Repeated jumps from Northwest Africa to the European continent: the case of peripheral populations of an annual plant

Running title: Peripheral populations of *Scrophularia arguta*

Francisco Javier Valtueña[1,*], Mario Fernández-Mazuecos[2], Tomás Rodríguez-Riaño[1], Josefa López[1], and Ana Ortega-Olivencia[1]

[1]Área de Botánica, Facultad de Ciencias, Universidad de Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain

[2]Real Jardín Botánico de Madrid, Consejo Superior de Investigaciones Científicas (RJB–CSIC), Plaza de Murillo 2, 28014 Madrid, Spain

*Corresponding author:

Francisco Javier Valtueña

Área de Botánica, Facultad de Ciencias, Universidad de Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain. E-mail: fjvaltu@unex.es

ABSTRACT

Peripheral populations (i.e. those occurring on the edge of a species' distribution range) can have different origins and genetic characteristics, and they may be critical for conservation of genetic diversity. We investigated European peripheral populations of *Scrophularia arguta*, a widespread, annual plant distributed from Arabia to Northwest Africa and Macaronesia. Only two small disjunct population groups of this species occur in Europe, specifically in West-Central and Southeast Iberia. To disclose the origin of these populations and determine their importance for conservation of *S. arguta* genetic diversity, we analysed DNA sequences from two nuclear and two plastid regions and AFLP markers in populations sampled mainly across the western distribution range of the species, and modelled the species distribution under present and late Quaternary conditions. The analyses revealed the presence of three distinct lineages of *S. arguta* in Europe, as a result of multiple colonisation waves at different times in the Quaternary. Two of these lineages, occurring in Southeast Iberia, are the result of more or less recent dispersal from Northwest Africa. In contrast, West-Central Iberian populations are strongly differentiated from the remaining range of *S. arguta*, and can be considered as peripheral relict populations. Our study is the first to demonstrate the occurrence of at least three colonisations of the European continent from Africa by a native plant species. The diverse histories and genetic makeup of resulting populations confirm the importance of peripheral populations, and particularly of ancient relict populations, for conservation of global genetic diversity in widespread species.

Key words: Disjunct populations, genetic diversity, long-distance dispersal, migration routes, phylogeography, *Scrophularia*.

1. INTRODUCTION

Peripheral populations are defined as those populations occurring on the edge of a species' distribution range. Usually, they are smaller, more isolated and highly fragmented, have low effective population sizes and reduced gene flow, and the incidence of selfing in them is greater (Lawton, 1993; Young *et al*., 1996; Pujol *et al*., 2009). The central-marginal hypothesis predicts that peripheral populations have lower genetic diversity and higher differentiation compared to populations in the core of the species' range due to genetic drift, founder effects, inbreeding and bottlenecks (Antonovics *et al.*, 2001; Cole, 2003; Arnaud-Haond *et al*., 2006; Eckert *et al*., 2008; Frankham *et al*., 2010). However, some studies have revealed exceptions to this pattern, including instances of similar (e.g., Putz *et al*., 2015), and even higher levels of genetic diversity in peripheral populations dependent on population sizes (Lázaro-Nogal *et al*., 2017). Beyond the general prediction of the central-marginal hypothesis, the genetic diversity and differentiation of peripheral populations can be the result of a complex history, and they may vary greatly depending on the causes behind population origin and on the degree of gene flow with other populations. For example, recent peripheral populations of a species that is currently expanding its range should have low genetic diversity but not show differentiation due to founder effects. Conversely, peripheral populations that have been isolated for a long time (peripheral relict populations) usually exhibit clear genetic differentiation that may be accompanied by a reduced level of genetic variation. In the latter case, the high differentiation makes these populations important for conservation of genetic diversity of the species, and the lower genetic diversity may affect their capacity to adapt to and survive any changes in the evolutionary pressures they experience (Lammi *et al*., 1999; Cassel-Lundhagen, 2010).

In the case of peripheral populations that are isolated from the remaining populations by long distances, they may be the result of vicariance (fragmentation of an ancient continuous range, caused by the emergence of barriers to gene flow or by the extinction of intermediate populations; Rosen, 1978; Wiley, 1988) or long-distance dispersal (LDD) (Wiley, 1988; Milne, 2006; Cox and Moore, 2016). While LDD is obvious in peripheral populations occurring on oceanic islands, it may be harder to distinguish vicariance from LDD in the case of continental populations separated by wide distances without a clear geographic barrier. The study of phylogenetic relationships and among-population genetic diversity by using molecular markers can be used to distinguish between these possibilities (Hewitt, 1999; Avise, 2000; Kropf *et al*., 2006).

The Iberian Peninsula, where European and North African taxa converge, is an area of high biodiversity (Valdés, 2006; Buira *et al*., 2017). Iberian plants related to the broader European flora have been studied extensively. Their relationship to other European taxa stems from the fact that the Iberian Peninsula was a refugial area for temperate plants during the Quaternary ice ages and a source area for subsequent northward colonisation (Hewitt, 1999). In contrast, studies focused on species distributed in both Iberia and North Africa and the causes of such distributions are much scarcer (Caujapé-Castells and Jansen, 2003; Pardo *et al*., 2008; Terrab *et al*., 2008; García-Aloy *et al*., 2017; Villa-Machío *et al*., 2018) in spite of the fact that the floras of these two regions show clear similarities (Valdés, 2006; Fennane and Ibn Tattou, 2012). Dispersal across the Strait of Gibraltar has been demonstrated to be responsible for the distribution of some Ibero–North African taxa (Migliore *et al*., 2012; Lavergne *et al*., 2013; García-Aloy *et al*., 2017) even though this geographical feature has apparently acted as a barrier to migration or gene flow in many other taxa (e.g. Escudero *et al*., 2008; Terrab *et al*., 2008; Casimiro-Soriguer *et al*., 2010). North Africa is also known to have been the source

area for colonisation of the Iberian Peninsula by taxa originating in the Middle East or distributed in the Mediterranean region (Lumaret *et al*., 2002; Pardo *et al*., 2008; Pérez-Collazos *et al*., 2009).

The genus *Scrophularia* includes approximately 270 species (Ortega-Olivencia and Devesa, 1993) of primarily Holarctic distribution. The genus originated during the Miocene (Navarro-Pérez *et al*., 2013, 2015), probably in Southwest Asia, its primary centre of diversity (Scheunert and Heubl, 2017). In Europe, the most important centre of diversity is the Iberian Peninsula, where 12 endemic species and subspecies are present (Ortega-Olivencia and Devesa, 1993; Ortega-Olivencia, 2009). Iberian endemics are closely related to those of Morocco (Ortega-Olivencia and Devesa, 2002; Ibn Tattou, 2007) and Macaronesia (Dalgaard, 1979), where five and ten species, respectively, are also endemic. Phylogenetically, Macaronesian lineages are closely related to western Mediterranean ones (Scheunert and Heubl, 2014; Navarro-Pérez *et al*., 2015). One of the most interesting species is *Scrophularia arguta* Sol. in Aiton, an annual, widespread plant inhabiting Macaronesia, south-western Europe, North Africa, the Horn of Africa, Arabia and Socotra Island (Ortega-Olivencia, 2009) (Fig. 1) that is the only known amphicarpic species in the genus. In Europe, *S. arguta* is restricted to the Iberian Peninsula, where it is located in two disjunct areas of Spain (Southeast and West-Central).

The collection history of *S. arguta* in Europe is unusual, as the number of discovered populations has increased in the last decades after a long time (nearly 140 years) without any reference or specimen collected in Europe. The first recorded specimen of this species in Europe was collected by E. Bourgeau in Sierra de Gádor (Almería Province) on 7 May 1851 and is housed in the herbarium of the Natural History Museum in Paris (Pl. d'Esp. Exsicc. 1388a, P 89/66P) (see Gay, 1856; Ortega-Olivencia and Devesa, 1993). The species was considered practically extinct in Spain until Molero (1989) collected it in

Águilas (Murcia Province). It was later rediscovered in Sierra de Gádor (Mota *et al*., 2005). Afterwards, three new populations were found, two in West-Central Spain (Sierra de Santiago de Alcántara, north-western Cáceres Province; Ortega-Olivencia *et al*., 2006) and one in Southeast Spain (southern Granada Province; Cabezudo *et al*., 2010). Finally, a few more populations have been discovered in the last few years (see Fig. 1), but always in Southeast or West-Central Spain. The distance between the westernmost population (Santiago de Alcántara) and the easternmost population (Mazarrón) is 560 km.

All these European populations are peripheral in the context of the widespread distribution of *S. arguta*. After discovering the West-Central Iberian populations, Ortega-Olivencia *et al*. (2006) hypothesised that the current disjunct distribution of peripheral populations of *S. arguta* in the Iberian Peninsula is the result of either vicariance or LDD. Climate forecasts for the southern half of the Iberian Peninsula predict a future increase in mean temperature and a decrease in precipitation (Brunet *et al*., 2009) that could favour the expansion of Ibero-North African species such as *S. arguta*.

Populations in the western part of the range of *S. arguta*, together with a small sample of eastern populations, were analysed by Valtueña *et al*. (2016) using two nuclear and two plastid DNA regions. In that study, two Iberian populations of *S. arguta* (one from the West-Central region and the other from the Southeast) were included. These peripheral populations had different haplotypes and did not group together in the same clade, which suggested that they originated from independent colonisation events. Nevertheless, a more exhaustive study (including more Iberian populations, additional genetic markers and modelling approaches) is required to determine the origins of peripheral European populations and estimate their levels of genetic diversity, differentiation and gene flow. Indeed, the approach of combining genetic data and species

distribution modelling is a promising strategy to reconstruct the history of species (e.g. Fernández-Mazuecos & Vargas, 2013; Leipold *et al*., 2017; see Morris and Shaw, 2018).

Given this background, the main aims of the present study were to: (1) use phylogeographic analyses (based on nuclear and plastid DNA sequences and AFLP markers) and species distribution modelling to estimate the temporal and geographical origin of the peripheral European populations of *S. arguta* and therefore determine the causes behind their disjunct distribution (vicariance *vs.* LDD), and (2) analyse the genetic diversity of *S. arguta* populations in the western part of its range to ascertain the importance of European populations in the context of the global genetic diversity of the species. With respect to the first objective, we aimed to determine the number, timing and geographical origin (Northwest Africa and/or Macaronesia) of colonisation events that gave rise to the Iberian populations. The second objective will provide useful information to inform conservation policies aimed at maintaining the genetic diversity and evolutionary potential of this species.

2. MATERIALS AND METHODS

2.1. Study species

Scrophularia arguta is one of the few annual species and the only amphicarpic species (i.e., having aerial, chasmogamous flowers as well as basal and/or underground, cleistogamous ones) in the genus. It is distributed in northern Africa and the Horn, Arabia and Macaronesia, with a few small, peripheral populations found in south-western Europe (Iberian Peninsula; Fig. 1), where it is rare. In the Iberian Peninsula, the ecology of south-eastern populations is different to that of West-Central (Extremadura) populations: the first set of populations inhabits xeric soils on calcareous or dolomite substrates at low altitudes (< 400 m asl), whereas the Extremaduran populations inhabit quartzitic rocky

areas over 500 m asl. In both cases, plants are found on somewhat nitrified soils at the foot of shaded walls or between the cracks of these walls.

2.2. Sampling strategy

We studied 32 *S. arguta* populations representing most of this species' western distribution (Fig. 1). The sampling included five populations from the Iberian Peninsula (i.e., two from Extremadura [West-Central] and one each from the south-eastern provinces of Granada, Almería and Murcia), 12 from Morocco and 15 from the Canary Islands (Table 1 and Fig. 1). Three of the Iberian populations were newly sampled for this study (IB2, IB4 and IB5), while the remaining 29 populations were sampled for previous studies based on external transcribed spacer (ETS), internal transcribed spacer (ITS) and ptDNA sequences (Valtueña *et al*., 2016, 2017). During several collection trips, we searched for further Iberian populations without success. In addition, four populations from north-eastern Africa and the Arabian Peninsula were sampled to represent the eastern part of the species distribution. Twenty-four of the 32 western populations (Table 1) were selected for the generation of amplified fragment length polymorphism (AFLP) markers, which have not been used to date in this species. Depending on population abundance, young leaves were collected from one to 20 individuals per population and stored in silica gel until analysis.

2.3. DNA extraction and sequencing

Genomic DNA was extracted using a Qiagen DNeasy Plant Mini kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocol.

In a previous study (Valtueña *et al*., 2016), two nuclear (ITS/ETS) and two plastid (*psb*A-*trn*H and *psb*J-*pet*A) regions were sequenced from 59 individuals and 32 populations of *S. arguta* and one individual of the outgroup species *S. megalantha* (GenBank accession numbers: KU945636–KU945694 for ETS, KC692563 and

KU926629–KU926686 for ITS, KU945695–KU945754 for *psb*A-*trn*H and KU945755–KU945815 for *psb*J-*pet*A). For this study, the same four regions were sequenced in two individuals of each newly sampled population (i.e., 24 new sequences). Amplification of these regions was performed as described by Valtueña *et al*. (2016). Sequencing was carried out in both strand directions by the Service of Applied Techniques for Biosciences (SATB) of Extremadura University (Badajoz, Spain). Sequences were manually checked and edited using Sequencher v4.10 (GeneCodes, Ann Arbor, MI, USA) and aligned by eye with MacClade v4.08 (Maddison and Maddison, 2005). All new sequences obtained in this study were submitted to GenBank (see Table S1).

2.4. Sequences analyses

In the previous investigation (Valtueña *et al*., 2016), sequences were analysed using statistical parsimony in TCS (Clement *et al*., 2000) to obtain a haplotype network, and using Bayesian inference in BEAST 1.8.2 (Drummond *et al*., 2012) to determine phylogenetic relationships of populations. These analyses were repeated using the same settings and calculation procedures but including the three newly sampled Iberian populations.

The absence of resolution at the base of the *S. arguta* tree and the radial pattern of the haplotype network obtained in the previous study (Valtueña *et al*., 2016) were interpreted as the result of demographic expansion, and the date of differentiation was estimated to be c. 3.28 Ma ago. In the present study, we therefore performed a mismatch distribution analysis (MDA) of combined ptDNA sequences of all sampled *S. arguta* populations to confirm the demographic expansion. To further verify the expansion model, we compared the raggedness index (HRag; Harpending, 1994) and goodness of fit based on the sum of squared deviations (SSD) between observed and expected mismatch distributions. Small values of HRag support sudden population expansion whereas high values suggest

stationary or bottlenecked populations (Harpending *et al*., 1993; Harpending, 1994). The date of a confirmed expansion (t) can be estimated according to the formula $t = \tau/2\mu k$, where τ is the expansion parameter, μ is the mutation rate per site and generation, and k is the sequence length. The analysis was performed in Arlequin 3.5.2.2 (Excoffier and Lischer, 2010) with the same ptDNA matrix used in the TCS analysis after removing the outgroup taxon *S. megalantha* (to avoid overestimation of the number of mutations and the value of τ, complex indels were treated as single mutations). The assumed mutation rate based on plastid noncoding regions was 1.52×10^{-9} substitutions per site per year (Yamane *et al*., 2006), and the sequence length was 1142 bp.

We reconstructed historical migration routes of *S. arguta* using Bayesian discrete phylogeographic analysis (DPA) (Lemey *et al*., 2009). ITS/ETS and ptDNA sequences were analysed separately. Analyses were conducted in BEAST v1.8.4. (Drummond *et al*., 2012). We defined four areas (Canary Islands, Iberian Peninsula, Northwest Africa and Northeast Africa) and assigned each *S. arguta* sample to one of them. *S. megalantha* was included as outgroup (with undetermined distribution), and all ingroup (*S. arguta*) sequences were constrained as a monophyletic group. Substitution models were selected using jModeltest (Darriba *et al*., 2012), and areas were mapped under an asymmetric substitution model. To simplify the analyses and facilitate convergence, a strict molecular clock (with rates as specified earlier) and a constant-size coalescent tree prior were implemented. For each analysis, two MCMC analyses were run for 10 million generations, sampling every 1000 generations. Analysis with Tracer (Rabaut and Drummond, 2007) confirmed adequate sample sizes. The two chains were combined using LogCombiner (Drummond *et al*., 2012) after discarding the first 10% of sampled generations as burn-in, and trees were summarized in a MCC tree using TreeAnnotator (Drummond *et al*., 2012).

2.5. AFLPs

For the AFLP analysis, we used 305 individuals belonging to 24 populations (populations with only one to three sampled individuals were excluded; Table 1). All reactions were performed simultaneously for all samples in three steps. Restriction-ligation and preselective amplification steps followed the protocol of Kropf *et al*. (2002), except that 30 cycles were used instead of 20 cycles in the preselective PCR program. The selective amplification followed that of Trybusch *et al*. (2006). Adapters and primers used in each step are listed in Table S2. AFLP products were separated as a multiplex using primers labelled with different fluorescent dyes (6-Fam, Hex and 56-TAMN) by the SATB. Reproducibility of the results was confirmed using replicate extracts of 14 samples.

AFLP fragments were scored manually with GeneMarker 1.5 (SoftGenetics, State College, PA, USA) for the presence/absence of fragments between 70 and 450 bp in size. Fragments that could not be scored unambiguously were excluded. The results of the scoring were converted to a presence/absence (i.e., 1/0) matrix and used for further analysis. Because all mismatched fragments among sample replicates were removed from the matrix, the matrix error rate was zero.

The genetic structure of the dataset was analysed using BAPS v6.0 (Corander *et al*., 2008). The analysis was run 20 times for $K = 2$ to 24, where K is the maximum number of populations assumed to be present in the sample. Admixture analyses (Corander and Marttinen, 2006) were run with 100 iterations to estimate admixture coefficients for individuals, with 200 reference individuals from each population and 20 iterations used to estimate admixture coefficients for reference individuals. In addition to this analysis, a spatial genetic mixture analysis of groups (Cheng *et al*., 2013) was performed in which the populations and population coordinates were incorporated into the analysis and K was set to 24. We additionally implemented Bayesian clustering in STRUCTURE (Pritchard

et al., 2000; Falush *et al*., 2003) with a number of clusters $K = 1$ to 24 (number of sampled populations). For each *K*, ten independent chains were run with 1,000,000 MCMC replicates after 100,000 burn-in replicates, admixture model and independent allele frequencies among populations. Post-processing was carried out with CLUMPAK (Kopelman *et al*., 2015) to compare *K* values, using Delta *K* (Evanno *et al*., 2005) and Prob(*K*), and to summarise results. In addition, to confirm genetic groups found in the BAPS and STRUCTURE analyses, a principal coordinates analysis (PCoA), a non-hierarchical grouping technique not dependent on prior knowledge of the source location of the sampled individuals, was conducted in GenAlEx v6.502 (Peakall and Smouse, 2006).

Various diversity parameters (number of fragments, number of private fragments, percentage of polymorphic loci and heterozygosity [H_E]) were calculated in GenAlEx at both population and group levels (groups being phylogenetic clades based on the combined analysis in BEAST –which are the same clades found in the DPA analysis of the plastid regions– and main haplogroups found in the TCS analysis). Gene diversity (total gene diversity [H_T], within-population gene diversity [H_S] and genetic differentiation [F_{ST}]) at the group level (phylogenetic clades and main haplogroups) were calculated in PopGene 32 (Yeh *et al*., 2000). The distribution of genetic variation in *S. arguta* was assessed by analysis of molecular variance (AMOVA) as implemented in Arlequin 3.5.2.2 (Excoffier and Lischer, 2010). The AMOVA was run with 1000 permutations considering the whole dataset (24 populations), with two and three hierarchical levels (i.e., considering the five phylogenetic clades and the seven main haplogroups).

2.6. Species distribution modelling

We used species distribution modelling (SDM) to evaluate the potential range of *S. arguta* in the present and to project it to late Quaternary conditions, including the last interglacial (LIG, c. 120–140 ka), the last glacial maximum (LGM, c. 22 ka) and the mid-Holocene (MH, c. 6 ka). The analysis was focused on our study region, including the Iberian Peninsula, Northwest Africa and the Canary Islands (26º to 45º N; 20º W to 5º E). Highly accurate geographic coordinates of 35 localities of *S. arguta* in this region were obtained from our sampled populations and herbarium specimens. The scarcity of occurrence data in eastern Africa prevented us from including this area. To reduce sampling bias, the dataset was filtered by randomly removing points that were within a buffer of 0.2º, resulting in a final dataset of 23 occurrences.

Layers of 19 bioclimatic variables under present conditions at a resolution of 30” were obtained from WorldClim 1.4. To avoid collinearity, we excluded variables displaying a high correlation with other variables ($|r| > 0.7$) in the study region. As a result, a set of seven variables were selected: bio1, bio3, bio4, bio6, bio12, bio14 and bio15. SDM was performed using the maximum entropy algorithm, implemented in Maxent v3.4 (Phillips *et al*., 2006). We used 80% of occurrences for model training, and 20% for model evaluation. Ten subsample replicates were run, and a mean model was calculated. The model was projected to past conditions using LIG layers from Otto-Bliesner *et al*. (2008) and LGM and MH layers based on the CMIP5 project (three global climate models [GCMs]: CCSM4, MIROC-ESM and MPI-ESM-P). Resolution was 30” for LIG and MH layers, and 2.5’ for LGM layers. For LGM and MH, an average of the three GCMs was calculated.

3. RESULTS

3.1. Sequences analyses

Iberian populations did not cluster together in the combined nuclear and plastid Bayesian tree (Fig. S1). Five main clades were identified in the tree (posterior probability = 0.99–1.00), but their relationships were unsupported. Two of these clades included Canarian populations only (Fig. S1): one comprising populations from the three westernmost islands sampled (western Canarian clade, WCa), and the other including populations from the two easternmost islands (eastern Canarian clade, ECa). Iberian populations were included in the remaining three clades. The two West-Central Iberian populations (IB1 and IB2) constituted their own clade (Extremaduran clade, Ex). The two south-eastern populations (IB3 and IB4, from Murcia and Almeria provinces, respectively) were included in a clade together with populations from Morocco (located on the Atlantic coast and the central Moroccan interior) and eastern populations (Arabia, Sudan and Socotra Island; widespread clade, Wid clade). Finally, the south-eastern Iberian population closest to Morocco (IB5, Granada Province) grouped with northern Moroccan populations located along the Mediterranean coast (North Morocco-Iberian clade, Mo-Gr).

Two of the three newly analysed Iberian populations had the same haplotypes as their nearest previously studied Iberian neighbour. In particular, population IB2 shared haplotype C with population IB1, and population IB4 shared haplotype A6 with population IB3 (Table 1 and Fig. 1). In contrast, population IB5 showed a haplotype that was not found in our previous study. This haplotype (F2) was only one mutational step away from haplotype F1 (haplotype F in our previous study). Haplotype F1 was only found in population MO3 (Table 1 and Fig. 1), located near the Moroccan coast across from population IB5.

The results of the MDA supported a sudden expansion model (HRag = 0.01062, p = 0.518; SSD = 0.00768, p = 0.417). The value of τ was 12.34378, with the expansion accordingly dated to 3.55 Ma.

DPAs for ptDNA and ITS/ETS sequences recovered some uncertainty on the location of the most recent common ancestor of all *S. arguta* samples, with similar posterior probabilities for the Canary Islands and Northwest Africa, and lower values for the remaining two areas (Figs. 2 and S2). In the ptDNA analysis (Fig. 2), five major clades were recovered, matching those of previous analyses. Two of these clades had a most probable ancestral location in the Canary Islands, two in Northwest Africa and one in the Iberian Peninsula. Within one of the ancestrally Northwest African clades (Wid), the analysis inferred migrations to Northeast Africa (two events), the Canary Islands (one event) and the Iberian Peninsula (one event, populations IB3 and IB4). A second probable migration to the Iberian Peninsula from Northwest Africa was inferred in the Mo-Gr clade (population IB5). A third migration to the Iberian Peninsula was suggested (populations IB1 and IB2), but its origin remained inconclusive. The analysis estimated that all migration events to the Iberian Peninsula occurred during the Quaternary, specifically in the last 2 million years. The ITS/ETS analysis was hardly informative due to the low probabilities of most ancestors (Fig. S2), but it was consistent with multiple origins of Iberian populations, including a supported Northwest African origin of population IB5.

3.2. AFLPs

The six AFLP primers generated 492 fragments. The mean number of fragments per individual was 122.0 ± 16.2 (range: 66–157).

The best partition uncovered in the BAPS analysis was at $K = 13$. However, at $K = 9$, the rate of change in the log probability of the data between successive K values was stabilised, which is considered to be a better indicator of the number of clusters (Fig. S3). In both cases ($K = 13$, Fig. 3A; and $K = 9$, Fig. S4), Iberian populations were included in three different clusters. Populations IB1 and IB2 formed a clearly differentiated cluster that appeared at $K = 3$ (Fig. S4). Populations IB3 and IB4 formed an exclusive cluster at

K = 13 (Fig. 3A), but they were clustered with MO8 at lower values of *K* (Fig. S4). Finally, population IB5 clustered with population MO3 and a few individuals of population MO4 at *K* = 13 (Fig. 3A), with population MO2 also included in this group at *K* = 9 (Fig. S4). In all Iberian populations, all individuals from a given population were assigned to the same cluster, as confirmed by the admixture analysis (Fig 3B). The best partition obtained from spatial clustering of groups of individuals was at *K* = 11 (log[marginal likelihood] = –42963.2343). Populations IB1 and IB2 formed a cluster, as did populations IB3 and IB4 (Fig. 3C). Population IB5 shared a cluster with populations MO2 and MO3 (Fig. 3C). In STRUCTURE analyses, Prob(*K*) suggested *K* =19 as the optimal clustering (Prob[*K*] = 1), while Delta *K* displayed the highest value at *K* = 3, with a further small peak at *K* = 9 (Fig. S3C). Results for *K* = 9, which were considered the most informative for our purposes, are shown in Fig. 3D whereas values from *K* = 2 to *K* = 12 are shown in Fig S5. The clustering was similar to that recovered from BAPS, but with a higher degree of admixture. Iberian populations were again recovered in three different clusters. Populations from West-Central Iberian Peninsula (IB1 and IB2) formed a cluster on their own, while populations from Southeast Iberia were clustered with North African populations. Specifically, populations IB3 and IB4 were clustered with a population from southern Morocco (MO8) while population IB5 was clustered with Moroccan populations located just across the Mediterranean Sea (MO2, MO3 and MO4) (Fig. 3D).

In the PCoA analysis, the first two axes explained 27.2% of the variance, with this percentage increasing to 33.4% when the third axis was considered. These results confirmed the clear differentiation of populations IB1 and IB2 with respect to the remaining populations (Fig. 4). When only the first two axes were considered, individuals from populations IB3 and IB4 clustered together with individuals from population MO8,

whereas individuals from population IB5 overlapped with those of populations MO2, MO3 and MO4 (Fig. 4A).

Values of genetic parameters calculated from the AFLP data at population, phylogenetic clade and main haplogroup levels are shown in Table 2. At the population level, the number of fragments varied between 173 (MO3) and 304 (MO10). The percentage of polymorphic loci ranged between 19.51% (LA7) and 59.55% (MO10), while H_E varied between 0.059 (LA7) and 0.152 (MO10). Only four populations showed a private fragment (IB1, TE1, TE2 and MO2). Iberian populations had intermediate values of the three parameters, with population IB1 displaying the highest number of fragments and IB4 the highest percentage of polymorphic loci and heterozygosity. When the phylogenetic clades uncovered by the Bayesian analysis were considered, the lowest number of fragments (276) was found in the Ex and WCa clades, with the former also exhibiting the lowest number of polymorphic loci (46.34%) and heterozygosity (0.120).

Gene diversity and differentiation values at the level of species, phylogenetic clades and main haplogroups are shown in Table 3. The H_S index (reflecting within-population gene diversity) at the species level was 0.108 and the H_T index (total gene diversity) was 0.222, while genetic differentiation among populations was very high (F_{ST} = 0.515). When phylogenetic clades were considered, H_S varied between 0.089 (ECa clade) and 0.124 (Wid clade), and H_T ranged from 0.119 (Ex clade) to 0.185 (Wid clade). The highest genetic differentiation was found in the ECa clade (F_{ST} = 0.400) and the lowest in the Ex clade (F_{ST} = 0.075). With respect to haplogroups, H_S varied between 0.089 (haplogroup D) and 0.128 (E), and H_T ranged from 0.119 (C) to 0.193 (E). The highest genetic differentiation was found in haplogroup D (F_{ST} = 0.400) and the lowest in haplogroup C (F_{ST} = 0.075).

According to the results of the AMOVA when populations were not grouped into clusters, genetic variation was similar among and within populations (53.6% *vs.* 46.4%, respectively; Table 4). When phylogenetic clades and main haplogroups were considered, the genetic variation found within populations was similar (43.7% *vs.* 44.5%). The genetic variation among groups and among populations was almost the same for haplogroups (27.9% *vs.* 27.6%); in the case of phylogenetic clades, however, the genetic variation found among groups was clearly higher than among populations (30.4% *vs.* 25.9%).

3.3. Species distribution modelling

Areas climatically suitable for *S. arguta* in the present were inferred mostly in western Morocco, north-eastern Morocco, north-western Algeria and the Canary Islands (Fig. 5). Narrow suitable areas were also inferred in the southern coast of the Iberian Peninsula, including known localities in Southeast Spain. However, localities in West-Central Spain (Extremadura) displayed low suitability. Projections to the past (Fig. 5) revealed stable suitable areas throughout the LIG, LGM and MH in western Morocco, the Canary Islands and, to a lesser degree, north-eastern Morocco. Low suitability in the Iberian Peninsula was inferred for the LIG, but suitable areas appear to have been present in south-eastern Spain at least since the LGM. Low suitability was estimated in West-Central Spain for all periods.

4. DISCUSSION

4.1. Genetic structure of western *Scrophularia arguta* populations

In a previous study, the last common ancestor of *Scrophularia arguta* lineages was dated to the late Pliocene (c. 3.28 Ma; Valtueña *et al*., 2016) based on two secondary calibration points. In the current study, mismatch distribution analysis confirmed this result but gave a slightly older date (3.55 Ma). The DNA sequences (Figs. 1, 2, S1) and AFLP data (Figs. 3, 4; Tables 2-4) revealed several genetic lineages in this species that mainly

corresponded to different geographical regions. The same result was obtained by Valtueña *et al*. (2016), who used ITS/ETS and plastid DNA sequences only, and was explained by the life-history traits of *S. arguta*: annual habit and reproductive success largely dependent on the production of cleistogamous flowers. The westernmost portion of the range of this species encompasses Macaronesia, Northwest Africa and the Iberian Peninsula, where several genetic lineages with distinct origins are found (see below).

With respect to Macaronesia, Valtueña *et al*. (2016) using plastid markers and focusing on the Canary Islands uncovered two ancient lineages with distinct distributions on the islands. Those findings were confirmed in the present study by the AFLP clustering analysis. The presence of ancient endemic lineages is also consistent with the long-term persistence of suitable areas on the Canary Islands in the Quaternary, as indicated by SDM (Fig. 5). The present genetic analyses additionally suggested a close relationship between some populations of Lanzarote (ECa clade) and Atlantic Moroccan coastal populations at low values of *K*. This suggests the existence of long-distance gene flow between the two areas (Atlantic Moroccan coast and Lanzarote). Long-distance gene flow has been previously proposed for other species exhibiting clear differentiation based on plastid DNA (Besnard *et al*., 2007; Migliore *et al*., 2012). Gene flow between both areas is supported by the STRUCTURE results, where most of the individuals from population MO1 show an important probability to belong to a genetic group constituted by Lanzarote populations at $K = 9$ (Fig. 3D). This result indeed implies a directionality of gene flow from the island to mainland. Gene flow may have been favoured during the LGM by marine regression that reduced the distance between both areas (Fernández-Palacios *et al.*, 2011).

Phylogenetic analysis indicated the existence of two clades in Northwest Africa, one of which included populations from the eastern range of the species (Wid clade; Figs. 2,

S1). The exact biogeographic relationship between eastern and western populations is still unclear. While the haplotype network suggests that haplotype A colonised Morocco from the east and was the founder of the Mo-Gr clade (see Valtueña *et al*., 2016), DPA analysis estimates that eastern populations most probably originated from western populations (Fig. 2). In any case, the high genetic diversity of *S. arguta* in Morocco seems to be due to *in situ* differentiation probably as result of climatic fluctuations over the last three million years (Lavergne *et al*., 2013) that favoured differentiation in isolated areas rather than multiple waves of colonisation from the eastern range of the species or recolonisation from Europe/Macaronesia. This is consistent with the long-term climatic suitability in Northwest Africa estimated by SDM (Fig. 5) and supports the idea that this area acted as a refuge for numerous taxa during the Pleistocene (Rodríguez-Sánchez *et al*., 2008; Désamoré *et al*., 2011; García-Aloy *et al*., 2017). The major lineages of *S. arguta* most likely arose by differentiation in isolated refugial areas during unfavourable periods when populations experienced range reductions. This contrasts with the situation in Macaronesia and Europe, where the multiple genetic groups of *S. arguta* are explained by several independent colonisation events.

4.2. Three independent origins of peripheral European populations

Both the nucleotide and AFLP data (Figs. 1-4) support at least three jumps onto the Iberian Peninsula involving three different geographic areas: (a) the province of Granada, (b) Southeast Iberia (Murcia and Almería provinces) and (c) West-Central Iberia (Extremadura). In addition, the results of the AFLP analysis demonstrate the absence of gene flow among the three areas. Gene flow was not detected even between the two geographically closest areas (196 km between IB4 and IB5).

Following the discovery of the Extremaduran populations, the disjunct distribution of *S. arguta* in the Iberian Peninsula was suggested to be a consequence of either vicariance

or LDD (Ortega-Olivencia *et al*., 2006). Valtueña *et al*. (2016) supported the hypothesis of two different colonisation events (two populations analysed, one from each Iberian area, fell into different clades). Our latest results, including genetic analyses and SDM, reject the vicariance hypothesis, and indicate that three colonisation events, rather than two, have taken place. Although repeated waves of colonization have been frequently demonstrated in other geographical contexts (e.g. Reisch, 2008; Escudero *et al.*, 2010), only two herbaceous species are currently known to have colonised the Iberian Peninsula from Northwest Africa more than once: the annual *Hypochaeris glabra* and the perennial *H. radicata* (Ortíz *et al*., 2008, 2009). In both of these cases, two colonisation events were inferred. Our study is the first to demonstrate the occurrence of at least three different colonisation events. Although *S. arguta* does not apparently have a syndrome for LDD, this species exhibits some life-history traits (annual habit, self-compatible breeding system, cleistogamous flowers, small seeds [dust-like in chasmogamous flowers] and a preference for somewhat disturbed habitats) that are compatible with this mode of dispersal (Lavergne *et al*., 2013; García-Aloy *et al*., 2017).

The Granadian population (IB5) is closely related to coastal Mediterranean populations of Morocco, as these populations formed a clade in the phylogenetic analysis and appeared in the same genetic group in the AFLP clustering analyses. Both results and potential areas estimated by SDM support the establishment of the IB5 population as a result of a relatively recent dispersal. The absence of AFLP differentiation from Mediterranean Moroccan populations suggests a late Quaternary dispersal from this area, probably the most recent dispersal event involving the Iberian Peninsula (see also DPA results in Fig. 2). Being located at the edge of the sea, containing few individuals and being restricted to a few meters of limestone-dolomitic rocky coast, the Granadian population is in a precarious situation: it may become extinct if the area suffers any

disturbances due to local anthropic activity or to the sea level rise induced by global climate change that is obviously occurring in the Mediterranean Sea as well as worldwide (Galassi and Spada, 2014; Zerbini *et al*., 2017).

Interestingly, our analyses revealed that the Murcian–Almerian populations (IB3, IB4) are closely related to those from West-Central Morocco. Two hypotheses might explain this relationship: (1) LDD from West-Central Morocco or (2) dispersal from northern Morocco (LDD or stepping stone dispersal), with the ancestral population of IB3 and IB4 subsequently becoming extinct, overrun by individuals with other genotypes, or not sampled by us. This second hypothesis is more likely, and the extinction or displacement of the ancestral population is supported by the fact that the Northwest African populations have suffered waves of contraction and expansion in response to climatic fluctuations in the Quaternary, as shown by SDM results (Fig. 5; Médail and Diadema, 2009; Hussemann *et al*., 2014). Under either scenario, the colonisation of the south-eastern peninsula by individuals with haplogroup A must have predated the colonisation that gave rise to the Granadian population (IB5), as IB3 and IB4 populations are genetically different from their most closely related Moroccan populations. This colonisation event could have used the large island remnants of a volcanic arc in the East Alboran basin existing until the Early Pleistocene (located to the East of the Strait of Gibraltar; Booth-Rea *et al*., 2018) as stepping stones between northern Morocco and south-eastern Spain.

According to the AFLP data (Table 2), the IB3 population has a lower number of fragments, percentage of polymorphic loci and heterozygosity, possibly a consequence of a founder effect if that population has recently originated from IB4. This low genetic variation may alternatively indicate that IB3 is suffering a bottleneck. This is supported by the low number of individuals present, i.e., an average of 33 mature individuals per

year (BORM, 2015) with annual fluctuations between 4 and 85. In 2012, when samples were collected, fewer than 20 individuals were present.

4.3. A peripheral relict population in West-Central Iberia

The Extremaduran populations (IB1, IB2) represent a well-differentiated lineage within *S. arguta* based on four lines of evidence. First, they are not grouped with any other populations in the phylogenetic tree (Figs. 2, S1). Second, they possess a unique plastid haplotype that differs by at least seven mutational steps from the closest haplotypes in any other population (Fig. 1). Third, they constitute an independent genetic group based on the AFLP clustering analyses, in which they formed the first group to be differentiated (beginning at $K = 3$; Figs. 3, 4, S4). And fourth, some degree of ecological differentiation exists in these populations, as their distribution is not predicted by SDM for any of the analysed periods (Fig. 5). In addition, the two populations have similar genetic characteristics; they share most identified AFLP loci and show low differentiation (F_{ST} = 0.075; Table 3). Considering these results together with the close proximity between the two populations (0.628 km), these seem to behave as a single population or be the result of recent fragmentation of a larger population. The high genetic differentiation between these plants and the remaining range of *S. arguta* for all molecular markers supports that the dispersal event that originated these populations is the oldest involving the Iberian Peninsula, predating the other two colonisations of Iberia uncovered in the present study (see Fig. 2).

Our results provide support for the idea that the Extremaduran populations constitute an evolutionary significant unit. In addition, given the location of these populations at the northern limit of the species range, they can be considered peripheral relict populations (Cassel-Lundhagen, 2010). In the haplotype network (Fig. 1), the haplotype of Extremaduran populations and those of the eastern Canarian populations share a common

ancestor, which is directly connected to the ancestral haplotype of the Wid clade (A1). However, given the low support for phylogenetic relationships of the Ex clade, the DPA was unable to confidently infer the origin of Extremaduran populations (Fig. 2), with the Canary Islands and Northwest Africa being the most likely options. The latter hypothesis seems more likely, as it implies a smaller dispersal distance and is consistent with the waves of contraction and expansion of Northwest African populations that may have led to the disappearance of ancestral genotypes, similar to the scenario described for the ancestor of Murcian–Almerian populations. The present-day range of *S. arguta* in West-Central Iberia is very restricted, and its habitat is completely different to that of other *S. arguta* populations, as supported by the low suitability in this area estimated by the SDM analysis for all periods. Their genetic parameters, however, do not show any evidence of negative effects caused by a reduced distribution and the lack of gene flow with other populations. This may be due to the fact that the two populations collectively have a large number of individuals (c. 4591 in 2006, at least 1285 in 2017 and c. 1000 in 2018), an atypical situation for this species.

4.4. The importance of peripheral populations for conservation

Interestingly, the history of European populations of *S. arguta* is very diverse, and it is not possible to refer to them as a single set. Each group of populations found in our study corresponds to a different type of peripheral population. The Granadian population (IB5) would be a peripheral population of an expanding species, and it has recently been originated by dispersal. As a result, genetic characteristics of this population are modelled by founder effect (low genetic diversity and no genetic differentiation from source populations; Schwaegerle and Schaal, 1979; Arnaud-Haond, 2006). Murcian–Almerian populations (IB3 and IB4) have been isolated from their source populations for a longer period of time, which has allowed the accumulation of mutations leading to some genetic

differentiation. In this case, a founder effect is not detected, and low genetic diversity is probably due to bottlenecks. Finally, Extremaduran populations (IB1 and IB2) are peripheral relict populations, displaying a strong differentiation from the rest of the populations of the species and a similar genetic diversity (Cassel-Lundhagen, 201**0;** Plenk *et al*., 2017). Overall, the European populations of *S. arguta* are a clear example of the evolution of peripheral populations in absence of gene flow, with population IB5 corresponding to the first stages after dispersal, populations IB1 and IB2 to populations isolated for a prolonged period of time, and populations IB3 and IB4 to an intermediate situation.

Therefore, the conservation value of the Iberian populations of *S. arguta* is unquestionable, given the idiosyncrasies of the three population groups. For example, populations in the south of the Iberian Peninsula probably contain alleles related to tolerance to water stress, while the Extremaduran populations may contain alleles related to adaptation to more mesic environments. Extremaduran plants are particularly relevant for conservation, since they make up one of the five major intraspecific clades on their own, and their extinction would mean the disappearance of genetic variants that are found nowhere else in the distribution range of *S. arguta*. The singularity of these populations reflects the importance of ancient peripheral populations for the global genetic diversity of species and highlights the need to understand the history of peripheral populations in order to conserve intraspecific diversity.

ACKNOWLEDGEMENTS

We thank the Environment Directorate (Murcia Region), especially M.Á. Carrión, for permission to collect samples (research permit AUT/FL/UND/021/2012) and also the environmental agents who accompanied us in the field. Appreciation is expressed to Prof. Enrique Rico for his interest in this paper. The English text of an early draft was edited

and improved by Dr. B. Goodson (Edanz Group Global Ltd.). This work was supported by the Spanish Ministry of Science and Innovation (CGL2011-24140) and the European Regional Development Fund. M.F.-M. was supported by a Juan de la Cierva fellowship (Spanish Ministry of Economy, Industry and Competitiveness, reference IJCI-2015-23459). We are grateful for comments and constructive criticism provided by two anonymous reviewers. The authors have no conflict of interest.

LITERATURE CITED

Antonovics J, Newman TJ, Best BJ. 2001. Spatially explicit studies on the ecology and genetics of population margins. In: Silvertown J, Antonovics J, eds. *Integrating Ecology and Evolution in a Spatial Context*. Oxford Blackwell, 91–116.

Arnaud-Haond S, Teixeira S, Massa SI, Billot C, Saenger P, Coupland G, Duarte CM, Serrão EA. 2006. Genetic structure at range edge: Low diversity and high inbreeding in Southeast Asian mangrove (*Avicennia marina*) populations. *Molecular Ecology* **15**: 3515–3525.

Avise JC. 2000. *Phylogeography: The history and formation of species*, Cambridge: Harvard University Press.

Besnard G, Christin PA, Baali-Cherif D, Bouguedoura N, Altheme F. 2007. Spatial genetic structure in the Laperrine's olive (*Olea europaea* subsp. *laperrinei*), a long-living tree from the central Saharan mountains. *Heredity* **99**: 649–657.

Booth-Rea G, Ranero CR, Grevemeyer I. 2018. The Alboran volcanic-arc modulated the Messinian faunal exchange and salinity crisis. *Scientific Reports* **8**: 13015.

BORM. 2015. Corrección de error en el Decreto nº. 244/2014, de 19 de diciembre, por el que se aprueban los planes de recuperación de las especies jara de Cartagena, brezo blanco, sabina de dunas, narciso de Villafuerte y *Scrophularia arguta*. *Boletín Oficial de la Región de Murcia* 54: 9580–9664.

Brunet M, Casado JM, de Castro M. *et al*. 2009. *Generación de Escenarios Regionalizados de Cambio Climático para España*. Madrid: AEMET.

Buira A, Aedo C, Medina L. 2017. Spatial patterns of the Iberian and Balearic endemic vascular flora. *Biodiversity and Conservation* **26**: 479–508.

Cabezudo B, Pérez Latorre AV, Casimiro-Soriguer F, Gavira O, García JA, Pavón M. 2010. Nota sobre *Abies pinsapo* Clemente ex Boiss. y otras plantas andaluzas. *Acta Botanica Malacitana* **35**: 195–204.

Casimiro-Soriguer R, Talavera M, Balao F, Terrab A, Herrera J, Talavera S. 2010. Phylogeny and genetic structure of *Erophaca* (Leguminosae), an East-West Mediterranean disjunct genus from Tertiary. *Molecular Phylogenetics and Evolution* **56**: 441–450.

Cassel-Lundhagen A. 2010. Peripheral relict populations of widespread species; evolutionary hotspots or just more of the same? In: Habel JC, Assmann T, eds. *Relict Species: Phylogeography and Conservation Biology*. Heidelberg Springer, 267–275.

Caujapé-Castells J, Jansen RK. 2003. The influence of the Miocene Mediterranean desiccation on the geographical expansion and genetic variation of *Androcymbium gramineum* (Cav.) McBride (Colchicaceae). *Molecular Ecology* **12**: 1515–1525.

Cheng L, Connor TR, Sirén J, Aenensen DM, Corander J. 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution* **30**: 1224–1228.

Clement M, Posada D, Crandall KA 2000. TCS, a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657–1660.

Cole CT. 2003. Genetic variation in rare and common plants. *Annual Review of Ecology, Evolution, and Systematics* **43**:213–237.

Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* **15**: 2833–2843.

Corander J, Martinen P, Siren J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**: 539.

Cox CB, Moore P D. 2016. *Biogeography: An Ecological and Evolutionary Approach*, 9th ed. Chichester: Willey Blackwell.

Dalgaard V. 1979. Biosystematics of the Macaronesian species of *Scrophularia*. *Opera Botanica* **51**: 1–64.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2, more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.

Désamore A, Laenen B, Devos N *et al*. 2011. Out of Africa, north-westwards Pleistocene expansion of the heather *Erica arborea*. *Journal of Biogeography* **38**: 164–176.

Drummond A J, Ho SYW, Rawlence N, Rambaut A.2007. *A Rough Guide to BEAST 1.4* https://code.google.com/archive/p/beast-mcmc/downloads. Date of last successful access (11 Jan. 2018).

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**: 1307–1320.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**:1969–1973.

Eckert CG, Samis KE, Lougheed SC. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* **17**: 1170–1188.

Escudero M, Vargas P, Arens P, Ouborg NJ, Luceño M. 2010. The east−west−north colonization history of the Mediterranean and Europe by the coastal plant *Carex extensa* (Cyperaceae). *Molecular Ecology* **19**: 352–370.

Escudero M, Vargas P, Valcárcel V, Luceño M. 2008. Strait of Gibraltar: an effective gene-flow barrier for wind-pollinated *Carex helodes* (Cyperaceae) as revealed by DNA sequences, AFLP, and cytogenetic variation. *American Journal of Botany* **95**: 745–755.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564–567.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Farris JS, Källersjö M, Kluge AG, Bult C. 1994. Testing significance of incongruence. *Cladistics* **10**: 315–319.

Fennane M, Ibn Tattou M. 2012. Statistiques et commentaires sur l'inventaire actuel de la flore vasculaire du Maroc. *Bulletin de l'Institut Scientifique, Rabat* **34**: 1–9.

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.

Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García-del-Rey E, Arévalo JR, Whittaker RJ. 2011. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography* **38**: 226–246.

Frankham R, Ballou JD, Briscoe DA. 2010. *Introduction to conservation genetics*. Cambridge: Cambridge University Press.

Galassi G, Spada G. 2014. Sea-level rise in the Mediterranean Sea by 2050: roles of terrestrial ice melt, steric effects and glacial isostatic adjustment. *Global and Planetary Change* **123**: 55–66.

García-Aloy S, Vitales D, Roquet C *et al*. 2017. North-west Africa as a source and refuge area of plant biodiversity: a case study on *Campanula kremeri* and *Campanula occidentalis*. *Journal of Biogeography* **44**: 2057–2068.

Gay MJ. 1856. IV–.M. Durieu de Maisonneuve a fait sur la végétation du Scrofularia argute Ait. des observations très curieuses, dont il m'a donné connaissance en ces termes, par lettre du 1er août dernier. *Bulletin de la Société Botanique de France* **3**: 569–570.

Harpending HC, Sherry ST, Rogers AR, Stoneking M. 1993. The genetic structure of ancient human populations. *Current Anthropology* **34**: 483–496.

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

Hewitt GM. 2014. Postglacial recolonization of European biota. *Biological Journal of the Linnean Society* **68**: 87–112.

Hussemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC. 2014. Paleartic biogeography revisited: evidence for the existence of a North African refugium for Western Palaeartic biota. *Journal of Biogeography* **41**: 81–94.

Ibn Tattou M. 2007. Scrophulariaceae. In: Fennane M, Ibn Tattou M, Mathez J, Ouyahya A, El Oualidi J, eds. *Flore Pratique du Maroc. Manuel de Détermination des Plantes Vasculaires – Angiospermae (Leguminosae-Lentibulariaceae)*. Rabat: Institut Scientifique, Université Mohammed V, vol. 2: 514–520.

Kropf M, Comes HP, Kadereit JW. 2006. Long-distance dispersal *vs* vicariance: the origin and genetic diversity of alpine plants in the Spanish Sierra Nevada. *New Phytologist* **172**: 169–184.

Kropf M, Kadereit JW, Comes HP. 2002. Late Quaternary distributional stasis in the submediterranean mountain plant *Anthyllis montana* L. (Fabaceae) inferred from ITS sequences and amplified fragment length polymorphism markers. *Molecular Ecology* **11**: 447–463.

Lammi A, Siikamäki P, Mustajärvi K. 1999. Genetic diversity, population size, and fitness in central and peripheral populations of a rare plant *Lychnis viscaria*. *Conservation Biology* **13**: 1069–1078.

Lavergne S, Hampe A, Arroyo J. 2013. In and out of Africa: how did the Strait of Gibraltar affect plant species migration and local diversification? *Journal of Biogeography* **40**: 24–36.

Lázaro-Nogal A, Matesanz S, García-Fernández A, Traveset A, Valladares F. 2017. Population size, center–periphery, and seed dispersers' effects on the genetic diversity and population structure of the Mediterranean relict shrub *Cneorum tricoccon*. *Ecology and Evolution* **7**: 7231–7242.

Leipold M, Tausch S, Poschlod P, Reisch Ch. 2017. Species distribution modelling and molecular markers suggest longitudinal range shifts and cryptic northern refugia of the typical calcareous grassland species *Hippocrepis comosa* (horseshoe vetch). *Ecology and Evolution* **7**: 1919–1935.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5**: e1000520.

Lumaret R, Mir C, Michauld H, Raynal V. 2002. Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Molecular Ecology* **11**: 2327–2336.

Maddison DR, Maddison WP. 2005. *MacClade v. 4.08.* http://www.macclade.org/ Date of last successful access (11 Jan. 2018).

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36**: 1333–1345.

Migliore J, Baumel A, Juin M, Médail F. 2012. From Mediterranean shores to central Saharan mountains, key phylogeographical insights from the genus *Myrtus*. *Journal of Biogeography* **39**: 942–956.

Milne RI. 2006. Northern hemisphere plant disjunctions: a window on Tertiary land bridges and climate change? *Annals of Botany* **98**: 465–472.

Molero J. 1989. Aportaciones a la flora del sudeste ibérico. II. *Anales del Jardín Botánico de Madrid* **47**: 265–268.

Morris AB, Shaw J. 2017. Markers in time and space: A review of the last decade of plant phylogeographic approaches. *Molecular Ecology* **17**: 2317–2333.

Mota JF, Cueto M, Pérez-García FJ *et al.* 2005. Contribución al conocimiento de la flora de Andalucía: citas novedosas de la provincia de Almería, el elemento estenócoro. *Acta Botanica Malacitana* **30**: 227–231.

Navarro-Pérez ML, Vargas P, Fernández-Mazuecos M, López J, Valtueña FJ, Ortega Olivencia A. 2015. Multiple Windows of colonization to Macaronesia by the dispersal-unspecialized *Scrophularia* since the Late Miocene. *Perspectives in Plant Ecology, Evolution and Systematics* **17**: 263–273.

Navarro-Pérez ML, López J, Fernández-Mazuecos M, Rodríguez-Riaño T, Vargas P, Ortega-Olivencia A. 2013. The role of birds and insects in pollination shifts of *Scrophularia* (Scrophulariaceae). *Molecular Phylogenetics and Evolution* **69**: 239–254.

Ortega-Olivencia A. 2009. *Scrophularia* L. In: Benedí C, Rico E, Güemes J, Herrero A, eds. *Flora Iberica.* Madrid: Real Jardín Botánico, CSIC, vol. **13**: 97–122.

Ortega-Olivencia A, Devesa JA. 1993. *Revisión del género Scrophularia L. (Scrophulariaceae) en la Península Ibérica e Islas Baleares.* Madrid: Monografías del Real Jardín Botánico de Madrid, CSIC, *Ruizia,* **11**.

Ortega-Olivencia A, Devesa JA. 2002. *Scrophularia* L. In: Valdés B, Rejdali M, Achhal El Kadmiri A, Jury JL, Montserrat JM, eds. *Checklist of Vascular Plants of N Morocco with Identification Keys.* Madrid: CSIC, vol. 2: 560–565.

Ortega-Olivencia A, Santos M, Valtueña FJ, Rico E. 2006. *Scrophularia arguta* Aiton en el occidente de la Península Ibérica. *Acta Botanica Malacitana* **31**: 238–240.

Ortíz MA, Tremetsberger K, Stuessy TF, Terrab A, García-Castaño JL, Talavera S. 2009. Phylogeographic patterns in *Hypochaeris* section *Hypochaeris* (Asteraceae, Lactuceae) of the western Mediterranean. *Journal of Biogeography* **36**: 1384–1397.

Ortíz MA, Tremetsberger K, Terrab A *et al.* 2008. Phylogeography of the invasive weed *Hypochaeris radicata* (Asteraceae): from Moroccan origin to world-wide introduced populations. *Molecular Ecology* **17**: 3654–3667.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. 2006. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**: 1751–1753.

Pardo C, Cubas P, Tahiri H. 2008. Genetic variation and phylogeography of *Stauracanthus* (Fabaceae, Genisteae) from the Iberian Peninsula and northern Morocco assessed by chloroplast microsatellite (cpSSR) markers. *American Journal of Botany* **95**: 98–109.

Peakall R, Smouse P E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.

Pérez-Collazos E, Sánchez-Gómez P, Jiménez JF, Catalán P. 2009. The phylogeographical history of the Iberian steppe plant *Ferula loscosii* (Apiaceae): a test of the abundant-centre hypothesis. *Molecular Ecology* **18**: 848–861.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231–259.

Plenk K, Bardy K, Höhn M, Thiv M, Kropf M. 2017. No obvious genetic erosion, but evident relict status at the westernmost range edge of the Pontic-Pannonian steppe plant *Linum flavum* L. (Linaceae) in Central Europe. *Ecology and Evolution* **7**: 6527–6539.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pujol B, Shu-Rong Z, Sanchez Vilas J, Pannell JR. 2009. Reduced inbreeding depression after species range expansion. *Proceedings of the National Academy of Sciences of the United States* **106**: 15379–15383.

Putz CM, Schmid C, Reisch. 2015. Living in isolation – population structure, reproduction, and genetic variation of the endangered plant *Dianthus gratianopolitanus* (Cheddar pink). *Ecology and Evolution* **5**: 3610–3621.

Rambaut A, Drummond AJ. 2007. *Tracer v1.4. MCMC Trace File Analyser* http://tree.bio.ed.ac.uk/software/tracer. Date of last successful access (11 Jan 2018).

Reisch C. 2008. Glacial history of *Saxifraga paniculata* (Saxifragaceae): molecular biogeography of a disjunct arctic-alpine species from Europe and North America. *Biological Journal of the Linnean Society* **93**: 385–398.

Rodríguez-Sánchez F, Pérez-Barrales R, Ojeda F, Vargas P, Arroyo J. 2008. The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews* **27**: 2100–2117.

Scheunert A, Heubl G. 2017. Against all odds: reconstructing the evolutionary history of *Scrophularia* (Scrophulariaceae) despite high levels of incongruence and reticulate evolution. *Organisms, Diversity and Evolution* **17**: 323–349.

Scheunert A, Heubl G. 2014. Diversification of *Scrophularia* (Scrophulariaceae) in the Western Mediterranean and Macaronesia- Phylogenetic relationships, reticulate evolution and biogeographic patterns. *Molecular Phylogenetics and Evolution* **70**: 296–313.

Schwaegerle KE, Schaal BA. 1979. Genetic variability and founder effect in the pitcher plant, *Sarracenia purpurea* L. *Evolution* **33**: 1210–1218.

Swofford DL. 2002. *PAUP, Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4.0b10*. Sunderland: Sinauer Associates.

Terrab A, Schönswetter P, Talavera S, Vela E, Stuessy TF. 2008. Range-wide phylogeography of *Juniperus thurifera* L., a presumptive keystone species of

western Mediterranean vegetation during cold stages of the Pleistocene. *Molecular Phylogenetics and Evolution* **48**: 94–102.

Trybusch S, Hanley S, Cho K-H *et al*. 2006. Getting the most out of fluorescent amplified fragment length polymorphism. *Canadian Journal of Botany* **84**: 1347–1354.

Valdés B. 2006. BIOGEO, a Biogeographical study of Andalusian flora in relation to that of N Morocco. *Bocconea* **19**: 161–168.

Valtueña FJ, López J, Álvarez J, Rodríguez-Riaño T, Ortega-Olivencia A. 2016. *Scrophularia arguta*, a widespread annual plant in the Canary Islands: a single recent colonization event or a more complex phylogeographic pattern? *Ecology and Evolution* **6**: 4258–4273.

Valtueña FJ, Rodríguez-Riaño T, López J, Mayo C, Ortega-Olivencia A. 2017. Peripatric speciation in an endemic Macaronesian plant after recent divergence from a widespread relative. *Plos One* 12: e0178459. doi.org/10.1371/journal.pone.0178459.

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.

Wiley EO. 1988. Vicariance biogeography. *Annual Review of Ecology, Evolution, and Systematics* **19**: 513–542.

Yamane K, Yano K, Kawahara T.2006. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Research* **13**: 197–204.

Yeh FC, Yang R, Boyle TJ, Ye Z, Xiyan JM. 2000. *PopGene32, Microsoft Windows–based freeware for population genetic analysis, version 1.32.* https://sites.ualberta.ca/~fyeh/popgene.html. Date of last successful access (11 Jan 2018).

Zerbini S, Raicich F, Prati CM *et al*. 2017. Sea-level change in the Northern Mediterranean Sea from long-period tide gauge time series. *Earth-Science Reviews* **167**: 72–87.

Table 1. Studied populations of *Scrophularia arguta*. Population abbreviations (Pop), locations, haplotypes (HAP), phylogenetic clades (PHY), numbers of individuals used (N) in the AFLP analysis and BAPS groups considering $K = 13$ (Group) are indicated
–: no individuals used in the AFLP analysis

				AFLP	
Pop	Location	HAP	PHY	N	Group
Iberian Peninsula (Spain):					
IB1	Spain, Cáceres, Santiago de Alcántara	C	Ex	20	1
IB2	Spain, Cáceres, Santiago de Alcántara	C	Ex	14	1
IB3	Spain, Murcia, Águilas	A_6	Wid	8	2
IB4	Spain, Almería, Pulpí	A_6	Wid	14	2
IB5	Spain, Granada, Almuñécar	F_2	Mo-Gr	9	3
Canary Islands (Spain):					
FU1	Fuerteventura, Tetir	D_1	ECa	13	4
FU2	Fuerteventura, Tiscamanita	D_1	ECa		–
FU3	Fuerteventura, Joros	D_1	ECa		–
GC	Gran Canaria, La Isleta	A_2	Wid		–
GO	La Gomera, Barranco de Guarimiar	G_2	WCa		–
LA1	Lanzarote, Tahiche	D_2	ECa	18	5
LA2	Lanzarote, Punta de las Mujeres	D_2	ECa	14	5
LA3	Lanzarote, Jameos del Agua	D_3	ECa	10	4/5
LA4	Lanzarote, Orzola	D_1	ECa		–
LA5	Lanzarote, San Bartolomé	D_1	ECa	12	4
LA6	Lanzarote, Tinajo	D_1	ECa	18	4
LA7	Lanzarote, El Golfo	D_3	ECa	15	4
PA	La Palma, Santa Cruz	G_2	WCa		–
TE1	Tenerife, Güimar	G_1	WCa	13	6
TE2	Tenerife, Pal-Mar	G_2	WCa	13	6/7
Northwestern Africa:					
MO1	Morocco, Safi Cape	E_1	Wid	13	8/9
MO2	Morocco, Zegangane	E_3	Mo-Gr	13	10
MO3	Morocco, Hassi-Berkane	F_1	Mo-Gr	5	3
MO4	Morocco, Had-Rouadi	E_3	Mo-Gr	15	3/11
MO5	Morocco, Beni-Sidel	E_3	Mo-Gr	15	9/11/12
MO6	Morocco, Sidi-Bou-Othmane	A_5	Wid	4	12
MO7	Morocco, Oued El-Abid Gorges	E_2/E_1	Wid	10	11/12/13
MO8	Morocco, Ouzaghar	A_2	Wid	12	13
MO9	Morocco, Oued Assaka	A_2	Wid	12	11/12/13
MO10	Morocco, Beddouza	B	Wid	15	9/11/12
MO11	Morocco, Safi	E_1	Wid		–
MO12	Morocco, Jebel Agouti, Agadir Melloul	A_2	Wid		–
Northeastern Africa and Arabian Peninsula:					
SA1	Saudi Arabia, Jabal Hada	A_4	Wid		–
SA2	Saudi Arabia, Al-Baha	A_4	Wid		–
SO	Yemen: Socotra, Fiheri Park	A_1	Wid		–
SU	Sudan: Arkawit, Jebel Elsit	A_3	Wid		–

1

Table 2. Genetic parameters (number of alleles, NA; number of private alleles, NPA; percentage of polymorphic loci, PPL; and expected heterozygosity, H_E) of *Scrophularia arguta* at population, phylogenetic clade and main haplogroup levels. The number of populations and studied individuals constituting each phylogenetic clade or main haplogroup are indicated in parentheses. Population abbreviations are as in Table 1

	NA	NPA	PPL (%)	H_E
Populations				
IB1	242	1	36.59	0.107
IB2	238	0	36.38	0.112
IB3	187	0	25.20	0.084
IB4	232	0	42.28	0.122
IB5	201	0	32.93	0.097
FU1	189	0	21.75	0.062
LA1	230	0	32.93	0.101
LA2	239	0	35.47	0.102
LA3	234	0	36.99	0.114
LA5	200	0	32.11	0.102
LA6	210	0	29.47	0.081
LA7	184	0	19.51	0.059
TE1	190	1	29.07	0.090
TE2	233	1	42.48	0.123
MO1	273	0	49.80	0.140
MO2	180	1	23.58	0.073
MO3	173	0	22.56	0.082
MO4	256	0	47.56	0.143
MO5	285	0	54.07	0.143
MO6	178	0	24.39	0.090
MO7	269	0	51.02	0.141
MO8	233	0	42.07	0.125
MO9	277	0	55.08	0.134
MO10	304	0	59.55	0.152
Mean ± SD	226.5 ± 37.7	0.17 ± 0.38	36.79 ± 11.58	0.107 ± 0.027
Phylogenetic clades				
Ex clade (2/34)	276	7	46.34	0.120
Wid clade (8/88)	452	5	91.87	0.177
Mo-Gr clade (5/57)	382	8	76.42	0.165
ECa clade (7/100)	348	4	67.48	0.149
WCa clade (2/26)	276	2	53.25	0.143
Haplogroups				
A (5/50)	388	1	78.86	0.168
B (1/15)	304	0	59.55	0.152
C (2/34)	276	7	46.34	0.120
D (7/100)	348	4	67.48	0.149
E (5/66)	433	1	87.20	0.179
F (2/14)	239	0	44.11	0.125
G (2/26)	276	2	53.25	0.143

1

Table 3. AFLP total (H_T) and within-population (H_S) gene diversity, and genetic differentiation (F_{ST}) for the entire dataset, phylogenetic clades and main haplogroups with more than one population

	H_S	H_T	F_{ST}
Species (all populations)	0.108	0.222	0.515
Phylogenetic clades			
Ex clade	0.110	0.119	0.075
Wid clade	0.124	0.185	0.330
Mo-Gr clade	0.108	0.170	0.368
ECa clade	0.089	0.148	0.400
WCa clade	0.110	0.145	0.244
Haplogroups			
A	0.111	0.175	0.365
C	0.110	0.119	0.075
D	0.089	0.148	0.400
E	0.128	0.193	0.336
F	0.090	0.127	0.291
G	0.110	0.145	0.244

1

Table 4. Analysis of molecular variance of *Scrophularia arguta* AFLP data considering the entire dataset (24 populations) with two and three hierarchical levels (both for five phylogenetic clades and seven main haplogroups)

Source of variation	d.f.	Sum of squares	Variance components	% of variation	*P*
Populations not grouping					
Among populations	23	10976.435	477.236	53.6	0.001
Within populations	281	8596.909	30.594	46.4	0.001
Phylogenetic clades					
Among groups	4	6063.799	1515.950	30.4	0.000
Among populations	19	4869.234	256.275	25.9	0.000
Within populations	281	8640.312	30.748	43.7	0.000
Main haplogroups					
Among groups	6	6361.008	1060.168	27.9	0.000
Among populations	17	4572.025	268.943	27.6	0.000
Within populations	281	8640.312	30.748	44.5	0.000

1

Figure **1**. Distribution map of studied populations of *Scrophularia arguta* from the western range of the species and TCS statistical parsimony network of plastid DNA haplotypes. Colours on the map and the haplotype network correspond to the main haplogroups. Phylogenetic clades indicated on the haplotype network are represented on the map by the following symbols: circle, Widespread clade (Wid); diamond, Extremaduran clade (Ex); hexagon, western Canarian clade (WCa); square, North Moroccan–Granadian clade (Mo–Gr); triangle, eastern Canarian clade (ECa). White stars indicate Iberian populations not sampled in this study. Population abbreviations are the same as in Table 1. In the haplotype network, the dashed line indicates the connection to the outgroup (*S. megalantha*), and haplotypes included in Iberian populations are encircled by a solid line. Small circles represent inferred mutational steps (white, one step; grey, more than one step, as indicated). Haplotype abbreviations are as in Table 1. Inset: map showing global distribution of the species.

Figure **2.** Discrete phylogeographic analysis (DPA) of *Scrophularia arguta* based on ptDNA sequences. The maximum clade credibility tree is shown, with branches coloured according to the most probable ancestral range. Branch thickness indicates posterior probability (PP), with PPs>0.90 also shown as numbers above branches. Pie charts represent PP distributions of ancestral range at well supported (PP≥0.95) nodes.

Figure **3.** BAPS and STRUCTURE clustering analyses of a *Scrophularia arguta* AFLP dataset. (A) BAPS result for K = 13, (B) BAPS admixture result for K = 13, (C) BAPS result from the spatial clustering analysis, and (D) STRUCTURE result for K = 9. Population abbreviations are as in Table 1.

Figure **4.** Principal coordinates analysis of the AFLP dataset of *Scrophularia arguta* based on genetic distances. Percentages of total variance explained by the first three coordinates are shown on the respective axes. Sample abbreviations are as in Table 1.

Ellipses indicate supported clades in the phylogenetic analysis (clade abbreviations as in Fig. 1). Inset in B shows the eigen values for the first 10 axes.

Figure 5. Species distribution modelling (SDM) of *Scrophularia arguta*, including the average model fitted to present climatic conditions and average projections to the last interglacial (c. 120–140 ka), last glacial maximum (c. 22 ka) and mid-Holocene (c. 6 ka). The logistic output of the Maxent model is shown. Localities used to build the model are shown as red dots in the present map.