionary units.

At a smaller geographical scale we used multiple mito-nuclear markers to understand the genetic structure of *M. glacialis* along the European coast. Levels of genetic divergence

between Lineage I and II of *M. glacialis* were much higher than the ones observed for some echinoderms that show reproductive isolation (levels of divergence as low as 0.9 % for the COI gene, Landry *et al.,* 2003). Analyses based on nuclear markers in *M. glacialis* did not detect clustering of individuals from these two mitochondrial lineages, ruling out the effect of assortative mating within these populations (Calderón *et al.,* 2010). Hence, despite the large genetic divergence between these two lineages (3% for COI, 4.8% for CR and 1.2% for 16S), free interbreeding and random association seems to exist between them, as shown by the non-significant nuclear F_{ST} value between lineages. Therefore, all Northern Hemisphere linages of *M. glacialis* should be considered as a unique evolutionary unit. The inclusion of different genetic markers allowed an estimation of splitting time, which leads to robust conclusions about the origin of lineages. Changes in sea level during middle Pleistocene linked to cyclic ice ages restricted marine circulation between the Atlantic and Mediterranean basins, and Mediterranean sub-basins (Patarnello *et al.* 2007). This promoted gene flow interruptions at both sides of the Gibraltar Strait, a biogeographical breakpoint that has important implications on the genetic divergence of marine Atlantic populations (e.g. Domingues *et al.*, 2007; Patarnello *et al*. 2007, and references herein). Gene flow interruptions, between 1.239 and 0.957 Ma, likely resulted in high genetic divergence between Atlantic and Mediterranean populations of *Marthasterias*, followed by post-glacial expansion and recolonization of the Mediterranean Sea by the already formed Atlantic Lineage I. Secondary contact and colonization of the Mediterranean by Lineage I potentially promoted interbreeding within this basin. Post-glacial recolonization of the Mediterranean Sea by Atlantic lineages is not something exclusive of *Marthasterias*, and has been reported for other marine invertebrates (Pellerito *et al*., 2009; Taboada & Pérez-Portela, 2016). Our data indicated a recent demographic expansion in the South European Atlantic Shelf, also reflected on Lineage I in the Western Mediterranean. This pattern of postglacial expansion and secondary contact is

consistent with that reported in other systems (Wilson & Veraguth, 2010). Estimations of demographic expansion in the South European Atlantic Shelf and Lineage I in the Western Mediterranean dated that event during the last 25 Ky. This estimation should be taken with caution, as the accurate mutation rate of the studied genes for *Marthasterias* remains unknown, but might indicate of an expansion initiated after the Last Glacial Maximum. This is coherent with other shallow benthic groups that showed similar demographic patterns (Wilson & Veraguth, 2010; Pérez-Portela *et al.*, 2013). We nonetheless detected no signals of demographic expansion in the Celtic Sea ecoregion and Western Mediterranean Lineage II, probably due to the smaller number of samples included in these geographical areas and lineage.

The different geographical distribution of the mitochondrial lineages I and II, and the presence haplotypes restricted to certain biogeographic ecoregions, supports that mitochondrial DNA might retain signatures of past hydrography and major biogeographic breaks (Waters & Roy, 2004). Accordingly, the AMOVA test and pairwise Φ_{ST} of the COI detected genetic divergence among the four ecoregions analysed. Marked phylogeographical discontinuities, as is the absence of some ancestral haplotypes typically evolve in response to long-term extrinsic barriers to gene flow. However, the present day structure may have been affected by a redistribution of genetic diversity by contemporary gene flow. The moderate genetic divergence found between populations contrasts with the isolation by distance pattern of other starfish species with large dispersal potential (Zulliger *et al.*, 2009) and complete panmixia of other echinoderms (Wangensteen *et al.* 2012; Pérez-Portela *et al*., 2013) across the same geographical area.

The signatures of historical differentiation among ecoregions were not mirrored by the present day genetic structure of *M. glacialis* as shown by the microsatellite loci. Our analyses did not detect clear genetic differentiation across the most important biogeographical barriers,

as the Almeria-Oran Front and at the south of the English Channel. However, most populations presented significant differences in their genetic structure.

Analyses of gene flow between Atlantic and Mediterranean basins estimated higher migration rates from the Atlantic to the Mediterranean basin, according to the mitochondrial COI. A predominant gene flow from west to east across the Gibraltar Strait is in accordance with a later recolonization of the Mediterranean Sea. This migration pattern has also been observed in other marine species (see Penant *et al.* 2013, Taboada & Pérez-Portela, 2016, and references therein), and is congruent with the strong superficial current that flows from the Atlantic to the Mediterranean (Candela, 1991). The reverse current from the Mediterranean to the Atlantic is dominated by denser waters that flow at a depth of 400-800 m, where environmental conditions may be unsuitable for *Marthasterias* larvae (Villalobos *et al.*, 2006), and hence limiting the dispersal of the endemic Mediterranean Lineage II to the Atlantic. Most of the genetic divergence (according to microsatellite loci) in the Atlantic region was detected in Sagres, Sao Vicente and Madeira. This genetic structure, and the isolation of the Azores, was further supported by pairwise F_{ST}. Furthermore, microsatellites showed large differentiation between two nearby populations, only 10 km apart, Sagres and São Vicente. This divergence may be consistent with the disruptive effect of a well-known upwelling in the area of the Cape São Vicente (Relvas & Barton, 2005), where these two sites are located. Such genetic dissimilarity seemed to have recently been produced as the mitochondrial structure indicates common demographic history along the upwelling area and shared haplotypes (H_6 and H_12) (Pérez-Portela *et al.,* 2010). The inflow of larvae from distant genetic sources, particular those with dissimilar environmental conditions plus strong physical barriers due to water currents, might underlie this large genetic divergence in such a short distance.

The genetic isolation of the Azores-Canaries-Madeira ecoregion compared to other

Atlanto-Mediterranean regions may be as a result of the combined effect of historical and contemporary processes. Past isolation during the Pleistocene glaciations (Domingues *et al.*, 2007), the absence of suitable habitat acting as stepping stone corridors for migrants between the islands and continent, and the disruptive effect of the Portugal current, flowing along the east Atlantic coast from northern Portugal southwards to northern Africa can be the factors driving genetic divergence of these populations (Domingues *et al.*, 2007).

Our study stresses the importance of taking into account different geographical scales to fully understand the distribution of the genetic diversity in marine coastal species. The Plio-Pleistocene climatic history, the associated sea level fluctuations and marine circulation have generated a complex evolutionary scenario of genetic isolation, lineages sorting, demographic expansions, secondary contacts, and gene flow that characterises the phylogeography of *Marthasterias*. Different genetic signatures observed in the genetic structure between mitochondrial DNA and microsatellites may be due to the stochasticity of the coalescence processes for different types of markers. Nevertheless, the potential effect of exogenous selection due to different environmental pressures on the distribution of mitochondrial lineages of *M. glacialis* cannot be ruled out. Mitochondrial genome plays a functional role in the respiratory metabolism, nucleo-cytoplasmic coadaptation, and/ or adaptive introgression (Bazin *et al.* 2006, and references herein). The absence of the Mediterranean lineage (Lineage II) in the Atlantic may be related to different selection pressures between basins. Specific adaptations on the mitochondrial DNA may explain the prevalence of haplotypes from one lineage. In this study twelve non-synonymous substitutions were found across the whole dataset of COI but there were not fixed at any specific lineage or geographical group of haplotypes, and therefore the potential effect of selection on this mitochondrial gene could not be determined. In addition, endogenous selection may potentially generate discrepancy in genetic structure observed in *M. glacialis* between markers. Endogenous barriers to gene flow

478 due to different factors such as allelic incompatibility, assortative mating, asynchrony in
479 spawning and selection against hybrids, among others, can shape species' genetic structure.
480 We did not observed signals of assortative mating in our Mediterranean populations, but in
481 many cases exogenous and endogenous selection act together and endogenous barriers trend
482 to coincide with environmental limits (Bierne *et al*. 2011), so tension zones of hybridization
483 are sometimes difficult to detect. Future studies focussing on selection and exploring the
484 genetic structure of populations at both sides of the Gibraltar Strait would be needed to clarify
485 this important matter.

486 To conclude, our study demonstrates the benefits of using a comprehensive sampling
487 together with a multilocus dataset to understand past and present processes shaping the
488 phylogeography of widespread species. Studies that only use mitochondrial genes may not
489 provide a complete picture and ultimately affect our perception of biodiversity patterns. Our
490 results demonstrated the presence of genetic distant lineages, most likely as a result of
491 vicariance events and the persistence of gene flow barriers. Future studies using genomic-
492 wide sequencing would allow unravelling the effect of natural selection on the genetic
493 structure of this species, as well as the detailed information regarding how secondary contacts
494 shaped genetic admixture events.

495

496 **ACKNOWLEDGEMENTS**

497

498 We are grateful to Joana Micael, Joana Xavier, Xavier Mina, Bernat Pincton, Claire Goodwin,
499 Oscar Serrano, Mikel Becerro, Vitor Almada, Joana Robalo, Owen Wangensteen, Xavier
500 Turon, Cruz Palacín, Pablo López, and Patrick Erwin for providing samples and logistic
501 support during the sampling trips. We also thank three anonymous reviewers and the editor
502 for their helpful comments and suggestions, which greatly improved the manuscript. This

research was supported by a *Juan de la Cierva* Research Fellowship from the Spanish Government to RP-P, and also financially supported by the Spanish Government projects BENTHOMICS (CTM2010-22218) and CHALLENGEN (CTM2013-48163). This paper is a contribution of the Consolidated Research Group 2009SRG665 of the Generalitat of Catalunya.

REFERENCES

Avise, J.C. (2000) *Phylogeography: The history and formation of species*. Harvard University Press, Cambridge, Massachusetts, p 447.

Barker, M. F. & Nichols, D. (1983) Reproduction, recruitment and juvenile ecology of the sta rfish, *Asterias rubens* and *Marthasterias glacialis*. *Journal of the Marine Biological Asso ciation of the United Kingdom*, **63**, 745-765.

Bazin, E., Glémin, S. & Galtier, N. (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570-572.

Beheregaray, L. B. (2008) Twenty years of phylogeography: the state of the field and the chal lenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754-3774.

Bierne, N., Welch, J., Loire, E., Bonhomme, F. & David, P. (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044-2072.

Blanco-Bercial, L., Etvarez-Marqués, F. & Bucklin, A. (2011) Comparative phylogeography a nd connectivity of sibling species of the marine copepod *Clausocalanus* (Calanoida). *Jou rnal of Experimental Marine Biology and Ecology*, **404**, 108-115.

Boissin, E., Stöhr, S. & Chenuil, A (2011). Did vicariance and adaptation drive cryptic speciat ion and evolution of brooding in *Ophioderma longicauda* (Echinodermata: Ophiuroidea), a common Atlanto-Mediterranean ophiuroid? *Molecular Ecology*, **20**, 4737–4755.

529 Borrero-Pérez, G. H., González-Wangüemert, M., Marcos, C. & Pérez-Ruzafa, A (2011). Phy
530 logeography of the Atlanto-Mediterranean sea cucumber *Holothuria (Holothuria) mamm*
531 *ata*: the combined effects of historical processes and current oceanographical pattern. *Mol*
532 *ecular Ecology*, **20**, 1964–1975.
533 Calderón, I., Ventura, C. R. R., Turon, X. & Lessios, H. A. (2010) Genetic divergence and ass
534 ortative mating between colour morphs of the sea urchin *Paracentrotus gaimardi*. *Molecu*
535 *lar Ecology*, **19**, 484-493.
536 Candela, J. (1991) The Gibraltar Strait and its role in the dynamics of the Mediterranean Sea.
537 *Dynamics of Atmospheres and Oceans*, **15**, 267-299.
538 Clark, A.M. (1974) Notes on some echinoderms from Southern Africa. *Bulletin of the British*
539 *Museum Natural History. Zoology*, **26**, 421-487
540 Cunningham, C. & Collins, T. (1998) *Beyond area relationships: Extinction and*
541 *recolonization in molecular marine biogeography*. In: *Molecular Approaches to Ecology*
542 *and Evolution* (eds. DeSalle R, Schierwater B), Birkhlutionea relationships
543 Debiasse, M.B., Nelson, B. J. & Hellberg, M. E. (2014) Evaluating summary statistics used to
544 test for incomplete lineage sorting: mito-nuclear discordance in the reef sponge *Callyspon*
545 *gia vaginalis*. *Molecular Ecology*, **23**, 225-238.
546 Domingues, V., Almada, V., Santos, R., Brito, A. & Bernardi, G. (2007) Phylogeography and
547 evolution of the triplefin *Tripterygion delaisi* (Pisces, Blennioidei). *Marine Biology*, **150**,
548 509-519.
549 Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals
550 using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-
551 2620.
552 Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software
553 package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47

554 Garcia-Cisneros, A., Palacín, C., Khadra, Y. B. & Pérez-Portela, R. (2016). Low genetic
555 diversity and recent demographic expansion in the red starfish *Echinaster sepositus*
556 (Retzius 1816). *Scientific Reports*, **6**, 33269
557 Gérard, K., Bierne, N., Borsa, P., Chenuil, A. & Féral, J. P. (2008) Pleistocene separation of
558 mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres
559 and strong genetic differentiation among southern populations. *Molecular Phylogenetics*
560 *and Evolution,* **49**, 84-91.
561 Harper, F. M., Addison, J. A. & Hart, M. W. (2007) Introgression versus immigration in
562 hybridizing high-dispersal echinoderms. *Evolution,* **61**, 2410-2418.
563 Hewitt, G. (2004) The structure of biodiversity - insights from molecular phylogeography.
564 *Frontiers in Zoology,* **1**, 4.
565 Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., *et al.* (2010) Phylogeography's past,
566 present, and future: 10 years after. *Molecular Phylogenetics and Evolution,* **54**, 291-301.
567 Hunter, R. L. & Halanych, K. M. (2010) Phylogeography of the Antarctic planktotrophic brittle star
568 *Ophionotus victoriae* reveals genetic structure inconsistent with early life history. *Marine*
569 *Biology,* **157**, 1693–1704.
570 Jombart, T., Devillard, S., Balloux, F. (2010) Discriminant analysis of principal components:
571 a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 1.
572 Landry, C., Geyer, L. B., Arakaki, Y., Uehara, T. & Palumbi, S. R. (2003) Recent speciation i
573 n the Indo-West Pacific: rapid evolution of gamete recognition and sperm morphology in
574 cryptic species of the sea urchin. *Procedings of the Royal Society of London,* **270**, 1839-1
575 847.
576 Maggs, C. A., Castilho, R., Foltz, D., *et al.* (2008) Evaluating signatures of glacial refugia for
577 North Atlantic benthic marine taxa. *Ecology,* **89**, S108-S122.
578 Marlow, J. R., Lange, C. B., Wefer, G. & Rosell-Melu, A. (2000) Upwelling Intensification
579 As Part of the Pliocene-Pleistocene Climate Transition. *Science,* **290**, 2288-2291.

Mortensen, T. (1927) *Echinoderms of the British Isles*. VIII. Oxford University Press, Oxford, p 471.

Mortensen, T. (1933) *Echinoderms of South Africa (Asteroidea and Ophiuroidea).* Videnskabelige Meddelelser fra Dansk naturhistorisk Forening i København 93: p 215–400.

Patarnello, T., Volckaert, F. A. & Castilho, R. (2007) Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break?. *Molecular Ecology*, **16**, 4426-4444.

Penant, G., Aurelle, D., Feral, J. P. & Chenuil, A. (2013) Planktonic larvae do not ensure gene flow in the edible sea urchin *Paracentrotus lividus*. *Marine Ecology Progress Series,* **480**, 155-170.

Pérez-Portela, R., Almada, V. & Turon, X. (2013) Cryptic speciation and genetic structure of widely distributed brittle stars (Ophiuroidea) in Europe. *Zoologica Scripta,* **42**, 151-169.

Pérez-Portela, R., Villamor, A. & Almada, V. (2010) Phylogeography of the sea star *Marthaste rias glacialis* (Asteroidea, Echinodermata): deep genetic divergence between mitochondri al lineages in the north-western Mediterranean. *Marine Biology,* **157**, 2015-2028.

Petit, R.J., El Mousadik, A. & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology,* **12**, 844-855.

Quinteiro, J., Rodríguez-Castro, J. & Rey-Méndez, M. (2007) Population genetic structure of the stalked barnacle *Pollicipes pollicipes* (Gmelin, 1789) in the northeastern Atlantic: influence of coastal currents and mesoscale hydrographic structures. *Marine Biology,* **153**, 47-60.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Relvas, P. & Barton, E.D. (2005) A separated jet and coastal counterflow during upwelling relaxation off Cape São Vicente (Iberian Peninsula). *Continental Shelf Research,* **25**, 29-49.

Siegel, D. A., Mitarai, S., Costello, C. J., *et al.,* (2008). The stochastic nature of larval

606 connectivity among nearshore marine populations. *Proceedings of the National Academy*
607 *of Sciences USA*, **105**, 8974-8979.
608 Spalding, M. D., Fox, H. E., Allen, G. R., *et al.* (2007) Marine Ecoregions of the World: A Bi
609 oregionalization of Coastal and Shelf Areas. *BioScience*, **57**, 573-583.
610 Taboada, S. & Pérez-Portela R. (2016) Contrasted phylogeographic patterns on mitochondrial
611 DNA of shallow and deep brittle stars across the Atlantic-Mediterranean area. *Scientific*
612 *Reports, In press*.
613 Villalobos, F. B., Tyler, P. A. & Young, C. M. (2006) Temperature and pressure tolerance of e
614 mbryos and larvae of the Atlantic seastars *Asterias rubens* and *Marthasterias glacialis* (E
615 chinodermata: Asteroidea): potential for deep-sea invasion. *Marine Ecology Progress Seri*
616 *es*, **314**, 109-117.
617 Waltari, E. & Hickerson, M. J. (2013) Late Pleistocene species distribution modelling of
618 North Atlantic intertidal invertebrates. *Journal of Biogeography*, **40**, 249-260.
619 Wangensteen, O. S., Turon, X., Pérez-Portela, R. & Palacín, C. (2012) Natural or
620 Naturalized? Phylogeography suggests that the abundant sea urchin *Arbacia lixula* is a
621 recent colonizer of the Mediterranean. *PLoS ONE*, **7**, e45067.
622 Waters, J. M. & Roy M. S (2004) Phylogeography of a high-dispersal New Zealand sea-star: d
623 oes upwelling block gene-flow? *Molecular Ecology*, **13**, 2797-2806.
624 Weber, A.T., Mérigot, B., Valière, S. & Chenuil, A. (2015) Influence of the larval phase on
625 connectivity: strong differences in the genetic structure of brooders and broadcasters in
626 the *Ophioderma longicauda* species complex. *Molecular Ecology*, **24**, 6080-6094.
627 Weersing, K. & Toonen, R. J. (2009) Population genetics, larval dispersal, and connectivity in
628 marine systems. *Marine Ecology Progress Series*, **393,** 13,ne
629 Wilson, A. B. & Veraguth, E. (2010) The impact of Pleistocene glaciation across the range of
630 a widespread European coastal species. *Molecular Ecology*, **19**, 4535-4553.

631 Zulliger, D., Tanner, S., Ruch, M. & Ribi, G. (2009) Genetic structure of the high dispersal
632 Atlanto-Mediterreanean sea star *Astropecten aranciacus* revealed by mitochondrial DNA
633 sequences and microsatellite loci. *Marine Biology*, **156**, 597-610.

634

635 **DATA ACCESSIBILITY**

636 The new DNA sequences obtained in this study have been deposited in Genbank (Accession
637 numbers "*pending*")

638

639 **BIOSKETCHES**

640 **Rocío Pérez-Portela** and **Adriana Villamor** are broadly interested in genetic connectivity,
641 historical and demographic events involved in the distribution of genetic diversity and
642 speciation of marine benthic invertebrates. Rocío Pérez-Portela's most recent research focuses
643 on exploring the molecular bases of adaptation, and stress responses to environmental shifts.
644 **Marc Rius'** research focuses on understanding the ecological and evolutionary mechanisms
645 that shape species ranges.

646

647 Author contributions: RPP, AV and MR designed the study and collected most of the
648 samples. RPP conducted DNA sequencing and AV the genotyping of populations with
649 microsatellite loci. RPP and AV performed the genetic analyses. RPP wrote the first
650 manuscript draft, and AV and MR contributed to the writing of the manuscript. All authors
651 have revised the final version of this manuscript.

652

653 **EDITOR:** Prof Peter Linder

654

655

TABLES

Table1. Sampling collection of *Marthasterias* across the North and south-east Atlantic Ocean and Mediterranean Sea. Sampling locality, biogeographic province/ ecoregion, population code, coordinates and number of specimens analysed. N_{COI}, N_{mc}, N_{16S}, N_{CR} and N_{ITS1}: Individuals analysed for COI, microsatellites, 16S, CR and ITS1, respectively. Information in bold corresponds to data obtained in this study.
* Populations used for population genetic analyses.

Locality	*Province/ Ecoregion*	*Code*	*Coordinates*	N_{COI}	N_{mc}	N_{16S}	N_{CR}	N_{ITS1}
Scilly Island	North European Seas/ Celtic Sea	GBR*	49°51N 6°21' W	12	-	-	-	-
Donegal, Ireland	North European Seas/ Celtic Sea	IR	54°36'N 8°11'W	4	-	-	-	-
Plymouth	North European Seas/ Celtic Sea	UK*	Pérez-Portela *et al.* 2010	28	29	2	2	2
Pais Vasco	Lusitanian/ South European Atlantic Shelf	PV*	Pérez-Portela *et al.* 2010	30	30	-	-	-
Santander	Lusitanian/ South European Atlantic Shelf	ST*	Pérez-Portela *et al.* 2010	20	20	-	-	-
Galicia, Ares	Lusitanian/ South European Atlantic Shelf	G*	Pérez-Portela *et al.* 2010	28	29	3	3	1
Sao Vicente	Lusitanian/ South European Atlantic Shelf	SV*	Pérez-Portela *et al.* 2010	22	28	-	-	-
Sagres	Lusitanian/ South European Atlantic Shelf	SA*	Pérez-Portela *et al.* 2010	19	20	-	-	-
Cascais	Lusitanian/ South European Atlantic Shelf	CAS*	Pérez-Portela *et al.* 2010	31	31	-	-	-
Algarve	Lusitanian/ South European Atlantic Shelf	ALG*	37° 4'25' N 8°39'W	14	15	4	4	3
San Pedro	Lusitanian/ South European Atlantic Shelf	SP	Pérez-Portela *et al.* 2010	3	-	-	-	-
Sao Miguel, Azores	Lusitanian/ Azores-Canaries-Madeira	AZ*	Pérez-Portela *et al.* 2010	13	14	3	3	-
Funchal, Madeira	Lusitanian/ Azores-Canaries-Madeira	MAD*	32°36' N 16°58'W	7	7	3	3	1
Flores, Azores	Lusitanian/ Azores-Canaries-Madeira	FL	39°26'N 31°16'W	4	-	-	-	-
Costa Brava	Mediterranean/ Western Mediterranean	CB*	Pérez-Portela *et al.* 2010	31 + 8	40	10	10	4
Ischia	Mediterranean/ Western Mediterranean	ISC*	40°40'N 13°55'E	20	20	12	12	7
Xàbia	Mediterranean/ Western Mediterranean	XAB	38°46'N 0°12'E	1	-	1	1	-
Múrcia	Mediterranean/ Western Mediterranean	SUR	37°29'N 1°18'W	8	-	-	-	-
Sta Maria, Sal, Cape Verde	West African Transition/ Cape Verde	CV	16°35'N 22°56'W	1	-	-	-	-
Cap Town, South Africa	Agulhas/ Agulhas Bank	SDA	33°58'S 18°19'E	31	-	15	15	7
Storm River, South Africa	Agulhas/ Agulhas Bank	SDI	34°15'S 23°27'E	2	-	2	2	2
TOTAL				337	283	55	55	27

Table 2. Mitochondrial lineages of *Martasterias* and their geographical distribution (ecoregions) across the North and south-east Atlantic Ocean and Mediterranean Sea. Fragments sequenced with the length of the fragment in brackets; (N) number of sequences, (h) number of haplotypes or alleles, (*Hd*) haplotype diversity (± standard deviation) and (π) nucleotide diversity (± standard deviation).

		DNA fragments															
		COI (617 pb)				**16S (486 pb)**				**CR (870 pb)**				**ITS1 (405 pb)**			
Lineage	**Geographical distribution by ecoregions**	**N**	**h**	***Hd***	***π***	**N**	**h**	***Hd***	***π***	**N**	**h**	***Hd***	***π***	**N**	**h**	***Hd***	***π***
Lineage I	Celtic Sea, South European Atlantic Shelf, Azores-Canaries-Madeira and Western Mediterranean	286	85	0.897± 0.013	0.0714 ± 0.0003	22	5	0.519± 0.114	0.001± 0.0004	22	17	0.972± 0.0022	0.010± 0.0008	18	10	0.857± 0.0910	0.0024± 0.0005
Lineage II	Western Mediterranean (Spain and Italy)	17	8	0.860± 0.068	0.0365± 0.0006	16	6	0.542± 0.147	0.002± 0.0006	16	12	0.958± 0.0013	0.004± 0.0006	9			
Lineage III	Cape Verde and Agulhas Bank (South Africa)	34	20	0.894± 0.043	0.0058± 0.00058	17	5	0.507± 0.140	0.001± 0.0004	17	16	0.993± 0.0005	0.006± 0.0006	12	3	0.591± 0.1080	0.0017± 0.0004
	TOTAL	337	113	0.943± 0.008	0.0147± 0.0001	55	16	0.843± 0.028	0.012± 0.0006	55	45	0.992± 0.005	0.023± 0.0006	39	13	0.869± 0.0480	0.0078± 0.0008

Table 3. Divergence between lineages of *Marthasterias*. Divergence between the nuclear Clade A and Clade B for the ITS1, and between the mitochondrial lineages I, II and III for the COI, 16S and the putative CR (only the 434 pb), respectively.

ITS1		***Mitochondrial***			
	CLADE A		*Lineage I*	*Lineage II*	
CLADE A (Lineage I & II)	-	*Lineage II*	3.0 % 1.2 % 4.8 %	- - -	COI 16S CR
CLADE B (Lineage III)	1.5 %	*Lineage III*	4.1 % 1.6 % 5.1 %	3.0 % 2.2 % 3.0 %	COI 16S CR

Table 4. Genetic diversity of European populations of *Marthasterias glacialis*. Population code, number of individuals sequenced (*N*), number of haplotypes (*h*), haplotype richness after rarefaction *Hk[7]*, number of private haplotypes (*Np*), haplotype diversity (*Hd*), nucleotide diversity (π) and percentage of every mitochondrial lineage per population (Lineage I and II as LI and LII, respectively) are showed for COI (Numbers in brackets are standard deviations). Allelic richness (*Ra[7])*, number of private alleles (*Npa*), observed heterozygosity (*Ho*), expected heterozygosity (*He*) and the fixation index (F_{IS}) are also calculated from microsatellite data. * Significant at P <0.05, and ** P <0.01 when departure from the HW Equilibrium.

	COI							***Microsallite data***				
Code	***N***	***h***	***Hk[7]***	***Np***	***Hd***	***π***	***% Lineages***	***Ra[7]***	***Npa***	***Ho***	***He***	***F_{IS}***
GBR	12	8	3.606	3	0.909 (±0.065)	0.00898 (±0.001)	LI-100%					
UK	28	12	3.435	5	0.854 (±0.055)	0.00849 (± 0.001)	LI-100%	**4.749**	2	0.612	0.704	0.1335**
PV	30	14	3.550	7	0.869 (± 0.049)	0.00506 (± 0.001)	LI-100%	4.443	0	0.513	0.693	0.2630**
ST	20	10	3.370	5	0.853 (±0.063)	0.00658 (± 0.001)	LI-100%	**5.150**	3	0.490	0.725	0.3302**
G	28	17	**4.062**	10	0.923 (± 0.037)	0.00814 (± 0.001)	LI-100%	4.231	2	0.517	0.598	0.1372*
SV	22	11	3.761	9	0.909 (± 0.045)	0.00608 (± 0.001)	LI-100%	3.804	2	0.468	0.569	0.1798**
SA	19	7	2.563	3	0.749 (± 0.075)	0.00350 (± 0.001)	LI-100%	3.725	0	0.369	0.494	0.2618**
CAS	31	12	3.499	6	0.873 (±0.039)	0.00732 (± 0.001)	LI-100%	4.686	0	0.600	0.672	0.1086**
ALG	14	11	**4.308**	6	0.943 (±0.054)	0.00611 (±0.001)	LI-100%	4.212	2	0.544	0.627	0.1404*
AZ	13	8	4.050	1	0.919 (± 0.057)	0.00797 (± 0.001)	LI-100%	3.978	1	0.501	0.563	0.1139*
MAD	7	5	3.429	2	0.857 (±0.137)	0.00374 (± 0.001)	LI-100%	**5.000**	2	0.600	0.723	0.1818
CB	39	23	**4.349**	12	0.949 (±0.023)	0.01391 (± 0.002)	LI- 82%, LII- 18%	4.389	0	0.521	0.618	0.1572**
ISC	20	15	**4.482**	6	0.963 (±0.028)	0.01718 (± 0.001)	LI- 60%, LII- 40%	**4.716**	2	0.648	0.724	0.1090**

Table 5. AMOVA tests based on sequences of COI and microsatellite loci of European populations of *Marthasterias*. Populations appear grouped in four ecoregions.
** Significant when $P < 0.001$

Source of variation	**d.f**	**Variance component**	**Fixation index**	**% of variation**	***P*-value**
		COI			
Among ecoregions	3	0.395	Φ_{CT}= 0.132	13.25	0.000**
Among populations within ecoregions	9	0.043	Φ_{SC}= 0.016	1.43	0.000**
Within populations	269	2.541	Φ_{ST}= 0.149	85.31	0.000**
Total	281	2.979			
		Microsatellites			
Among ecoregions	3	-0.014	F_{CT}= -0.009	0.00	0.661
Among populations within ecoregions	8	0.105	F_{SC}= 0.063	6.39	0.000**
Within populations	271	0.262	F_{ST}= 0.168	15.92	0.000**
Within individuals	283	1.297	F_{IS}= 0.214	78.58	0.000**
Total	565	930.217			

Table 6. Pairwise comparisons between European populations *Marthasterias glacialis*. Φ_{ST} values for COI (below the diagonal) and F_{ST} values for microsatellite loci (above the diagonal).

	GBR	UK	G	PV	ST	CAS	SV	SA	ALG	AZ	MAD	CB	ISC
UK	0.011	-	0.030***	-0.002	-0.005	0.017*	0.106***	0.107***	0.018	0.051***	0.079***	0.022***	0.008
G	-0.031	0.066*	-	0.047***	0.031*	0.032***	0.080***	0.105***	0.038***	0.088***	0.139***	0.027***	0.066***
PV	0.075*	0.238***	0.032	-	0.005	0.029***	0.116***	0.142***	0.035*	0.047***	0.097***	0.042***	0.033**
ST	0.009	0.156***	-0.001	-0.001	-	0.004	0.098***	0.126***	0.008	0.047*	0.061*	0.025*	0.010
CAS	-0.025	0.092*	-0.024	0.022	-0.006	-	0.071***	0.125***	-0.001	0.036***	0.104***	0.039***	0.025*
SV	0.071	0.223***	0.044	0.006	0.003	0.041	-	0.190***	0.073***	0.112***	0.238***	0.096***	0.102**
SA	0.107*	0.275***	0.059*	-0.007	0.002	0.047	0.019	-	0.124***	0.195***	0.238***	0.121***	0.133***
ALG	0.223***	0.335***	0.189***	0.195***	0.188***	0.198***	0.086	0.244***	-	0.033*	0.108***	0.025*	0.014*
AZ	0.123	0.234***	0.115***	0.142***	0.117***	0.117***	0.062	0.175***	-0.009	-	0.135***	0.092***	0.075**
MAD	0.236***	0.347***	0.229***	0.293***	0.232***	0.233***	0.169***	0.324***	0.103	0.055	-	0.119***	0.068***
CB	0.022*	0.103***	0.043*	0.093***	0.047	0.049*	0.068*	0.089*	0.139***	0.091***	0.114	-	0.043***
ISC	0.173*	0.231***	0.224***	0.317***	0.243***	0.245***	0.281***	0.305***	0.328***	0.282***	0.282*	0.058	-

For COI: * significant at P >0.05, and *** significant at P >0.0098 after False Discovery Rate (FDR) correction.
For microsatellite loci: * significant at P >0.05, and *** significant at P >0.0104 after FDR correction.
(For GBR we only have the COI marker)

761 **FIGURE LEGENDS**

762 **Figure 1**. Map of the sampling sites of *Marthasterias*. **a**) Diamonds represent sites and colours
763 different ecoregions. Collection sites are abbreviated as in Table 1. Pie charts represent COI
764 haplotype or lineage frequencies per ecoregion. Size is proportional to the number of samples
765 analysed. Private haplotypes per ecoregion for Lineage I are filled in white, for Lineage II in
766 dark grey and for Lineage III in black. Underlined codes represent sites included in the
767 population genetic analyses. **b**) Frequency of the different clusters assigned to individuals by
768 Structure (K=3) based on microsatellite data. Individuals were assigned to a specific cluster
769 when showed ≥50% of probability to one cluster. Bright pick represents the frequency of
770 individuals with <50% of assignation probability to a specific cluster (mixed genotypes) (see
771 also Figure 3).

772 **Figure 2**. Networks and BI phylogenetic tree of *Marthasterias* haplotypes from COI and
773 ITS1fragments. a) Networks from the COI and ITS1 fragments: Area of the circles is
774 proportional to the number of individuals; lines between circles represent one mutational step
775 and additional perpendicular dashes indicate number of mutations when more than one; red
776 dots are missing haplotypes. For the COI network, partitions inside the circles are the
777 proportion of every haplotype per ecoregion. For the ITS1 network, partitions in different
778 colours represent the proportion of each mitochondrial lineage per nuclear allele. b) Simplified
779 version of the BI phylogenetic tree, including posterior probabilities, based on COI sequences.
780 Only the names of the most frequent haplotypes and from Cape Verde (CV) are presented.

781 **Figure 3**. Bayesian analysis results of European populations of *Marthasterias glacialis* from
782 Structure. a) Left: Results of the assignment of individual genotypes for K=2 and K=3; Right:
783 Delta K plot for the whole dataset; and b) Left: Assignment of individual genotypes within the
784 Mediterranean basin including the two populations (CB and ISC); Right: Delta K plot for the
785 two Mediterranean populations. The assignment of individual genotypes is represented by

vertical bar partitioned into K-coloured segments that represent its estimated membership fraction in each of the inferred groups. * on the (b) graph represent individuals from the mitochondrial Lineage II.

Figure 4. Genetic structure of European populations of *Marthasterias glacialis*. **a**. Multidimensional scaling (MDS) base on the Φ_{ST} values obtained from the COI and F_{ST} from microsatellite loci data in *Marthasterias glacialis*. Coloured diamonds represent localities from different ecoregions (Light green: Celtic Sea; white: South European Atlantic Shelf; pink: Azores-Canaries-Madeira; light grey: Western Mediterranean). **b**. DAPC analysis based on microsatellite loci.

Figure 5. Demographic analyses of *Marthasterias glacialis* based on Bayesian Skyline plots. Time is measured in million years (Ma). Black lines illustrate mean size estimations, and blue lines show 95 % confidence interval. For the Western Mediterranean ecoregion analyses were separately performed for Lineage I and Lineage II.

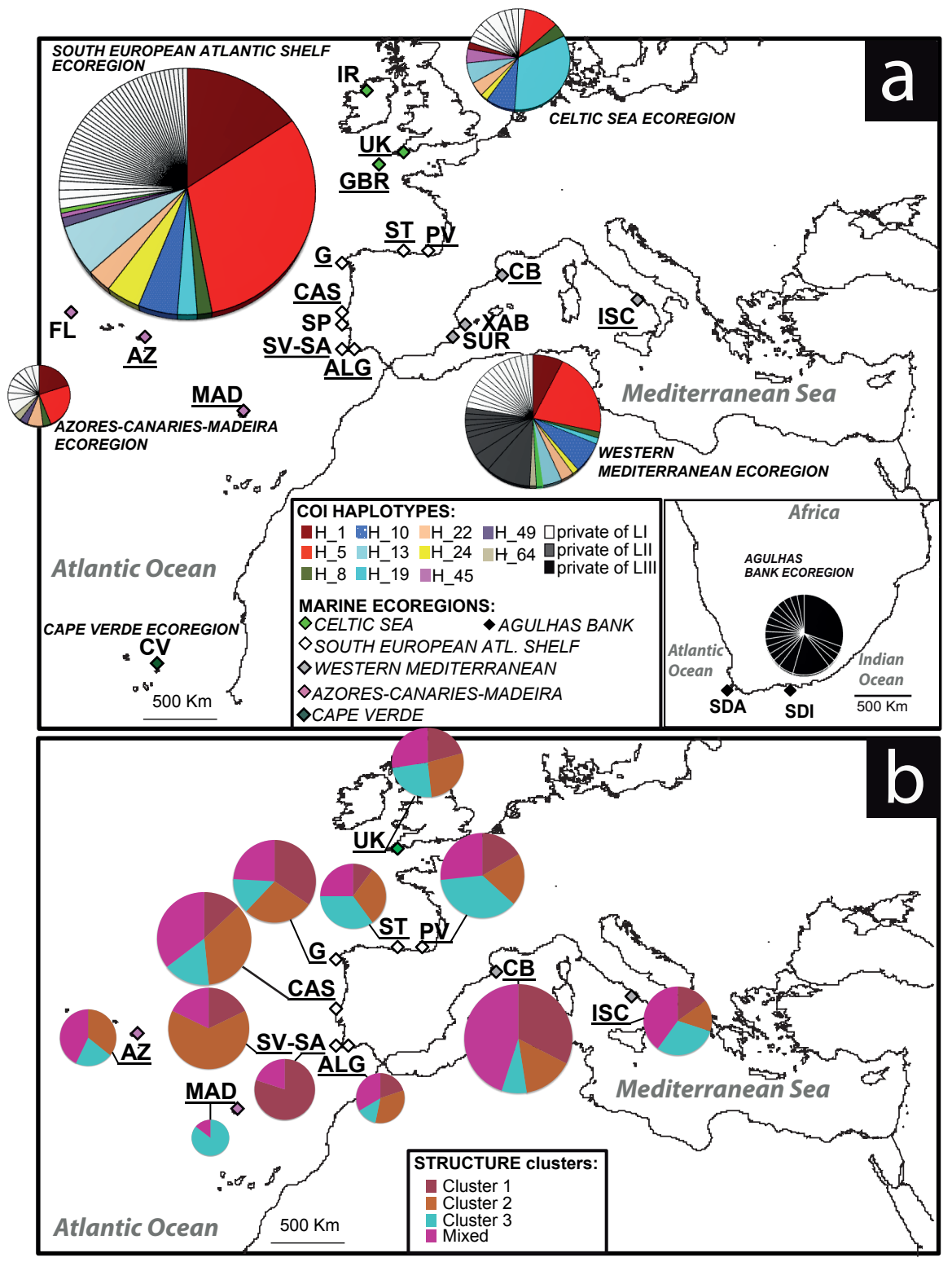
Figure 6. Gene flow between the Atlantic and Mediterranean populations of *Marthasterias glacialis*. Graphical representation from Lamarc outputs of historical migration based on coalescence between the South European Atlantic Shelf and Western Mediterranean ecoregions. The two graphs represented the final results for 3 different replicates.

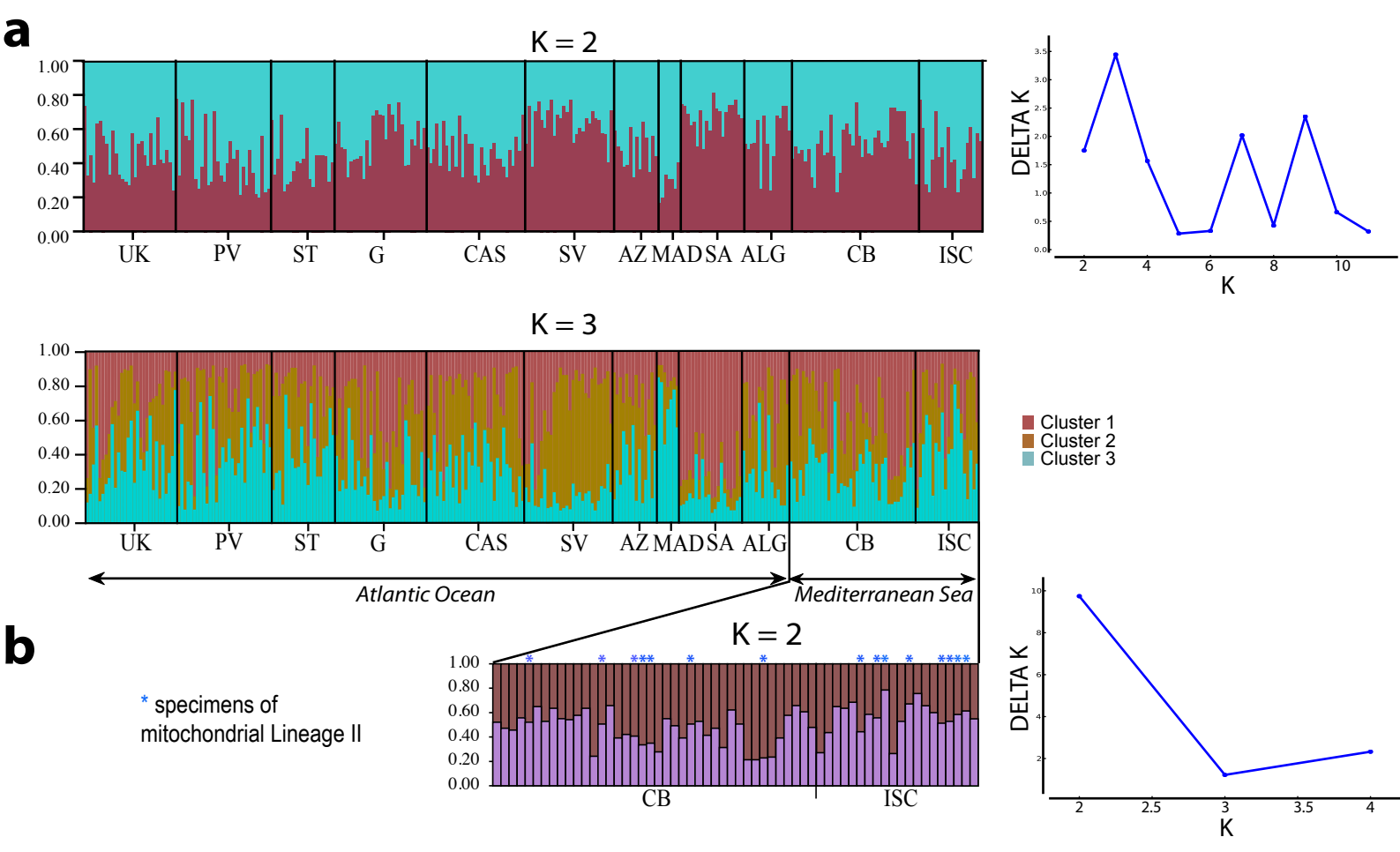
SUPPORTING INFORMATION

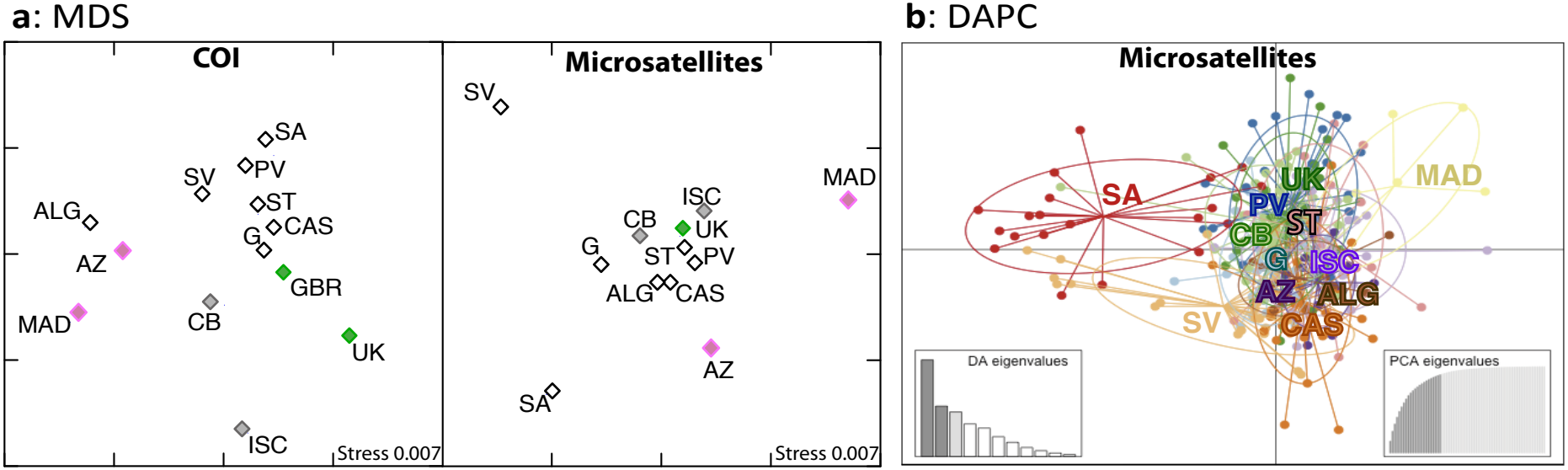
Appendix S1. Genetic analyses.

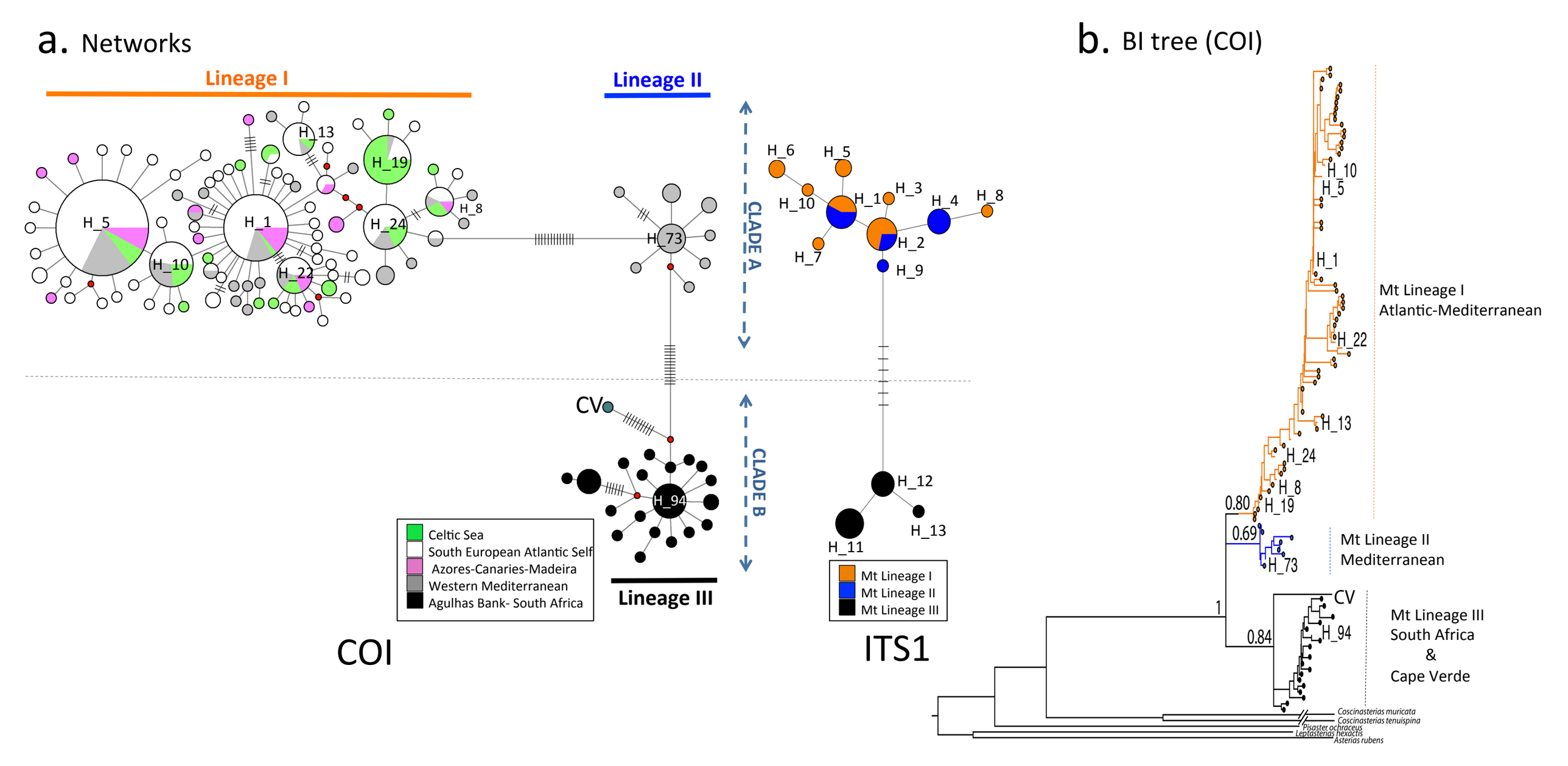
Appendix S2. Combined phylogenetic trees.

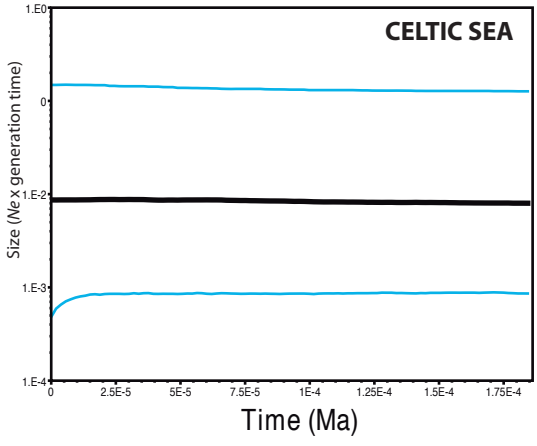
Appendix S3. Trees and networks: mitochondrial fragments.

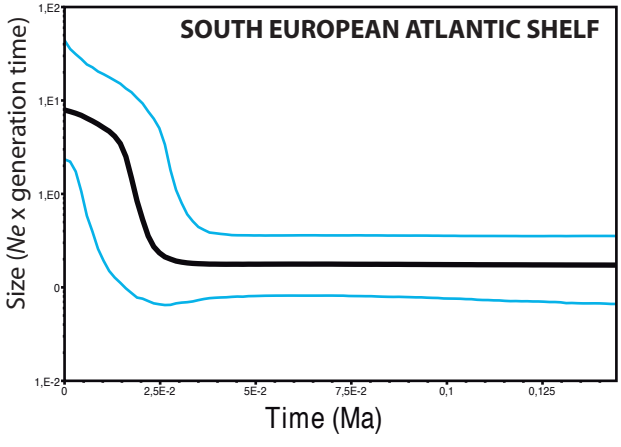

a
SOUTH EUROPEAN ATLANTIC SHELF ECOREGION
CELTIC SEA ECOREGION
IR
UK
GBR
ST
PV
G
CB
CAS
SP
XAB
SUR
ISC
SV-SA
ALG
FL
AZ
MAD
AZORES-CANARIES-MADEIRA ECOREGION
Mediterranean Sea
WESTERN MEDITERRANEAN ECOREGION
Atlantic Ocean
CAPE VERDE ECOREGION
CV
500 Km
COI HAPLOTYPES:
H_1 H_10 H_22 H_49 private of LI
H_5 H_13 H_24 H_64 private of LII
H_8 H_19 H_45 private of LIII
MARINE ECOREGIONS:
CELTIC SEA
AGULHAS BANK
SOUTH EUROPEAN ATL. SHELF
WESTERN MEDITERRANEAN
AZORES-CANARIES-MADEIRA
CAPE VERDE
Africa
AGULHAS BANK ECOREGION
Atlantic Ocean
Indian Ocean
SDA
SDI
500 Km
b
UK
ST
PV
G
CAS
CB
ISC
AZ
SV-SA
ALG
MAD
Mediterranean Sea
Atlantic Ocean
500 Km
STRUCTURE clusters:
Cluster 1
Cluster 2
Cluster 3
Mixed

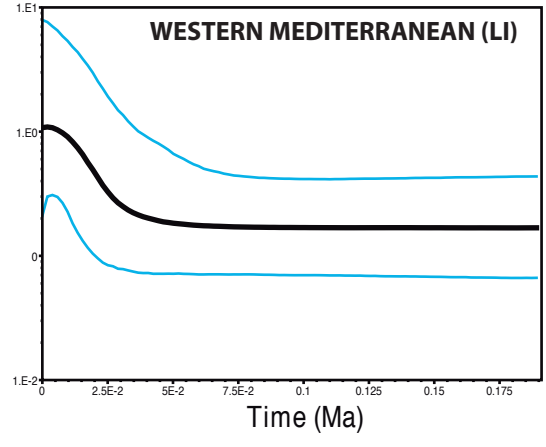

a
K = 2
1.00
0.80
0.60
0.40
0.20
0.00
UK
PV
ST
G
CAS
SV
AZ
MAD
SA
ALG
CB
ISC
DELTA K
K
K = 3
Cluster 1
Cluster 2
Cluster 3
Atlantic Ocean
Mediterranean Sea
b
* specimens of
mitochondrial Lineage II
K = 2
CB
ISC

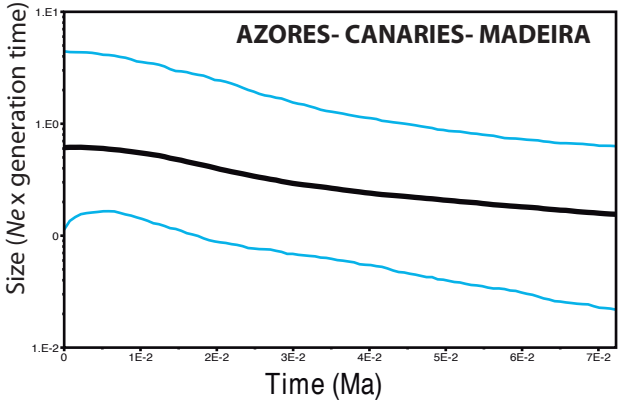

a: MDS
COI
SA
PV
SV
ST
ALG
CAS
G
AZ
GBR
MAD
CB
UK
ISC
Stress 0.007
Microsatellites
SV
MAD
ISC
CB
UK
G
ST
PV
ALG
CAS
AZ
SA
Stress 0.007
b: DAPC
Microsatellites
SA
UK
PV
ST
MAD
CB
G
ISC
AZ
ALG
SV
CAS
DA eigenvalues
PCA eigenvalues

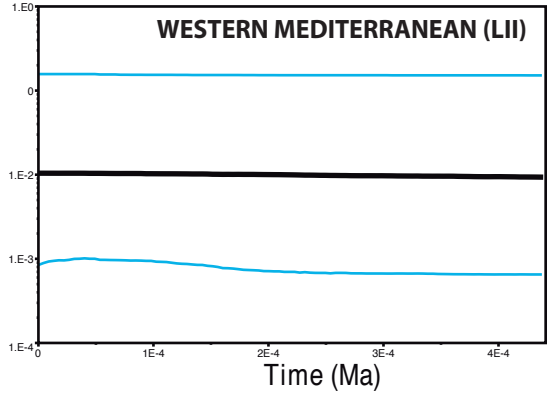

a. Networks
Lineage I
Lineage II
H_13
H_19
H_5
H_1
H_24
H_8
H_10
H_22
H_73
CLADE A
H_6
H_5
H_10
H_1
H_3
H_4
H_8
H_2
H_7
H_9
CV
H_94
CLADE B
H_12
H_13
H_11
Celtic Sea
South European Atlantic Self
Azores-Canaries-Madeira
Western Mediterranean
Agulhas Bank- South Africa
Lineage III
Mt Lineage I
Mt Lineage II
Mt Lineage III
COI
ITS1
b. BI tree (COI)
H_10
H_5
H_1
H_22
H_13
H_24
H_8
H_19
0.80
0.69
H_73
CV
H_94
1
0.84
Mt Lineage I
Atlantic-Mediterranean
Mt Lineage II
Mediterranean
Mt Lineage III
South Africa
&
Cape Verde
Coscinasterias muricata
Coscinasterias tenuispina
Pisaster ochraceus
Leptasterias hexactis
Asterias rubens

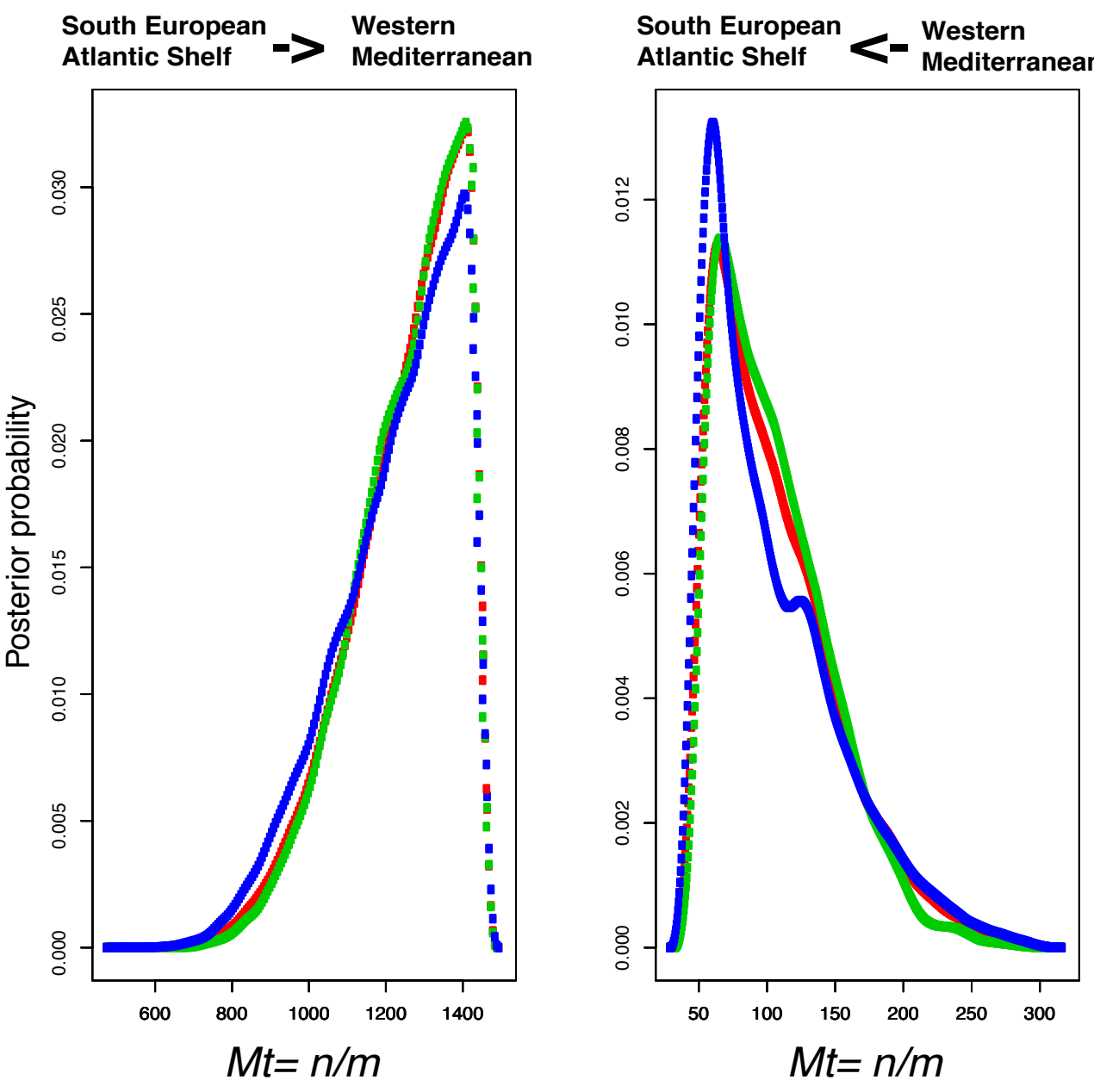

CELTIC SEA
Size (Ne x generation time)
1.E0
0
1.E-2
1.E-3
1.E-4
0
2.5E-5
5E-5
7.5E-5
1E-4
1.25E-4
1.5E-4
1.75E-4
Time (Ma)


SOUTH EUROPEAN ATLANTIC SHELF
Size (Ne x generation time)
1,E2
1,E1
1,E0
0
1,E-2
0
2,5E-2
5E-2
7,5E-2
0,1
0,125
Time (Ma)


WESTERN MEDITERRANEAN (LI)
1.E1
1.E0
0
1.E-2
0
2.5E-2
5E-2
7.5E-2
0.1
0.125
0.15
0.175
Time (Ma)


AZORES- CANARIES- MADEIRA
Size (Ne x generation time)
1.E1
1.E0
0
1.E-2
0
1E-2
2E-2
3E-2
4E-2
5E-2
6E-2
7E-2
Time (Ma)


WESTERN MEDITERRANEAN (LII)
1.E0
0
1.E-2
1.E-3
1.E-4
0
1E-4
2E-4
3E-4
4E-4
Time (Ma)


South European Atlantic Shelf
->
Western Mediterranean
South European Atlantic Shelf
<-
Western Mediterranean
Posterior probability
0.000
0.005
0.010
0.015
0.020
0.025
0.030
600
800
1000
1200
1400
Mt= n/m
0.000
0.002
0.004
0.006
0.008
0.010
0.012
50
100
150
200
250
300
Mt= n/m

Journal of Biogeography

SUPPORTING INFORMATION

Lineage splitting, secondary contacts and genetic admixture of a widely distributed marine invertebrate

Pérez-Portela R, Rius M, Villamor A

Appendix S1. Genetic analyses. Details of the genetic analyses performed in this study

DNA amplification and sequencing

PCR conditions were as described in Pérez-Portela et al. (2010), and annealing temperatures ranged between 40ºC and 65ºC (See the Table S1 below). Specific primers for the 16S fragment were designed with the software Primer3 0.4.0 (http://www.fokker. wi.mit.edu/primer3/input.htm) from an alignment of 16S gene sequences of starfish obtained from Genbank. Primers used for the CR amplified a fragment containing the final part of the 12SrRNA gene, the tRNA-Glu, the putative control region and a portion of the 16SrRNA gene, not overlapping with the fragment amplified with specific primers 16SMg_F and 16SMg_R. Details on the primers used are found in the Table S1 below. The same primer pairs were used for the sequencing reaction, and products were sequenced in an ABI Prism 377XL automated sequencer (Scientific and Technical Services of the University of Barcelona- STS). Sequences were edited and aligned using Bioedit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and Clustal X, and the alignments confirmed by eye.

We tested for signs of selection in the three mitochondrial markers with the Hudson-Kreitman-Aguadé test (HKA) as implemented in DNAsp 5.10 (http://www.ub.edu/dnasp/DnaSP_OS.html) and using *Coscinasterias tenuispina* (sequenced in this study), *C. muricata*, *Pisaster ochraceus*, *Leptasterias hexactis*, and *Asterias rubens* as the outgroup (Accession numbers for COI: "*pending*", EU869907.1, HM542340.1, HM542232.1, GU672414.1; 16S: AY706147.1, DQ297083.1, DQ297110.1 DQ297094.1, AY652504.1; and CR: "*pending*", AY220848.1, M25320.1, EF555661.1, EF179779.1). Intra-locus recombination was also tested for the nuclear intron ITS1 with the same software, and saturation for

the COI was analysed with the software Dambe 5.3.74 (https://it.stonybrook.edu/software/title/dambe).

Signals of saturation were not detected with the index of substitution saturation (I_{ss}) always under the critical index of substitution saturation ($I_{ss.c}$). HKA tests were not significant, indicating no signs of selection in any of the mitochondrial markers, nor recombination was detected in the ITS1 (estimates of R= 0.001, minimum number of recombination events Rm = 0).

Microsatellites isolation, amplification and genotyping

Genomic DNA from one individual of *M. glacialis* collected in Cabrera Island (NW Mediterranean; 39º08'N, 2º54'E) was extracted using QIAamp DNA Minikit columns (QIAGEN, www.qiagen.com). A genomic library enriched for four repeated motifs was constructed as follows: DNA was digested and ligated using the FIASCO protocol (Zane *et al.* 2002), hybridized to four 5' biotinylated probes $(CA)_{15}$, $(GA)_{15}$, $(CAA)_{10}$, $(GATA)_{10}$ (Sygma-Genosys, www.sigmaaldrich.com) and captured with streptavidin-coated magnetic particles (Promega). Recovered DNA was amplified in 20 cycles of PCR, purified and cloned using pGEM-T Easy Vector II (Promega). White colonies were transferred onto membranes and screened for positive clones using digoxigenin labeled probes. Positive clones were sequenced on an ABI Prism 3700 automated sequencer (Applied Biosystems) in the STS of the University of Barcelona. Sequences were screened for microsatellite motifs and primers were designed using the software Primer3 0.4.0. Twelve pairs were selected for amplification tests and nine consistently amplified genomic DNA. Four pairs were later rejected as they amplified either monomorphic fragments for the populations studied (Accession n^{o}: GU827388) either fragments with a pattern of multiple peaks, difficult to interpret.

Forward primers for each locus were labelled with fluorescent dyes (Applied Biosystems) (see table below) and used in a 7:3 proportion with unlabelled forward primer. PCR parameters were: 5 min denaturation at 95°C, 40 amplification cycles of 30 sec at 95°C, 30 sec at locus specific annealing temperature (See the table below), and 30 sec at 72°C, followed by a final extension cycle of 4 min at 72°C. Resulting PCR products were genotyped using an ABI Prism 377XL automated sequencer. Allele

size characterization was performed with an internal standard (LIZ, Applied Biosystems) using PeakScanner™ software 1.0 (Applied Biosystems).

Linkage disequilibrium between microsatellite loci was computed in Arlequin 3.1 software before population genetic analyses, and there was no evidence of linkage between any pair of them (P >0.01). Details about diversity levels of the five microsatellite loci are presented in Table S2.

Table S1: Details of the primer pairs used in this study. DNA fragment amplified, sequences of the primers used and Accession numbers (for microsatellites), temperature of annealing (T^a), reference of the work where those primers were published, and evolution model applied for phylogenetic reconstruction or repeat motif (for microsatellites). For microsatellite markers the dye used for each forward primer is also presented.

DNA fragment	**Primers**	**T^a**	**Reference**	**Evolution model**
Cytochrome c oxidase s ubunit I (COI)	MgCOI_F 5' TCTCATATT TGG GC TTGAG 3' MgCOI_R 5'TAGGTG TTGAAA GAGAAT GG 3'	40°- 45°C	Pérez-Portela *et al*. 2010	HKY+G G= 0.1450; ti/tv= 5.4956
16S rRNA (16S)	16SMg_F 5' GTGGAATAAGGAGTTCTG 3' 16SMg_R 5' AAGATTTTAATG GTCGAA 3'	50°-52°C	Specifically designed for this study	HKY ti/tv= 5.8560
Control Region (CR)	E12Sa 5' ACACATCGCCCG TCACTC TC 3' E16Sb 5' GACGACAAGACCCTATCGAGC C 3'	50°-52°C	Waters *et al*. 2004	K80 ti/tv= 8.6644
First internal transcribed spacer (ITS1)	ITS1 5' TCC TAGGTGAACCTGCG G 3' ITS2 5' GCTGCGTTCTTCATCGATGC 3'	65°-55°C touchdown	White *et al*. 1990	GTR+I+G I= 0.6680; G= 0.0290
Histone 3 (H3)	H3_F 5'ACAATGCGYCGYACYAAGCAGACAGC 3' H3_R 5' GTT GGATRTCYTTGGGCATGATGGT 3'	65°-55°C touchdown	Foltz & Mah 2009	-
Microsatellite name/ Acc. number	**(Dye) Primers**			**Repeat motif**
MG GATA 2.59/ GU734027	(6 FAM) F 5' ACAATAGAGTCCCCGGTTCC 3' R: GCCTTTGATGGAAAGTGAGAG 3'	53°-55°C	Specifically designed for this study	$(GT)^{10}GG\ (GT)^{3}$
MG AG4.10/ GU734028	(NED) 5' GATGACAGATTGTGGGTTTGA 3' R 5'AGTCTGAGATGGGCCAAAGA 3'	54°-55°C	Specifically designed for this study	$(CA)^{8}$
MG AC1.43/ GU734029	(VIC) F 5' TTGCTACGATGAAGATGCTGA 3' R 5' CAACCGAAGAAAGAGCGAAG 3'	55°C	Specifically designed for this study	$(GT)^{9}$
MG AC1.24/ GU734031	(VIC) F 5' GTTGGGAGGCTGAAAATCAA 3' R 5 'CATTTCGCGTTCAAATCCTC 3'	55°C	Specifically designed for this study	$(GT)^{8}$
MG AC5.52/ GU734033	(PET) F 5' GGCCTTAGAAGTGGGTCACA 3' R 5' TTGATAAATGCGCTGCCAAT 3'	55°C	Specifically designed for this study	$(GA)^{6}$

Table S2. Genetic diversity of *M. glacialis* microsatellite loci. Name of the marker (LOCUS), total number of alleles found in the whole dataset (Ntotal), number of alleles per population and locus (N), number of private alleles (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), and Fixation index (F_{IS}). * Significant at P <0.05, and*** Significant at P <0.01

		Plymouth					Galicia					Basque Country					Santander					Cascais				
LOCUS	Ntotal	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}
Margla1	17	10	1	0.41379	0.77737	0.4721***	10	0	0.51724	0.75197	0.3160*	9	0	0.6000	0.75989	0.2133	10	1	0.350000	0.83077	0.5850***	11	0	0.54839	0.78107	0.3014***
Margla2	16	8	0	0.75000	0.72143	-0.0404	8	1	0.51724	0.58137	0.1121	8	0	0.56667	0.70452	0.1984	7	0	0.45	0.80128	0.4448**	9	0	0.87097	0.81861	-0.0651
Margla3	14	6	0	0.82759	0.80883	-0.0236	8	1	0.72414	0.81549	0.1138*	8	0	0.63333	0.82429	0.2347*	11	2	0.70	0.85128	0.1815	9	0	0.74194	0.83025	0.1080
Margla4	14	12	1	0.58621	0.75136	0.2229*	6	0	0.41379	0.46098	0.1040	6	0	0.43333	0.66497	0.3522***	8	0	0.65	0.67692	0.0408	7	0	0.35484	0.44051	0.1971*
Margla5	7	2	0	0.48276	0.45977	-0.0509	3	0	0.41379	0.38052	-0.0891	4	1	0.33333	0.51356	0.3548*	2	0	0.30	0.46667	0.3631	2	0	0.48387	0.48916	0.0110

Sao Vicente					Sagres					Algarve					Azores					Madeira					Costa Brava					Ischia				
N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}	N	N_A	H_O	H_E	F_{IS}
8	2	0.48148	0.60377	0.2056**	5	0	0.2	0.27821	0.2864*	7	1	0.93333	0.82989	-0.1297	10	1	0.57143	0.78836	0.2828*	6	0	0.42857	0.76923	0.4627*	12	0	0.575	0.85222	0.3281**	7	0	0.5	0.83333	0.4062***
9	0	0.57143	0.74545	0.2367	6	0	0.35	0.68205	0.4933***	7	1	0.6	0.77241	0.2294**	7	0	0.5	0.71429	0.3080*	5	0	0.71429	0.76923	0.0769	11	0	0.55263	0.62737	0.1205	6	0	0.78947	0.77952	-0.0131
8	0	0.78571	0.82597	0.0496	8	0	0.65	0.83846	0.2293	8	0	0.78571	0.75132	-0.0476	6	0	0.71429	0.71958	0.0076	7	0	0.71429	0.82418	0.1429	9	0	0.75	0.78291	0.0426	7	1	0.85	0.71538	-0.1941
1	0			-	4	0	0.35	0.35513	0.0148	4	0	0.2	0.30345	0.3488	2	0	0.07692	0.07692	-0.0000	5	1	0.57143	0.72527	0.2258	8	0	0.55	0.50285	-0.0951	10	1	0.75	0.78077	0.0404
2	0	0.03571	0.10325	0.6582	4	1	0.29412	0.31551	0.0698	2	0	0.2	0.48046	0.5922*	3	0	0.64286	0.51587	-0.2581	2	0	0.57143	0.52747	-0.0909	3	0	0.17949	0.32334	0.4481**	2	0	0.35	0.51154	0.3214

Phylogenetic analyses, Beast and Lamarc

Phylogenetic analyses and divergence: The best-fit models of nucleotide substitution per genetic marker, COI, 16S, CR and ITS1, was selected by statistical comparison of 88 different models of evolution with the program Jmodeltest 2.1.3 (http://code.google.com/p/jmodeltest2/) with the Akaike Information Criterion (AIC). For the ML approach we used the software PhyML 3.0 (http://www.atgc-montpellier.fr/phyml/). We implemented the model calculated before, and 1,000 bootstrap replicates were run to assess the robustness of the nodes. We used as outgroups the species: *C. tenuispina*, *C. muricata*, *P. ochraceus*, *L. hexactis*, and *A. rubens*. For the nuclear ITS1 only the species *A. rubens* and *A. forbesi* were included as outgroups due to the difficulty to align sequences of this fragment (accession numbers AF346614.1, AF346615.1 and AF346610.1). For BI reconstruction and inference of divergence times between lineages we used the software BeaUti 2 and Beast 2.1.2. Divergences were estimated for the concatenated data set of the three mitochondrial markers. Unlinked substitution models and clock models were set for all three partitions. Substitution models and site heterogeneity model were set according to the best fit model as selected in jModeltest for each partition (see Table S1). Two independent runs were performed setting the COI clock model as fixed to 2.2% and 2.8%, according to the newest estimates for the closest relative of *Marthasterias* for which information is available (Foltz *et al*., 2013 and references herein). The other two partition clock models were estimated every time as unlinked lognormal relaxed clocks in order to allow for independent rates across different branches. Tree prior selected was Yule-Process. Length of chain was 10^7 and number of chains was 10 for all reconstructions. Tracer 1.6 was used to check the convergence to a stationary distribution of parameters; TreeAnnotator, with a burnin of 1,000 summarized all trees into a single target tree, which was visualized and edited in FigTree 1.4.

Demographic analysis: The coalescent-based Bayesian skyline plots (BSP) were used to estimate past demographic events for different ecoregions based on the COI. The BSP *priors* included the implementation of the substitution model, a strict clock model, and the constant skyline model in Beast. As explained before a clock model was fixed to 2.2% and 2.8%. Between 10 and 50 million MCMC generations were run per area, sampled every $1{,}000^{th}$–$5{,}000^{th}$ step, and a burnin of 1,000–10,000. We ensured that the runs were long enough and above 200 ESS by visualizing the result outputs

in Tracer. Tracer was also used to generate the evolution of effective population size under the skyline plot model, expressed as N_eT (T = generation time) over time.

Lamarc: A Bayesian search strategy applied for COI was used to estimate migration across the Gibraltar Strait, between the South European Atlantic Shelf and Western Mediterranean regions. Effective population size had to be also calculated before migration with the same software. The strategy included several "initial-runs" (a total of six) to find the most accurate *priors* for our dataset. The first run was performed with default parameters, and five additional runs were computed with most accurate parameters obtained from previous runs. The "initial-runs" included three different replicates with 5 initial chains and six final chains. Once accurate *priors* were obtained, they were implemented in a "definitive" run, with variation values of migration between 550 and 1,900 from the South European Atlantic Shelf to the Western Mediterranean, and 0 and 300 in the reverse way. The definitive run was based on three different replicates with 20 initial chains of 5,000 MCMC each, burning of 1,000, and six final chains of 50,000 MCMC each, with a burning period of 5,000. Two simultaneous heating searches (1 and 1.2) were performed per replicate. The best evolutionary model as inferred with jModelTest was implemented. To visualise whether the run was long enough reaching a plateaus of probability, and to confirm the existence of at least 200 independent simulations (Effective Sample Size-ESS) for each parameter, results were summarised per replicate and then combined for all replicates in Tracer. Migration rates (*Mt*) were expressed as the number of migrants per generation $Mt = n/m$, where n is the immigration rate per generation, and m the substitution rate. Effective population size (N) was expressed as the logarithm of *Theta (2Nm).* Distribution of the posterior probabilities for migration and *Theta* were then graphically represented in R.

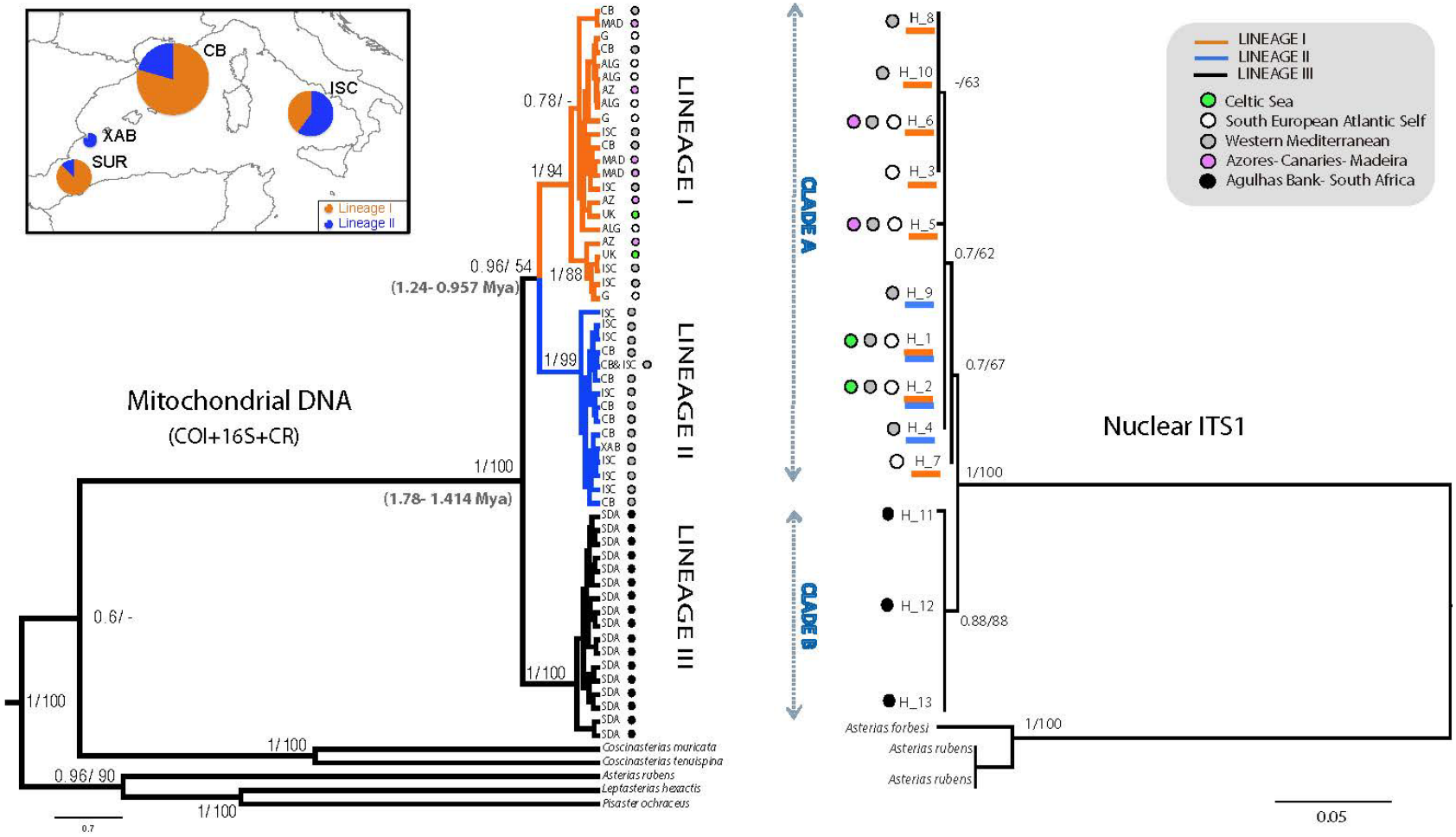
References

Foltz, D. W., Fatland, S. D., Eléaume, M., Markello, K., Howell, K. L., Neill, K. & Mah, C. L. (2013) Global population divergence of the sea star *Hippasteria phrygiana* corresponds to the onset of the last glacial period of the Pleistocene. *Marine Biology*, **160**, 1285-1296.

Pérez-Portela, R., Villamor, A. & Almada, V. (2010) Phylogeography of the sea star *Marthasterias glaci alis* (Asteroidea, Echinodermata): deep genetic divergence between mitochondrial lineages in the nor th-western Mediterranean. *Marine Biology,* **157**, 2015-2028.

143 Waters, J. M., O'Loughlin, P. M. & Roy, M. S. (2004). Cladogenesis in a starfish species complex from
144 southern Australia: evidence for vicariant speciation?. *Molecular Phylogenetics and Evolution*, **32**,
145 236-245.
146 White, T. J., Bruns, T., Lee, S. J. W. T. & Taylor, J. W. (1990). Amplification and direct sequencing of
147 fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and*
148 *applications*, **18**, 315-322.
149 Zane, L., Bargelloni, L. & Patarnello, T. (2002) Strategies for microsatellite isolation: a review.
150 *Molecular Ecology* **11**, 1–16.
151

Appendix S2. Combined phylogenetic trees. Bayesian Inference trees of *Marthasterias*, including Maximum Likelihood bootstrap supports and posterior probability on the nodes, generated from sequences of the concatenated mitochondrial fragments, and the ITS1 nuclear fragment. Values of support below 0.5 and 50% are not shown (-). The scale bar represents the number of nucleotide substitutions per site. On the mitochondrial tree divergence times between lineages, population codes and biogeographical origin are also presented. On the ITS1 tree, both biogeographical origin and mitochondrial lineage to which specimens belong are shown. A map of the Mediterranean area showing the distribution of Lineage I and II is also presented. On the map, pie charts represent mitochondrial lineage frequencies per locality, and the size is proportional to the number of samples analysed.



Appendix 3. Trees and networks: mitochondrial fragments. Bayesian inference trees, including posterior probabilities on the nodes, and networks separately reconstructed for the 16SrDNA and the Control Region (CR) fragments.