

Research Article

Phylogeographic sampling guided by species distribution modeling reveals the Quaternary history of the Mediterranean–Canarian *Cistus monspeliensis* (Cistaceae)

Alberto J. Coello^{1,2*} , Mario Fernández-Mazuecos¹ , Carlos García-Verdugo³ , and Pablo Vargas¹ 

¹Departamento de Biodiversidad y Conservación, Real Jardín Botánico (RJB-CSIC), Plaza de Murillo 2, Madrid 28014, Spain

²Escuela Internacional de Doctorado, Universidad Rey Juan Carlos, C/Tulipán s/n, Móstoles 28933, Spain

³Departamento de Biología, Universitat de les Illes Balears, Ctra. Valldemossa km 7,5, Palma 07122, Balearic Islands, Spain

*Author for correspondence. E-mail: ajcoello@rjb.csic.es

Received 22 July 2019; Accepted 17 January 2020; Article first published online 21 January 2020

Abstract Accurate inference in phylogeography requires appropriate sampling strategies. Complex questions demand a large sample size at both the population and genetic levels to obtain precise reconstructions. This is the case of the phylogeographic history of *Cistus monspeliensis*, a plant that displays low plastid (cpDNA) diversity in the Mediterranean Basin but high diversity in the Canary Islands. Here, we aimed to identify Mediterranean refugial areas and to accurately quantify inter-island colonization events in the Canaries. Using a previous study as starting point, we increased sample size in two ways: (i) additional sampling of plastid genetic markers (from 1041 to 1899 bp); and (ii) additional sampling of populations (from 47 to 69) in long-term persistence areas suggested by species distribution modeling (SDM). The synergy between SDM and extended population sampling helped find higher genetic diversity. Our deeper phylogeographic sampling of *C. monspeliensis* revealed the following: (i) potential refugia in long-term persistence areas with high cpDNA diversity in western Europe and the Canary Islands; and (ii) a significant increase (from 7 to 12) in the number of inferred inter-island colonization events across the archipelago. Our results stress the usefulness of SDM to identify the genetic signature associated with potential refugial areas. We herein propose a field sampling approach based on SDM that, in combination with a larger cpDNA sampling, can help answer a wide array of phylogeographic questions, such as the location of Quaternary refugia and number of colonizations across archipelagos.

Key words: Canary Islands, Cistaceae, *Cistus monspeliensis*, ecological niche modeling, Mediterranean Basin, phylogeography, Quaternary refugia, sampling strategy, species distribution modeling.

1 Introduction

Phylogeography is aimed at analyzing the geographic distribution of genetic lineages, thus bridging population genetics (microevolution) and systematics (macroevolution) (Avice et al., 1987; Avice, 2009). Phylogeographic analyses help discover climatic refugial areas (Provan & Bennett, 2008), reconstruct migration routes (Hewitt, 1996; Taberlet et al., 1998), and infer events of oceanic island colonization (Caccone et al., 2002; Vargas et al., 2015; Shaw & Gillespie, 2016). The ability to infer phylogeographic patterns is determined by some interrelated factors, such as the complexity of the evolutionary history of species and the geographic extent of distribution areas (Avice, 2009). Nevertheless, an unrepresentative sample frequently hinders the accurate inference of phylogeographic patterns, and only an appropriate sample size can render reliable results (Table 1).

The sampling effort needed to reach a suitable sample size can be addressed in two basic ways: (i) by increasing genetic sampling (length of sequenced DNA from the nuclear,

mitochondrial, or plastid genomes) and (ii) by increasing field sampling (number of individuals/populations) (Björklund & Bergek, 2009). The genetic sampling approach has been explored in the last decades, leading to increasingly large genetic datasets, particularly in recent years, owing to next-generation sequencing (NGS) techniques (Garrick et al., 2015). The great deal of data provided by NGS can accomplish goals not achievable with Sanger sequencing at a relatively low cost (Egan et al., 2012). However, NGS data are predominantly obtained from the nuclear genome, whereas phylogeography traditionally uses a single, variable locus lacking recombination (typically maternally inherited) and with a high mutation rate, such as the mitochondrial (mtDNA) and plastid (cpDNA) genomes (Avice et al., 1987). These characteristics are key to reconstruct phylogeographic patterns such as the number of inter-island colonization events across oceanic archipelagos (Vargas et al., 2015; García-Olivares et al., 2017).

Regarding field sampling, it is clear that covering the entire distribution of a species is the most appropriate approach to

Table 1 Phylogeographic studies of Canarian and Mediterranean species where sampling effort was increased with respect to a previous study of the same species

Species	Geography	Differences in sampling effort	Results after increased sampling
Plants			
<i>Astragalus edulis</i> (Fabaceae)	Canary Islands	Increase in population number	Compared to an early analysis (Peñas et al., 2016), an increased sample size, together with new statistical analyses, allowed Bobo-Pinilla et al. (2018) to infer ancient colonization of the Canary Islands, probably from western Morocco.
<i>Canarina canariensis</i> (Campanulaceae)	Canary Islands	Increase in genetic markers	Compared to a previous study focused on unveiling phylogenetic relationships in the genus (Mairal et al., 2015a), a subsequent study was focused on the Canary Islands and added genetic information (AFLPs), which helped detect a strong population genetic structure and determine the relevance of palaeo-islands as refugia (Mairal et al., 2015b).
<i>Cistus monspeliensis</i> (Cistaceae)	Mediterranean Region and Canary Islands	Increase in population number and genetic markers	Our increase in sampling effort with respect to the previous study (Fernández-Mazuecos & Vargas, 2011) helped discover a refugial area in western Iberia and detect nearly twice as many colonizations across the Canary Islands (see main text for further details).
<i>Euphorbia</i> sect. <i>Aphyllis</i> subsect. <i>Macaronesicæ</i> (Euphorbiaceae)	Macaronesia (and some locations in the Mediterranean Region)	Increase in the population (and species) number	A more intense sampling of populations by Barres et al. (2017) enabled the detection of two independent events of back-colonization (from the Canary Islands to the mainland) in two species of this group (<i>E. regis-jubae</i> and <i>E. pedroi</i>), one of which was previously unknown (Sun et al., 2016).
<i>Olea europaea</i> subsp. <i>guanchica</i> (Oleaceae)	Canary Islands	Increase in population number and genetic markers	García-Verdugo et al. (2010a) suggested recurrent seed dispersal (colonization) between islands in the Canarian archipelago. Subsequent intense sampling (García-Verdugo et al., 2010b) allowed the definition of genetic discontinuities among islands and populations despite high dispersal between islands.
<i>Olea europaea</i> subsp. <i>europaea</i> (Oleaceae)	Mediterranean Region	Increase in population number and genetic markers	Besnard et al. (2013) precisely defined three refugia and the cradle of the first domestication of the olive tree, unknown in the previous study (Besnard et al., 2002).
<i>Periploca laevigata</i> (Apocynaceae)	Canary Islands	Increase in population number [†] and genetic	Compared to the previous study (García-Verdugo et al., 2015), the expanded dataset of García-

Continued

Table 1 Continued

Species	Geography	Differences in sampling effort	Results after increased sampling
Animals			
<i>Pecten jacobaeus</i> (bivalve)	Mediterranean Region	Increase in population number	Verdugo et al. (2017) rejected the monophyly of Canarian populations (pattern of multiple colonizations).
<i>Stenosoma nadejda</i> (malacostracan)	Mediterranean Region	Increase in population number [†]	A first analysis (Saavedra & Peña, 2005) suggested some degree of eastern–western population differentiation in the Mediterranean, but the addition of new populations ruled out this differentiation (Telahigue et al., 2018). Compared to the original study (Xavier et al., 2009), the first increase in population number suggested vicariance in multiple glacial refugia (Xavier et al., 2011), and a further increase suggested long-distance dispersal events and allopatry in Mediterranean populations (Xavier et al., 2011, Xavier et al., 2016).
<i>Cyanistes teneriffae</i> (bird)	Canary Islands	Increase in genetic markers	The use of genomic data, together with traditional Sanger sequencing, helped Stervander et al. (2015) detect three independent colonizations of the Canarian archipelago and reject the previously hypothesized back-colonization to Africa (Päckert et al., 2013).
<i>Miniopterus schreibersii</i> (mammal)	Mediterranean Region	Increase in population number [‡]	Compared to an earlier analysis (Furman et al., 2010), the expanded dataset of Bilgin et al. (2016) redefined phylogeographic patterns of the species: a glacial refugium was located in a previously unconsidered area and a new pattern of expansion in the Mediterranean Region was proposed.
<i>Podarcis muralis</i> (sauropsid)	Mediterranean Region	Increase in population number and genetic markers	The expanded sampling (Salvi et al., 2013) helped propose Mediterranean and extra-Mediterranean glacial refugia for the species, beyond the previously proposed refugia in the Italian peninsula (Giovannotti et al., 2010).

Note: [†]Populations used in the study of García-Verdugo et al. (2017) were similar to those used in the previous study (García-Verdugo et al., 2015), but the number of individuals per population was increased; [‡]Although the number of genetic markers was also increased, addition of 28S in Xavier et al. (2011) with respect to Xavier et al. (2009) and addition of Elongation factor 1 α in Xavier et al. (2016) with respect to Xavier et al. (2009), only Cytochrome c oxidase subunit I gene (COI) sequences were used to interpret the effect of adding new populations; [§]Bilgin et al. (2016) also sequenced cytochrome b (cyt-b), in addition to the hypervariable region I (HV1, present in both studies); the main interpretation is based on HV1.

phylogeographic analysis, whereas local studies are less informative (Table 1). However, the high cost of extensive sampling complicates the study of widespread species.

As repeatedly reported in phylogeographic studies, the genetic diversity of a species usually displays a heterogeneous distribution across populations. As a general pattern, areas that have remained environmentally suitable for a long time (long-term persistence areas or refugia) display higher genetic diversity (Russell et al., 2014; Dalmaris et al., 2015), and thus the discovery of these refugia is one of the prime aims in phylogeography. In this regard, the Mediterranean Basin is one of the most studied regions of the world. A variety of Mediterranean refugia sheltering plant species during adverse periods (such as glaciations) have been suggested, with many of them being located in the three southern European peninsulas (Médail & Diadema, 2009). Nevertheless, refugia are not only found in the mainland, as oceanic archipelagos also display areas that have long remained climatically stable and suitable for plant survival, notably of relict lineages (Vargas, 2007). Although the complex dynamics of oceanic archipelagos (derived mainly from volcanic activity) complicates the identification of refugia, some areas that have remained stable in the long term seem to be particularly relevant as refugia, such as the palaeo-islands of the Canary archipelago (Mairal et al., 2015b).

Species distribution modeling (SDM) is a powerful tool to reconstruct potential distributions (Alvarado-Serrano & Knowles, 2014), and it has often been used, in combination with genetic results, to provide evidence of refugial areas (e.g., Migliore et al., 2018; Villa-Machío et al., 2018). However, to the best of our knowledge, SDM has not been routinely used to guide sampling strategies in phylogeographic studies with the aim of including potential refugial areas (see comments in Waltari et al., 2007).

In this context, *Cistus monspeliensis* (Fig. 1C) emerges as an ideal system to quantify the net contribution of increasing either genetic or field sampling and thus answer unsolved questions in widespread species. Interestingly, *C. monspeliensis* was the first plant species, simultaneously with *Radula lindenberghiana* (Laenen et al., 2011), where higher levels of genetic diversity were discovered on oceanic islands than in the continental region of origin (Fernández-Mazuecos & Vargas, 2011). Later, this pattern was found in additional plant species (Désamoré et al., 2012; Patiño et al., 2015; García-Verdugo et al., 2017). However, *C. monspeliensis* remains the most unbalanced example of genetic diversity distribution, which raises questions that remain unsolved. Fernández-Mazuecos & Vargas (2011) postulated that Quaternary climate changes in the Mediterranean Basin caused a massive extinction of mainland populations. This extinction may have

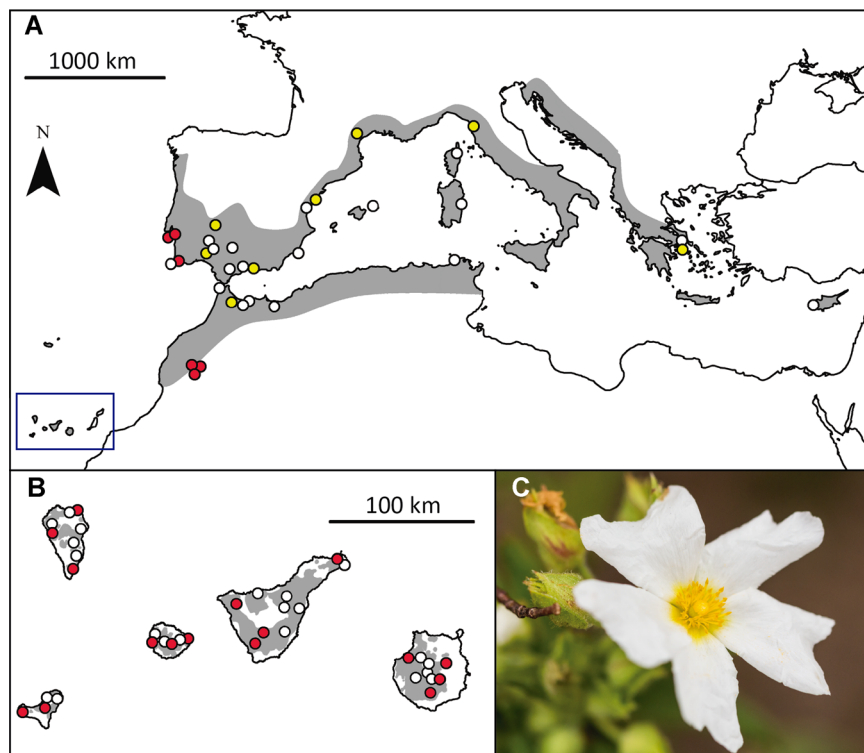


Fig. 1. Distribution range and sampling localities of *Cistus monspeliensis* in **A**, the Mediterranean Basin and **B**, the five westernmost islands of the Canary archipelago; and **C**, a photo of the species from Tenerife. Populations with sequence data taken from previous studies (Fernández-Mazuecos & Vargas, 2010, 2011) are shown as white dots; populations sampled in previous studies but with an increased number of individuals sequenced in the present study are shown as yellow dots; and populations newly sampled for this study are shown as red dots. The known distribution of *C. monspeliensis* is shaded in gray in the Mediterranean Basin (based on Martín & Guinea, 1949) and the Canary Islands (based on Atlantis 3.0.ob1, www.biodiversidadcanarias.es). The location of the Canary Islands is shown as a blue square in A. Photograph by A. J. Coello.

resulted in a dramatic genetic bottleneck, followed by recent expansion from some refugia, a pattern also described in other Mediterranean plants (Vendramin et al., 2008; Mateu-Andrés et al., 2015). Nevertheless, the location of mainland refugia remains unknown for *C. monspeliensis*. In contrast, the more stable climate of the Canary Islands helped preserve high levels of genetic diversity (Cronk, 1992; Médail & Quézel, 1999) and favored a complex history of inter-island colonization. The question remains as to whether the current estimate of inter-island colonization events in *C. monspeliensis* (see Vargas et al., 2015) would change after increasing sampling effort.

Here, we reanalyze the Quaternary history of *C. monspeliensis* in the Mediterranean Basin and the Canary Islands by implementing SDM and expanding population and genetic datasets. Our working hypothesis is that an increase in the sample size guided by SDM maximizes the opportunity of finding refugial areas and sources of inter-island colonization. This approach would lead to the detection of higher cpDNA diversity and a more reliable phylogeographic reconstruction than that of previous studies using a field sampling strategy based on geographic representativeness (Fernández-Mazuecos & Vargas, 2011). In the present study, we aim (i) to investigate the use of SDM in the design of field sampling strategies that include potential refugia; (ii) to evaluate if phylogeographic reconstruction is substantially improved by increasing field and genetic sampling effort in the Mediterranean Basin; and (iii) to contrast the effect of increased field and genetic sampling on the number of inferred colonizations across the Canarian archipelago.

2 Material and Methods

2.1 Study species

Cistus monspeliensis L. (Cistaceae) is a lowland shrub distributed in the Mediterranean Basin and the five westernmost islands of the Canarian archipelago (Fig. 1). It occurs from sea level to 1200 m on different types of substrates (acidic and calcareous), as part of the successional scrub of thermophilous, laurel, and pine forests (Martín & Guinea, 1949; Juhren, 1966; Demoly & Montserrat, 2005). This species is included in the white-flowered clade of the genus *Cistus* (Guzmán & Vargas, 2005), and it is genetically depauperate in the Mediterranean Basin compared to the Canary Islands, an archipelago colonized by the species <1 Mya (Fernández-Mazuecos & Vargas, 2010, 2011). In fact, the geographic distribution of cpDNA diversity allowed the inference of active inter-island colonization (at least seven colonizations) across the archipelago despite the absence of dispersal syndromes for *C. monspeliensis* diaspores (Fernández-Mazuecos & Vargas, 2011; Vargas et al., 2015).

2.2 Study area

The Mediterranean Basin comprises southern Europe, southwestern Asia, and northern Africa approximately between 27°–46°N and 10°W–38°E (Fig. 1A), and it has an annual drought of about three months in summer. The Canary Islands form an oceanic archipelago located between 27.5°–29.5°N and 13°–18.5°W (Fig. 1), which is less than 100 km from Africa and includes seven major islands dated between 1 and more than

20 Mya (Troll & Carracedo, 2016). The Canary Islands have a Mediterranean to subtropical climate influenced by proximity to the African coast, trade winds, and the wide altitudinal range of some islands.

2.3 Species distribution modeling

Higher levels of genetic diversity are expected in areas that have remained climatically favorable to *C. monspeliensis* throughout the Quaternary climatic cycles than in those areas of recent favorable conditions. To find areas with potentially high diversity, we used SDM based on 60 high-quality, spatially unbiased occurrences downloaded from the Global Biodiversity Information Facility (GBIF.org, 2016) (Figs. 2G, 2H; Doc. S1 and Table S1.1 in Doc. S1). We downloaded bioclimatic variables from WorldClim 1.4 (Hijmans et al., 2005) and selected seven that were not highly correlated in the study area (Doc. S1 and Table S1.2 in Doc. S1). The study area included the entire Mediterranean Basin, the Canary Islands, and Madeira. We included Madeira because there are unconfirmed historical records of *C. monspeliensis* for this archipelago (Short, 1994); however, the species is not currently considered to be part of the Madeiran flora (http://www4.uma.pt/gbm/checklist/lista_flora.php). To estimate a model under current conditions, we employed the maximum entropy algorithm, as implemented in Maxent 3.3.3k (Phillips et al., 2006), because it performs well with presence-only data (Elith et al., 2006) (see Doc. S1). The model was projected to the Mid-Holocene (MH, ca. 6 kya), Last Glacial Maximum (LGM; ca. 22 kya) and Last Interglacial (LIG; ca. 120–140 kya). Long-term persistence areas were estimated as the intersection of inferred presence areas since each period to the present: (i) long-term persistence since the LIG was the intersection of potential distributions in the LIG, LGM, MH, and present, (ii) long-term persistence since the LGM was the intersection of LGM, MH, and current distributions, and (iii) long-term persistence since the MH was the intersection of MH and current distributions.

2.4 Sampling strategies and DNA sequencing

With the aim of improving the resolution of phylogeographic patterns, we followed two strategies to increase sample size with respect to previous phylogeographic studies of *C. monspeliensis* (Fernández-Mazuecos & Vargas, 2010, 2011). For the increase of field sample size, first we overlapped long-term persistence maps (see SDM results) with the locations of populations sampled in the previous studies. We found that the two main long-term persistence areas (since the LIG) in the Mediterranean Basin were poorly sampled in previous studies (Fig. 1A). Therefore, we then sampled materials from these areas, in western Portugal and western Morocco (23 individuals from six populations). To obtain a balanced sample, we also sequenced 24 additional individuals from eight previously sampled populations across the Mediterranean distribution range (Fig. 1A; see Doc. S1). In the case of the Canary Islands, as long-term persistence areas had already been sampled in previous studies, we increased the sample size in all islands of the archipelago by sampling 56% of new populations from long-term persistence areas (Fig. 1A;

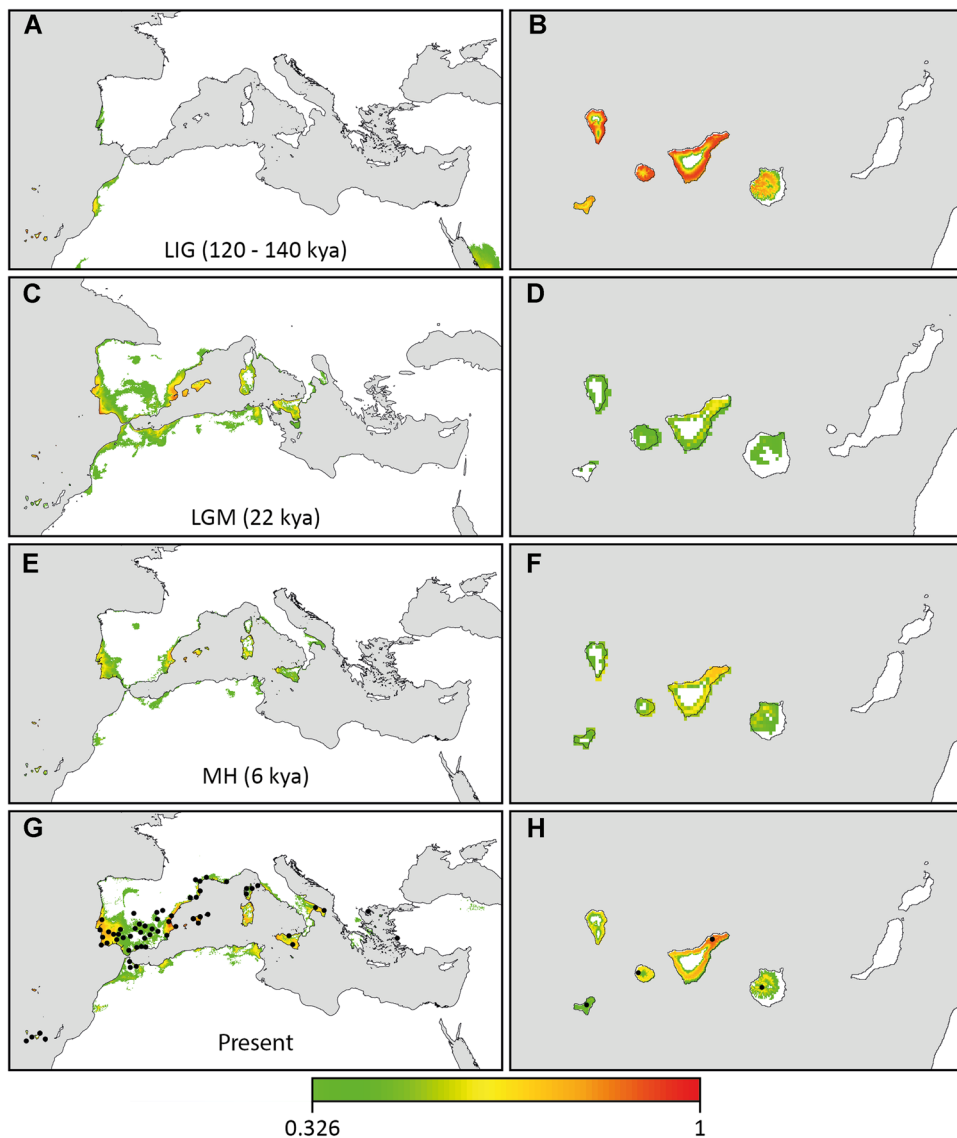


Fig. 2. Potential distribution of *Cistus monspeliensis* in the Mediterranean Basin (left), and zoom-in of the Canary Islands (right), estimated by species distribution modeling (Maxent) for **A, B**, the Last Interglacial (LIG), 120–140 kya; **C, D**, the Last Glacial Maximum (LGM), 22 kya; **E, F**, the Mid-Holocene (MH), 6 kya; and **G, H**, the present. Records used in species distribution modeling (SDM) are shown as dots (**G, H**). For each period, colors represent the probability of suitable conditions as shown in the legend. Values below the presence/absence threshold (0.326; see Section 2) are omitted. Distributions in LGM and MH represent average values across three climatic models with similar results (see Doc. S1 and Fig. S1.1 in Doc. S1) The Canarian coastline is taken from Norder et al. (2018).

see Doc. S1). In total, we added 84 individuals from 30 populations to the original dataset.

To increase genetic sampling, we added the cpDNA region that is most variable in angiosperms (*rpl32-trnL*) (Shaw et al., 2007) (see Doc. S1). In total, three cpDNA regions from 163 individuals and 69 populations of *C. monspeliensis* were used in this study (see Doc. S2). We used cpDNA because it displays maternal inheritance in *Cistus* (Guzmán & Vargas, 2009), it is haploid, and it lacks recombination. This set of markers help trace back seed movements reliably, and thus infer colonization events across islands (Vargas et al., 2015; see below).

2.5 Analysis of genetic diversity

All sequences were assembled in Geneious 10.2.2 (Kearse et al., 2012) and aligned using MAFFT 7.308 (Katoh et al., 2002). The resulting matrices were checked by visual inspection and corrected by introducing minor changes when necessary. After confirming topological congruence among the three cpDNA regions (data not shown), they were concatenated into a single alignment (Doc. S3). Genealogical relationships among haplotypes of *C. monspeliensis* were inferred using the statistical parsimony algorithm (Templeton et al., 1992), as implemented in TCS 1.21 (Clement et al., 2000), with parameters as shown in the study of Fernández-Mazuecos & Vargas (2011). We also

inferred relationships among haplotypes using five additional species of *Cistus* as the outgroup (Doc. S1) with one sequence representative of each haplotype (Doc. S4). Bayesian inference, as implemented in MrBayes 3.2.6 (Ronquist et al., 2012), was conducted using two runs of 10 million generations each, with a sample frequency of 1000 and a burn-in of 10%. Models of sequence evolution, selected with jModelTest (Darriba et al., 2012), were GTR + I for *trnS-trnG*, HKY + I for *psbK-trnS*, and GTR + G for *rpl32-trnL*.

To describe the geographic distribution of genetic diversity, we calculated the haplotypic diversity in different long-term persistence areas (estimated by SDM, see Section 3) and outside them, for both the Mediterranean Basin and the Canary Islands. Haplotypic diversity (H) was calculated as described in Nei & Tajima (1981). Values were calculated using the whole set of populations sampled in the present study. In addition, we calculated haplotypic diversity (H), the number of haplotypes (h), and the number of private haplotypes (ph) for the entire Mediterranean Region and the Canary Islands (each island separately and the whole archipelago).

2.6 Bayesian phylogeographic analysis

To reconstruct migration routes, with a focus on the Canary Islands, we performed a Bayesian discrete phylogeographic analysis (DPA; Lemey et al., 2009) in BEAST 1.10.4 (Drummond & Rambaut, 2007) using all sequences available in this study. We constrained two groups as monophyletic (Mediterranean haplotypes and Canarian haplotypes) and defined six areas (Mediterranean Basin, La Palma, El Hierro, La Gomera, Tenerife, and Gran Canaria). Nucleotide substitution model GTR + G was selected with jModelTest for the entire alignment. The reconstruction of ancestral areas followed an asymmetric substitution model, and we used a Bayesian stochastic search variable selection (BSSVS) to infer statistically supported migration routes. We implemented an uncorrelated relaxed clock with a lognormal distribution for the DNA partition and a strict clock for the area partition, with a simple constant size coalescent tree prior to facilitate

convergence. The root age was calibrated using a uniform prior distribution from 0.2 to 0.93 Mya (Fernández-Mazuecos & Vargas, 2011), with an initial age of 0.565 Mya. We performed two runs of 10 million generations each, sampled every 1000 generations, and applied a 10% burn-in. Adequacy of sample sizes was confirmed with Tracer; the two runs were combined with LogCombiner (discarding the burn-in) and trees were summarized in a maximum clade credibility tree using TreeAnnotator. Finally, we used SpreaD3 0.9.6 (Bielejec et al., 2016) to calculate Bayes factors (BF) and identify strongly supported migration routes as those connections between areas with $BF > 10$ (Kass & Raftery, 1995).

2.7 Sample size and number of colonization events

To explore the effect of increasing field and genetic sample size, we prepared four datasets (Table 2): (i) the original dataset of Fernández-Mazuecos & Vargas (2011), with a total of 53 individuals from 21 populations and two cpDNA regions; (ii) a dataset with increased genetic sampling (after adding the *rpl32-trnL* region), in which populations and individuals remained the same as in the original dataset; (iii) a dataset with increased field sampling (after adding 37 individuals from 16 populations) and the two original plastid regions; and (iv) a dataset with both increased field and genetic sample size, including all available data (total of 90 individuals from 37 populations, three cpDNA regions). We inferred numbers of haplotypes and colonization events between islands for all these datasets as indicators of phylogeographic signal. The inference of colonization events between islands followed the method proposed by Vargas et al. (2015), in which the distribution of haplotypes indicates the minimum number of inter-island colonizations as

$$C = \sum_i (p_i - 1)$$

where C is the number of inter-island colonizations and p_i is the number of islands where haplotype i occurs.

Table 2 The effect of sample size on the number of detected haplotypes and the number of inferred colonizations between islands in Canarian populations of *Cistus monspeliensis*

Dataset	Populations (individuals)	cpDNA regions	Insular haplotypes [†]	Inferred colonizations [†]
1 - Original dataset	21 (53)	2 (<i>trnS-trnG</i> , <i>psbK-trnS</i>)	10	7
2 - Expanded genetic sampling	21 (53)	3 (<i>trnS-trnG</i> , <i>psbK-trnS</i> , <i>rpl32-trnL</i>)	11	7
3 - Expanded field sampling	37 (90)	2 (<i>trnS-trnG</i> , <i>psbK-trnS</i>)	13	12
4 - Full dataset	37 (90)	3 (<i>trnS-trnG</i> , <i>psbK-trnS</i> , <i>rpl32-trnL</i>)	16	11

Note: The number of individuals sampled, the number of populations, and the number of cpDNA regions used are indicated. Dataset 1 was the original dataset from Fernández-Mazuecos & Vargas (2011). Dataset 2 included an additional DNA region, whereas dataset 3 included additional populations and individuals. Dataset 4 included both the additional cpDNA region and additional populations and individuals; [†]The number of insular haplotypes was taken from Fernández-Mazuecos & Vargas (2011) and the number of inferred colonizations from Vargas et al. (2015).

3 Results

3.1 Potential species distribution and long-term persistence areas

The inferred potential distribution of *Cistus monspeliensis* under current conditions (Figs. 2G, 2H) was mainly found in the Mediterranean Basin, the Canary Islands, and Madeira, with a high predictive accuracy (AUC = 0.939). Precipitation of wettest and driest quarters, temperature seasonality, and mean temperature of the driest quarter were the variables that contributed most to the model (Table S1.2 in Doc. S1). Projection to the past showed a restricted distribution in the LIG, with a few potential areas mainly in the Atlantic coast of north-western Africa and the Iberian Peninsula, Arabian Peninsula, Madeira, and the five westernmost islands of the Canary archipelago (Figs. 2A, 2B). The potential area (values above the 10 percentile training presence threshold) during the LIG was 76.6% smaller than the potential area under current conditions. In the LGM, potential areas were located across the western and central Mediterranean Basin, from the Atlantic coast of Iberia and north-western Africa to the south-eastern Italian Peninsula, and also on Madeira and the five westernmost islands in the Canary archipelago (Figs. 2C, 2D). The potential area in the LGM was 6.2% larger than that inferred for present conditions. By contrast, in the MH, potential habitats for *C. monspeliensis* were more restricted than those available under current conditions (54.3% smaller) and mainly spanned areas nearby the Atlantic and Mediterranean coasts across the western and central Mediterranean Basin, as well as the five westernmost Canary Islands and Madeira (Figs. 2E, 2F). Long-term persistence areas since the LIG were detected on the five westernmost islands of the Canary archipelago, Madeira, and the western end of the Mediterranean Basin, mainly including Atlantic coastal areas of Iberia (Portugal) and the western High Atlas (southwestern Morocco) (Fig. 3). Long-term persistence areas since the LGM and since the MH were larger than long-term persistence areas since the LIG.

3.2 Phylogenetic and phylogeographic analyses

Combination of *trnS-trnG*, *psbK-trnS*, and *rpl32-trnL* sequences of *C. monspeliensis* resulted in an aligned sequence length of 1899 bp (674 bp for *trnS-trnG*, 367 bp for *psbK-trnS*, and 858 bp for *rpl32-trnL*). A total of 20 substitution-based haplotypes were observed, forming a single network with no loops and few (5) missing haplotypes (Fig. 4C). The Mediterranean Basin and the Canary Islands did not share haplotypes, and three missing haplotypes separated the Mediterranean and Canarian haplogroups. In the Bayesian phylogenetic tree (inset in Fig. 5), all *C. monspeliensis* haplotypes formed a monophyletic group with a posterior probability (PP) of 1.00, which supports single ancestry. Two sister clades were inferred: one endemic to the Mediterranean Basin with four haplotypes (M1–M4) and the other endemic to the Canary Islands with 16 haplotypes (C1–C16).

In the Mediterranean Basin, haplotype M1 was found in the majority of individuals and populations (Fig. 4A; Doc. S2). The other Mediterranean haplotypes were only observed in Portugal and in two Spanish populations. In the Canary Islands (Fig. 4B), we observed 10 haplotypes private to single islands (Table 3). The most abundant and widespread

Canarian haplotype was C13. The haplotype network (Fig. 4C) and the phylogenetic analysis (inset in Fig. 5) revealed that La Palma has an exclusive haplotype lineage (1.00 PP) connected to the rest of the network by two mutational steps. The Canarian haplotype C1, found exclusively in north-eastern Tenerife (Anaga peninsula), was connected to the Mediterranean haplotype M1 through four mutational steps.

The Canary Islands showed a significantly higher number of haplotypes (16 vs. 4) and higher haplotypic diversity ($H = 0.887$ vs. 0.309) than the Mediterranean Basin (Table 3). Within the archipelago, Tenerife was the most diverse island. In the Mediterranean Basin, H was higher in older long-term persistence areas (Fig. 6). The highest haplotypic diversity was found in persistence areas since the LIG, whereas the lowest value was found outside persistence areas. By contrast, in the Canary Islands, H was similar in long-term persistence areas and outside them.

The Bayesian discrete phylogeographic analysis (Fig. 5) showed a high uncertainty for the ancestral area of all Canarian lineages, with very similar probabilities for Tenerife and La Palma. However, Tenerife was clearly the most probable ancestral area for most statistically supported Canarian subclades. The BSSVS showed three strongly supported routes of migration across the Canary Islands, with all of them departing from Tenerife (Fig. 5). The route to

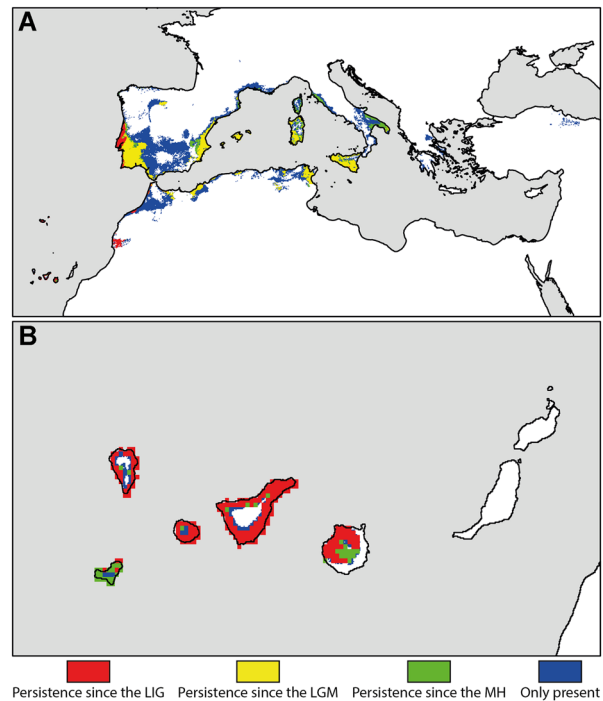


Fig. 3. Long-term persistence areas of *Cistus monspeliensis* since each analyzed period to the present in **A**, the Mediterranean Basin and **B**, zoom-in of the Canary Islands. Long-term persistence since the Last Interglacial (LIG; 120–140 kya to present) in red; since the Last Glacial Maximum (LGM, 22 kya to present) in yellow; since the Mid Holocene (MH; 6 kya to present) in green; and only present distribution in dark blue. Canarian coastline was taken from Norder et al. (2018).

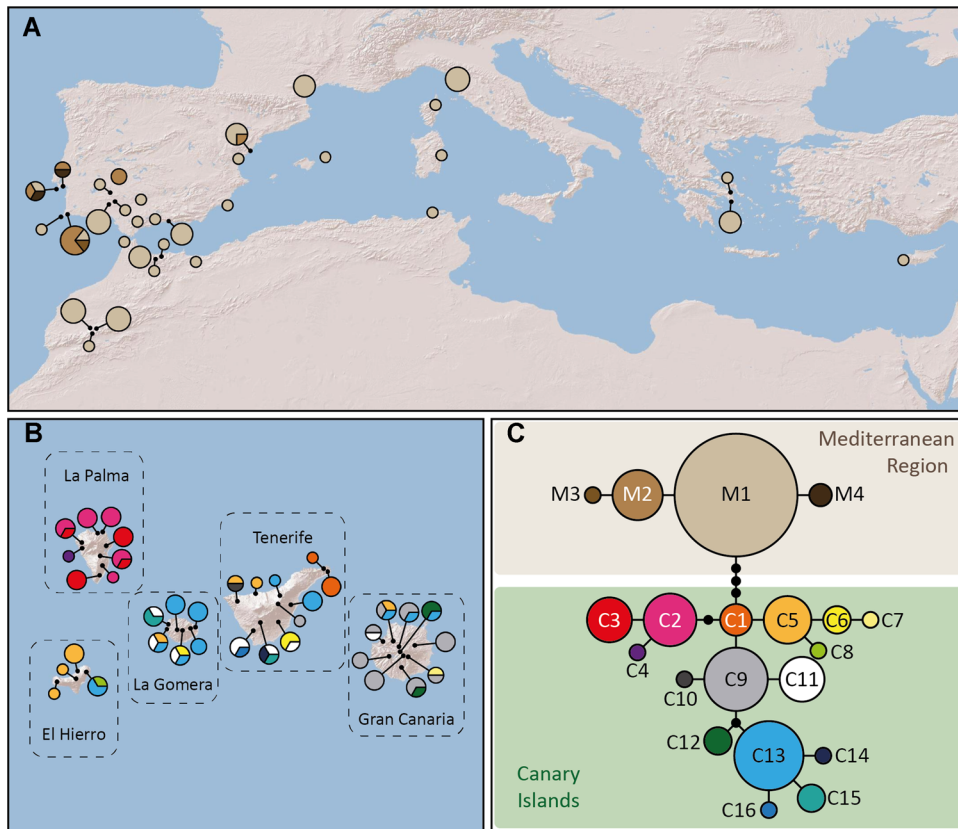


Fig. 4. Phylogeographic analysis of *Cistus monspeliensis* based on three cpDNA regions (*trnS-trnG*, *psbK-trnS*, and *rpl32-trnL*) and all individuals available in this study (Canarian populations correspond to dataset 4 in Table 3). Geographic distribution of haplotypes in **A**, the Mediterranean Basin and **B**, the Canary Islands; circles sizes are proportional to the number of individuals sampled in each site. Base maps from World Shaded Relief by Esri. **C**, Statistical parsimony network of *C. monspeliensis* haplotypes; lines represent single substitutions, dots indicate missing haplotypes (extinct or not found), and circle sizes are proportional to haplotype frequencies. A detailed list of haplotypes occurring in each population is available in Doc. S2.

La Gomera was the best supported (BF = 143.92), followed by routes to Gran Canaria (BF = 48.09) and El Hierro (BF = 11.09) (see Table S1.4 in Doc. S1).

3.3 Effect of sample size

The effects of sample size increase are summarized in Table 2. More haplotypes were detected as sample size increased at both the genetic and field levels. Nevertheless, we detected a more significant increase when expanding field sampling rather than genetic sampling. In particular, the increase in the number of haplotypes was higher when increasing the number of individuals (ca. 41%) (Table 2) than when adding a similar percentage (ca. 45%) of cpDNA base pairs. Similarly, we only inferred additional colonizations between islands when increasing field sampling (from 7 to 12). Furthermore, we detected one less colonization when using the full dataset compared to the dataset with increased field sampling only, an artefact caused by the method used for the inference of colonization events (Vargas et al., 2015) when finding more nucleotide variation (Coello et al., unpublished data).

4 Discussion

The two strategies to increase sample size displayed considerable differences regarding their effect on the phylogeographic patterns of *Cistus monspeliensis*. On the one hand, increased field sampling, guided by SDM, resulted in the finding of a higher haplotypic diversity over previous studies (Fernández-Mazuecos & Vargas, 2011) and helped disclose previously unknown refugial areas across the Mediterranean Basin. On the other hand, expanded genetic sampling resulted in a smaller increase in phylogeographic information than expanded field sampling. Due to the crucial information provided by SDM, we argue that explicit sampling strategies need to be designed and justified before starting any phylogeographic study.

4.1 Western Quaternary refugia in the Mediterranean

Contrary to expectations, the LGM did not show a strong reduction of habitat suitability in the Mediterranean Basin in comparison to the current distribution of *C. monspeliensis* (Figs. 2C, 2D). This result is congruent with some other Mediterranean plants such as *Olea europaea* (Besnard et al., 2013), *Ceratonia siliqua* (Viruel et al., 2019), and

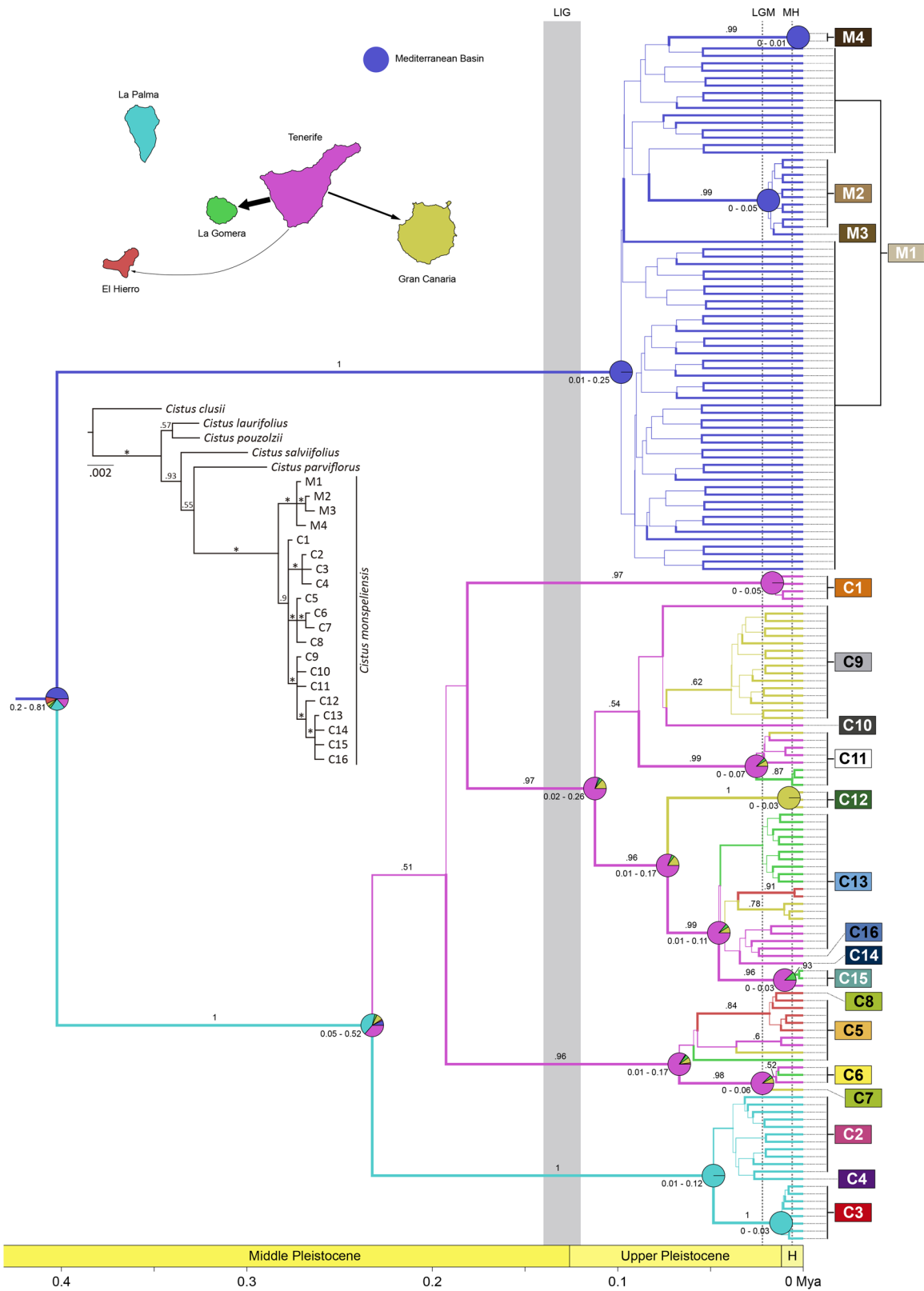


Fig. 5. Continued

Table 3 Genetic diversity of *Cistus monspeliensis* in the Mediterranean Basin and the Canary Islands (in bold), based on three cpDNA regions (*trnS-trnG*, *psbK-trnS*, and *rpl32-trnL*)

Area	<i>n</i>	<i>h</i>	<i>ph</i>	<i>H</i>
Mediterranean	73	4	4	0.309
Canary Islands	90	16	16	0.887
El Hierro	8	3	1	0.607
La Palma	20	3	3	0.563
La Gomera	17	5	0	0.640
Tenerife	21	10	4	0.905
Gran Canaria	24	6	2	0.598

Note: *n*, number of sampled individuals; *h*, number of substitution-based haplotypes; *ph*, number of private haplotypes; *H*, haplotypic diversity.

Lavatera maritima (Villa-Machío et al., 2018). The oldest long-term persistence areas (since LIG) in the Mediterranean Basin were mainly estimated along the Atlantic shores of western Iberia (Portugal) and north-western Africa (Morocco) (Fig. 3A). These areas are congruent with previously hypothesized refugial areas for Mediterranean plants (Médail & Diadema, 2009). The climate of these refugial areas has been markedly buffered by oceanic influence, a pattern found in some other Mediterranean plants (Fernández-Mazuecos & Vargas, 2013; García-Verdugo et al., 2013). As a result, western Iberia and north-western Africa emerge as potential Quaternary refugia for angiosperms (Gómez & Lunt, 2007; Bobo-Pinilla et al., 2018). In addition, these two areas can be considered centers of diversity involved in the recent expansion of *C. monspeliensis* across the Mediterranean Basin during the Late Quaternary, as the divergence of Mediterranean haplotypes probably occurred after the LIG (Fig. 5).

Remarkably, our new phylogeographic reconstruction, based on a sampling strategy guided by geographic representativeness and SDM (Figs. 2, 3), confirmed a Quaternary refugium that harbors high plastid diversity in western Iberia (Fig. 4A) (see Provan & Bennett, 2008). The remaining Mediterranean Basin populations displayed a subset of the haplotypes found in western Iberian populations (Fig. 4A), and haplotypic diversity decreased in areas that became suitable more recently (Fig. 6). This expected result (Russell et al., 2014; Dalmaris et al., 2015) is likely a consequence of a recent post-bottleneck expansion across the Mediterranean Basin from Quaternary refugia located in western Iberia (see Fernández-Mazuecos &

Vargas, 2010, 2011). Nevertheless, not all long-term persistence areas suggested by SDM can be considered refugia for *C. monspeliensis*. A relatively large, long-term persistence area was inferred in south-western Morocco (Fig. 3), but the species seems to be rare in this region, and we found low genetic diversity in populations sampled in the vicinity (Fig. 4A). This fact may be related to several factors that limited our power to detect refugial areas, such as biotic interactions and stochasticity of the dispersal processes. For this reason, although SDM based on bioclimatic variables helped disclose long-term persistence areas (in combination with phylogeographic data), non-climatic factors also influence the phylogeographic history of species.

4.2 Long-term persistence and active colonization across the Canary Islands

Reciprocal monophyly of Canarian and Mediterranean haplotypes is consistent with an early colonization of the Canary Islands from the mainland after species formation in the Quaternary (Fig. 5; Fernández-Mazuecos & Vargas, 2010, 2011). A single colonization of the Canarian archipelago has left a clear signature in the plastid genome of *C. monspeliensis*; however, we cannot rule out additional colonizations that have not left a signature in the plastid genome (see Caujapé-Castells et al., 2017). Interestingly, the Canarian ancestral haplotype was only found in the Anaga Peninsula (north-eastern Tenerife) (Figs. 4B, 4C), one of the oldest parts of the island (Troll & Carracedo, 2016). This result, similar to that found in *Canarina canariensis* (Mairal et al., 2015b), is probably related to the greater geological stability in the palaeo-islands of Tenerife in comparison with more volcanically active areas. Our results underline the biogeographic importance of Tenerife as a source island for inter-island colonization (Fig. 5). Indeed, Tenerife is the island that hosts the highest floristic diversity in the archipelago (Arechavaleta et al., 2010) and the highest diversity of *Cistus* species (Guzmán & Vargas, 2010). In addition to the western Iberian refugium, our results support the Canary Islands as the most significant Quaternary refugium for *C. monspeliensis* (Fernández-Mazuecos & Vargas, 2011). We estimated high habitat suitability over time (Fig. 2) and concomitant long-term persistence areas since the LIG on the five islands where the species currently occurs (Fig. 3B). Consistently, the Canarian archipelago shows higher levels of plastid genetic diversity than the Mediterranean Basin (Table 3). Climatic stability of oceanic islands has been postulated to be responsible for survival of numerous plant lineages that

Fig. 5. Discrete phylogeographic analysis (DPA) of *Cistus monspeliensis* based on three cpDNA regions (*trnS-trnG*, *psbK-trnS*, and *rpl32-trnL*). The maximum clade credibility tree is shown, in which branch thickness represents Bayesian posterior probabilities and values above 0.5 are indicated; branch colors correspond to the most probable ancestral area (see map) and pie charts represent posterior probability distributions of ancestral areas for well-supported nodes (PP > 0.95). The 95% highest posterior density intervals for ages of the same nodes are also shown. Time slices used to project the distribution model of *C. monspeliensis* (see Fig. 2) are highlighted in the phylogeny: Mid-Holocene (MH; 6 kya), Last Glacial Maximum (LGM; 22 kya), and Last Interglacial (LIG; 120–140 kya). The phylogeny in the inset represents the relationships among *C. monspeliensis* haplotypes and five additional *Cistus* species (used as outgroup) inferred in MrBayes (PPs > 0.95 are indicated as asterisks). The map of the five westernmost islands of the Canarian archipelago includes migration routes with Bayes factor (BF) > 10 (thickness of arrows are proportional to BF values).

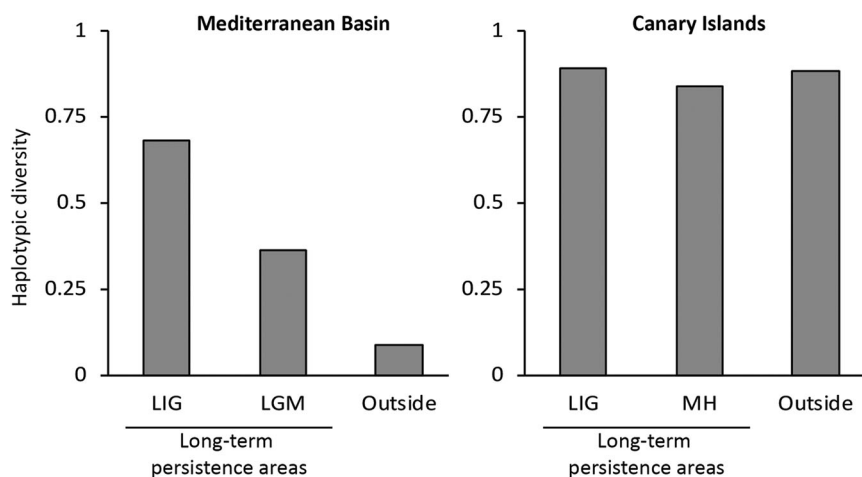


Fig. 6. Haplotypic diversity of *Cistus monspeliensis* in long-term persistence areas and outside them, calculated using three cpDNA regions (*trnS-trnG*, *psbK-trnS*, and *rpl32-trnL*). Results are shown for the Mediterranean Basin and the Canary Islands.

ferred from the dramatic changes of the continents (Vargas, 2007). Contrary to expectations (Russell et al., 2014; Dalmaris et al., 2015), the Canary Islands do not show increased haplotypic diversity in long-term persistence areas of each island (Fig. 6). We suggest that the relatively small size of the islands and active inter-island colonization (see Table 2) may have blurred the effect of habitat suitability through time. Furthermore, the low resolution of SDMs in these islands relative to their size, together with other factors not considered herein, such as topography (Harter et al., 2015), makes estimations of habitat suitability less accurate. In fact, given the complex and dynamic history of oceanic archipelagos, including the Canary Islands (Troll & Carracedo, 2016), and the role of palaeo-islands as potential refugia given their geological stability (Mairal et al., 2015b), we recommend an assessment of geological history prior to fieldwork in oceanic archipelagos to guarantee the sampling of areas of potential relevance for genetic diversity.

As previously described (Fernández-Mazuecos & Vargas, 2011; Vargas et al., 2015), *C. monspeliensis* displays a high colonization capacity with numerous colonization events between islands (Fig. 5; Table 2) despite the absence of long-distance dispersal syndromes in the genus (Guzmán & Vargas, 2009). Our analysis of colonization patterns using haplotypes as dispersal units, though simpler, revealed a colonization history similar to that inferred by Bayesian discrete phylogeographic analysis (Fig. 5). At least 12 independent colonization events were estimated using our strategy of sampling additional populations (versus seven colonizations inferred earlier in Vargas et al., 2015). This high number of inter-island colonizations, together with the high habitat availability in the archipelago over time indicated by SDM (Fig. 2), supports an important role of establishment (ecology) rather than dispersal mechanisms (Guzmán & Vargas, 2009; Viana et al., 2016; Carvajal-Endara et al., 2017). Nevertheless, an extremely favorable climate does not necessarily ensure colonization in cases of failure of dispersal (van der Pijl, 1982). This may be the case of *C. monspeliensis* in the archipelago of Madeira. Despite the high habitat

availability in Madeira in all time periods analyzed (Fig. 2) and the detection of large potential areas of long-term persistence (Fig. 3A), the checklist of Madeira (http://www4.uma.pt/gbm/checklist/lista_flora.php) has no records of *C. monspeliensis*. In addition, limited dispersal may have led to striking genetic isolation of La Palma populations, as shown by the presence of a single sub-lineage composed of three endemic haplotypes (Figs. 4B, 4C, 5; Table 3). Our results stress the importance of both dispersal capabilities and habitat suitability for island colonization.

4.3 Effect of sample size on phylogeographic patterns

Sample size is critical for any reliable study in phylogeography and population genetics (Björklund & Bergeck, 2009). In this study, our strategy of extending both genetic and population sampling revealed a significant increase of cpDNA diversity. In particular, we detected more haplotypes in both insular populations (16 vs. 10) and Quaternary refugia of the Mediterranean Basin (4 vs. 1; Fig. 4; Table 2) than a previous study (Fernández-Mazuecos & Vargas, 2011). Nevertheless, subsets of our insular dataset showed that extending the number of populations helped detect a higher number of new haplotypes (three more) than expanding the genetic dataset (one more) (Table 2). As a result, we detected a considerable increase in the number of colonizations among islands when population sampling was increased (12 events), but failed when only genetic sampling was increased (seven events; see Table 2). In fact, although the general trend is to increase genetic data in phylogeographic studies (e.g., using NGS techniques; Garrick et al., 2015), our study stresses the importance of considering a field sampling strategy guided by SDM to target potential areas of long-term persistence in the context of a geographically representative sampling. An increase in population number is observed in some studies of Mediterranean and Canarian species (as reviewed in Table 1), which improved the phylogeographic reconstruction of a species compared to previous analyses. For this reason, we suggest that, prior to sampling in the field, SDM analyses can help detect areas with potentially high levels of genetic

diversity that otherwise could have been overlooked (Figs. 4, 6) (Waltari et al., 2007). In sum, we call attention to the importance of not only extending DNA sequencing with new molecular techniques in phylogeographic studies but also sampling a high number of populations and individuals in the field with the guidance of SDM techniques. This is particularly true for addressing specific phylogeographic questions for species with low genetic diversity and widespread distribution (Fernández-Mazuecos et al., 2014; Forrest et al., 2017). Although our results are based on maternally inherited, nonrecombinant cpDNA sequences due to our focus on colonization patterns (Vargas et al., 2015), the use of nuclear DNA can provide additional insights into the phylogeographic history of species. In particular, future studies of nuclear markers in *C. monspeliensis* will add to our understanding of the geographic structure of genetic diversity in this species.

5 Conclusions

Our results underscore the importance of field sampling design for the accurate inference of phylogeographic reconstructions, particularly when very recent events of extinction–expansion in the evolutionary history of a species can complicate the description of phylogeographic patterns. Specifically, SDM is a powerful tool to plan field sampling strategies aimed at detecting areas with potentially high genetic diversity, which may improve the reconstruction of phylogeographic patterns. This approach helped us find a previously undetected refugial area (western Iberia) that may have been the source of population expansion of *Cistus monspeliensis* across the Mediterranean Basin after the LIG. In parallel, the Canary Islands have remained climatically favorable for *C. monspeliensis* throughout the late Quaternary, and thus long-term survival is postulated as the main reason for the high level of cpDNA diversity in the archipelago and the numerous colonization events inferred among islands. In future phylogeographic studies, SDM analyses, as well as the sampling of a high number of geographically representative populations depending upon species distribution ranges, should be routinely considered prior to fieldwork. Furthermore, studies of complex and dynamic areas, such as oceanic islands, need to incorporate a sample that includes areas less affected by volcanism and that considers the occurrence of palaeoislands by the time the earliest colonist arrived in the archipelago.

Acknowledgements

The authors thank Emilio Cano, Alberto Herrero, Lucía Sastre, José González, Olga Popova, Yolanda Ruiz, and Ying Tan for laboratory assistance; Daniel Romero for field assistance; Anass Terrab for samples from Morocco; Yurena Arjona, Ana Otero, María Leo, and Nagore G. Medina for advise on data processing and analyses; Noelia Fernández-Prado, Michelle M. Howard, Ricarda Riina, and Marcos Méndez for their comments to improve the text; and the MA herbarium for plant material. We are also very grateful to two anonymous reviewers for their comments that helped improve the scientific quality of the manuscript. This study is part of the project CGL2015–67865-P, funded by the Spanish Ministry of

Economy, Industry and Competitiveness. A.J.C. was supported by the Spanish Ministry of Education, Culture and Sport through a FPU fellowship (FPU16/05681). M.F.-M. was supported by the Spanish Ministry of Economy, Industry and Competitiveness through a Juan de la Cierva contract (IJCI-2015–23459). C.G.-V. was supported by a “Vicenç Mut” postdoctoral fellowship (Conselleria d’Innovació, Recerca i Turisme, Govern de les Illes Balears and European Social Fund).

Data sharing

Occurrences used in the SDM analyses are available in Table S1.1 in Doc. S1. DNA sequences are available in the GenBank database (accessions are indicated in Doc. S2). Alignments used in this study are available in Docs. S3 and S4.

References

- Alvarado-Serrano DF, Knowles LL. 2014. Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Molecular Ecology Resources* 14: 233–248.
- Arechavaleta M, Rodríguez S, Zurita N, García A. 2010. *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres 2009*. Gobierno de Canarias.
- Avice JC. 2009. Phylogeography: Retrospect and prospect. *Journal of Biogeography* 36: 3–15.
- Avice JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18: 489–522.
- Barres L, Galbany-Casals M, Hipp AL, Molero J, Vilatersana R. 2017. Phylogeography and character evolution of *Euphorbia* sect. *Aphyllis* subsect. *Macaronesicae* (Euphorbiaceae). *Taxon* 66: 324–342.
- Besnard G, Khadari B, Baradat P, Bervillé A. 2002. *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics* 104: 1353–1361.
- Besnard G, Khadari B, Navascues M, Fernandez-Mazuecos M, El Bakkali A, Arrigo N, Baali-Cherif D, Brunini-Bronzini de Caraffa V, Santoni S, Vargas P, Savolainen V. 2013. The complex history of the olive tree: From Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences* 280: 20122833.
- Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. 2016. SpreaD3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Molecular Biology and Evolution* 33: 2167–2169.
- Bilgin R, Gürün K, Rebelo H, Puechmaile SJ, Maracı Ö, Presetnik P, Benda P, Hulva P, Ibáñez C, Hamidovic D, Fressel N, Horáček I, Karataş A, Karataş A, Allegri B, Georgiakakis P, Gazaryan S, Nagy ZL, Abi-Said M, Lučan RK, Bartonička T, Nicolau H, Scaravelli D, Karapandža B, Uhrin M, Paunović M, Juste J. 2016. Circum-Mediterranean phylogeography of a bat coupled with past environmental niche modeling: A new paradigm for the recolonization of Europe? *Molecular Phylogenetics and Evolution* 99: 323–336.

- Björklund M, Bergek S. 2009. On the relationship between population differentiation and sampling effort: Is more always better? *Oikos* 118: 1127–1129.
- Bobo-Pinilla J, Peñas de Giles J, López-González N, Mediavilla S, Martínez-Ortega MM. 2018. Phylogeography of an endangered disjunct herb: Long-distance dispersal, refugia, and colonization routes. *AoB Plants* 10: ply047.
- Caccone A, Gentile G, Gibbs JP, Fritts TH, Snell HL, Betts J, Powell JR. 2002. Phylogeography and history of giant Galápagos tortoises. *Evolution* 56: 2052–2066.
- Carvajal-Endara S, Hendry AP, Emery NC, Davies TJ. 2017. Habitat filtering not dispersal limitation shapes oceanic island floras: Species assembly of the Galápagos archipelago. *Ecology Letters* 20: 1–10.
- Caujapé-Castells J, García-Verdugo C, Marrero-Rodríguez Á, Fernández-Palacios JM, Crawford DJ, Mort ME. 2017. Island ontogenies, syngameons, and the origins and evolution of genetic diversity in the Canary endemic flora. *Perspectives in Plant Ecology, Evolution and Systematics* 27: 9–22.
- Clement M, Posada D, Crandall KA. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.
- Cronk QCB. 1992. Relict floras of Atlantic islands: Patterns assessed. *Biological Journal of the Linnean Society* 46: 91–103.
- Dalmaris E, Ramalho CE, Poot P, Veneklaas EJ, Byrne M, Cj Y. 2015. A climate change context for the decline of a foundation tree species in south-western Australia: Insights from phylogeography and species distribution modelling. *Annals of Botany* 116: 941–952.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772.
- Demoly JP, Montserrat P. 2005. *Cistus*. In: Castroviejo S, Acedo C, Cirujano S, Laínz M, Montserrat P, Morales R, Muñoz Garmendia F, Navarro C, Paiva JAR, Soriano C eds. *Flora Ibérica*. Madrid: Real Jardín Botánico, CSIC.
- Désamoré A, Laenen B, González-Mancebo JM, Jaén Molina R, Bystrakova N, Martínez-Klimova E, Carine MA, Vanderpoorten A. 2012. Inverted patterns of genetic diversity in continental and island populations of the heather *Erica scoparia* s.l. *Journal of Biogeography* 39: 574–584.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.
- Egan AN, Schlueter J, Spooner DM. 2012. Applications of next-generation sequencing in plant biology. *American Journal of Botany* 99: 175–185.
- Elith J, Graham CH, Anderson PR, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM McC, Townsend Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Fernández-Mazuecos M, Jiménez-Mejías P, Rotllan-Puig X, Vargas P. 2014. Narrow endemics to Mediterranean islands: Moderate genetic diversity but narrow climatic niche of the ancient, critically endangered *Naufraga* (Apiaceae). *Perspectives in Plant Ecology, Evolution and Systematics* 16: 190–202.
- Fernández-Mazuecos M, Vargas P. 2010. Ecological rather than geographical isolation dominates Quaternary formation of Mediterranean *Cistus* species. *Molecular Ecology* 19: 1381–1395.
- Fernández-Mazuecos M, Vargas P. 2011. Genetically depauperate in the continent but rich in oceanic islands: *Cistus monspeliensis* (Cistaceae) in the Canary Islands. *PLoS One* 6: e17172.
- Fernández-Mazuecos M, Vargas P. 2013. Congruence between distribution modelling and phylogeographical analyses reveals Quaternary survival of a toadflax species (*Linaria elegans*) in oceanic climate areas of a mountain ring range. *New Phytologist* 198: 1274–1289.
- Forrest A, Escudero M, Heuertz M, Wilson Y, Cano E, Vargas P. 2017. Testing the hypothesis of low genetic diversity and population structure in narrow endemic species: The endangered *Antirrhinum charidemi* (Plantaginaceae). *Botanical Journal of the Linnean Society* 183: 260–270.
- Furman A, Öztunç T, Postawa T, Çoraman E. 2010. Shallow genetic differentiation in *Miniopterus schreibersii* (Chiroptera: Vespertilionidae) indicates a relatively recent re-colonization of Europe from a single glacial refugium. *Acta Chiropterologica* 12: 51–59.
- García-Olivares V, López H, Patiño J, Alvarez N, Machado A, Carracedo JC, Soler V, Emerson BC. 2017. Evidence for megalandslides as drivers of island colonization. *Journal of Biogeography* 44: 1053–1064.
- García-Verdugo C, Calleja JA, Vargas P, Silva L, Moreira O, Pulido F. 2013. Polyploidy and microsatellite variation in the relict tree *Prunus lusitanica* L.: How effective are refugia in preserving genotypic diversity of clonal taxa? *Molecular Ecology* 22: 1546–1557.
- García-Verdugo C, Forrest AD, Balaguer L, Fay MF, Vargas P. 2010a. Parallel evolution of insular *Olea europaea* subspecies based on geographical structuring of plastid DNA variation and phenotypic similarity in leaf traits. *Botanical Journal of the Linnean Society* 162: 54–63.
- García-Verdugo C, Forrest AD, Fay MF, Vargas P. 2010b. The relevance of gene flow in metapopulation dynamics of an oceanic island endemic, *Olea europaea* subsp. *guanchica*. *Evolution* 64: 3525–3536.
- García-Verdugo C, Mairal M, Monroy P, Sajeva M, Caujapé-Castells J. 2017. The loss of dispersal on islands hypothesis revisited: Implementing phylogeography to investigate evolution of dispersal traits in *Periploca* (Apocynaceae). *Journal of Biogeography* 44: 2595–2606.
- García-Verdugo C, Sajeva M, La Mantia T, Harrouni C, Msanda F, Caujapé-Castells J. 2015. Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology* 24: 726–741.
- Garrick RC, Bonatelli IAS, Hyseni C, Morales A, Pelletier TA, Perez MF, Rice E, Satler JD, Symula RE, Thomé MTC, Carstens BC. 2015. The evolution of phylogeographic data sets. *Molecular Ecology* 24: 1164–1171.
- GBIF.org. 2016. GBIF Occurrence Download. The Global Biodiversity Information Facility [online]. Available from <https://doi.org/10.15468/dl.cnif4m> [accessed 7 July 2016].
- Giovannotti M, Nisi-Cerioni P, Caputo V. 2010. Mitochondrial DNA sequence analysis reveals multiple Pleistocene glacial refugia for *Podarcis muralis* (Laurenti, 1768) in the Italian Peninsula. *Italian Journal of Zoology* 77: 277–288.
- Gómez A, Lunt DH. 2007. Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N eds. *Phylogeography of Southern European Refugia*. Dordrecht: Springer Netherlands.
- Guzmán B, Vargas P. 2005. Systematics, character evolution, and biogeography of *Cistus* L. (Cistaceae) based on ITS, trnL-trnF, and

- matK sequences. *Molecular Phylogenetics and Evolution* 37: 644–660.
- Guzmán B, Vargas P. 2009. Long-distance colonization of the Western Mediterranean by *Cistus ladanifer* (Cistaceae) despite the absence of special dispersal mechanisms. *Journal of Biogeography* 36: 954–968.
- Guzmán B, Vargas P. 2010. Unexpected synchronous differentiation in Mediterranean and Canarian *Cistus* (Cistaceae). *Perspectives in Plant Ecology, Evolution and Systematics* 12: 163–174.
- Harter DEV, Irl SDH, Seo B, Steinbauer MJ, Gillespie R, Triantis KA, Fernández-Palacios J-M, Beierkuhnlein C. 2015. Impacts of global climate change on the floras of oceanic islands – Projections, implications and current knowledge. *Perspectives in Plant Ecology, Evolution and Systematics* 17: 160–183.
- Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Juhren MC. 1966. Ecological observations on *Cistus* in the Mediterranean vegetation. *Forest Science* 12: 415–426.
- Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* 90: 773–795.
- Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.
- Laenen B, Désamoré A, Devos N, Shaw AJ, González-Mancebo JM, Carine MA, Vanderpoorten A. 2011. Macaronesia: A source of hidden genetic diversity for post-glacial recolonization of western Europe in the leafy liverwort *Radula lindenbergiana*. *Journal of Biogeography* 38: 631–639.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Computational Biology* 5: e1000520.
- Mairal M, Pokorny L, Aldasoro JJ, Alarcón M, Sanmartín I. 2015a. Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: The case of the Rand Flora genus *Canarina* (Campanulaceae). *Molecular Ecology* 24: 1335–1354.
- Mairal M, Sanmartín I, Aldasoro JJ, Culshaw V, Manolopoulou I, Alarcón M. 2015b. Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: The case of the widespread endemic *Canarina canariensis* (Campanulaceae). *Molecular Ecology* 24: 3944–3963.
- Martín M, Guinea E. 1949. *Jarales y jaras (Cistografía hispánica)*. Madrid: Ministerio de Agricultura; Dirección General de Montes, Caza y Pesca Fluvial; Instituto Forestal de Investigaciones y Experiencias.
- Mateu-Andrés I, Ciurana M-J, Aguilera A, Boisset F, Guara M, Laguna E, Currás R, Ferrer P, Vela E, Puche MF, Pedrola-Monfort J. 2015. Plastid DNA homogeneity in *Celtis australis* L. (Cannabaceae) and *Nerium oleander* L. (Apocynaceae) throughout the Mediterranean basin. *International Journal of Plant Sciences* 176: 421–432.
- Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* 36: 1333–1345.
- Médail F, Quézel P. 1999. The phylogeographical significance of S.W. Morocco compared to the Canary Islands. *Plant Ecology* 140: 221–244.
- Migliore J, Baumel A, Leriche A, Juin M, Médail F. 2018. Surviving glaciations in the Mediterranean region: An alternative to the long-term refugia hypothesis. *Botanical Journal of the Linnean Society* 187: 537–549.
- Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145–163.
- Norder SJ, Baumgartner JB, Borges PAV, Hengl T, Kissling WD, van Loon EE, Rijdsdijk KF. 2018. A global spatially explicit database of changes in island palaeo-area and archipelago configuration during the late Quaternary. *Global Ecology and Biogeography* 27: 500–505.
- Päckert M, Martens J, Hering J, Kvist L, Illera JC. 2013. Return flight to the Canary Islands – The key role of peripheral populations of Afrocanarian blue tits (*Aves: Cyanistes teneriffae*) in multi-gene reconstructions of colonization pathways. *Molecular Phylogenetics and Evolution* 67: 458–467.
- Patiño J, Carine M, Mardulyn P, Devos N, Mateo RG, Gonzalez-Mancebo JM, Shaw AJ, Vanderpoorten A. 2015. Approximate bayesian computation reveals the crucial role of oceanic islands for the assembly of continental biodiversity. *Systematic Biology* 64: 579–589.
- Peñas J, Barrios S, Bobo-Pinilla J, Lorite J, Martínez-Ortega MM. 2016. Designing conservation strategies to preserve the genetic diversity of *Astragalus edulis* Bunge, an endangered species from western Mediterranean region. *PeerJ* 4: e1474.
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- Provan J, Bennett KD. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* 23: 564–571.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Russell J, van Zonneveld M, Dawson IK, Booth A, Waugh R, Steffenson B. 2014. Genetic diversity and ecological niche modelling of wild barley: Refugia, large-scale post-LGM range expansion and limited mid-future climate threats? *PLoS One* 9: e86021.
- Saavedra C, Peña JB. 2005. Nucleotide diversity and Pleistocene population expansion in Atlantic and Mediterranean scallops (*Pecten maximus* and *P. jacobaeus*) as revealed by the mitochondrial 16S ribosomal RNA gene. *Journal of Experimental Marine Biology and Ecology* 323: 138–150.
- Salvi D, Harris DJ, Kaliontzopoulou A, Carretero MA, Pinho C. 2013. Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: Phylogeographic insights from the common wall lizard. *BMC Evolutionary Biology* 13: 147.
- Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany* 94: 275–288.
- Shaw KL, Gillespie RG. 2016. Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. *Proceedings of the National Academy of Sciences USA* 113: 7986–7993.
- Short MJ. 1994. Cistaceae. In: Press JR, Short MJ eds. *Flora of Madeira*. London: The Natural History Museum.

- Stervander M, Illera JC, Kvist L, Barbosa P, Keehnen NP, Pruischer P, Bensch S, Hansson B. 2015. Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes* spp.) – phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. *Molecular Ecology* 24: 2477–2494.
- Sun Y, Li Y, Vargas-Mendoza CF, Wang F, Xing F. 2016. Colonization and diversification of the *Euphorbia* species (sect. *Aphyllis* subsect. *Macaronesicae*) on the Canary Islands. *Scientific Reports* 6: 34454.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.
- Telahigue K, Hajji T, El Cafsi M, Saavedra C. 2018. Genetic structure and demographic history of the endemic Mediterranean scallop *Pecten jacobaeus* inferred from mitochondrial 16s DNA sequence analysis. *Animal Biodiversity and Conservation* 41: 61–73.
- Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.
- Troll VR, Carracedo JC. 2016. *The Geology of the Canary Islands*. Amsterdam: Elsevier.
- van der Pijl L. 1982. *Principles of dispersal in higher plants*. Berlin: Springer.
- Vargas P. 2007. Are Macaronesian islands refugia of relict plant lineages? A molecular survey. In: Weiss S, Ferrand N eds. *Phylogeography of Southern European Refugia*. Dordrecht: Springer Netherlands.
- Vargas P, Rumeu B, Heleno RH, Traveset A, Nogales M. 2015. Historical isolation of the Galápagos carpenter bee (*Xylocopa darwini*) despite strong flight capability and ecological amplitude. *PLoS One* 10: 1–13.
- Vendramin GG, Fady B, González-Martínez SC, Hu FS, Scotti I, Sebastiani F, Soto Á, Petit RJ. 2008. Genetically depauperate but widespread: The case of an emblematic Mediterranean pine. *Evolution* 62: 680–688.
- Viana DS, Gangoso L, Bouten W, Figuerola J. 2016. Overseas seed dispersal by migratory birds. *Proceedings of the Royal Society B: Biological Sciences* 283: 20152406.
- Villa-Machío I, Fernández de Castro AG, Fuertes-Aguilar J, Nieto Feliner G. 2018. Out of North Africa by different routes: Phylogeography and species distribution model of the western Mediterranean *Lavatera maritima* (Malvaceae). *Botanical Journal of the Linnean Society* 187: 441–455.
- Viruel J, Le Galliot N, Pironon S, Nieto Feliner G, Suc J-P, Lakhal-Mirleau F, Juin M, Selva M, Bou Dagher Kharrat M, Ouahmane L, La Malfa S, Diadema K, Sanguin H, Médail F, Baumel A. 2019. A strong east–west Mediterranean divergence supports a new phylogeographic history of the carob tree (*Ceratonia siliqua*, Leguminosae) and multiple domestications from native populations. *Journal of Biogeography*. doi: 10.1111/jbi.13726.
- Waltari E, Hijmans RJ, Peterson AT, Nyári ÁS, Perkins SL, Guralnick RP. 2007. Locating Pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. *PLoS One* 2: e563.
- Xavier R, Branco M, dos Santos AM. 2016. Using a phylogeographic approach to investigate the diversity and determine the distributional range of an isopod (Crustacea: Peracarida), *Stenosoma nadejda* (Rezig, 1989) in the Atlantic-Mediterranean region. *Hydrobiologia* 768: 315–328.
- Xavier R, Santos AM, Lima FP, Branco M. 2009. Invasion or invisibility: Using genetic and distributional data to investigate the alien or indigenous status of the Atlantic populations of the peracarid isopod, *Stenosoma nadejda* (Rezig 1989). *Molecular Ecology* 18: 3283–3290.
- Xavier R, Zenboudji S, Lima FP, Harris DJ, Santos AM, Branco M. 2011. Phylogeography of the marine isopod *Stenosoma nadejda* (Rezig, 1989) in North African Atlantic and western Mediterranean coasts reveals complex differentiation patterns and a new species. *Biological Journal of the Linnean Society* 104: 419–431.

Supplementary Material

The following supplementary material is available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12570/supinfo>:

Doc. S1. Expanded methodology.

Doc. S2. Biological material used in this study (including accession numbers of Genbank sequences).

Doc. S3. Concatenated alignment of *trnS-trnG*, *psbK-trnS* and *rpl32-trnL* sequences from all individuals of *C. monspeliensis* used in this study.

Doc. S4. Concatenated alignment of *trnS-trnG*, *psbK-trnS* and *rpl32-trnL* haplotypes detected for *C. monspeliensis* and the outgroup (*Cistus* spp.) used for phylogenetic analysis (see inset in Fig. 5).