Phylogeography of a gypsum endemic plant across its entire distribution range in the western Mediterranean

Mario Blanco-Sánchez[1*], Michael J. Moore[2], Marina Ramos-Muñoz[1], Beatriz Pías[3], Alfredo García-Fernández[1], María Prieto[1], Lidia Plaza[1], Ignacio Isabel[1], Adrián Escudero[1] and Silvia Matesanz[1]

[1] Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos. C/ Tulipán s/n, 28933, Móstoles, Spain.

[2] Department of Biology, Oberlin College, Oberlin, Ohio 44074, U.S.A.

[3] Departamento de Biodiversidad, Ecología y Evolución. Universidad Complutense de Madrid. C/José Antonio Nováis 2, 28040, Madrid, Spain.

*Author for correspondence: mario.blanco@urjc.es

Manuscript received _____; revision accepted _____.

Running head: Phylogeography of a gypsum endemic plant

ABSTRACT

Premise of the study: Gypsum soils in the Mediterranean Basin house large numbers of edaphic specialists that are adapted to stressful environments. The evolutionary history and standing genetic variation of these taxa have been influenced by the geological and paleoclimatic complexity of this area and the long-standing effect of human activities. However, little is known about the origin of Mediterranean gypsophiles and the factors affecting their genetic diversity and population structure.

Methods: Using phylogenetic and phylogeographic approaches based on microsatellites and sequence data from nuclear and chloroplast regions, we evaluated the divergence time, genetic diversity and population structure of 27 different populations of the widespread Iberian gypsophile *Lepidium subulatum* throughout its entire geographic range.

Results: *Lepidium subulatum* diverged from its nearest relatives ~3 Mya, and the ITS and *psbA/matK* trees supported the monophyly of the species. These results suggest that both geological and climatic changes that occurred in the region around the Plio-Pleistocene promoted its origin, compared to other evolutionary processes. We found high genetic diversity in both nuclear and chloroplast markers, but a greater population structure in the chloroplast data. This suggests that while seed dispersal is limited, pollen flow may be favored by the presence of numerous habitat patches that enhance the movement of pollinators.

Conclusions: Despite being an edaphic endemic, *L. subulatum* possesses high genetic diversity probably related to its relatively old age and high population sizes across its range, and highlights the value of using different markers to fully understand the phylogeographic history of plant species.

Keywords: Phylogeography, gypsophiles, genetic diversity, population structure, nuclear microsatellites, cpDNA, pollen flow, seed dispersal, *Lepidium subulatum*

INTRODUCTION

The genetic diversity of plant populations and how it is distributed geographically across species ranges depends on processes that operate at different spatial and temporal scales. The Mediterranean Basin has experienced a highly complex geological and paleoclimatic history. Past changes in its geological, climatic and ecological conditions, especially during the Pliocene and Pleistocene (5.33-0.01 Mya), have been decisive in shaping the genetic composition of Mediterranean plants (Blondel et al., 2010; Nieto Feliner, 2014). More recently, humans have profoundly transformed Mediterranean ecosystems through long-standing, yet dynamic activities (Blondel et al., 2010; Nieto Feliner, 2014), further contributing to modulate the genetic diversity and structure of plant populations in this region (Thompson, 2005).

A particular example within Mediterranean taxa are gypsophiles, defined as plants that are restricted to gypsum (calcium sulfate dihydrate) soils (Meyer, 1986). In the Mediterranean Basin, these soils harbor rich plant communities with large proportions of endemic species adapted to arid and semiarid conditions (Escudero et al., 2015). Iberian gypsum outcrops have been dated to as old as the Cambrian, but more than two thirds of the gypsum soils in this area appeared in the Cenozoic, mostly during the Neogene (Escavy et al., 2012). Different events that occurred in this period favored the formation of gypsum outcrops. First, geological events like the Alpine Orogeny allowed the accumulation of salts and sediments in basins (Escavy et al., 2012). Furthermore, the tectonic uplift of the Gibraltar arc reduced water flow from the Atlantic Ocean to the Mediterranean Sea, resulting in the Messinian Salinity Crisis (~6 - 5.3 Mya). This, together with global changes in sea level, produced the desiccation of the Mediterranean Sea by evaporation processes that favored gypsum precipitation (Garcia-Castellanos and Villaseñor, 2011). Second, changes in the paleoclimatic conditions of the Mediterranean region further accelerated evaporation by rainfall reduction and prevented the loss of precipitated gypsum by leaching (Parsons, 1976). The progressive aridification and

seasonality of precipitation that started 9.5-8 Mya (van Dam, 2006) led to the appearance of the Mediterranean climate 3.2 Mya, characterized by high seasonality and marked summer drought (Suc, 1984). Although both the availability of gypsum soils and the increasingly drier climatic conditions of the late Miocene and Pliocene likely determined the origin of Iberian gypsophiles, it is yet not established whether these species originated in gypsum environments or, alternatively, in other stressful habitats from which they colonized gypsum soils (Escudero et al., 2015).

A remarkable feature of gypsum environments is their discontinuous spatial configuration. Not only are gypsum soils naturally fragmented into island-like outcrops surrounded by other substrates (Escudero et al., 2015), but also anthropogenic practices like agriculture and livestock grazing have exacerbated the natural patchiness of these habitats in the Mediterranean region for centuries (Pueyo et al., 2008). Both natural and human-induced fragmentation may affect the genetic diversity and structure of gypsophile populations due to neutral processes such as genetic drift, demographic changes, inbreeding and reduced gene flow (Aguilar et al., 2008). This unique spatial configuration may be even more critical in species that lack effective seed dispersal mechanisms, as is the case in most widely-distributed gypsophiles (Escudero et al., 2015). However, livestock practices like transhumance and grazing could enhance gene flow between populations if they promote seed movement (Pueyo et al., 2008; Azcárate et al., 2013). Consequently, genetic diversity and population structure of gypsophiles may be determined by a complex interaction between landscape configuration and land use, among others.

Phylogeography provides a useful framework to assess the origin and evolutionary history of species and closely related species groups (Avise, 2000). Combining markers with different mutation rates enables phylogeographic studies to elucidate how past and present processes have modulated the genetic diversity and structure of populations (Wang, 2011).

Furthermore, the use of markers with different modes of transmission such as chloroplast DNA (maternally inherited and dispersed only by seeds in most angiosperms) and nuclear DNA (biparentally inherited and dispersed by both seeds and pollen) allows for the quantification of the relative contribution of seed and pollen flow to the genetic structure of populations (Ennos et al., 1999; Petit et al., 2005).

In this study, we used a phylogeographic approach based on nuclear microsatellite loci and chloroplast and nuclear sequence data to assess the origin, genetic diversity and population structure of the gypsophile *Lepidium subulatum* L. (Brassicaceae) throughout its entire distribution range. *Lepidium subulatum* is a regionally dominant gypsophile endemic to the Iberian Peninsula and North Africa and is the most geographically widespread gypsophile in the western Mediterranean (Romão and Escudero, 2005). Because of its high substrate specificity, dominance and life-history traits common to other gypsophiles, *L. subulatum* provides a compelling study system to evaluate the genetic diversity, structure and date of origin of gypsophiles. We studied 27 different populations that represent the current geographical and climatic distribution of the species, to address the following questions: 1) When did the evolutionary divergence of *L. subulatum* occur and how was it influenced by the complex geological and paleoclimatic history of the Mediterranean Basin? 2) Do populations of the species show different levels of genetic diversity? 3) Are populations genetically structured, and if so, is this structure explained by their geographical location and/or by historic demographic changes? and 4) Is genetic variation and population structure inferred by either microsatellites or chloroplast markers different, and if so, how does it relate to pollen and seed flow? This is the first study to estimate the date of origin of this species and the distribution of genetic diversity across its entire geographical and climatic range. We expect that the complex historical events experienced by Mediterranean plants had a major role in the origin and evolution of *L. subulatum*. We also hypothesize that *L. subulatum* populations show substantial

genetic structure due to the spatial configuration of gypsum soils and the reproductive attributes of the species.

MATERIALS AND METHODS

Habitat and species description—

Gypsum plant communities in the Iberian Peninsula are mostly composed by chamaephytes and ephemeral annual plants, with a large proportion of endemic species. In these systems, plants form discrete patches immersed in a matrix of bare ground and biological soil crusts (BSC) formed by cyanobacteria, lichens and mosses (Escudero et al., 2015).

The genus *Lepidium* L. is one of the largest in the Brassicaceae, with approximately 175 widespread plant species. Most of them are edaphic generalists, but two species, *Lepidium subulatum* and *Lepidium cardamines*, are restricted to the gypsum soils of the Iberian Peninsula. *Lepidium subulatum* L. (Brassicaceae) is one of the most common and widely distributed gypsophiles in Iberian gypsum habitats (Romão and Escudero, 2005). It is a non-clonal perennial shrub (20–60-cm high) endemic to the Iberian Peninsula and North Africa. This species is mainly outcrossing with partial self-compatibility, as supported by both field experiments (Gómez et al., 1996) and low inbreeding coefficients inferred from molecular markers (Gómez-Fernández et al., 2016; Matesanz et al., 2018). It has entomophilous pollination, being pollinated by a very rich community of generalist species from seven different orders of insects (Santamaría et al., 2018). Seeds are released from very numerous small fruits (silicles), lack obvious long-distance dispersal mechanisms and have a mucilage that enhances seed adhesion to the soil (Romão and Escudero, 2005).

Population sampling—

We sampled 27 populations in the Iberian Peninsula and North Africa, spanning the worldwide geographic and climatic distribution of *L. subulatum* (Table 1 and Fig. 1). Each population was assigned to one of five different geographic zones that roughly match different river basins in the Iberian Peninsula (Fig. 1) and are related to the main gypsum vegetation habitats described in the region (Mota et al., 2011). Elevations of sampled populations varied from 219 (ALF) to 1157 m asl (TOP). The closest sampled populations (SMV and CHI) were 15 km apart and the furthest populations (ARGL and BAL) were separated by 972 km. At each population, fresh leaves of 20 individuals were collected and stored in paper bags, except for the Moroccan population (MAR, 10 individuals), and the Peralta population (PER, 14 individuals). Leaves were air-dried and stored until DNA extraction. Voucher specimens (one per sampled population) were deposited at the herbarium of the Universidad Rey Juan Carlos (Móstoles, Spain; Appendix S1, see Supplemental Data with this article). Additionally, 14 samples from the same locality in Algeria (Chott Ech Chergui region) dating from 1884 to 1952 were obtained from herbarium specimens (Muséum National d'Histoire Naturelle, Paris, France; Appendix S1). A total of 508 samples were included in the study. At each site, sampled individuals were collected in a $\approx 20 \times 20$ m area at S – SE aspect to homogenize microenvironmental conditions experienced by individuals. The south-oriented slopes of gypsum hills in the study region receive more insolation and have lower water availability compared to north-oriented slopes. Furthermore, gypsophiles are dominant and more abundant on slopes with S-SE aspects. All populations had moderate to large size, from several hundred to several thousand individuals. Individuals within populations were separated at least one meter from each other, to avoid sampling closely-related individuals.

Climatic information of each population was extracted from CHELSA Bioclim layers (Karger et al., 2017) using ArcMap 10.2.2 (ArcGIS Desktop, ESRI, Redlands, California, USA). A 2 km buffer around each population was created to account for the within-site climatic

heterogeneity. Sampled populations spanned a wide climatic range: mean annual temperature ranged from 11.4 to 16.9 ºC and mean annual precipitation ranged from 254.7 to 647.8 mm (Table 1).

DNA extraction, microsatellite markers, cpDNA markers and PCR conditions—

Genomic DNA was extracted from 30 mg of air-dried leaf tissue, using a commercial kit (DNeasy Plant Minikit; QIAGEN, California, USA) with minor changes to the manufacturer's extraction protocol to improve the process. DNA extraction success was checked using 1% agarose gels stained with GreenSafe Premium (NZYTech, Lisbon, Portugal). Ten species-specific, nuclear polymorphic microsatellite markers previously described in Martínez-Nieto, Merlo, Mota, Salmerón-Sánchez, & Segarra-Moragues (2012) were used to assess neutral genetic diversity. Detailed information concerning microsatellite markers used and PCR reactions is found in Appendix S2 and Appendix S3, respectively. Amplified DNA was analyzed using an ABI 3730 (Applied Biosystems, Madrid, Spain) at "Unidad de Genómica y Proteómica" of Universidad Complutense (Madrid, Spain), employing the GS500 size standard.

For phylogeographic analyses, we performed a preliminary screening with ten nuclear, chloroplast, and mitochondrial loci widely used in phylogeographic studies (Appendix S4). From this screening we selected the chloroplast *matK* gene and the *psbA-trnH* intergenic spacer region because they showed relatively high variability at the population level. We sequenced *matK* and *psbA-trnH* (henceforth referred to simply as *psbA*) from 8-10 individuals of each study population. Additionally, we also sequenced the same regions in four individuals from two different populations (Orusco de Tajuña and Portalrubio de Guadamejud, in the center of the Iberian Peninsula) of *Lepidium cardamines* L., an Iberian gypsophile species that is a close relative of *L. subulatum* (Mummenhoff et al., 2009), to evaluate whether the two species share haplotypes indicative of processes like hybridization, chloroplast capture, and/or incomplete

lineage sorting that might confound interpretation of phylogeographic data (Schaal et al., 1998). Finally, to test the monophyly and to date the origin of *L. subulatum*, we sequenced the nuclear internal transcribed spacer (ITS) region and the chloroplast *trnT-trnL*, *trnL* intron and *trnL-trnF* regions (Appendix S4), respectively, of one individual from four different *L. subulatum* populations (BAL, ECZ, TDL, SMV) and one individual from one population (Orusco de Tajuña) of *L. cardamines*. Detailed information concerning PCR conditions is found in Appendix S3. Amplified DNA was sequenced at Macrogen DNA Sequencing Service (Madrid, Spain).

Microsatellite genotyping and alignment of chloroplast sequences—
Microsatellite scoring was performed using GeneMarker v2.2.0 (SoftGenetics, State College, Pennsylvania, USA). Each sample was manually checked by three different researchers to guarantee a robust scoring process. Different sizes of the amplified DNA fragments were considered as different alleles. Every microsatellite locus exhibited polymorphic patterns, yielding one (homozygous) or two alleles (heterozygous) per individual at each locus, consistent with the ploidy level of the species. We repeated five percent of the samples to ensure the repeatability of the scoring process. For population ARGL, only four individuals were successfully genotyped. Therefore, this population was excluded from analyses of microsatellite genetic diversity and population structure. We only considered in our analyses the individuals for which at least 9 of 10 loci were successfully genotyped, representing 98.6% of all individuals.

DNA sequences were manually trimmed, edited and cleaned using SEQUENCHER 5.4.6 (Gene Codes Corporation, Ann Arbor, Michigan, USA). A total of 207 individuals were successfully sequenced for *matK* and 219 for *psbA*. We were able to concatenate *matK* and

psbA regions from 204 individuals, which were used in all downstream analyses. Sequence alignment was performed in AliView (Larsson, 2014), with manual adjustments.

Statistical analyses—

Phylogenetic analyses—

To assess the evolutionary relationships between *L. subulatum* and other species of *Lepidium* and to test the monophyly of *L. subulatum* (see other phylogenies in Beilstein, Nagalingum, Clements, Manchester, & Mathews, 2010; Mummenhoff et al., 2009), we estimated phylogenies of *Lepidium* using newly generated sequences of *L. subulatum* and *L. cardamines* as well as publicly available sequences of other species of *Lepidium*, for the nuclear ITS region and the chloroplast *matK* gene and *psbA* spacer region. For ITS, we included one individual each from 4 populations of *L. subulatum* (BAL, ECZ, TDL, SMV) that covered the entire geographic range of the species. We also included one individual from one population (Orusco de Tajuña) of *L. cardamines*. We added these to all *Lepidium* ITS sequences available on GenBank (Clark et al., 2016), which yielded 90 species from the genus in total (including *L. subulatum* and *L. cardamines*). As outgroups, we downloaded GenBank ITS sequences for seven *Arabidopsis* species and three *Cardaria* species. The total data set included 408 accessions, with 1-59 individuals per species (Appendix 1). Sequences were aligned as described above, excluding the ambiguous regions for downstream analyses. Maximum Likelihood (ML) analyses were conducted in RAxML (Stamatakis, 2014) using the CIPRES Science Gateway v. 3.3 (Miller et al., 2010), selecting 1000 replicates, the GTRCAT model, and rapid bootstrapping. We undertook similar ML analyses on the *matK* and *psbA* data generated for phylogeographic analyses (described above). However, sequences of other taxa of *Lepidium* beyond *L. subulatum* and *L. cardamines* were not available for inclusion.

To understand the temporal divergence of *L. subulatum*, we performed a molecular dating analysis using the *trnT-trnL*, *trnL* intron and *trnL-trnF* regions for 68 different species of *Lepidium* and 12 outgroup species (*Brassica napus*, *Cochlearia pyrenaica*, and ten species of *Arabidopsis*; see Appendix 1). Sequence alignment was performed in Aliview, excluding the ambiguous regions for further analysis. The dating analysis was performed in BEAST v1.10.4 (Suchard et al., 2018) using the CIPRES Science Gateway v. 3.3 (Miller et al., 2010). We selected three different unlinked partitions with the HKY substitution model (Hasegawa et al., 1985) for each partition. We used an uncorrelated lognormal relaxed clock model, which allows uncorrelated rates of molecular evolution across the tree, and a birth-death process as tree prior (Gernhard, 2008). We calibrated the tree at the basal node (the split point *Lepidium-Arabidopsis*), using the date obtained by Guo et al., (2017) for the crown clade "A": 16.9-20.3-24 Mya, constraining the calibration point with a normal distribution with mean = 20.3 and standard deviation = 2.0. Then, we ran a relaxed log-normal clock with default priors to estimate prior distributions to be used in a second analysis that was used to estimate priors for the final analysis. BEAST analyses were run for 40 million generations, logging parameters and trees every 1000 generations. Convergence, mixing, and effective sample sizes (ESS) of parameters were checked using Tracer v1.5.0 (Rambaut and Drummond, 2009). A burn-in of 1000 trees was removed from each analysis. The remaining trees were used to generate a maximum clade credibility tree with TreeAnnotator v1.8.2 (Rambaut and Drummond, 2014).

Phylogeographic analyses—

To evaluate phylogeographic patterns within *L. subulatum*, a haplotype network using the concatenated *psbA* and *matK* sequences of each individual was estimated using PopART (Leigh and Bryant, 2015) and employing the TCS method, which is appropriate to estimate genealogies among populations (Clement et al., 2002). We also performed a ML phylogeny estimated from

RAxML (Stamatakis, 2014), using 10000 replicates, the GTRCAT model, and rapid bootstrapping (Appendix S5).

To test the existence of historical demographic changes, Tajima's *D* (Tajima, 1989), Fu and Li's *F** (Fu and Li, 1993) and Fu's F_S (Fu, 1997) statistics were calculated for each population using DnaSP6 (Rozas et al., 2017). These tests were originally designed to assess the neutrality of markers, but their combination is also useful to test departures from population equilibrium due to historical demographic changes, bottlenecks or genetic hitchhiking (Fu, 1997). Thus, these tests allow for distinguishing the relative role of demographic changes or other processes (like gene flow or mutation) in shaping the allele frequencies of populations. While significant and positive Tajima's *D* values can inform us about the admixture of two different populations, significant and negative Tajima's *D* values indicate a recent bottleneck in a population (Tajima, 1989; Aris-Brosou and Excoffier, 1996). Fu's F_S is also used to test for demographic expansion and it is described as more sensitive to the growth of the populations than Tajima's *D* (Chávez-Pesqueira and Núñez-Farfán, 2016). These tests may be performed only if the populations possess more than one haplotype.

Intrapopulation genetic diversity—

We checked the presence of null alleles and genotyping errors such as allele dropouts or false positive alleles due to stuttering in the nuclear microsatellites dataset, using Micro-Checker 2.2.3 (Van Oosterhout et al., 2004). No genotyping errors or null alleles were detected. For each population, we calculated the following genetic diversity indices: *P*, proportion of polymorphic loci; *A*, allele richness (mean number of alleles per locus); A_{rare}, mean number of rarefied alleles per locus; A_e, mean number of effective alleles; H_o, observed heterozygosity ($H_o = 1 - \sum_k \sum_i Pkii/np$, Nei, 1987); H_e, expected heterozygosity ($\mathrm{H_e} = \tilde{\mathrm{n}}/(\tilde{\mathrm{n}} - 1)[1 - \sum_{\mathrm{i}} \bar{p}_i^2 - \mathrm{H}_o/2\tilde{\mathrm{n}}]$, Nei, 1987); F_{IS}, inbreeding coefficient; β, neutral genetic differentiation between

populations (from 0 to 1; Weir & Hill, 2002); the number of private alleles and the number of multilocus genotypes. A, A_{rare}, H_o, H_e, F_{IS} (and their confidence intervals) and β were calculated using the functions *nb.alleles*, *allelic.richness*, *basic.stats*, *boot.ppfis* and *betas*, respectively, from the package hierfstat (Goudet and Jombart, 2015) as implemented in R (R Core Team, 2018). Rarefaction allowed for calculating the mean number of alleles per locus (A_{rare}) considering equal sample sizes in all populations (note that MAR only had 10 individuals sampled). A_e was calculated using the function *genetic_diversity* (package gstudio; Dyer, 2016), the number of private alleles was calculated using the function *private_alleles* and the number of multilocus genotypes was calculated using function *poppr*, both in package poppr (Kamvar et al., 2014).

Genetic diversity of chloroplast markers was assessed using DnaSP6 (Rozas et al., 2017). For each population, we calculated the number of segregating sites, the number of haplotypes, haplotype diversity (Hd) and nucleotide diversity (π).

Population structure—

To assess population differentiation we calculated a pairwise F_{ST} matrix based on microsatellite markers using the *genet.dist* function (package hierfstat, Goudet & Jombart, 2015). The matrix of pairwise Nei's (D) differences between populations from chloroplast markers was calculated using the *pairnei* function (package haplotypes; Aktas, 2015). We also calculated a Euclidean geographical distance matrix between populations, performed with ecodist package (Goslee and Urban, 2007), using the UTM coordinates of each population.

To assess the distribution of genetic variation of microsatellite and chloroplast markers across regions and populations, we used Analysis of Molecular Variance (AMOVA, Excoffier, Smouse, & Quattro, 1992). AMOVAs were performed using the *poppr.amova* function (package poppr, Kamvar et al., 2014), with 99999 permutations and excluding within-

313 individual variation. We performed two different AMOVAs: 1) Non-hierarchical AMOVA,
314 considering all populations within the same region; 2) Hierarchical AMOVA, assigning each
315 population to each of five geographical zones (Fig. 1).

316 To assess whether closer populations are more genetically similar, we tested for
317 isolation by distance (IBD). We performed two different Mantel correlograms (Legendre and
318 Legendre, 2012), using the pairwise genetic distance matrix calculated from microsatellite
319 markers and from chloroplast markers and the pairwise geographical distance matrix between
320 populations. While Mantel tests show the overall relationship between the genetic and the
321 geographic matrix, a Mantel correlogram compares the pairwise genetic distance matrix (F_{ST} in
322 our case) and the pairwise geographical distance matrix (Euclidean distance), which allow for
323 finding significant correlations between them at different distance classes. Each distance class
324 includes all pairs of points that are included within a specific distance. A correlation index
325 (Mantel statistic, rM) between genetic and geographical distance matrices is calculated for each
326 distance class. The size and number of distance classes was set using Sturge's rule (Legendre
327 and Legendre, 2012). Significance was tested using 99999 permutations. Mantel correlograms
328 were generated using the *mantel.correlog* function (package vegan, Oksanen et al., 2019).

329 Population genetic structure from microsatellite markers was further evaluated using the
330 Bayesian clustering algorithm in STRUCTURE v. 2.3. (Pritchard et al., 2000). This method
331 evaluates the membership of each individual to a specific genetic cluster (*K*). We performed 10
332 independent runs for each *K* (from $K = 1$ to $K = 30$), with a burn-in period of 10^5 iterations and
333 10^6 MCMC iterations after the burn-in period, using the admixture model, where individuals
334 from different *K* values could have a common ancestry (Falush et al., 2003), as recommended
335 for microsatellites. We ran STRUCTURE assuming correlated and independent allele
336 frequencies (Pritchard et al., 2000; Falush et al., 2003) and both methods provided very similar
337 clustering results. STRUCTURE results were extracted using Structure Harvester (Earl and

vonHoldt, 2012), which were then used to generate CLUMPP input files. Then, using CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007), results from 10 runs of each *K* were combined, using the Greedy algorithm. Membership of each individual to a specific genetic cluster was visualized using DISTRUCT 1.1 (Rosenberg, 2004). To ensure the assignment performed by STRUCTURE, we repeated the clustering assignment with rMavericK (Verity and Nichols, 2016), obtaining virtually the same assignment results.

Some recent work has drawn attention to the problems related to determining the appropriate number of genetic clusters (*K*) (Meirmans, 2015; Janes et al., 2017). To determine this, we first considered the average log probability (L_K) of the data for each *K*, and determined the value of *K* for which this probability is maximized (Pritchard et al., 2000). We also calculated the optimum value of *K* using the Evanno method (Evanno et al., 2005) implemented in Structure Harvester (Earl and vonHoldt, 2012). This *ad hoc* method is based on changes in the mean values of log probability of data at successive *K* values. The Evanno method and L_K simplify model assumptions because these methods obtain the value of *K* for each assignment model, so estimating the optimum value of *K* requires comparison between models, which is not straightforward (Verity and Nichols, 2016). Thus, we also calculated *K* using rMavericK. This software uses generalized thermodynamic integration (GTI), which has been hypothesized to be more accurate and precise (Verity and Nichols, 2016). Therefore, *K* was calculated using rMavericK, although L_K and Evanno methods provided similar results (Appendix S6).

RESULTS

Analyses of sequence data—

Information on lengths and sequence variation for all sequence alignments is provided in Appendix S7. In both the ITS (Appendix S8) and chloroplast (*trnT-trnL*, *trnL* intron and *trnL-trnF*; Fig. 2) trees, *Lepidium subulatum* was sister to the Iberian gypsophile *L. cardamines* (ITS

bootstrap support = 97%; chloroplast Bayesian posterior probability = 95%). In the ITS tree, all sequences of *L. subulatum* formed a clade with high support (bootstrap support = 98%). In the *matK/psbA* tree (Appendix S5), haplotypes of *L. subulatum* and *L. cardamines* were relatively distant from each other and none were shared between the two species. The molecular dating analysis based on the chloroplast loci dated the evolutionary divergence of *L. subulatum* from 5.08 – 1.33 Mya (mean = 3.01 Mya; Fig. 2). Furthermore, the divergence of the gypsophile clade of *L. subulatum* and *L. cardamines* was dated to 5.96 – 2.05 Mya (mean = 3.86 Mya).

Haplotype analyses recovered 22 different haplotypes and 19 segregating sites (*S*). Total nucleotide (π) and total haplotype diversity (Hd) across populations was 0.0038 and 0.747, respectively. Twelve populations possessed more than one haplotype, while 15 populations possessed one fixed haplotype for all sampled individuals (Table 2).

The haplotype network showed that *L. subulatum* was connected to its closest relative *L. cardamines* by three mutation steps, with no shared haplotypes between the two species (Fig. 3). The network was complex, with one loop and three extinct or unsampled haplotypes. Despite the complexity of the network, we identified four common haplotypes. The most frequent haplotype (haplotype A, in blue), was found in 16 populations (in nine of them it was the only haplotype present) and was distributed broadly across the Iberian Peninsula (in the north-west, the center and the south). The second most common haplotype (haplotype B, in red) was separated from haplotype A by one mutational step and was found mainly in 4 populations of the eastern Iberian Peninsula: CAB, PDG, VAL and YEB, and one individual from VY population. The third most common haplotype was haplotype C (in purple), which was restricted to the Ebro River Valley. This haplotype was the most divergent, being separated by many mutational steps from all other main haplotypes. The fourth most common haplotype was haplotype D (in yellow), which was restricted to North Africa (ARGL and MAR) and was the

only haplotype found in this region. Overall, the center of the Iberian Peninsula showed the highest haplotype diversity.

We obtained very similar results from the phylogenetic tree based on *psbA* and *matK*. Although some of the groups were identical between the tree and the network, the low support values of some branches in the tree showed the uncertainty of relationships among some haplotypes (*e.g.*, see green haplotypes in Fig. 3 and Appendix S5).

The Mantel Correlogram based on the chloroplast markers showed that the closest populations (first distance class, 62.15 km) were significantly similar (Fig. 4a; $R_M = 0.181$, p-value = 0.010), and populations separated by ~350 km were statistically different (Fig. 4a; $R_M = -0.176$, p-value = 0.041), confirming the presence of isolation by distance.

The non-hierarchical AMOVA of chloroplast loci performed with all individuals and populations showed a variation of 46.08% among populations and 53.92% within them (p-value < 0.001; Table 3). In the hierarchical AMOVA with populations grouped by their geographic location, 32.82% of the variation was explained by the geographic region (p-value < 0.001; Table 3).

In the populations with more than one haplotype, overall Tajima's *D*, Fu and Li's *F** and Fu's F_S were not statistically different from 0 (p-value > 0.05), except for SMV, which showed a significantly positive Tajima's *D* value (i.e., a higher average pairwise differences observed than expected; Table 2). Therefore, our results suggest the admixture of two distant populations in SMV population; and we did not detect demographic changes in the other populations.

Microsatellite analyses—

We found 145 different alleles among the 504 individuals, for an average of 14.5 alleles per locus. The number of alleles per locus ranged from six (Locus 10 and 11) to 24 (Locus 4). Microsatellite genetic diversity was high for all populations (Table 4). Most populations

possessed 100% polymorphic loci, except for AGR, SMV, SPP and YEB, with 90% polymorphic loci. Expected heterozygosity ranged from 0.452 (SEG) to 0.681 (BAZ; Table 4). Observed heterozygosity varied from 0.415 (SPP) to 0.718 (PER; Table 4). The fixation index F_{IS} was low for all populations, ranging from -0.193 (PER, showing a heterozygote excess) to 0.237 (BAZ) and none of the populations had a F_{IS} statistically different from 0. Population-specific F_{ST} (β) varied from 0.064 (BAZ) to 0.379 (SEG; Table 4). The average number of alleles per locus (*A*) ranged from 3.3 (SEG) to 7.1 (BAZ), with an overall average of 5.15 alleles per locus. The rarefied mean number of alleles per locus (A_{rare}) ranged from 2.62 (CAB population) to 5.65 (BAZ population), with an overall average of 4.40 alleles per locus. The mean number of effective alleles per locus (*Ae*) varied from 1.93 (SEG) to 3.89 (BEL), with an overall average of 3.06 effective alleles per locus. The number of multilocus genotypes matched the number of individuals sampled in each population, except for ARA and SPP, where there were two individuals with the same genotype. We found a total of 23 private alleles in 15 of the 26 populations, ranging from one to three per population.

Microsatellite population structure—

Pairwise F_{ST} values were generally low, ranging from very low (0.030) between populations CHI and SMV to high (0.440) between populations SEG and SPP (Appendix S9). Results from rMavericK clearly supported the presence of three different genetic clusters (K=3). L_K and the Evanno method supported K=2 but also K=3 (Appendix S6). Thus, we selected K=3 that allowed a clearer interpretation of the data. Based on the K=3 solution, most of the populations included admixed individuals assigned to more than one genetic cluster. Four populations from the Tajo river basin (ARA, BEL, SPP and YEB) and one from the Guadalquivir-Júcar-Segura basins (CAB) contained individuals that were mostly assigned to one genetic cluster (blue, Fig. 5). Individuals from AGR, BAL, ECZ, PDG, TDL, TOR and VAL were mostly assigned to a

different genetic cluster (yellow, Fig. 5). Individuals from APG, BAZ, PER, TOP and VY were mostly assigned to the magenta genetic cluster (Fig. 5). Finally, individuals from ALF, AZQ, CHI, GEL, MAR, PEÑ, SEG, SMV and TER belonged to two or even three different genetic clusters (frequently a mixture of the magenta and yellow genetic clusters; Fig. 5).

The non-hierarchical AMOVA with all individuals and populations showed that 81.37% of variation was found within populations and 18.63% between populations (p-value < 0.001; Table 3). There was a small but significant population structure explained by the geographical location of the populations. In the geographic AMOVA, 2.95% of the variation was explained by the geographic region (p-value < 0.001; Table 3).

The Mantel correlogram based on microsatellites did not show evidence of isolation by distance (IBD). Only the closest populations (first distance class, 58.55 km) were significantly similar (Fig. 4b; $R_M = 0.151$, p-value = 0.023).

DISCUSSION

Our molecular dating results suggest that the Iberian gypsophilic clade composed by *Lepidium subulatum* and *L. cardamines* originated ~ 3.86 Mya (5.96-2.05 Mya) and the stem lineage of *L. subulatum* diverged ~3.01 Mya (5.08-1.33 Mya). Thus, it is likely that the specialization to gypsum soils (gypsophily) in this group appeared at some point from the latest Miocene to the early Pleistocene, in the ancestor of both species. Furthermore, these dates for the divergence of the study species in the Plio-Pleistocene also suggest that the paleoclimatic and geological events that occurred in the Mediterranean Basin around this period could be associated with the origin and further expansion of this gypsophile. First, the massive emergence at the surface of gypsum soils during the Neogene consequence of evaporitic processes in the region (Escavy et al., 2012) increased the probability of colonizing a novel edaphic habitat by chance (chance dispersal *sensu* Rajakaruna, 2017; Escudero et al., 2015; Moore & Jansen, 2007), likely

facilitating the evolution of gypsum-restricted taxa. Second, the progressive aridification of the Mediterranean basin before and during the Messinian salinity crisis (~6-5 Mya) not only favored the creation of gypsum soils, but also probably acted as an evolutionary force promoting the evolution of *L. subulatum* and other gypsophiles in the new climatic conditions (Thompson, 2005). It has been hypothesized that certain gypsophiles may have been preadapted to the global aridification that started in the mid-Miocene that subsequently colonized gypsum soils (Escudero et al., 2015 and references therein). However, the availability of gypsum soils in the Iberian Peninsula prior to our estimated date of origin (Escavy et al., 2012) suggests that this is not the case for *L. subulatum*. The relatively old date of origin of the species would provide enough time to colonize isolated gypsum patches and is congruent with its widespread distribution in the Iberian Peninsula, even moreso when its inefficient dispersal ability is considered (Escudero, Iriondo, Olano, Rubio, & Somolinos, 2000; see Moore & Jansen 2007 for similar patterns). Our results also agree with the estimated age of other Iberian and non-Iberian gypsophiles. The clade that includes *Helianthemum squamatum* (L.) Dum. Cours. started its diversification 4.37 Mya (8.57 – 1.65 Mya; Aparicio et al., 2017) and the clade formed by *Ferula loscosii* (Lange) Willk. and its sister species diverged from their common ancestor 4 Mya (6.4 – 1.6 Mya; Pérez-Collazos et al., 2009). Other North American gypsophiles such as *Tiquilia hispidissima* (Torr. & A. Gray) A.T. Richardson split from its nearest relatives in the early/mid-Pliocene (5 – 3.5 Mya; Moore and Jansen, 2007).

The evolutionary distinctiveness of *Lepidium subulatum* and *L. cardamines* in both the haplotype network and the ITS phylogeny is important because it reinforces the idea that past edaphic and climatic changes could be important in the origin of *L. subulatum*, compared to other evolutionary processes. The two species did not share haplotypes, which is consistent with a lack of hybridization between both species that could have resulted in chloroplast capture (Schaal et al., 1998). Nevertheless, it is important to note that existing population sampling of

L. cardamines is limited and additional sampling may reveal shared haplotypes between the two species. Some authors have noted the importance of hybridization in the origin of edaphic specialists (Rajakaruna, 2017 and references therein), but Ellstrand, Whitkus, & Rieseberg, (1996) reported that Brassicaceae taxa are not particularly prone to natural hybridization. Several aspects, including differences in their reproductive phenology (Hernández Bermejo and Clemente, 1993) may have served to minimize potential hybridization between them.

Based on the haplotype analysis, *L. subulatum* may have originated in the center of the Iberian Peninsula. This region shows the highest haplotype diversity (see also individual *psbA* and *matK* haplotype networks in Appendix S10), suggesting that populations in this region have had enough time to reach such high diversity. Furthermore, gypsum outcrops of the Tajo Valley present the greatest climatic variation of the entire distribution range, which could also explain the high genetic diversity found in this region. The populations of the Ebro Valley (purple shades in Fig. 3) possessed the most distantly related haplotypes, which indicates that gene flow via seeds between the Ebro Valley and the rest of the Iberian Peninsula has likely been limited during the evolutionary history of *L. subulatum* (see Fig. 1 and Appendix S10).

Interestingly, North Africa populations (MAR and ARGL) are fixed for a single haplotype that is closely related to the most common one. It is thus likely that the colonization of North Africa occurred via a recent, long-distance dispersal event from the Iberian Peninsula. Several pieces of evidence support this claim. First, our analysis estimated the mean date of origin of the species after the Messinian Salinity Crisis, when the Iberian Peninsula and North Africa were disconnected again by the Mediterranean Sea. Second, if *L. subulatum* had been isolated in North Africa for at least 6-5 My (during the Messinian Salinity Crisis, when the Mediterranean Sea was desiccated) we would expect a greater haplotype diversity in North Africa or, alternatively, only one, much more divergent haplotype. Our results match those of other studies that have found that populations from both sides of the Mediterranean Sea were

closely related as a consequence of long-distance dispersal events between the Iberian Peninsula and North Africa (Terrab et al., 2008), suggesting that these events have not been rare within the Mediterranean region (see Nieto Feliner, 2014 and references therein).

In our analyses, we did not detect significant demographic changes at the species level, although 15 populations showed a fixed haplotype. These fixed populations likely experienced bottlenecks caused by founder effects, likely reflecting the poor seed dispersal ability of the species. However, *L. subulatum* showed high chloroplast genetic diversity at the species level. This high haplotype diversity observed across populations in the chloroplast markers was also coupled with high overall genetic diversity in microsatellite markers. Furthermore, *L. subulatum* also exhibited high microsatellite intrapopulation diversity in all populations. These high values of genetic diversity are congruent with the current high number of individuals at each population, which may reach up to several thousand plants (personal observation). The effective population size of organelle genes is lower than that of nuclear genes (Petit et al., 2005), which could explain the slightly higher values of genetic diversity found in microsatellites markers in some populations. Some authors have reported that edaphic specialists may be composed of genetically depauperate populations due to the specialization to the substrate (see Rajakaruna, 2017 for a deeper discussion), and as such, they may constitute evolutionary dead-ends. However, our results for this species show that this is not necessarily the case, and agree with other studies that also found high levels of genetic diversity at the landscape level in both Iberian (Matesanz et al., 2019) and non-Iberian gypsophiles (Aguirre-Liguori et al., 2014).

Even though genetic variation was high regardless of the type of marker, we found contrasting results for the population genetic structure inferred by microsatellites and chloroplast markers. We observed significant genetic structure in both markers, but greater geographic structure in chloroplast loci. We are aware that comparing markers with different

number of alleles and/or different mutation rates (e.g. nuclear microsatellites and chloroplast sequence data) could bias the comparison of genetic differentiation among populations (Meirmans, 2006; Jost, 2008; Verity & Nichols 2014). Jost (2008) proposed D_{est} as a nearly unbiased estimator to assess genetic differentiation between populations accounting for different allele numbers. The calculated D_{est} values of nuclear and chloroplast markers for our populations are virtually identical to the computed F_{ST} values (see Appendix S11), suggesting that the large difference in population structure is not an artifact due to the choice of markers, but rather, it is due to eco-evolutionary processes. Because chloroplast DNA is maternally inherited and nuclear DNA is biparentally inherited in our species, it is likely that the greater genetic structure observed in the chloroplast data indicates that gene flow via pollen is higher than via seeds in *L. subulatum*. Indeed, using the indices of population differentiation (F_{ST} and G_{ST}) calculated for both markers types, and applying Ennos' equation (see Appendix S12), we estimated an effective gene flow via pollen between ~2-10 times higher than via seeds, agreeing with studies reporting that pollen flow is usually higher than seed flow (Petit et al., 2005).

However, we did not expect such large restrictions to the movement of seeds among populations in this system. In a field study assessing the role of grazing in gypsum plant communities, Pueyo et al., (2008) found that livestock act as effective seed dispersal agents between fragments. Accordingly, livestock practices, which often involve the movement of cattle across different geographical regions in the Iberian Peninsula (Azcárate et al., 2013), could have favored the movement of seeds between different *L. subulatum* populations, reducing the high genetic structure found in chloroplast markers. However, our results show limited seed dispersal, particularly between geographical regions. Specifically, the populations from the Ebro Valley and North Africa were strongly different from all other populations, as shown by the isolation by distance among different regions, suggesting that animals are likely not playing a key role in the movement of seeds in our system, at least when long distances are

562 considered. Despite the importance of transhumance in the Iberian Peninsula, drove roads of
563 the Ebro Valley never have been connected to all others main drove roads (see Fig. 1 in
564 Manzano & Casas, 2010), which could have increased the differences between this region and
565 the rest of the Iberian Peninsula. Furthermore, several nearby populations in the same
566 geographic region did not share haplotypes (see Tajo Valley in Fig. 3), which shows limited
567 seed dispersal even across short distances. These results may also be explained by the fact that,
568 similar to other gypsophiles, seeds of *L. subulatum* lack obvious long-distance dispersal
569 mechanisms (Escudero et al., 2000; Moore and Jansen, 2007). Therefore, our results also
570 suggest that seed movement between distant areas may only be possible by chance long-
571 distance dispersal events. The Iberian Mountain Range, which separates the Ebro Basin
572 populations from all others, could restrict seed movement between the populations from the
573 Ebro Basin and all the other populations, accounting for the high genetic differences observed
574 between this region and the rest. Similarly, the presence of the Mediterranean Sea may also
575 block seed dispersal from the Iberian Peninsula to North Africa, explaining the distinctive
576 haplotype in these populations.

577 Conversely, pollen movement does not appear to have been strongly limited between
578 populations or geographical regions, as shown by the assignment of individuals from
579 populations from different geographical regions to the same genetic cluster in microsatellite
580 analyses (Fig. 5). High pollen flow among populations and regions could be favored by the
581 presence of numerous patches of gypsum habitat among populations that would increase their
582 connectivity, allowing an efficient movement of different pollinators between populations
583 (Santamaría et al., 2018; Matesanz et al., 2019). *Lepidium subulatum* presents an advanced
584 phenology compared to other species of gypsum ecosystems (Hernández Bermejo and
585 Clemente, 1993; Matesanz et al., 2018) and it is possible that pollinators could actively seek
586 the flowering plants at this early season, facilitating pollen flow to further distances.

Interestingly, in a few instances populations within the same region (sometimes located less than 50 km apart) possessed individuals that were assigned to different genetic clusters (Fig. 5). Although we cannot pinpoint the exact processes that modulate this complex pattern, several factors, including uneven pollen flow between populations, differential barriers to pollen flow at small scales, differences in connectivity among populations and population size could be responsible for this pattern (Aguilar et al., 2008).

CONCLUSIONS

Our results show how paleoclimatic and geological changes in Plio-Pleistocene could be important in the origin and evolution of *L. subulatum*. The contrasting pattern of genetic structure found in the nuclear and chloroplast markers, suggesting lower seed flow among populations compared to pollen flow, also highlight the importance of using both maternally and biparentally inherited markers to fully understand the phylogeography of plant species. Furthermore, the species exhibited high values of genetic diversity in both markers, especially in microsatellites. Our results suggest that regionally dominant gypsophiles like *L. subulatum* have had broad distributions and maintained high effective population sizes during their evolutionary history, suggesting that these gypsophilic taxa are relatively old. Although the markers used in this study inform us about the neutral genetic diversity of the populations, if neutral genetic diversity and quantitative genetic diversity were correlated in populations of *L. subulatum*, our results would suggest the existence of adaptive potential to cope with changing conditions. In this context, further studies should focus on the levels of quantitative genetic variation of populations and whether it is influenced by the geographical location or the evolutionary history of the populations.

ACKNOWLEDGEMENTS

This study was funded by the Spanish Ministerio de Economía y Competitividad (grant GYPSEVOL, CGL2016-75566-P). This study was also supported by the Community of Madrid (grant Remedinal3-CM, S2013/MAE-2719) and by the European Union (grant GYPWORLD H2020-MSCA-RISE-2017). We sincerely thank the editor, an anonymous reviewer, and Kent Holsinger for their helpful comments that improved the manuscript. We thank Miguel Ballesteros, José Miguel Castillejo, Rocío Chaves, Pablo Ferrandis, Alfredo García, Roberto Lázaro, Juan Lorite, Arantzazu L. Luzuriaga, Hilario Matesanz, Gabriel Monserrat, Alicia Montesinos, Juan F. Mota, Laura Ortiz, Sara Palacio, Esteban Salmerón and Ana M. Sánchez, for their help locating populations of the study species and their expertise in gypsum ecosystems and geology. We are profoundly grateful to Mathieu Chambouleyron (Emirates Center for Wildlife Propagation, ECWP, Missour, Morocco) for collecting leaf samples of the Moroccan population.

AUTHOR CONTRIBUTIONS

M.B.-S., A.E., and S.M. conceived and designed the study; M.B.-S., M.R.-M., B.P., A.G.-F. and S.M. collected leaf samples; M.B.-S., M.R.-M., L.P. and I.I. performed the laboratory work. M.B.-S., M.J.M., A.G.-F., M.P. and S.M. performed the statistical analysis and contributed to the interpretation of the results; M.B.-S. led the writing of the manuscript with input from all other authors.

DATA AVAILABILITY

Raw DNA data, microsatellite scoring, final chloroplast DNA sequence assembly and climatic data are available on Figshare (https://doi.org/10.6084/m9.figshare.c.5132882). New generated DNA sequences of *L. subulatum* and *L. cardamines* from different markers used in this study have been submitted to GenBank (*psbA L. subulatum*: MW044696 – MW044914; *psbA L.*

637 *cardamines*: MW048746 – MW048748; *matK L. subulatum*: MW119359 – MW119565; *matK*
638 *L. cardamines*: MW048759 – MW048761; ITS *L. subulatum*: MW067154 – MW067157; ITS
639 *L. cardamines*: MW058062; *trnL* intron *L. subulatum*: MW048753 – MW048755; *trnL* intron
640 *L. cardamines*: MW048749 – MW048750; *trnL-trnF L. subulatum*: MW048756 – MW048758;
641 *trnL-trnF L. cardamines*: MW048751 – MW048752).
642

643 ## SUPPORTING INFORMATION

644 Additional Supporting Information may be found online in the supporting information section
645 at the end of the article.
646 APPENDIX S1. Voucher accessions of used specimens.
647 APPENDIX S2. Detailed information of microsatellite loci used.
648 APPENDIX S3. Detailed information of PCR reactions.
649 APPENDIX S4. Molecular markers used in the preliminary screening and selected for our
650 study.
651 APPENDIX S5. Maximum Likelihood phylogenetic tree based on concatenated *matK* and *psbA*
652 sequences.
653 APPENDIX S6. Graphics of the tests used to select the appropriate number of genetic clusters.
654 APPENDIX S7. Final length of the alignments used in this study.
655 APPENDIX S8. Maximum Likelihood phylogenetic tree of *Lepidium* based on ITS sequences.
656 APPENDIX S9. Pairwise F_{ST} matrix from nuclear microsatellites for all sampled populations.
657 APPENDIX S10. Haplotype networks individually performed for *psbA* and *matK* loci.
658 APPENDIX S11. D_{est} values of nuclear and chloroplast markers and their comparison with F_{ST}
659 values.
660 APPENDIX S12. Population differentiation indices of chloroplast and nuclear markers and
661 ratio pollen flow-seeds flow.

LITERATURE CITED

Aguilar, R., M. Quesada, L. Ashworth, Y. Herrerias-Diego, and J. Lobo. 2008. Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches. *Molecular Ecology* 17: 5177–5188.

Aguirre-Liguori, J. A., E. Scheinvar, and L. E. Eguiarte. 2014. Gypsum soil restriction drives genetic differentiation in Fouquieria shrevei (Fouquieriaceae). *American Journal of Botany* 101: 730–736.

Aktas, C. 2015. Haplotypes: haplotype inference and statistical analysis of genetic variation. *R package version* 1.

Aparicio, A., S. Martín-Hernanz, C. Parejo-Farnés, J. Arroyo, S. Lavergne, E. B. Yeşilyurt, M. L. Zhang, et al. 2017. Phylogenetic reconstruction of the genus Helianthemum (Cistaceae) using plastid and nuclear dna-sequences: Systematic and evolutionary inferences. *Taxon* 66: 868–885.

Aris-Brosou, S., and L. Excoffier. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution* 13: 494–504.

Avise, J. 2000. John C. Avise - Phylogeography_ The History and Formation of Species-Harvard University Press (2000).pdf. 455.

Azcárate, F. M., I. Robleño, J. Seoane, P. Manzano, and B. Peco. 2013. Drove roads as local biodiversity reservoirs: Effects on landscape pattern and plant communities in a Mediterranean region. *Applied Vegetation Science* 16: 480–490.

Beilstein, M. A., N. S. Nagalingum, M. D. Clements, S. R. Manchester, and S. Mathews. 2010. Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* 107: 18724–18728.

Blondel, J., J. Aronson, J.-Y. Bodiou, and G. Boeuf. 2010. The Mediterranean Region - Biological Diversity in Space and Time. *Oxford University Press*: 401.

Chávez-Pesqueira, M., and J. Núñez-Farfán. 2016. Genetic diversity and structure of wild populations of Carica papaya in Northern Mesoamerica inferred by nuclear microsatellites and chloroplast markers. *Annals of Botany* 118: 1293–1306.

Clark, K., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2016. GenBank. *Nucleic Acids Research* 44: D67–D72.

Clement, M. J., Q. Snell, P. Walker, D. Posada, and K. A. Crandall. 2002. TCS: estimating gene genealogies. ipdps, 184.

van Dam, J. A. 2006. Geographic and temporal patterns in the late Neogene (12-3 Ma) aridification of Europe: The use of small mammals as paleoprecipitation proxies. *Palaeogeography, Palaeoclimatology, Palaeoecology* 238: 190–218.

Dyer, R. J. 2016. gstudio: Tools related to the spatial analysis of genetic marker data. *R package version* 1.

Earl, D. A., and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Ellstrand, N. C., R. Whitkus, and L. H. Rieseberg. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences* 93: 5090–5093.

Ennos, R. A., A. Langdon, X.-S. Hu, and W. F. Sinclair. 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. *Molecular systematics and plant evolution*: 1–19.

710 Escavy, J. I., M. J. Herrero, and M. E. Arribas. 2012. Gypsum resources of Spain: Temporal
711 and spatial distribution. *Ore Geology Reviews* 49: 72–84.
712 Escudero, A., J. M. Iriondo, J. M. Olano, A. Rubio, and R. C. Somolinos. 2000. Factors
713 affecting establishment of a gypsophyte: The case of Lepidium subulatum
714 (Brassicaceae). *American Journal of Botany* 87: 861–871.
715 Escudero, A., S. Palacio, F. T. Maestre, and A. L. Luzuriaga. 2015. Plant life on gypsum: a
716 review of its multiple facets. *Biological Reviews* 90: 1–18.
717 Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals
718 using the software STRUCTURE: a simulation study. *Molecular ecology* 14: 2611–
719 2620.
720 Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred
721 from metric distances among DNA haplotypes: application to human mitochondrial
722 DNA restriction data. *Genetics* 131: 479–491.
723 Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of Population Structure Using
724 Multilocus Genotype Data : Linked Loci and Correlated Allele Frequencies. 1587: 1567–
725 1587.
726 Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth,
727 hitchhiking and background selection. *Genetics* 147: 915–925.
728 Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693–
729 709.
730 Garcia-Castellanos, D., and A. Villaseñor. 2011. Messinian salinity crisis regulated by
731 competing tectonics and erosion at the Gibraltar arc. *Nature* 480: 359–363.
732 Gernhard, T. 2008. The conditioned reconstructed process. *Journal of Theoretical Biology*
733 253: 769–778.
734 Gómez-Fernández, A., I. Alcocer, and S. Matesanz. 2016. Does higher connectivity lead to
735 higher genetic diversity? Effects of habitat fragmentation on genetic variation and
736 population structure in a gypsophile. *Conservation Genetics* 17: 631–641.
737 Gómez, J. M., R. Zamora, J. A. Hódar, and D. García. 1996. Experimental study of
738 pollination by ants in Mediterranean high mountain and arid habitats. *Oecologia* 105:
739 236–242.
740 Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of
741 ecological data. *Journal of Statistical Software* 22: 1–19.
742 Goudet, J., and T. Jombart. 2015. Estimation and tests of hierarchical F-statistics. R Core
743 Team.
744 Guo, X., J. Liu, G. Hao, L. Zhang, K. Mao, X. Wang, D. Zhang, et al. 2017. Plastome
745 phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18.
746 Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a
747 molecular clock of mitochondrial DNA. *Journal of molecular evolution* 22: 160–174.
748 Hernández Bermejo, J. E., and M. Clemente. 1993. Lepidium (L.) R. Br. *Flora iberica* 4:
749 311–327.
750 Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation
751 program for dealing with label switching and multimodality in analysis of population
752 structure. *Bioinformatics* 23: 1801–1806.
753 Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, and
754 R. L. Andrew. 2017. The K = 2 conundrum. *Molecular Ecology* 26: 3594–3602.
755 Jost, L. 2008. GST and its relatives do not measure differentiation. *Molecular Ecology* 17:
756 4015–4026.
757 Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic
758 analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:
759 e281.

760 Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E.
761 Zimmermann, et al. 2017. Climatologies at high resolution for the earth's land surface
762 areas. *Scientific Data* 4: 1–20.
763 Larsson, A. 2014. AliView: a fast and lightweight alignment viewer and editor for large
764 datasets. *Bioinformatics* 30: 3276–3278.
765 Legendre, P., and L. F. J. Legendre. 2012. Numerical ecology. Elsevier.
766 Leigh, J. W., and D. Bryant. 2015. POPART: Full-feature software for haplotype network
767 construction. *Methods in Ecology and Evolution* 6: 1110–1116.
768 Manzano, P., and R. Casas. 2010. Past, present and future of Transhumancia in Spain:
769 Nomadism in a developed country. *Pastoralism: research, policy and practice* 1: 72–90.
770 Martínez-Nieto, M. I., M. Encarna Merlo, J. F. Mota, E. Salmerón-Sánchez, and J. G.
771 Segarra-Moragues. 2012. Microsatellite loci in the gypsophyte Lepidium subulatum
772 (Brassicaceae), and transferability to other Lepidieae. *International journal of molecular*
773 *sciences* 13: 11861–11869.
774 Matesanz, S., A. García-Fernández, A. Limón-Yelmo, A. Gómez-Fernández, and A.
775 Escudero. 2018. Comparative landscape genetics of gypsum specialists with naturally-
776 patchy distributions reveal their resilience to anthropogenic fragmentation. *Perspectives*
777 *in Plant Ecology, Evolution and Systematics* 34: 1–9.
778 Matesanz, S., M. Ramos-Muñoz, M. Blanco-Sánchez, A. García-Fernández, A. M. Sánchez,
779 and A. Escudero. 2019. Migración, variabilidad genética y plasticidad fenotípica en
780 especies de yesos y su papel en la respuesta al cambio climático. *Revista Ecosistemas* 28:
781 48–59.
782 Meirmans, P. G. 2006. Using the Amova Framework To Estimate a Standardized Genetic
783 Differentiation Measure. *Evolution* 60: 2399.
784 Meirmans, P. G. 2015. Seven common mistakes in population genetics and how to avoid
785 them. *Molecular Ecology* 24: 3223–3231.
786 Meyer, S. E. 1986. The ecology of gypsophile endemism in the eastern Mojave Desert.
787 *Ecology* 67: 1303–1313.
788 Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for
789 inference of large phylogenetic trees. *2010 Gateway Computing Environments*
790 *Workshop, GCE 2010*.
791 Moore, M. J., and R. K. Jansen. 2007. Origins and Biogeography of Gypsophily in the
792 Chihuahuan Desert Plant Group Tiquilia subg. Eddya (Boraginaceae). *Systematic Botany*
793 32: 392–414.
794 Mota, J. F., P. Sánchez-Gómez, and J. S. Guirado. 2011. Diversidad vegetal de las yeseras
795 ibéricas. *El reto de los archipiélagos edáficos para la biología de la conservación.*
796 *ADIF-Mediterráneo Asesores Consultores, Almería*.
797 Mummenhoff, K., A. Polster, A. Mühlhausen, and G. Theißen. 2009. Lepidium as a model
798 system for studying the evolution of fruit development in Brassicaceae. *Journal of*
799 *Experimental Botany* 60: 1503–1513.
800 Nei, M. 1987. Molecular evolutionary genetics. Columbia university press.
801 Nieto Feliner, G. 2014. Patterns and processes in plant phylogeography in the Mediterranean
802 Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics* 16: 265–278.
803 Oksanen, A. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R.
804 Minchin, et al. 2019. Package ' vegan '.
805 Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-
806 CHECKER: Software for identifying and correcting genotyping errors in microsatellite
807 data. *Molecular Ecology Notes* 4: 535–538.
808 Parsons, R. F. 1976. Gypsophily in plants-a review. *American Midland Naturalist*: 1–20.
809 Pérez-Collazos, E., P. Sánchez-Gómez, J. F. Jiménez, and P. Catalán. 2009. The

phylogeographical history of the Iberian steppe plant Ferula loscosii (Apiaceae): a test of the abundant-centre hypothesis. *Molecular Ecology* 18: 848–861.
Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini, and G. G. Vendramin. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14: 689–701.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
Pueyo, Y., C. L. Alados, O. Barrantes, B. Komac, and M. Rietkerk. 2008. Differences in gypsum plant communities associated with habitat fragmentation and livestock grazing. *Ecological Applications* 18: 954–964.
R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna.
Rajakaruna, N. 2017. Lessons on Evolution from the Study of Edaphic Specialization. *Botanical Review*: 1–40.
Rambaut, A., and A. J. Drummond. 2009. Tracer v1. 5.0.
Rambaut, A., and A. J. Drummond. 2014. TreeAnnotator v1. 8.2.
Romão, R. L., and A. Escudero. 2005. Gypsum physical soil crusts and the existence of gypsophytes in semi-arid central Spain. *Plant Ecology* 181: 127–137.
Rosenberg, N. A. 2004. DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.
Rozas, J., A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E. Ramos-Onsins, and A. Sánchez-Gracia. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution* 34: 3299–3302.
Santamaría, S., A. M. Sánchez, J. López-Angulo, C. Ornosa, I. Mola, and A. Escudero. 2018. Landscape effects on pollination networks in Mediterranean gypsum islands. *Plant Biology* 20: 184–194.
Schaal, B. A., D. A. Hayworth, K. M. M. Olsen, J. T. Rauscher, and W. A. Smith. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7: 465–474.
Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
Suc, J. P. 1984. Origin and evolution of the mediterranean vegetation and climate in Europe. *Nature* 307: 429–432.
Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4: vey016.
Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
Terrab, A., P. Schönswetter, S. Talavera, E. Vela, and T. F. Stuessy. 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*.
Thompson, J. D. 2005. Plant evolution in the Mediterranean. Oxford University Press on Demand.
Verity, R., and R. A. Nichols. 2014. What is genetic differentiation, and how should we measure it - G ST, D, neither or both? *Molecular Ecology* 23: 4216–4225.
Verity, R., and R. A. Nichols. 2016. Estimating the number of subpopulations (K) in structured populations.
Wang, I. J. 2011. Choosing appropriate genetic markers and analytical methods for testing landscape genetic hypotheses. *Molecular Ecology* 20: 2480–2482.

860 Weir, B. S., and W. G. Hill. 2002. Estimating F-Statistics. *Annual Review of Genetics* 36:
861 721–750.

862 **Table 1**: Population code, location, geographical coordinates, elevation, and geographical region of the 27 populations of *Lepidium subulatum* L. used in this
863 study.

864

Geographic region	Population code	Population location	Geographical coordinates		Altitude (m asl)	T. mean (ºC)	T. min. (ºC)	T. max. (ºC)	Prec. (mm)
Duero Basin (DB)	BAL	Los Balbases (Burgos, Spain)	42º 13' 20.3" N	4º 4' 30.9" W	851	11.4	4.2	19.8	467.7
	PEÑ	Peñafiel (Valladolid, Spain)	41º 35' 25.0" N	4º 6' 30.1" W	815	12.6	4.5	22.0	432.3
	SEG	Vallelado (Segovia, Spain)	41º 24' 48.5" N	4º 25' 30.3" W	818	12.7	4.7	22.2	510.7
	TOR	Torquemada (Palencia, Spain)	42º 2' 26.6" N	4º 20' 49.3" W	833	12.1	4.6	20.8	443.6
Ebro Basin (EB)	ALF	Alfajarín (Zaragoza, Spain)	41º 37' 25.5" N	0º 41' 52.3" W	219	15.7	7.7	25.2	363.2
	GEL	Gelsa (Zaragoza, Spain)	41º 27' 5.3" N	0º 22' 24.6" W	254	15.7	7.6	25.3	367.3
	PER	Peralta (Navarra, Spain)	42º 23' 22.5" N	1º 48' 38.5" W	385	13.6	6.1	21.9	589.9
	TDL	Tamarite de Litera (Huesca, Spain)	41º 53' 9.5" N	0º 24' 58.5" E	418	13.8	6.0	23.0	490.0
	TER	Villalba Baja (Teruel, Spain)	40º 25' 9.7" N	1º 4' 48.2" W	954	12.0	4.2	21.5	339.7
Guadalquivir and Júcar-Segura Basins (GJSB)	AGR	Agramón (Albacete, Spain)	38º 24' 51.2" N	1º 37' 56.1" W	388	16.9	8.9	26.4	300.6
	APG	Altiplano granadino (Granada, Spain)	37º 33' 23.5" N	3º 2' 47.9" W	738	15.7	6.8	25.9	523.5
	BAZ	Hoya de Baza (Granada, Spain)	37º 38' 0.8" N	2º 34' 37.1" W	903	14.4	6.0	24.2	459.9
	CAB	Cabezo Redondo (Alicante, Spain)	38º 38' 32.9" N	0º 53' 33.5" W	533	15.7	8.2	24.4	370.7
	ECZ	Escúzar (Granada, Spain)	37º 3' 20.2" N	3º 44' 41.5" W	927	14.2	6.0	23.7	520.3
	TOP	Topares (Almería, Spain)	37º 52' 18.4" N	2º 11' 22.0" W	1157	12.4	4.4	21.9	395.1
	VAL	Valdeganga (Albacete, Spain)	39º 8' 10.4" N	1º 44' 26.7" W	632	15.2	7.1	25.1	346.7
	VY	Venta de Yesos (Almería, Spain)	37º 5' 2.3" N	2º 17' 7.3" W	539	16.3	9.4	24.7	254.7
Tajo Basin (TB)	ARA	Aranjuez (Madrid, Spain)	40º 1' 51.5" N	3º 32' 54.4" W	595	15.8	6.4	26.8	406.9
	AZQ	Aranzueque (Guadalajara, Spain)	40º 30' 23.7" N	3º 6' 47.1" W	742	13.6	5.3	24.2	414.9
	BEL	Belinchón (Cuenca, Spain)	40º 4' 43.5" N	3º 4' 3.7" W	706	15.1	5.8	26.0	419.2
	CHI	Chinchón (Madrid, Spain)	40º 10' 13.2" N	3º 25' 59.4" W	676	15.1	5.9	26.0	465.3
	PDG	Portalrubio de Guadamejud (Cuenca, Spain)	40º 16' 15.8" N	2º 35' 14.7" W	794	13.9	5.2	24.5	508.6
	SMV	San Martín de la Vega (Madrid, Spain)	40º 13' 19.2" N	3º 35' 3.3" W	551	15.6	6.4	26.6	376.3
	SPP	San Pedro Palmiches (Cuenca, Spain)	40º 25' 51.9" N	2º 23' 51.1" W	850	13.6	5.0	24.0	647.8
	YEB	Yebra (Guadalajara , Spain)	40º 20' 43.0" N	2º 56' 27.2" W	718	14.3	5.6	25.1	419.8
North Africa (NA)	ARGL	Chott Ech Chergui (Algeria)	34° 17' 59.3" N	0° 40' 33.5" E	989	16.4	6.8	27.9	257.3
	MAR	Yerada (Morocco)	34º 13' 29.4" N	2º 7' 21.2" W	944	16.5	8.1	26.5	282.2

865 **Table 2**: Genetic diversity indices based on concatenated chloroplast *matK* and *psbA* regions for the 27 populations of *Lepidium subulatum*. Significant
866 values for Tajima's *D*, Fu and Li's *F** and Fu's F_S tests are in bold.

Population code	**N. of sequences**	**N. of segregating sites**	**N. of haplotypes**	**Haplotype diversity (Hd)**	**Nucleotide diversity (π)**	**Tajima's *D***	**Fu and Li's *F****	**Fu's F_S**
AGR	7	0	1	0	0	-	-	-
ALF	7	1	2	0.571	0.00080	1.342	1.102	0.856
APG	8	0	1	0	0	-	-	-
ARA	8	0	1	0	0	-	-	-
ARGL	2	0	1	0	0	-	-	-
AZQ	7	0	1	0	0	-	-	-
BAL	7	0	1	0	0	-	-	-
BAZ	7	0	1	0	0	-	-	-
BEL	8	2	2	0.250	0.00070	-1.310	-1.514	0.762
CAB	8	0	1	0	0	-	-	-
CHI	7	1	2	0.286	0.00040	-1.237	-1.374	0.856
ECZ	8	0	1	0	0	-	-	-
GEL	8	6	4	0.750	0.00334	0.215	0.881	0.869
MAR	10	0	1	0	0	-	-	-
PDG	8	3	3	0.607	0.00185	0.585	0.401	0.723
PEÑ	8	2	2	0.536	0.00150	1.449	1.297	2.083
PER	8	2	2	0.571	0.00160	1.794	1.384	2.216
SEG	8	0	1	0	0	-	-	-
SMV	8	3	2	0.571	0.00239	**1.982**	1.541	3.149
SPP	7	1	2	0.286	0.00040	-1.237	-1.374	0.856
TDL	8	1	2	0.250	0.00035	-1.310	-1.514	0.762
TER	8	0	1	0	0	-	-	-
TOP	7	0	1	0	0	-	-	-
TOR	8	0	1	0	0	-	-	-
VAL	8	0	1	0	0	-	-	-
VY	8	2	3	0.464	0.00070	-1.310	-1.514	-0.999
YEB	8	2	2	0.250	0.00070	-1.310	-1.514	0.762
Overall	204	19	22	0.747	0.00382	-0.399	-1.640	-5.931

867 **Table 3**: Results of two different AMOVA tests for microsatellite and chloroplast markers: 1) non-hierarchical AMOVA; 2) hierarchical AMOVA considering
868 the geographic location (regions) of the populations; df = degrees of freedom.

869

		Nuclear DNA					Chloroplast DNA				
AMOVA type	Source of variation	df	Sum of squares	Variance	Percentage of variation	p-value	df	Sum of squares	Variance	Percentage of variation	p-value
1) Non – hierarchical AMOVA	Among populations	25	1544.746	1.419	18.627 %	< 0.001	26	1389.277	6.129	46.080 %	< 0.001
	Within populations	982	6074.267	6.201	81.373 %		177	1269.484	7.172	53.920 %	
	Total	1007	7619.012	7.621	100 %		203	2658.760	13.302	100 %	
2) Populations grouped by geographic location	Among regions	4	368.845	0.227	2.955 %	< 0.001	4	831.532	4.691	32.815 %	< 0.001
	Among populations within regions	21	1175.901	1.245	16.230 %		22	557.745	2.432	17.015 %	
	Within populations	982	6074.267	6.201	80.815 %		177	1269.484	7.172	50.170 %	
	Total	1007	7619.012	7.673	100 %		203	2658.760	14.296	100 %	

Table 4: Genetic diversity indices of the 26 populations (excluding ARGL) of *Lepidium subulatum* using 10 microsatellite loci. *N*: Number of individuals sampled; *N eff.*: Effective number of individuals sampled; *P*: percentage of polymorphic loci; *A*: Mean number of alleles per locus; A_{rare}: Rarefied number of alleles per locus (10 individuals, 20 genes); *Ae*: Mean number of effective alleles per locus; H_o: Observed heterozygosity; H_e: Expected heterozygosity; F_{IS}: Inbreeding coefficient; β: Population-specific F_{ST} coefficient.

Population code	***N***	***N eff.***	***P***	***A***	A_{rare}	A_e	H_o	H_e	F_{IS}	β	**Nb. of private alleles**	**Nb. of genotypes**
AGR	20	18.9	90.0%	3.8	3.47	2.54	0.528	0.548	0.060	0.245	0	20
ALF	20	19.7	100.0%	5.4	4.68	3.35	0.587	0.605	0.065	0.169	0	20
APG	20	19.9	100.0%	6.1	4.87	3.50	0.574	0.580	0.024	0.203	2	20
ARA	20	19.8	100.0%	5.9	4.91	3.29	0.621	0.618	-0.002	0.150	2	19
AZQ	20	19.9	100.0%	6.0	5.11	3.80	0.669	0.662	-0.009	0.090	2	20
BAL	20	19.5	100.0%	4.8	4.11	2.56	0.463	0.515	0.171	0.292	0	20
BAZ	20	19.8	100.0%	7.1	5.65	3.86	0.520	0.681	0.270	0.064	1	20
BEL	20	19.9	100.0%	7.0	5.57	3.89	0.538	0.647	0.178	0.111	1	20
CAB	20	19.6	100.0%	3.4	2.92	2.08	0.500	0.497	0.026	0.317	0	20
CHI	20	19.9	100.0%	6.8	5.54	3.71	0.557	0.627	0.111	0.138	1	20
ECZ	20	19.1	100.0%	4.5	3.76	2.32	0.470	0.505	0.074	0.306	0	20
GEL	20	19.9	100.0%	5.6	4.78	3.41	0.594	0.608	0.049	0.165	0	20
MAR	10	9.7	100.0%	4.5	4.50	3.05	0.479	0.602	0.258	0.149	1	10
PDG	20	19.6	100.0%	5.5	4.61	3.32	0.572	0.631	0.145	0.132	0	20
PEÑ	20	20	100.0%	6.0	5.10	3.53	0.570	0.667	0.144	0.084	1	20
PER	14	13.8	100.0%	4.0	3.79	2.79	0.718	0.601	-0.093	0.164	2	14
SEG	20	19.7	100.0%	3.3	3.01	1.93	0.471	0.452	0.000	0.379	0	20
SMV	20	19.3	90.0%	5.5	4.78	3.38	0.514	0.599	0.204	0.176	1	20
SPP	20	19.7	90.0%	4.1	3.49	2.49	0.415	0.483	0.153	0.335	1	19
TDL	20	20	100.0%	5.1	4.15	2.47	0.485	0.493	0.019	0.323	3	20
TER	20	19.5	100.0%	5.3	4.41	3.07	0.561	0.602	0.066	0.172	1	20
TOP	20	19.8	100.0%	4.8	4.17	2.89	0.548	0.594	0.134	0.184	0	20
TOR	20	20	100.0%	5.6	4.83	3.38	0.640	0.635	0.008	0.127	0	20
VAL	20	19.9	100.0%	4.4	4.03	3.14	0.663	0.619	-0.043	0.150	0	20
VY	20	19.8	100.0%	5.1	4.39	3.15	0.479	0.585	0.194	0.196	2	20
YEB	20	19.8	90.0%	4.2	3.75	2.64	0.479	0.521	0.094	0.283	2	20
Overall	504	19.096	98.5%	5.15	4.40	3.06	0.547	0.584	0.089	0.196	23	502

APPENDIX 1

List of: a) GenBank accession numbers for ITS sequences used in this study (individuals with only one accession number included sequence for both ITS regions). b) GenBank accession numbers for *trnT-trnL*, *trnL* intron and *trnL-trnF* regions used in this study (hyphens indicate missing sequences).

a) ***Arabidopsis arenicola,*** GQ922906; ***Arabidopsis arenosa* 1,** AAU52182; ***Arabidopsis arenosa* 2,** AAU43231; ***Arabidopsis arenosa* 3,** AAU43230; ***Arabidopsis arenosa* 4,** AAU43232; ***Arabidopsis arenosa* 5,** AAU43233; ***Arabidopsis arenosa* 6,** AAU43229; ***Arabidopsis arenosa* 7,** AAU52181; ***Arabidopsis croatica* 1,** DQ528930; ***Arabidopsis croatica* 2,** DQ528949; ***Arabidopsis croatica* 3,** DQ528826; ***Arabidopsis croatica* 4,** DQ528825; ***Arabidopsis halleri* 1,** DQ528887; ***Arabidopsis halleri* 2,** DQ528882; ***Arabidopsis halleri* 3,** DQ528881; ***Arabidopsis halleri* 4,** DQ528884; ***Arabidopsis halleri* 5,** DQ528883; ***Arabidopsis halleri* 6,** DQ528885; ***Arabidopsis halleri* 7,** DQ528886; ***Arabidopsis lyrata* 1,** DQ528819; ***Arabidopsis lyrata* 2,** DQ528815; ***Arabidopsis lyrata* 3,** DQ528820; ***Arabidopsis lyrata* 4,** DQ528814; ***Arabidopsis lyrata* 5,** DQ528817; ***Arabidopsis lyrata* 6,** DQ528816; ***Arabidopsis lyrata* 7,** DQ528821; ***Arabidopsis lyrata* 8,** DQ528818; ***Arabidopsis pedemontana,*** DQ914842; ***Arabidopsis thaliana* 1,** KM892649; ***Arabidopsis thaliana* 2,** DQ528813; ***Cardaria chalepensis,*** AJ628275, AJ628276; ***Cardaria draba,*** AJ628277, AJ628278; ***Cardaria pubescens,*** AJ628279, AJ628280; ***Lepidium affghanum,*** DQ780948; ***Lepidium africanum,*** AJ582441, AJ582498; ***Lepidium aletes,*** FM178548, FM178549; ***Lepidium alluaudii,*** AJ582436, AJ582493; ***Lepidium alyssoides* 1,** KX646435; ***Lepidium alyssoides* 2,** KF022714; ***Lepidium angustissimum,*** KC174369; ***Lepidium apetalum* 1,** AJ582466, AJ582514; ***Lepidium apetalum* 2,** JF976762; ***Lepidium apetalum* 3,** JF976761; ***Lepidium apetalum* 4,** JF976760; ***Lepidium apetalum* 5,** JF976759; ***Lepidium apetalum* 6,** JF976758; ***Lepidium apetalum* 7,** JF976757; ***Lepidium apetalum* 8,** JF976756; ***Lepidium apetalum* 9,** JF976755; ***Lepidium apetalum* 10,** JF976770; ***Lepidium apetalum* 11,** JF976767; ***Lepidium apetalum* 12,** MF785672; ***Lepidium apetalum* 13,** FJ980405; ***Lepidium apetalum* 14,** JF976769; ***Lepidium apetalum* 15,** JF976754; ***Lepidium apetalum* 16,** DQ310525; ***Lepidium apetalum* 17,** KM892613; ***Lepidium apetalum* 18,** JF976768; ***Lepidium apetalum* 19,** JF976766; ***Lepidium apetalum* 20,** JF976765; ***Lepidium apetalum* 21,** JF976764; ***Lepidium apetalum* 22,** JF976763; ***Lepidium arbuscula,*** AJ582451, AJ582517; ***Lepidium armoracia,*** AJ582454, AJ582502; ***Lepidium aschersonii,*** AJ582426, AJ582483; ***Lepidium aucheri* 1,** AJ582443, AJ582525; ***Lepidium aucheri* 2,** KF850569; ***Lepidium austrinum,*** AJ582467, AJ582515; ***Lepidium banksii* 1,** AJ582433, AJ582490; ***Lepidium banksii* 2,** KC109332; ***Lepidium banksii* 3,** KC109331; ***Lepidium bidendatum,*** AJ582468, AJ582516; ***Lepidium bipinnatifidum,*** AJ582446, AJ582522; ***Lepidium biplicatum,*** FM178550, FM178551; ***Lepidium bonariense* 1,** AJ582458, AJ582506; ***Lepidium bonariense* 2,** HM134831; ***Lepidium campestre* 1,** AJ582412, AJ582469; ***Lepidium campestre* 2,** AF055197; ***Lepidium capense,*** AJ582452, AJ582500; ***Lepidium capitatum,*** FM178552, FM178553; ***Lepidium cardamines* 1,** FM178554, FM178555; ***Lepidium cardamines* 2,** MW058062; ***Lepidium chalepense,*** KX646446; ***Lepidium crenatum,*** KX646437; ***Lepidium davisii* 1,** KX774365; ***Lepidium davisii* 2,** FJ541491; ***Lepidium davisii* 3,** FJ541492; ***Lepidium davisii* 4,** FJ541493; ***Lepidium davisii* 5,** FJ541494; ***Lepidium densiflorum,*** KX646438; ***Lepidium desertorum,*** AJ582453, AJ582501; ***Lepidium desvauxii* 1,** AJ582429, AJ582486; ***Lepidium desvauxii* 2,** KC109334; ***Lepidium dictyotum,*** AJ582415, AJ582472; ***Lepidium didymum* 1,** KM892610; ***Lepidium didymum* 2,** KM892632; ***Lepidium didymum* 3,** KM892647; ***Lepidium divaricatum,*** AJ582437, AJ582494; ***Lepidium draba* 1,** KJ623487; ***Lepidium draba* 2,** FM164554, FM164555; ***Lepidium draba* 3,** EF367913; ***Lepidium draba* 4,** KU746329; ***Lepidium draba* 5,** KX774361; ***Lepidium draba* 6,** KX646439; ***Lepidium draba* 7,** KX646440; ***Lepidium draba* 8,** KX646441; ***Lepidium draba* 9,** KX646444; ***Lepidium draba* 10,** KX646445; ***Lepidium draba* 11,** KF022715; ***Lepidium fasciculatum,*** AJ582428, AJ582485; ***Lepidium ferganense* 1,** AJ582449, AJ582519; ***Lepidium ferganense* 2,** KM892614; ***Lepidium flavum,*** AJ582444, AJ582524; ***Lepidium flexicaule* 1,** AJ582430, AJ582487; ***Lepidium flexicaule* 2,** AF100685; ***Lepidium flexicaule* 3,** KC109335; ***Lepidium flexicaule* 4,** KC109337; ***Lepidium flexicaule* 5,** KC109336; ***Lepidium foliosum* 1,** KC109339; ***Lepidium foliosum* 2,** KC109338; ***Lepidium fremontii,*** AJ582456, AJ582504; ***Lepidium fremontii* subsp. *fremontii,*** KX646447; ***Lepidium graminifolium,*** FN821616; ***Lepidium heterophyllum,*** KX646448; ***Lepidium hirtum subsp. hirtum,*** AJ582413, AJ582470; ***Lepidium huberi,***

KX646451; ***Lepidium hyssopifolium,*** AJ582435, AJ582492; ***Lepidium kirkii,*** EF109738, EF109739; ***Lepidium lacerum* 1,** FN821519; ***Lepidium lacerum* 2,** FN821675; ***Lepidium lacerum* 3,** FN821676; ***Lepidium lasiocarpum,*** AJ582455, AJ582503; ***Lepidium latifolium,*** AJ582447, AJ582521; ***Lepidium latipes,*** AJ582416, AJ582473; ***Lepidium lyratum,*** AJ582448, AJ582520; ***Lepidium meyenii* 1,** AJ582445, AJ582523; ***Lepidium meyenii* 2,** KX646452; ***Lepidium montanum* 1,** AJ582457, AJ582505; ***Lepidium montanum* 2,** EF367921; ***Lepidium montanum* 3,** EF367922; ***Lepidium montanum* 4,** EF367923; ***Lepidium montanum* 5,** EF367924; ***Lepidium montanum* 6,** EF367925; ***Lepidium montanum* 7,** EF367926; ***Lepidium montanum* 8,** EF367927; ***Lepidium montanum* 9,** EF367928; ***Lepidium montanum* 10,** EF367929; ***Lepidium montanum* 11,** EF367930; ***Lepidium montanum* 12,** KX646453; ***Lepidium montanum* 13,** EF367931; ***Lepidium montanum* 14,** EF367932; ***Lepidium montanum* 15,** EF367933; ***Lepidium montanum* 16,** EF367934; ***Lepidium montanum* 17,** EF367935; ***Lepidium montanum* 18,** EF367936; ***Lepidium montanum* 19,** EF367937; ***Lepidium montanum* 20,** EF367938; ***Lepidium montanum* 21,** EF367939; ***Lepidium montanum* 22,** EF367940; ***Lepidium montanum* 23,** EF367914; ***Lepidium montanum* 24,** EF367941; ***Lepidium montanum* 25,** EF367942; ***Lepidium montanum* 26,** EF367943; ***Lepidium montanum* 27,** EF367944; ***Lepidium montanum* 28,** EF367945; ***Lepidium montanum* 29,** EF367946; ***Lepidium montanum* 30,** EF367947; ***Lepidium montanum* 31,** EF367948; ***Lepidium montanum* 32,** EF367949; ***Lepidium montanum* 33,** EF367950; ***Lepidium montanum* 34,** EF367915; ***Lepidium montanum* 35,** EF367951; ***Lepidium montanum* 36,** EF367952; ***Lepidium montanum* 37,** EF367953; ***Lepidium montanum* 38,** EF367954; ***Lepidium montanum* 39,** EF367955; ***Lepidium montanum* 40,** EF367956; ***Lepidium montanum* 41,** EF367957; ***Lepidium montanum* 42,** EF367958; ***Lepidium montanum* 43,** EF367959; ***Lepidium montanum* 44,** EF367960; ***Lepidium montanum* 45,** EF367916; ***Lepidium montanum* 46,** EF367961; ***Lepidium montanum* 47,** EF367962; ***Lepidium montanum* 48,** EF367963; ***Lepidium montanum* 49,** EF367964; ***Lepidium montanum* 50,** EF367965; ***Lepidium montanum* 51,** EF367966; ***Lepidium montanum* 52,** EF367967; ***Lepidium montanum* 53,** EF367968; ***Lepidium montanum* 54,** EF367969; ***Lepidium montanum* 55,** EF367970; ***Lepidium montanum* 56,** EF367917; ***Lepidium montanum* 57,** EF367918; ***Lepidium montanum* 58,** EF367919; ***Lepidium montanum* 59,** EF367920; ***Lepidium muelleri ferdinandi,*** AJ582427, AJ582484; ***Lepidium myriocarpum,*** AJ582442, AJ582499; ***Lepidium naufragorum* 1,** AJ582422, AJ582479; ***Lepidium naufragorum* 2,** AF100686; ***Lepidium navasii* 1,** KU213888; ***Lepidium navasii* 2,** KU213880; ***Lepidium navasii* 3,** KU213881; ***Lepidium navasii* 4,** KU213882; ***Lepidium navasii* 5,** KM201470; ***Lepidium navasii* 6,** KM201471; ***Lepidium navasii* 7,** KM201472; ***Lepidium navasii* 8,** KM201473; ***Lepidium navasii* 9,** KM201474; ***Lepidium navasii* 10,** KU213883; ***Lepidium navasii* 11,** KU213884; ***Lepidium navasii* 12,** KM201477; ***Lepidium navasii* 13,** KU213885; ***Lepidium navasii* 14,** KU213886; ***Lepidium navasii* 15,** KU213887; ***Lepidium navasii* 16,** KM201478; ***Lepidium navasii* 17,** KM201479; ***Lepidium navasii* 18,** KM201480; ***Lepidium navasii* 19,** KM201481; ***Lepidium navasii* 20,** KM201482; ***Lepidium navasii* 21,** KU213889; ***Lepidium navasii* 22,** KU213890; ***Lepidium navasii* 23,** KM201465; ***Lepidium navasii* 24,** KU213891; ***Lepidium navasii* 25,** KU213892; ***Lepidium navasii* 26,** KU213893; ***Lepidium navasii* 27,** KU213894; ***Lepidium navasii* 28,** KU213895; ***Lepidium navasii* 29,** KU213896; ***Lepidium navasii* 30,** KU213897; ***Lepidium navasii* 31,** KU213898; ***Lepidium navasii* 32,** KU213899; ***Lepidium navasii* 33,** KU213900; ***Lepidium navasii* 34,** KM201466; ***Lepidium navasii* 35,** KU213901; ***Lepidium navasii* 36,** KU213902; ***Lepidium navasii* 37,** KU213903; ***Lepidium navasii* 38,** KM201475; ***Lepidium navasii* 39,** KM201476; ***Lepidium navasii* 40,** KU213904; ***Lepidium navasii* 41,** KU213905; ***Lepidium navasii* 42,** KU213906; ***Lepidium navasii* 43,** KU213907; ***Lepidium navasii* 44,** KU213908; ***Lepidium navasii* 45,** KM201467; ***Lepidium navasii* 46,** KU213909; ***Lepidium navasii* 47,** KU213910; ***Lepidium navasii* 48,** KU213911; ***Lepidium navasii* 49,** KU213912; ***Lepidium navasii* 50,** KU213913; ***Lepidium navasii* 51,** KM201468; ***Lepidium navasii* 52,** KM201469; ***Lepidium navasii* 53,** KU213878; ***Lepidium navasii* 54,** KU213879; ***Lepidium nesophilum* 1,** KC109342; ***Lepidium nesophilum* 2,** KC109343; ***Lepidium nitidum,*** AJ582414, AJ582471; ***Lepidium oblongum* 1,** AJ582462, AJ582510; ***Lepidium oblongum* 2,** KX646454; ***Lepidium obtusatum,*** KC109344; ***Lepidium obtusum* 1,** MH507026; ***Lepidium obtusum* 2,** MH507027; ***Lepidium obtusum* 3,** MH507028; ***Lepidium obtusum* 4,** MH507029; ***Lepidium obtusum* 5,** MH507030; ***Lepidium obtusum* 6,** MH507031; ***Lepidium oleraceum* 1,** AJ582434, AJ582491; ***Lepidium oleraceum* 2,** KC109352; ***Lepidium oleraceum* 3,** KC109359; ***Lepidium oleraceum* 4,** KC109347; ***Lepidium oleraceum* 5,** KC109349; ***Lepidium oleraceum* 6,** KC109361; ***Lepidium oleraceum* 7,** KC109345; ***Lepidium oleraceum* 8,** KC109355; ***Lepidium oleraceum* 9,** KC109360; ***Lepidium oleraceum* 10,** KC109356; ***Lepidium oleraceum* 11,** KC109357; ***Lepidium oleraceum* 12,** AF100687; ***Lepidium oleraceum* 13,** KC109354; ***Lepidium oleraceum* 14,** KC109350; ***Lepidium oleraceum* 15,** KC109351; ***Lepidium oleraceum* 16,** KC109358; ***Lepidium oleraceum* 17,** KC109348;

Lepidium oleraceum* 18,** KC109353; ***Lepidium oleraceum* 19,** KC109346; ***Lepidium oxycarpum, AJ582417, AJ582474; ***Lepidium oxytrichum,*** AJ582424, AJ582481; ***Lepidium paniculatum,*** FM164556, FM164557; ***Lepidium papilliferum* 1,** FJ541497; ***Lepidium papilliferum* 2,** EF367976; ***Lepidium papilliferum* 3,** EF367977; ***Lepidium papilliferum* 4,** EF367978; ***Lepidium papilliferum* 5,** EF367979; ***Lepidium papilliferum* 6,** EF367980; ***Lepidium papilliferum* 7,** EF367981; ***Lepidium papilliferum* 8,** EF367982; ***Lepidium papilliferum* 9,** EF367983; ***Lepidium papilliferum* 10,** EF367984; ***Lepidium papilliferum* 11,** EF367985; ***Lepidium papilliferum* 12,** FJ541498; ***Lepidium papilliferum* 13,** EF367986; ***Lepidium papilliferum* 14,** EF367987; ***Lepidium papilliferum* 15,** EF367988; ***Lepidium papilliferum* 16,** EF367989; ***Lepidium papilliferum* 17,** EF367990; ***Lepidium papilliferum* 18,** EF367991; ***Lepidium papilliferum* 19,** EF367992; ***Lepidium papilliferum* 20,** EF367993; ***Lepidium papilliferum* 21,** EF367994; ***Lepidium papilliferum* 22,** EF367995; ***Lepidium papilliferum* 23,** FJ541495; ***Lepidium papilliferum* 24,** EF367996; ***Lepidium papilliferum* 25,** EF367997; ***Lepidium papilliferum* 26,** EF367998; ***Lepidium papilliferum* 27,** EF367999; ***Lepidium papilliferum* 28,** EF368000; ***Lepidium papilliferum* 29,** EF368001; ***Lepidium papilliferum* 30,** EF368002; ***Lepidium papilliferum* 31,** EF368003; ***Lepidium papilliferum* 32,** EF368004; ***Lepidium papilliferum* 33,** EF368005; ***Lepidium papilliferum* 34,** FJ541496; ***Lepidium papilliferum* 35,** EF368006; ***Lepidium papilliferum* 36,** EF367971; ***Lepidium papilliferum* 37,** EF367972; ***Lepidium papilliferum* 38,** EF367973; ***Lepidium papilliferum* 39,** EF367974; ***Lepidium papilliferum* 40,** EF367975; ***Lepidium papillosum,*** AJ582425, AJ582482; ***Lepidium perfoliatum* 1,** DQ399120; ***Lepidium perfoliatum* 2,** EF368007; ***Lepidium perfoliatum* 3,** KJ623470; ***Lepidium perfoliatum* 4,** KJ623469; ***Lepidium perfoliatum* 5,** KJ623472; ***Lepidium perfoliatum* 6,** KJ623471; ***Lepidium perfoliatum* 7,** JF976773; ***Lepidium perfoliatum* 8,** JF976772; ***Lepidium perfoliatum* 9,** JF976771; ***Lepidium phlebopetalum* 1,** FM178556, FM178557; ***Lepidium phlebopetalum* 2,** AY254528; ***Lepidium pinnatifidum,*** AJ582464, AJ582512; ***Lepidium pinnatum,*** AJ582439, AJ582496; ***Lepidium platypetalum,*** DQ780949; ***Lepidium pseudohyssopifolium,*** AJ582431, AJ582488; ***Lepidium pseudopapillosum,*** AJ582423, AJ582480; ***Lepidium pseudotasmanicum,*** AJ582432, AJ582489; ***Lepidium quitense,*** AJ582463, AJ582511; ***Lepidium rotundum,*** DQ780950; ***Lepidium rubtzovii* 1,** FN821677; ***Lepidium rubtzovii* 2,** FN821520; ***Lepidium ruderale* 1,** AJ582465, AJ582513; ***Lepidium ruderale* 2,** KX646455; ***Lepidium ruderale* 3,** JF976777; ***Lepidium ruderale* 4,** KJ623528; ***Lepidium ruderale* 5,** KJ623529; ***Lepidium ruderale* 6,** KJ623527; ***Lepidium ruderale* 7,** JF976776; ***Lepidium ruderale* 8,** JF976775; ***Lepidium ruderale* 9,** JF976774; ***Lepidium sativum* 1,** AJ582459, AJ582507; ***Lepidium sativum* 2,** AF283494, AF283495; ***Lepidium sativum* 3,** AY662279; ***Lepidium sativum* 4,** LC090011; ***Lepidium schinzii,*** AJ582440, AJ582497; ***Lepidium serra,*** AJ582450, AJ582518; ***Lepidium sisymbrioides* 1,** DQ997559; ***Lepidium sisymbrioides* 2,** DQ997564; ***Lepidium sisymbrioides* 3,** DQ997560; ***Lepidium sisymbrioides* 4,** DQ997568; ***Lepidium sisymbrioides* 5,** DQ997562; ***Lepidium sisymbrioides* 6,** DQ997561; ***Lepidium sisymbrioides* 7,** DQ997570; ***Lepidium sisymbrioides* 8,** DQ997565; ***Lepidium sisymbrioides* 9,** DQ997569; ***Lepidium sisymbrioides* subsp. *matau,*** AJ582418, AJ582475; ***Lepidium sisymbrioides* subsp. *kawarau* 1,** AJ582419, AJ582476; ***Lepidium sisymbrioides* subsp. *kawarau* 2,** AF100688; ***Lepidium sisymbrioides* subsp. *sisymbrioides,*** AJ582420, AJ582477; ***Lepidium solandri* 1,** DQ997567; ***Lepidium solandri* 2,** DQ997566; ***Lepidium solandri* 3,** DQ997553; ***Lepidium solandri* 4,** DQ997556; ***Lepidium solandri* 5,** DQ997558; ***Lepidium solandri* 6,** DQ997557; ***Lepidium solandri* 7,** DQ997554; ***Lepidium solandri* 8,** DQ997555; ***Lepidium solandri* 9,** DQ997563; ***Lepidium solandri* 10,** DQ997571; ***Lepidium spinescens,*** AJ582461, AJ582509; ***Lepidium spinosum* 1,** AJ582460, AJ582508; ***Lepidium spinosum* 2,** KX646456; ***Lepidium subcordatum,*** FN821674; ***Lepidium subulatum* 1,** MW067154; ***Lepidium subulatum* 2,** MW067155; ***Lepidium subulatum* 3,** MW067156; ***Lepidium subulatum* 4,** MW067157; ***Lepidium tenuicaule,*** AJ582421, AJ582478; ***Lepidium tiehmii,*** FM164558, FM164559; ***Lepidium trifurcum,*** AJ582438, AJ582495; ***Lepidium vesicarium,*** KX646458; ***Lepidium virginicum* 1,** AF283496, AF283497; ***Lepidium virginicum* 2,** AY662280; ***Lepidium virginicum* 3,** LC090012; ***Lepidium virginicum* 4,** HM134830; ***Lepidium virginicum* 5,** AF128109; ***Lepidium virginicum* 6,** KM892658; ***Lepidium virginicum* 7,** GQ478095; ***Lepidium virginicum* 8,** KP214507;

b) ***Arabidopsis arenicola,*** DQ914838, GQ244583, - ; ***Arabidopsis croatica,*** DQ529064, AY665580, - ; ***Arabidopsis lyrata,*** DQ529095, GQ244585, - ; ***Arabidopsis neglecta,*** FJ477707, LN610061, - ; ***Arabidopsis pedemontana,*** KF547407, KF547039, - ; ***Arabidopsis petrogena,*** DQ529090, DQ313520, - ; ***Arabidopsis suecica,*** LN610047, AY167921, - ; ***Arabidopsis suecica,*** LN610047, AY167921, - ; ***Arabidopsis thaliana,*** KP191402, KX668047, - ; ***Arabidopsis umezawana,*** LN610051, LN610063, - ; ***Brassica napus,*** EF426775, -, - ; ***Cochlearia pyrenaica,*** HQ268698, -, - ; ***Lepidium africanum,*** AY015921, AY015833, AY015703;

Lepidium alluaudii, AY015922, AY015834, AY015706; ***Lepidium apetalum,*** DQ821406, -, - ; ***Lepidium arbuscula,*** AY015924, AY015836, AY015707; ***Lepidium armoracia,*** AY015925, AY015837, AY015709; ***Lepidium aschersonii,*** AY015926, AY015838, AY015711; ***Lepidium aucheri,*** AY015927, AY015839, AY015713; ***Lepidium austrinum,*** AY015928, AY015840, AY015715; ***Lepidium banksii,*** AY015929, AY015841, AY015717; ***Lepidium bipinnatifidum,*** AY015931, AY015843, AY015721; ***Lepidium bonariense,*** MK261665, AY015844, AY015723; ***Lepidium capense,*** AY015933, AY015846, AY015728; ***Lepidium cardamines,*** -, MW048749, MW048751; ***Lepidium desertorum,*** AY015934, AY015847, AY015730; ***Lepidium desvauxii,*** AY015935, KC109371, AY015731; ***Lepidium dictyotum,*** AY015936, AY015849, AY015734; ***Lepidium echinatum,*** AY015937, AY015850, AY015735; ***Lepidium ferganense,*** AY015938, AY015851, AY015738; ***Lepidium flavum,*** AY015908, AY015852, AY015739; ***Lepidium flexicaule,*** AY015939, AY015853, AY015741; ***Lepidium fremontii,*** AY015940, AY015854, AY015815; ***Lepidium heterophyllum,*** AY015941, AY015855, AY015816; ***Lepidium hirtum subsp. calycotrichum,*** AY015942, AY015856, AY015817; ***Lepidium hirtum subsp. dhayense,*** AY015943, AY015857, AY015818; ***Lepidium hirtum subsp. hirtum,*** AY015819, AY015858, - ; ***Lepidium hirtum subsp. nebrodense,*** AY015945, AY015859, AY015820; ***Lepidium hirtum subsp. petrophilum,*** AY015946, AY015860, AY015821; ***Lepidium hyssopifolium,*** AY015947, AY015861, AY015743; ***Lepidium lasiocarpum,*** AY015948, EF367912, AY015745; ***Lepidium latifolium,*** MH507041, MH507043, AY015747; ***Lepidium latipes,*** AY015950, AY015864, AY015749; ***Lepidium leptopetalum,*** AY01595, AY015865, AY015751; ***Lepidium linifolium,*** AY015952, AY015866, AY015753; ***Lepidium lyratum,*** AY015953, AY015867, AY015755; ***Lepidium montanum,*** FJ541504, EF367772, AY015760; ***Lepidium muelleri-ferdinandi,*** AY015956, AY015870, AY015761; ***Lepidium myriocarpum,*** AY015957, AY015871, AY015764; ***Lepidium naufragorum,*** AY015958, AY015872, AY015765; ***Lepidium nitidum,*** AY015959, AY015873, AY015767; ***Lepidium oblongum,*** AY015960, AY015874, AY015769; ***Lepidium oleraceum,*** AY015961, AY015875, AY015771; ***Lepidium oxycarpum,*** AY015962, AY015876, AY015773; ***Lepidium oxytrichum,*** AY015963, AY015877, AY015776; ***Lepidium papillosum,*** AY015964, AY015878, AY015777; ***Lepidium pedicellosum,*** AY015965, AY015879, AY015779; ***Lepidium perfoliatum,*** KJ623396, KJ623328, - ; ***Lepidium phlebopetalum,*** AY015966, AY015881, AY015783; ***Lepidium pholidogynum,*** AY015967, AY015882, AY015785; ***Lepidium pinnatifidum,*** AY015968, AY015883, AY015787; ***Lepidium pinnatum,*** AY015969, AY015884, AY015827; ***Lepidium pseudohyssopifolium,*** - , AY015885, AY015789; ***Lepidium pseudopapillosum,*** AY015971, AY015886, - ; ***Lepidium pseudotasmanicum,*** AY015972, AY015887, AY015826; ***Lepidium quitense,*** AY015973, AY015888, AY015794; ***Lepidium rigidum,*** AY015974, AY015889, AY015828; ***Lepidium ruderale,*** KJ623452, KJ623383, AY015795; ***Lepidium schinzii,*** AY015976, AY015892, AY015797; ***Lepidium serra,*** AY015977, AY015893, AY015799; ***Lepidium sisymbrioides,*** DQ997054, - , - ; ***Lepidium sisymbrioides subsp. kawarau,*** AY015978, AY015894, AY015801; ***Lepidium sisymbrioides subsp. matau,*** AY015979, AY015895, AY015803; ***Lepidium sisymbrioides subsp. sisymbrioides,*** AY015980, AY015896, AY015805; ***Lepidium spinescens,*** AY015981, AY015897, AY015807; ***Lepidium spinosum,*** AY015914, AY015898, AY015824; ***Lepidium subulatum,*** - MW048753, MW048756; ***Lepidium trifurcum,*** AY015983, AY015900, AY015811; ***Lepidium villarsii,*** AY015916, AY015901, AY015825.

Figure captions

Figure 1: a) Map of the Iberian Peninsula and North Africa showing all sampled populations and their assignment to regions for analyses of population structure (these regions are related to the main gypsum vegetation habitats described; Mota, Sánchez-Gómez, & Guirado, 2011); b), c) and d) Gypsum environments at ARA, SPP and TOP, respectively; e) Individual of *Lepidium subulatum* at the end of its fruiting period (mid-June); f) Flowers of *L. subulatum.*

Figure 2: Maximum clade credibility (MCC) tree obtained from the BEAST analysis based on concatenated *trnT-trnL, trnL* intron and *trnL-trnF* sequences. Blue bars show highest posterior densities (HPD) credibility intervals and numbers above branches show mean estimated divergence time (Mya). HPD and dates for *L. subulatum* are in bold and dark blue (Bayesian posterior probability of this node= 95%).

Figure 3: Haplotype network for 27 populations of *Lepidium subulatum*, based on concatenated *matK* and *psbA* sequences of 204 individuals. The main groups (A, B, C and D) are shown in the haplotype network. Three missing haplotypes (extinct or unsampled) are represented by small black dots in the haplotype network. The size of the different haplotypes in the network is proportional to the number of individuals with each haplotype. The size of the pie charts in the map is proportional to the number of samples in each population. Note that the location of the ARGL population is approximate. See Appendix S10 for haplotype networks for each locus.

Figure 4: Mantel Correlograms calculated from a) microsatellite and b) chloroplast markers. In both cases, solid squares indicate that the Mantel statistic is different from zero at the 95%

confidence level. Distance classes were calculated using Sturge's rule (Legendre & Legendre, 2012).

Figure 5: Population structure (*K*=3) inferred by Bayesian cluster analyses (STRUCTURE) for 504 *L. subulatum* individuals from 26 populations. Each individual is represented by a vertical bar in each population. The size of the boxes is proportional to the number of individuals sampled in each population.

Figure 1

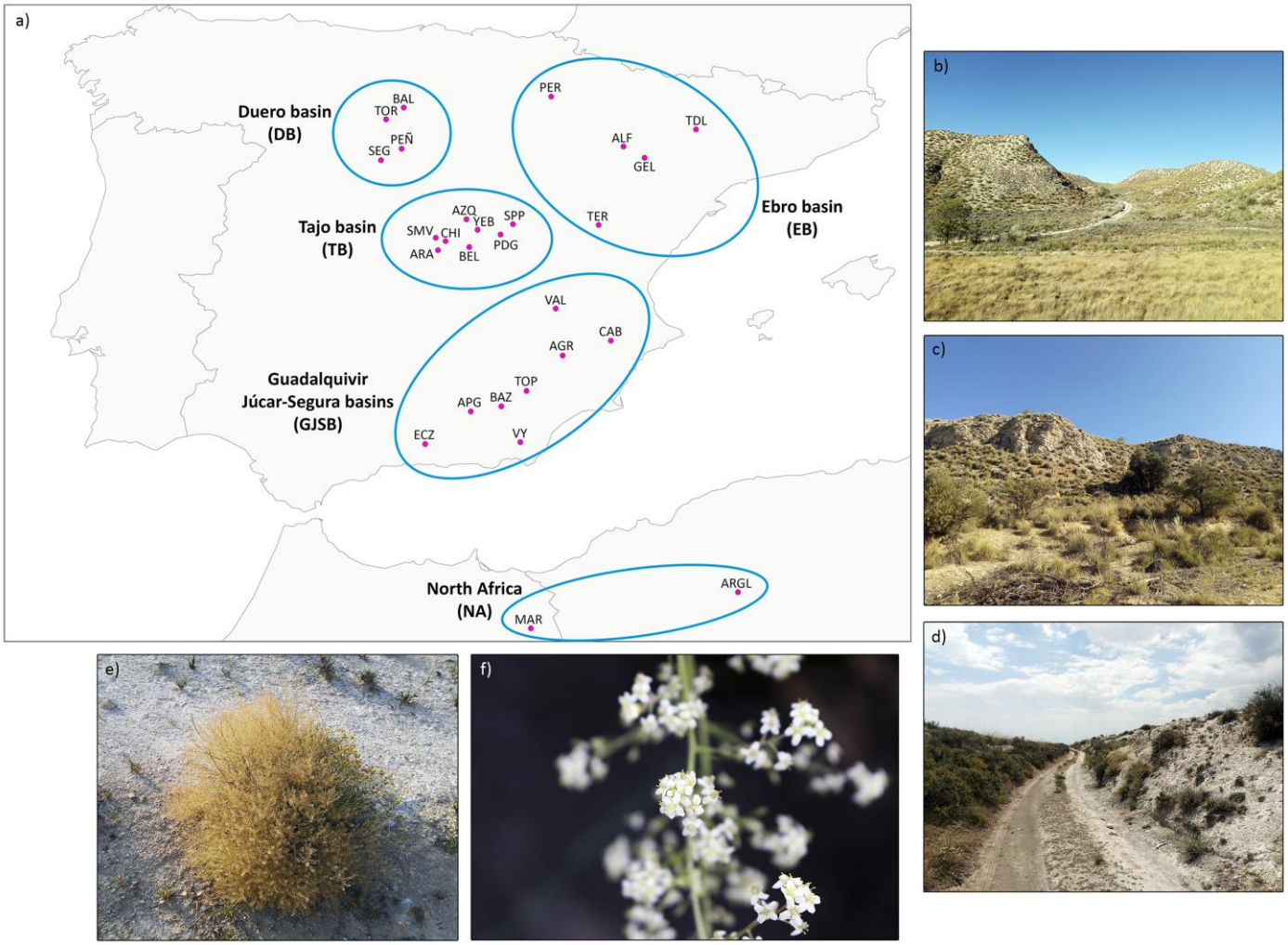

a)
Duero basin
(DB)
BAL
TOR
PEÑ
SEG
Tajo basin
(TB)
AZO
YEB
SPP
SMV
CHI
PDG
ARA
BEL
Ebro basin
(EB)
PER
TDL
ALF
GEL
TER
Guadalquivir
Júcar-Segura basins
(GJSB)
VAL
CAB
AGR
TOP
APG
BAZ
ECZ
VY
North Africa
(NA)
ARGL
MAR
b)
c)
d)
e)
f)

Figure 2

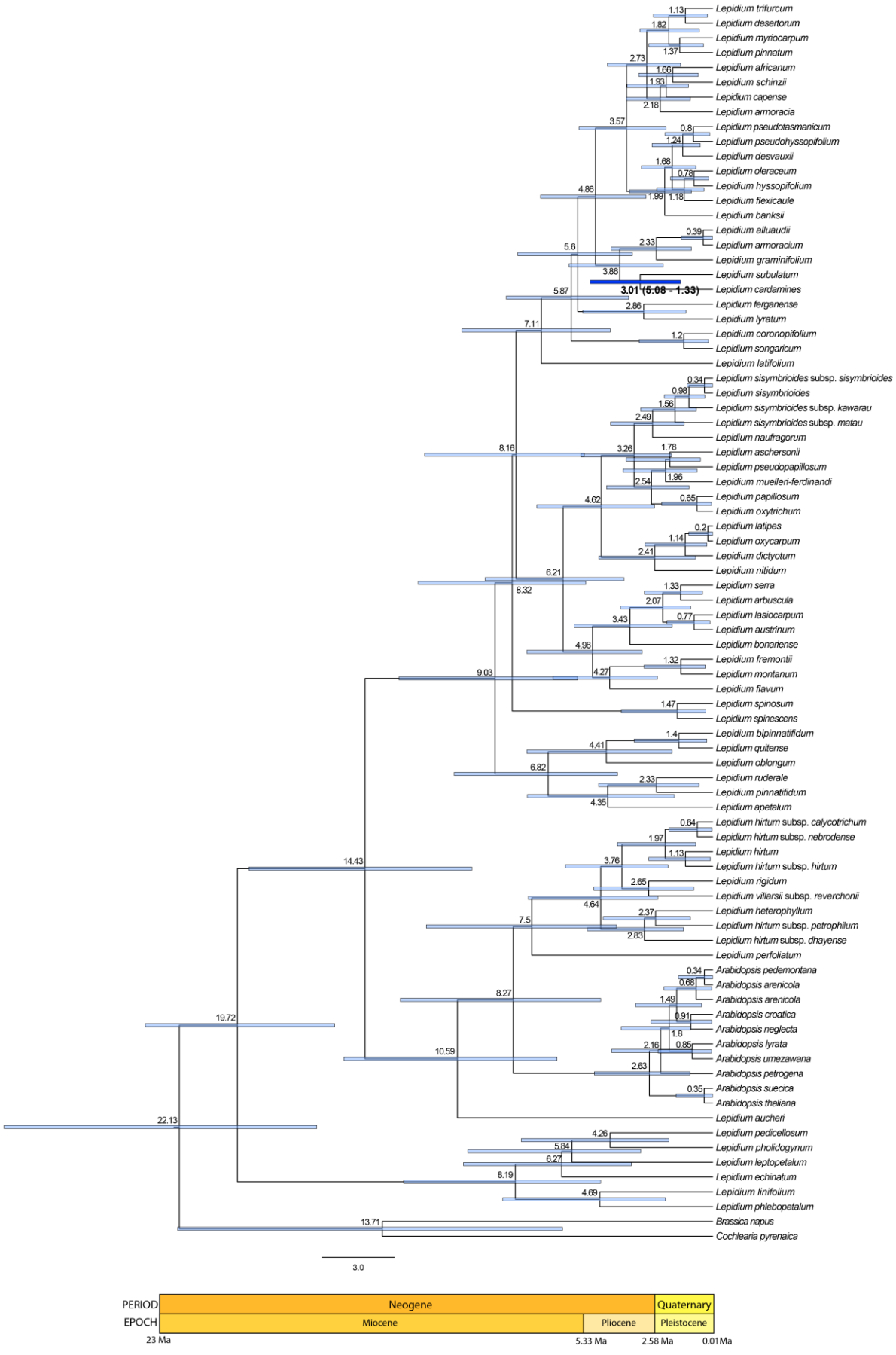

Lepidium trifurcum
Lepidium desertorum
Lepidium myriocarpum
Lepidium pinnatum
Lepidium africanum
Lepidium schinzii
Lepidium capense
Lepidium armoracia
Lepidium pseudotasmanicum
Lepidium pseudohyssopifolium
Lepidium desvauxii
Lepidium oleraceum
Lepidium hyssopifolium
Lepidium flexicaule
Lepidium banksii
Lepidium alluaudii
Lepidium armoracium
Lepidium graminifolium
Lepidium subulatum
Lepidium cardamines
3.01 (5.08 - 1.33)
Lepidium ferganense
Lepidium lyratum
Lepidium coronopifolium
Lepidium songaricum
Lepidium latifolium
Lepidium sisymbrioides subsp. sisymbrioides
Lepidium sisymbrioides
Lepidium sisymbrioides subsp. kawarau
Lepidium sisymbrioides subsp. matau
Lepidium naufragorum
Lepidium aschersonii
Lepidium pseudopapillosum
Lepidium muelleri-ferdinandi
Lepidium papillosum
Lepidium oxytrichum
Lepidium latipes
Lepidium oxycarpum
Lepidium dictyotum
Lepidium nitidum
Lepidium serra
Lepidium arbuscula
Lepidium lasiocarpum
Lepidium austrinum
Lepidium bonariense
Lepidium fremontii
Lepidium montanum
Lepidium flavum
Lepidium spinosum
Lepidium spinescens
Lepidium bipinnatifidum
Lepidium quitense
Lepidium oblongum
Lepidium ruderale
Lepidium pinnatifidum
Lepidium apetalum
Lepidium hirtum subsp. calycotrichum
Lepidium hirtum subsp. nebrodense
Lepidium hirtum
Lepidium hirtum subsp. hirtum
Lepidium rigidum
Lepidium villarsii subsp. reverchonii
Lepidium heterophyllum
Lepidium hirtum subsp. petrophilum
Lepidium hirtum subsp. dhayense
Lepidium perfoliatum
Arabidopsis pedemontana
Arabidopsis arenicola
Arabidopsis arenicola
Arabidopsis croatica
Arabidopsis neglecta
Arabidopsis lyrata
Arabidopsis umezawana
Arabidopsis petrogena
Arabidopsis suecica
Arabidopsis thaliana
Lepidium aucheri
Lepidium pedicellosum
Lepidium pholidogynum
Lepidium leptopetalum
Lepidium echinatum
Lepidium linifolium
Lepidium phlebopetalum
Brassica napus
Cochlearia pyrenaica
3.0
PERIOD
EPOCH
Neogene
Quaternary
Miocene
Pliocene
Pleistocene
23 Ma
5.33 Ma
2.58 Ma
0.01Ma

Figure 3

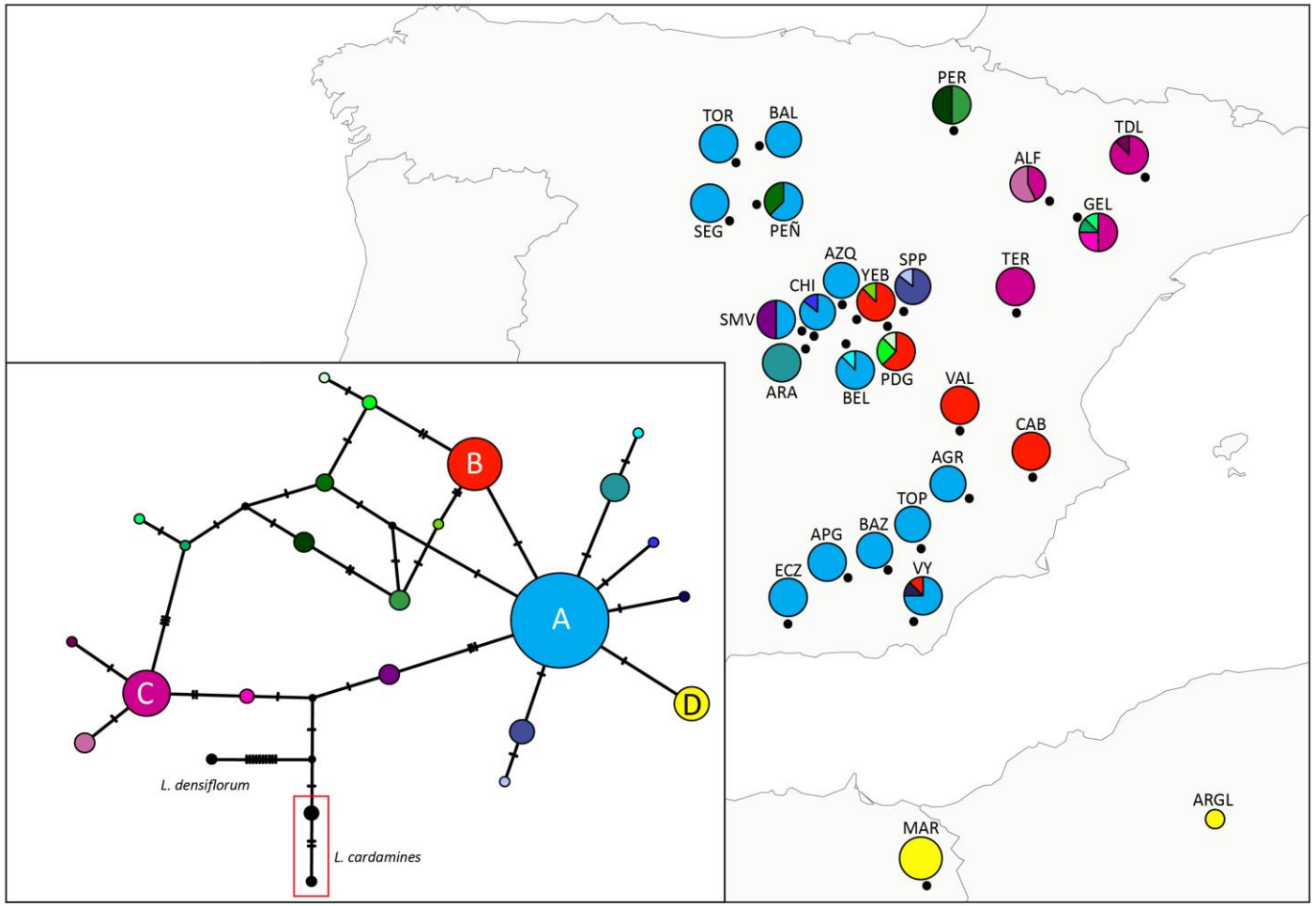

PER
TOR
BAL
TDL
ALF
GEL
SEG
PEÑ
AZQ
SPP
TER
CHI
YEB
SMV
ARA
BEL
PDG
VAL
CAB
AGR
TOP
APG
BAZ
ECZ
VY
B
A
C
D
L. densiflorum
L. cardamines
ARGL
MAR

Figure 4

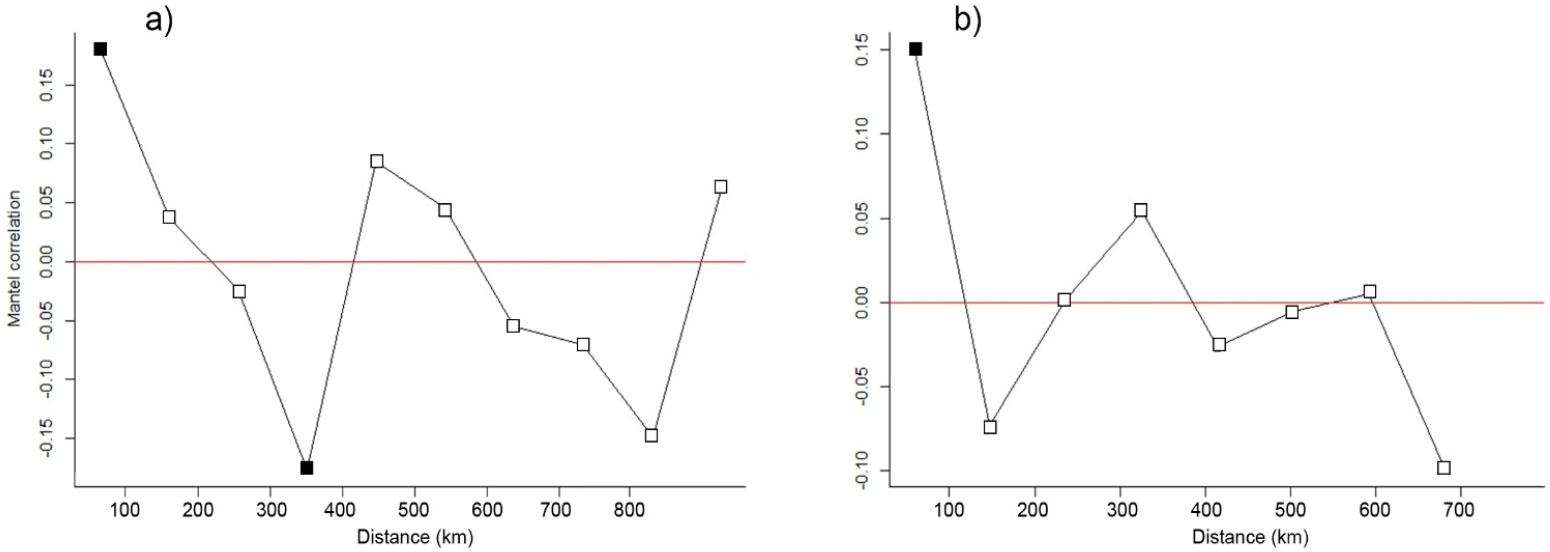
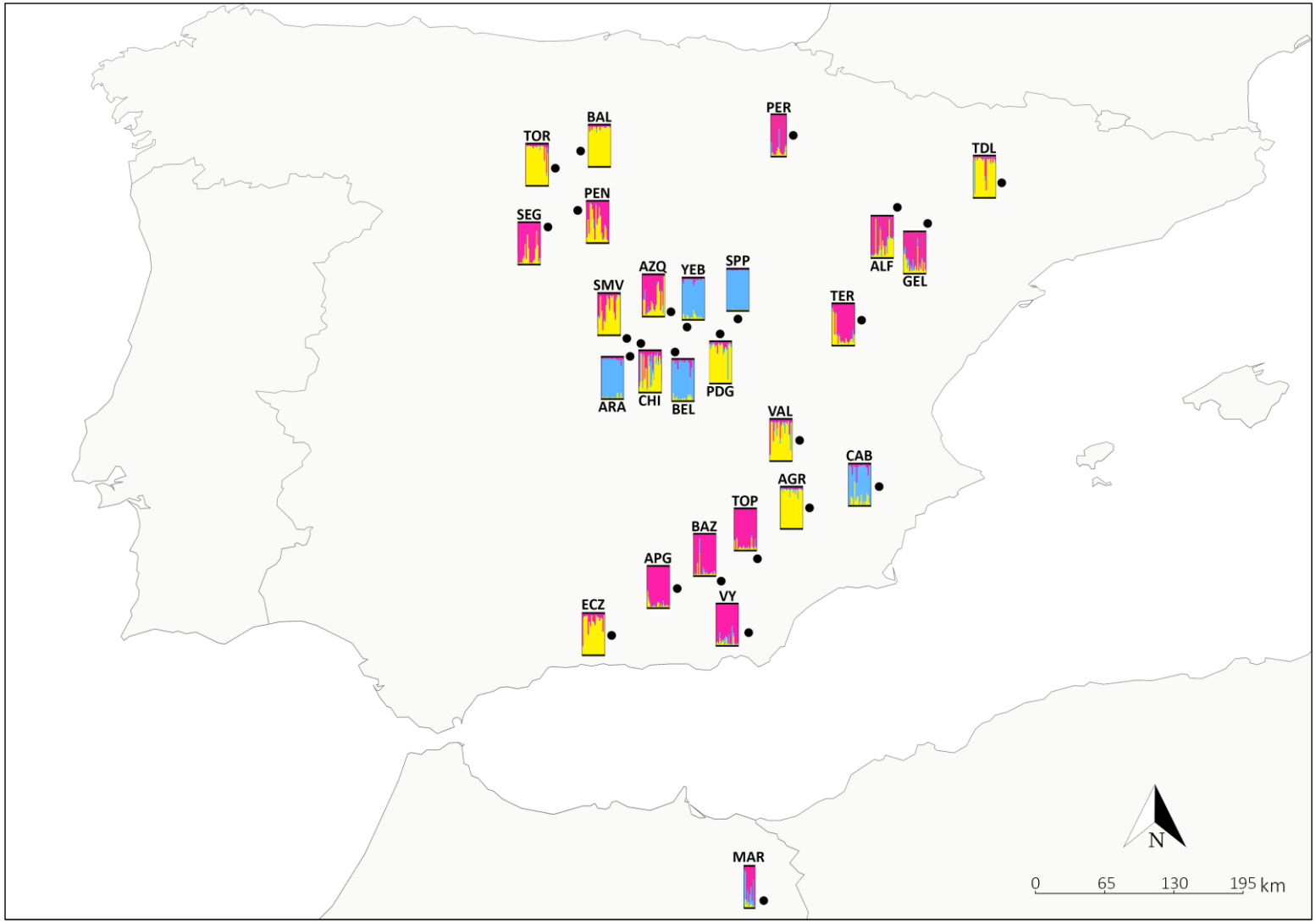

a)
Mantel correlation
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
100
200
300
400
500
600
700
800
Distance (km)
b)
-0.10
-0.05
0.00
0.05
0.10
0.15
100
200
300
400
500
600
700
Distance (km)

Figure 5


TOR
BAL
PER
TDL
SEG
PEN
ALF
GEL
AZQ
YEB
SPP
SMV
TER
ARA
CHI
BEL
PDG
VAL
CAB
AGR
TOP
BAZ
APG
VY
ECZ
MAR
N
0
65
130
195 km