Original article

Title: North-west Africa as a source and refuge area of plant biodiversity: a case study on *Campanula kremeri* and *C. occidentalis*

Sara García-Aloy*[1], Daniel Vitales[1], Cristina Roquet[2], Isabel Sanmartín[3], Pablo Vargas[3], Julián Molero[4], Peris Kamau[5], Juan José Aldasoro[1,6†] and Marisa Alarcón [1†]

[1]Institut Botànic de Barcelona (IBB-CSIC-ICUB), Barcelona, Spain

[2]Laboratoire d'Écologie Alpine, Université Grenoble Alpes, Grenoble, France

[3]Real Jardín Botánico (RJB-CSIC), Madrid, Spain.

[4]Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain

[5]National Museums of Kenya, Botany Department, Nairobi, Kenya

[6]Universidad Rey Juan Carlos, Móstoles, Spain

† equal contributions

[*]Correspondence: Sara García-Aloy, Institut Botànic de Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s/n, Parc de Montjuïc, E-08038 Barcelona, Spain.

E-mail: sara.garcia.aloy@gmail.com

Short running head: The phylogeography of *Campanula kremeri* and *C. occidentalis*

Word count: 6997

Estimate number of journal pages: 12 pages

ABSTRACT

Aim North-west Africa, due to its unique position at the crossroads between Macaronesia and the Iberian Peninsula, has played an important role on the emergence and maintenance of Mediterranean plant diversity. Here, we reconstruct the phylogeographic history of a lineage of bellflowers comprising the north African-south Iberian species *Campanula kremeri* and the Canarian *C. occidentalis* (*Azorina*-group), to investigate the genetic imprints left by past climatic and palaeogeographical events on the northern African flora.

Location North-west Africa, Iberian Peninsula, and the Canary Islands.

Methods We reconstructed the biogeographic history of the *Azorina*-group in order to provide a phylogenetic background. We then investigated phylogeographic patterns within *C. kremeri* and *C. occidentalis* using AFLP and sequence data. We integrated these results with past species-distribution modelling to understand current biodiversity patterns within this lineage.

Results The ancestor of *C. kremeri*-*C. occidentalis* diverged in the Late Miocene/Early Pliocene. Nuclear data supported species monophyly, whereas plastid data suggested that *C. kremeri* is paraphyletic. Maghrebian populations of *C. kremeri* showed high genetic diversity, while Iberian ones and those of *C. occidentalis* exhibited lower values.

Main conclusions Repeated expansion-retraction events associated with Pleistocene climatic changes in North Africa facilitated gene flow across Maghrebian ranges in *C. kremeri*. Mountain massifs in north-west Africa likely acted as refugia for Mediterranean plants during interglacial periods, whereas range expansion in cooler periods triggered dispersal to neighbouring regions. The range of *C. kremeri* expanded to the Iberian Peninsula by long-distance dispersal across the Strait of Gibraltar during the Pleistocene. The relatively old

age inferred for *C. occidentalis* together with its low genetic diversity point to a recent colonization of the Canary Islands from north-west Africa followed by extinction in the mainland.

Keywords: AFLP, annual plants, Campanulaceae, dispersal, genetic admixture, Iberian Peninsula, Macaronesia, phylogeography, species distribution modelling, Strait of Gibraltar.

INTRODUCTION

The Mediterranean region has long been recognized as a plant biodiversity hotspot with a complex evolutionary history (Myers & Haines, 2000). Though several recent studies have focused recently on the evolution of plants from the Mediterranean islands and the southern European peninsulas (Médail & Diadema, 2009; Molina-Venegas *et al.*, 2013), studies on northern African lineages are much scarcer in comparison, and this flora has been little explored to date. This is particularly striking because the northern African flora is very diverse, comprising both narrow endemics and widespread taxa, and a high proportion of annual and autogamous taxa. The latter is probably a consequence of long-lasting disturbance regimes, drought events, and climatic fluctuations in the last three million years (Lavergne *et al*., 2013).

The few studies focusing on northern African plant diversity support the north-west African region as a refuge and a cradle of genetic diversity (e.g. Rodríguez-Sánchez *et al.*, 2008; Désamoré *et al.*, 2011). The heterogeneous relief of the Maghreb region, combined with climatic stability in certain areas, have favoured a high degree of genetic distinctiveness and diversity (Rodríguez-Sánchez *et al*., 2008; Husemann *et al*., 2014). For example, the Baetic-Riffian ranges account for up to 18% of plant species richness in the Mediterranean Basin (Molina-Venegas *et al*., 2013). Some studies focusing on annual plants (Ortiz *et al.*, 2009; Valtueña *et al.*, 2016) have found a phylogeographic pattern in which genetic structure and diversity peak in the northern Maghreb followed by a gradient of decreasing diversity towards southern Europe; others (Fernández-Mazuecos & Vargas, 2011; García-Verdugo *et al.*, 2015) show genetic diversity centres in southern Europe

and Macaronesia (Table S1.1). North-west Africa has been intermittently connected to Southwest Europe through the Strait of Gibraltar. At the end of the Miocene, the Messinian Salinity Crisis (c. 5.9 Ma; Duggen *et al*., 2003) allowed biotic exchange between Europe and Africa across this strait (Fiz-Palacios & Valcárcel, 2013). The Strait of Gibraltar was re-flooded during the Zanclean (5.33-3.60 Ma), but during Pleistocene glacial cycles (2.6-0.10 Ma) sea-level fluctuations (Bintanja *et al*., 2005) likely favoured connections between Africa and Europe through reduced geographic distances. Thanks to their colonization abilities, annual plant species are the best candidates to have repeatedly dispersed between South Europe and North Africa (Lavergne *et al*., 2013). Given this long-history of connection and disconnection events across the Strait, one would expect to see it reflected in multiple waves of colonization and retraction.

In addition to this connection with Southwest Europe, the north-west African flora shows important affinities with the Macaronesian flora. The north-west African mainland has been likely the source of colonization events in many Macaronesian lineages (Sanmartín *et al*., 2010). Macaronesia has also probably played a role as a refuge during dry periods; consequently patterns of both regions are often interconnected (Mairal *et al*., 2015).

Here, we reconstruct phylogeographic patterns within a lineage of bellflowers (*Campanula*, Campanulaceae), the sister-group formed by the western Atlanto-Mediterranean species *Campanula kremeri* and *C. occidentalis*, in order to investigate phylogeographical connections among the Maghreb region, the Iberian Peninsula and Macaronesia. To provide a phylogenetic background for this lineage, we first reconstruct phylogenetic relationships and the biogeographic history of the more inclusive *Azorina*-group (Alarcón *et al*., 2013), a lineage of *Campanula* that includes c. 23 species, most of them distributed in north Africa and including *C. kremeri*-*C. occidentalis*. We then use DNA sequences, AFLP, and species distribution modelling to reconstruct phylogeographic patterns in two annual sister-species: *C. kremeri*, endemic to southern Iberian Peninsula and north-west Africa, and *C. occidentalis*, occurring in the Canary Islands. Our aims are to: (1) Infer the imprints left by past climatic and palaeogeographical events on the genetic structure

of *C. kremeri*-*C. occidentalis* as a case study of the north-western African flora. (2) Detect geographic barriers, colonization events and potential refugia that shaped current genetic diversity patterns in this lineage. (3) Examine whether the Strait of Gibraltar acted as an effective geographic barrier to gene flow for *C. kremeri*, with special attention to possible waves of colonization across this strait.

MATERIALS AND METHODS

Study group and sampling

To clarify phylogenetic relationships within the *Azorina*-group, 29 species were included in the dataset: 23 species of the *Azorina*-group, plus three Asian relatives (*C. cashmeriana, C. dimorphantha, C. lehmanniana*), two species of the subgenus *Roucela* (*C. creutzburgii* and *C. drabifolia*), and *C. sibirica*, which were used as outgroups according to Olesen *et al*. (2012).

In order to investigate phylogeographic patterns in *C. kremeri* and *C. occidentalis*, we sampled 16 populations (134 individuals in total), spanning the geographical range of each species: for *C. kremeri*, we sampled three populations in the Iberian Peninsula and eight in the Maghreb region (three in the Rif and five in the Atlas); for *C. occidentalis*, one population in Lanzarote, one in Fuerteventura and three in Tenerife (Tables 1 & S1.2).

DNA sequencing

Total DNA was extracted from silica gel-dried plant tissue using the "DNeasy Plant Mini Kit" (QIAGEN Inc., California, USA) according to the manufacturer's instructions. To reconstruct phylogenetic relationships within the *Azorina*-group, four highly variable cpDNA regions were sequenced: *pet*B-*pet*D (885 bp), *rpl32*-*trn*L (484 bp), *trn*S-*trn*G (737 bp) and *trn*L–*trn*F (892 bp). We also sequenced ITS (579 bp) and the low-copy-nuclear gene PPR11 (639 bp). The phylogeographic history of *C. kremeri*-*C occidentalis* was reconstructed with the same markers, employing one individual per population. Finally, a within-population study was

performed in this lineage using only plastid haplotypes; for this, we chose the two most variable regions, *rpl32-trn*L and *trn*S-*trn*G, generating 134 sequences for each marker. In total, we generated 392 cpDNA and 91 nrDNA sequences. Sources of material, location of vouchers, amplification parameters, GenBank accessions and full references are detailed in Appendix S1 (Tables S1.2 & S1.3).

Phylogenetic and biogeographic analyses in the *Azorina* clade

Sequences were aligned with MAFFT (Katoh *et al*., 2005) and checked by eye. JMODELTEST 2.2 (Posada, 2008) was used to determine the best-fitting model of sequence evolution using the Akaike information criteria (AIC): GTR + Γ for plastid dataset and GTR for ITS and PPR11. Phylogenetic analyses were performed on separate and concatenated matrices of the four plastid and two nuclear regions. Divergence times between lineages were estimated with BEAST 1.7.5 (Drummond & Rambaut, 2007), using a relaxed clock method with uncorrelated rates drawn from a lognormal distribution; a calibration time of 21 Ma was assigned to the split between *Campanula sibirica* and the *Azorina*-group (root node), based on Olesen *et al*. (2012), and using a normal-distributed prior (mean = 21 Ma; Standard Deviation = 2). The plastid dataset and the nuclear markers (ITS and PPR11) were treated as three separate partitions, with the substitution and clock models unlinked. A birth-death prior was used as the speciation model, with four runs of $5x10^7$ generations each, sampling every 1000th generation. Resulting posterior distributions for parameter estimates were checked in TRACER 1.4.1 (Drummond & Rambaut, 2007) and maximum credibility (MCC) trees were calculated after removing a burn-in of 20% with TREE ANNOTATOR 1.6. The MCC tree inferred from BEAST was used as input to reconstruct the spatio-temporal evolution of the group, using the R package 'BioGeoBEARS' (Matzke, 2014): eight distribution areas (Azores Islands; Asia and Balkans; Canary Islands; Cape Verde Islands; Central Africa; East Africa, Arabia and Socotra; Iberian and Italian Peninsulas, Sicily and Balearic Islands; north-west Africa) were defined; all available models in BioGeoBEARS were fitted and compared based on AIC; final biogeographic reconstruction was performed with the model that yielded the lowest AIC (DEC+J).

145

146 **Phylogenetic study of *C. kremeri-C. occidentalis* lineage**

147 Phylogenetic relationships within the *C. kremeri-C. occidentalis* lineage were reconstructed with nuclear and
148 plastid markers. As the nuclear and plastid markers were incongruent for this clade (ILD test, Farris *et al*.,
149 1995, *p*-value < 0.05; ITS was congruent with PPR11, $P = 0.49$), we performed independent BEAST analyses
150 for the plastid and nuclear datasets sampling one individual per population of *C. kremeri* and *C. occidentalis*
151 (11 and 5 populations, respectively). Parameters employed were the same as in the *Azorina*-group BEAST
152 analysis, but we used a calibration time of 13 Ma (SD = 1.5) assigned to the split between *Campanula*
153 *dimorphantha* and the *Azorina*-group (the root-node in this analysis). Some species of *Campanula* can
154 hybridize (Nyman, 1991) and we obtained a different topology from the nuclear and the plastid datasets. We
155 consequently built a rooted phylogenetic network using DENDROSCOPE 3 (Huson & Scornavacca, 2012) to
156 search for hybridization events. We used two sets of 1000 trees randomly sampled from the BEAST analyses of
157 the nrDNA and cpDNA datasets and computed a level-k network, minimizing the number of reticulations in
158 any biconnected component of the network. Reticulations were limited to the branches present in 95% of the
159 trees of each dataset, i.e., 47.5% threshold. Additionally, we used *BEAST (Heled & Drummond, 2010) and a
160 similar approach to Blanco-Pastor *et al*. (2012) to explore the incongruence found between cpDNA and
161 nrDNA phylogenies. We constructed a multi-labelled species tree to retrieve the origin of the ancestor
162 lineages of clades affected by reticulation processes; this multi-labelled tree was assembled by assigning
163 assumed sequences from the putative hybrids to separate labels: the nuclear sequences to one label (N), and
164 the plastid ones to the other (P). Hence, two labels of a potential hybrid (N and P) were treated as different
165 species in the *BEAST analysis; this ensured that putative reticulations were analysed without violating the
166 assumptions of the multispecies coalescent *BEAST model. Four Markov chain Monte Carlo (MCMC)
167 analyses were run for 10^8 generations, sampling every 1000^{th} generation, using the same settings as above,

and with *C. dimorphantha* and *Azorina vidalii* as outgroups. A DENSITREE plot (Bouckaert, 2010) was used to summarize all possible topologies.

Haplotype and DPA analyses

The evolutionary history of the *C. kremeri*-*C. occidentalis* lineage was further investigated using a concatenated haplotype dataset including only the regions *rpl32-trn*L and *trn*S-*trn*G sequenced for all individuals of these two species. Haplotype diversity (Hd), nucleotide diversity (π), Tajima D, Fu's Fs and other genetic parameters were calculated with DNASP 5.1 (Librado & Rozas, 2009) without considering gaps (Tables 1 & S1.4). Relationships among plastid haplotypes were inferred in TCS 1.21 (Clement *et al*., 2000) with gaps as missing data. The Bayesian discrete phylogeographic analysis (DPA) of Lemey *et al*. (2009), implemented in BEAST, was used to trace the history of migration events. DPA uses a continuous-time Markov Chain process, in which the discrete states correspond to geographic locations of sequences and transition rates between states to migration rates between areas (Mairal *et al*., 2015). Three areas were defined (Iberian Peninsula, North Africa and Canary Islands), with between-areas migration rates and the geodispersal rate scaler modelled using default gamma-prior distributions. We used the BSSVS method and SPREAD 1.0.6 (Bielejec *et al*., 2011) to identify the most likely diffusion routes. Bayes factors comparisons (BFs) with a cut-off value of 3 were used, and results visualized into a KML file. For the two nuclear markers, we inferred median networks using NETWORK 4.2.0.143 (Bandelt *et al*., 1999), employing the same individuals as in the previous phylogenetic analyses.

AFLP genotyping and analyses

To explore the population genetic structure raised by the nuclear compartment, we carried out AFLP genotyping analyses (Vos *et al*., 1995). We used the AFLP Plant Mapping Kit (Applied Biosystems); genomic DNA was digested with the enzymes EcoRI and MseI and linked to adaptors. Thirty-two combinations of

selective primers were tested, and four were retained that showed clear and evenly distributed bands and polymorphic profiles: 1- EcoRI6-FAM-ACC/ MseI- CCT; 2- EcoRI6-FAM-ACT/ MseI- CAC; 3- EcoRIVIC-AGG/ MseI-CAA, and 4- EcoRIVIC-AGG/ MseI-CAC. For each sample, 0.3 μL of 6-FAM-labelled and VIC-labelled selective PCR products were combined with 0.5 μL of GeneScan 500 LIZ and 13.5 μL of formamide. Fragment electrophoresis was conducted at PCM (Spain) using ABI 3730 capillary sequencer.

Amplified fragments were analysed using GENEMAPPER 3.7 (Applied Biosystems), and peaks ranging between 100 and 500 base pairs were recorded. We estimated error rates with AFLP SCORER (Whitlock *et al*., 2008), and fixed them to 5% for each primer combination. A total of 796 fragments were scored. Data reliability was assessed by comparison of duplicates (26 tests). The reproducibility obtained was 91–100%, with a mean of 95.8 %. Based on the AFLP presence/absence matrix, the number of private fragments per population or group of populations was recorded (Table 1).

We estimated Nei's gene diversity (Hj), F_{ST}, the percentage of polymorphic fragments per individual (P) (Nei & Li, 1979) with AFLPSURV 1.0 (Vekemans, 2002), assuming partial self-fertilisation and Hardy-Weinberg equilibrium. We estimated allelic frequencies with a Bayesian method, employing non-uniform prior distribution. We calculated F_{ST} with 10,000 permutations. Neighbour-nets were inferred using SPLITSTREE 4.10 (Huson & Bryant, 2006). We quantified the amount of genetic differentiation of population groups using a hierarchical analysis of molecular variance (AMOVA) using ARLEQUIN 3.0 (Excoffier *et al.*, 2005; see Table S1.5).

We assessed the structure of populations with STRUCTURE 2.2 (Pritchard *et al*., 2000), assuming admixture and uncorrelated allele frequencies between groups. We ran 500,000 generations (burn-in of 100,000), for K values from one to six, with ten repetitions, considering only those runs with the highest likelihoods values and we used the LnP (D) measure for the successive decomposition of groups (Evanno *et al*., 2005). BARRIER 2.2 (Manni & Guérard, 2004) was used to identify possible geographic locations acting as major genetic barriers; significance was tested with 1000 bootstrapped distance matrices. To test the effect of

the spatial distance in the genetic structure found in the AFLP analysis, we used the Mantel test (NTSYS 2.1; Rohlf, 1998) to correlate the genetic (as F_{ST}) and spatial distances within the main lineages derived from our results.

Species distribution modelling

Species distribution modelling was performed to infer the potential distributions of *Campanula kremeri* and *C. occidentalis*, under present climatic conditions and late Quaternary conditions (Last Inter-glacial period LIG, and Last Glacial Maximum LGM). The occurrence datasets comprised 29 localities for *C. kremeri* and 15 localities for *C. occidentalis* (Table S1.6). We employed the maximum entropy algorithm as implemented in MAXENT 3.3 (Phillips, 2006). We retrieved 19 bioclimatic variables from the WorldClim website (Hijmans *et al*., 2005) which were clipped to cover the Iberian Peninsula, Maghreb and the Canary Islands. Highly correlated variables ($r > 0.7$) were reduced to seven uncorrelated variables used as predictors to calibrate the distribution models in MAXENT. To test model predictive performance, we split localities into training (75%) and test data (25%), with ten subsample replicates. The distribution model under current conditions was projected to two time slices of the late Quaternary: the LIG (c. 80 ka), model of Otto-Bliesner *et al*. (2006), and the LGM (c. 21 ka) under two models: the Community Climate System Model (CCSM; Collins *et al*., 2006) and the Model for Interdisciplinary Research on Climate (MIROC; Hasumi & Emori, 2004).

RESULTS

Dated phylogeny and biogeography of the *Azorina*-group

The combined plastid-nuclear dataset resulted in a sequence alignment of 4436 bp (*pet*B–*pet*D, *rpl32*–*trn*L, *trn*S–*trn*G, *trn*L–*trn*F, *ITS*, and *PPR*11). Six Asian species were reconstructed as outgroups to the *Azorina*-group, formed by *Azorina vidalii* and two major subclades: clade I formed by *C. kremeri* and *C. occidentalis;* and clade II containing the remaining species (Figs 1, S2.1, S2.2). The split between *A. vidalii* and the

ancestor of clades I and II was estimated at 8.2 Ma (95%HPD 7.52–12.46), while the divergence time of clades I and II was dated at 6.9 Ma (95%HPD 4.96–9.51). The biogeographic reconstruction with the DEC+J model supported an ancestral Asian *Campanula* that dispersed towards Africa (Fig. 1). For the ancestor of clades I and II, the reconstruction with the highest probability corresponds to north-west Africa, followed by the colonization by several lineages of Macaronesia, Arabia and the southern European Peninsulas (Fig. 1). The most-recent-common-ancestor of clade I (*C. kremeri* and *C. occidentalis*) was distributed in north-west Africa (Fig. 1), further diverging into three main lineages divided by their geographic distribution: (1) Canary Islands; (2) eastern Rif and Atlas; and (3) western Rif and Iberian Peninsula (Fig. 2a).

Phylogenetic study of the lineage *C. kremeri-C. occidentalis*

The phylogenetic analyses with an extended population sampling within *C. kremeri-C. occidentalis* group depicted significant incongruence among the plastid and the nuclear compartments. According to the nuclear dataset, this lineage split into two clades that match the delimitation of the two species (*PP* > 0.99; Figs 2a & S2.1). In contrast, the plastid dataset led to a different topology, showing *C. kremeri* as paraphyletic, with C. *occidentalis* nested within (*PP* > 0.99; Figs 2a & S2.2). The DENDROSCOPE k-level phylogenetic network showed one significant reticulation event within *C. kremeri* (Fig. 2b), between the lineage comprising Atlas populations and the lineage leading to the western Rif populations. The multilabelled *BEAST analysis, with plastid and nuclear sequences from the Atlas labelled as different lineages, resulted in a tree topology in which all posterior probabilities were higher than 0.95 (Fig. 2c). All possible topologies as recovered by the DENSITREE plot (Fig. S2.3) suggested this same reticulation event.

Phylogeographic patterns in *C. kremeri* and *C. occidentalis*

The plastid haplotype network analysis (Fig. 3b) showed differences in 40 nucleotide positions. We detected 11 haplotypes, ten within *C. kremeri* and one in *C. occidentalis*. Haplotypes 1–2 were exclusive of

264 populations ER1 and AT1; haplotypes 3–6 were found only in AT2-5. Haplotype 3 had the highest number of
265 connections. Haplotype 7 was exclusive to IP2, while haplotype 8 was shared by all populations in the Iberian
266 Peninsula and the western Rif (WR2). Haplotypes 9–10 were exclusive to the western Rif. Haplotype 11 was
267 the only one found in the Canary Islands. Our data rendered a Hd considerably higher in North Africa than in
268 the Iberian Peninsula and the Canary Islands (Tables 1 & S1.4). For some populations (e.g. AT2 & ER1;
269 Table 1), we could not obtain sequences for all sampled individuals. This unevenness in the molecular data
270 might have reduced the power to detect intra-population variability within those populations. Nevertheless, the
271 results from DNA sequencing were consistent with those from the AFLP analysis (see below).

272 The DPA analysis based on *rpl32–trn*L and *trn*S–*trn*G (Fig. 3a inside the square) showed a first
273 divergence event, corresponding to the separation between two clades: 1) all accessions from the eastern Rif
274 and the Atlas (*PP* = 1); and 2) with a lower support (*PP* = 0.88), the accessions from the Canary Islands,
275 Iberian Peninsula, and western Rif. Three migration routes were supported by the BF test: one colonization
276 event from North Africa to the Iberian Peninsula, one possible re-colonization from the Iberian Peninsula to
277 Africa, and one dispersal event from continental Africa to the Canary Islands (Fig. 3a). In contrast, the nuclear
278 network showed a central network formed by the eastern Rif, the Atlas and the western Rif, with two long
279 branches: one connecting with the Iberian haplotype, and a second connecting the two Canarian haplotypes
280 (Fig. 3c).

281 For the AFLP analysis, estimates of the Nei's gene diversity (Hj), the percentage of polymorphic
282 fragments per individual (P), and the number of private fragments (Np) were high in populations of eastern
283 Rif and Atlas, and considerably lower in the Iberian Peninsula and Canary Islands (Table 1). STRUCTURE
284 indicated that the number of optimal groups was K = 3 (Fig. S2.4), clustering the populations into three
285 groups: 1) Canary Islands; 2) Atlas and eastern Rif; and 3) western Rif and Iberian Peninsula (Fig. 4). These
286 groups were highly congruent with the BARRIER analysis, revealing two major boundaries separating the same
287 groups. The results for K = 4 and K = 5 were very similar (Fig. S2.5), though the Atlas cluster is further

288 divided into two clusters. F_{ST} values were consistent with the results obtained with the chloroplast-only
289 dataset, showing the same genetic cohesions between populations. The Mantel test showed a significant
290 correlation between genetic and geographic distance only for the Baetic-western Rif (r = 0. 734, $^{**}P$ = 0.002)
291 and the Atlas-eastern Rif areas (r = 0.577, $^{**}P$ = 0.018), but not for *C. occidentalis* (r = 0. 018, P= 0.46).
292 Hierarchical AMOVA showed the largest proportion of genetic variation for groups 3 and 5 (Table S1.5); the
293 3-group results were consistent with those of STRUCTURE (Figs 4). The Neighbor-net diagram obtained from
294 AFLP data (Fig. 4) was in line with the results of the median network based on nuclear DNA sequences (Fig.
295 3c): five major clusters, with Iberian and Canarian samples sorted into more differentiated groups, while
296 north-west African populations showed considerable admixture among them.
297
298 **Species distribution modelling**
299 *Campanula kremeri* is distributed in wet-eroded sites at the foothills of western Maghrebian and southern
300 Iberian ranges, while *C. occidentalis* grows in rather similar habitats in the Canaries. The first species is
301 abundant in the Rif and Middle Atlas and infrequent in the High Atlas (Table S1.6). We obtained distribution
302 models with high predictive accuracy for both species according to the area under the curve (AUC = $0.952 \pm$
303 0.051 and 0.994 ± 0.003, respectively). The main predictor variable for *C. kremeri* was the precipitation of
304 the wettest quarter (bio16), whereas for *C. occidentalis*, it was the temperature annual range (bio7). The
305 predicted current distribution of both species was consistent with their known distribution range (Fig. S2.6)
306 and similar to the potential distribution during LIG. In contrast, the LGM projections (CCSM and MIROC,
307 Fig. S2.6) revealed a larger potential distribution of both species.
308
309 **DISCUSSION**
310 **North-west Africa as a hub of diversification in Mediterranean plants**

311 According to our results, the *Azorina*-group within *Campanula* originated from an ancestor that dispersed
312 from West Asia to North Africa in the Late Miocene (c. 8.7–13.3 Ma) which agrees with previous studies
313 (Roquet *et al*., 2009; Alarcón *et al*., 2013). Therefore, Afro-Macaronesian bellflowers would have needed to
314 adapt to the incipient aridification of North Africa during this period (Zhang *et al*., 2014), which might have
315 been possible thanks to a combination of physiological and morphological features, such as autogamous or
316 facultative reproductive systems, short lifespan, easily dispersible seeds, and the ability to cope with a variety
317 of disturbed habitats and substrates. Importantly, biogeographic analyses suggest that most basal divergence
318 events within the *Azorina*-group involved north-west Africa as ancestral area, and date back to the Late
319 Miocene-Pliocene. From this area, several dispersal events to nearby adjacent regions (Central Africa; Eastern
320 Africa, Arabia and Socotra; Macaronesia; and southern European Peninsulas; Fig. 1) took place. This role of
321 north-west Africa as a source area of dispersal events and a "hub of diversification" for Western
322 Mediterranean plants mirrors the pattern found in animals (Husemann *et al*., 2014) and plants (Valtueña *et al*.,
323 2016). One explanation for this role is the high topographic complexity, with the Atlas and Rif mountain
324 ranges allowing genetic isolation among populations and allopatry. Another is range-shifts during Pleistocene
325 glacial cycles, which would have favoured both secondary contacts and subsequent isolation.
326
327 **Evolutionary origins of *Campanula kremeri* and *C. occidentalis***
328 The evolutionary origins of *Campanula kremeri* and sister-species *C. occidentalis* are an interesting case study
329 to investigate plant evolutionary dynamics in north-west Africa. Phylogeographic analyses suggest that the
330 disjunct distribution of this lineage originated from dispersal events out of north-west Africa to nearby
331 regions, the Iberian Peninsula and Canary Islands. Specifically, the plastid network presents one haplotype
332 (H8) shared by the Iberian populations of *C. kremeri* and those inhabiting the western Rif Mountains (Figs
333 3a–b), while AFLP data supports a link between these two groups of populations (Fig. 4). Thus, the disjunct
334 distribution across the Strait of Gibraltar observed in *C. kremeri* is probably the result of relatively recent

colonization from the Maghrebian massifs during Pleistocene climatic oscillations. Low genetic diversity and the small number of private fragments found in the Iberian populations of *C. kremeri* also agree with this hypothesis. Regular wet winds (Dorman *et al.,* 1995) and sea level drops (c.150 m, Lambeck *et al.,* 2002) could have facilitated the dispersal of the dust like seeds of *C. kremeri* between the Rif and the Iberian Peninsula. Our study thus agrees with the hypothesis that the Strait of Gibraltar is not an impermeable barrier for annual species, in contrast to perennial ones, which show stronger genetic breaks (Rodríguez-Sánchez *et al*., 2008).

In addition to the role of Maghrebian massifs as a source for dispersal events, phylogeographic analysis of plastid and nuclear markers indicate that they could have acted as climatic refuges preserving population genetic diversity during past climatic fluctuations. According to the central-marginal hypothesis (Eckert *et al*., 2008), genetic diversity and structure should be higher in areas that constitute refugia because of the preservation of genotypes that went extinct in other areas and the long-term persistence of populations (Hewitt, 2000). The number of private fragments (an indicator of population persistence in isolation) and our haplotype diversity measures are higher for the Atlas and Rif populations of *C. kremeri* (Table 1). In these high-elevation regions, climatic shifts could have been compensated by vertical migration. Species distribution modelling for northern African populations of *C. kremeri* predicts changes in geographic range between glacial and inter-glacial periods during the Pleistocene (Fig. S2.6), i.e., a general range contraction during warm inter-glacial phases that did not affect the mountain ranges close to the Strait of Gibraltar. Pleistocene range contraction-expansions in the Atlas and Rif mountain ranges have also been suggested to explain the maintenance of a geographic genetic structure in other north-west African taxa (Médail & Diadema 2009; Husemann *et al*., 2014), and in agreement with the “refugia within refugia" hypothesis (Gómez & Lunt, 2007). In sum, our study suggests that the topographic heterogeneity of north-west Africa played a key role for both the emergence and maintenance of genetic biodiversity in these lineages. The most ancient range cores for *C. kremeri* – the mountains of the Rif and Middle Atlas – likely acted as climatic

359 refuges in the successive climatic crisis that followed the Late Tertiary climate cooling (Médail & Diadema,
360 2009; Molina-Venegas *et al*., 2013), preserving population genetic diversity and constituting the source of
361 founder events to nearby regions.
362
363 **Dispersal to the Canary Islands and climatic extinction**
364 Divergence between *C. kremeri* and *C. occidentalis* is dated c. 3.7 Ma (2.3–7.4; Fig. 1), indicating an early
365 dispersal event in the Late Miocene-Pliocene. Yet, this old stem-age contrasts with the surprisingly low
366 genetic diversity values found among populations within this species. A possible explanation – suggested for
367 other North Africa-Macaronesian disjunct bellflowers (Mairal *et al*., 2015) – is that the ancestor of the current
368 populations of *C. occidentalis* became isolated in parts of the Atlantic coast of Maghreb with a Macaronesian
369 type of climate after the Late Miocene aridification of the continent (Fig. S2.6; Médail & Diadema, 2009), and
370 they would have dispersed only recently to the Canarian archipelago. These putative continental ancestors
371 would have later gone extinct (Mairal *et al*., 2015), or may still persist in unknown locations in these poorly
372 explored regions. This hypothesis is in agreement with our SDM reconstructions of the potential range of *C.*
373 *occidentalis* during the LGM: two areas of the Atlantic coast relatively close to the Canary Islands were
374 depicted as harbouring high climate suitability for this species during the Pleistocene LGM.
375 The strikingly lower genetic diversity found in the Canarian *C. occidentalis* compared to the North
376 African populations of *C. kremeri* (Table 1) disagrees with other plant population studies, reporting higher
377 genetic diversity in Macaronesia than in north-west Africa (e.g. García-Verdugo *et al*., 2015). Depauperate
378 genetic diversity in island plants compared to mainland species has been explained by multiple factors,
379 including biological traits such as poor dispersal capacity and long generation times, physical characteristics
380 of the archipelagos, or shorter times between founder events (Stuessy *et al*., 2014). In contrast, both *C.*
381 *kremeri* and *C. occidentalis* show efficient dispersal mechanisms and a high-selfing capacity (unpublished
382 data), which could have contributed to the successful colonization of the Iberian Peninsula and the Canaries.

Though we found no significant signature of a bottleneck in *C. occidentalis* (Table S1.4), these traits could have allowed species to recover from genetic bottlenecks derived from founder events, permitting the subsequent accumulation of genetic variability (Stuessy *et al*., 2014).

An interesting result of our analysis is the conflicting phylogenetic signal between the nuclear and plastid genomes for the position of *C. occidentalis*. Whereas the nuclear phylogeny shows an early split between a monophyletic *C. occidentalis* and the *C. kremeri* clade, the plastid markers show *C. occidentalis* embedded within *C. kremeri* (Fig. 2a). Given the low level of gene flow inherent in plant plastid transmission (Wolfe *et al*., 1987), phylogenetic history explained by the plastid compartment could predate the signal of an admixed nuclear compartment. Assuming maternal inheritance of plastids, the maternal progenitor lineage of the populations of *C. kremeri* from Atlas-eastern Rif could be the ancestor of all remaining populations–including the Canarian *C. occidentalis*. In contrast, the paternal progenitor could be affected by the reticulation of *C. kremeri* from Atlas-eastern Rif with *C. kremeri* from the western Rif (Fig. 2b). The haplotype networks (Figs 3b–c) and the AFLP data suggest also introgression among the Atlas, eastern Rif, and western Rif populations (Figs 4 & S2.5). Though we cannot exclude incomplete lineage sorting (ILS) in the nuclear compartment as alternative explanation, it should be noted that the AFLP results support also the introgression hypothesis; this technique is presumed to be more robust to ILS because of the numerous independently transmitted loci (Avise, 2004). Interestingly, these results (together with the species distribution modelling) indicate that genetic barriers among populations in different North African mountain ranges of *C. kremeri* have not always been completely impermeable, likely favouring a higher genetic diversity, which in turn might have mitigated the risk of genetic bottlenecks during retraction periods.

CONCLUSIONS

Our study shows that the mountain ranges of the Rif and Middle Atlas acted as climatic refugia for *Campanula kremeri*, from where this species colonized other areas in north-west Africa, the Iberian Peninsula,

and the Canary Islands. Repeated expansion–retraction cycles favoured gene flow across north-west African ranges and led to local accumulation of genetic variability; and the Strait of Gibraltar acted as a semi-permeable geographic barrier for *C. kremeri*. The relatively old age and low genetic diversity found in the Canarian endemic *C. occidentalis* suggest a recent dispersal origin from an ancestral population in the north-west African coast, further supported by a predicted range reduction during the LGM in species distribution models. Comparative phylogeographic studies on other endemic Mediterranean species with distribution in north-west Africa are needed to corroborate the conclusions reached here: (1) the role of the mountain massifs in north-west Africa as both climatic refugia for Mediterranean and Macaronesian plants and sources of colonization events during Pleistocene climatic oscillations; and (2) the partial permeability of the Strait of Gibraltar for annual plant species, which likely favoured genetic differentiation .

ACKNOWLEDGEMENTS

This research was supported by the Spanish government through projects CGL2009-13322-C03-01, CGL2012-40129-C02-01 and CGL2015-67849-P (MINECO/FEDER) to I. Sanmartín, and CGL2009-10031 to P. Vargas and GREB. For laboratory assistance, we thank E. Cano and F. Durán (RJB-CSIC), and M. I. García (Unidad de Genómica, Parque Científico de Madrid). We are especially grateful for their help with sampling to L. Saéz, A. Rovira, R. Vilatersana, J. Fuertes, A. Hipold, S. Sholtz, and the herbaria of BCN, BC, EA, MA, and W. M. Alarcón was funded by the JAE-Doc program (CSIC/FSE).

REFERENCES

Alarcón, M., Roquet, C., García-Fernández, A., Vargas, P. & Aldasoro, J.J. (2013) Phylogenetic and phylogeographic evidence for a Pleistocene disjunction between *Campanula jacobaea* (Cape Verde Islands) and *C. balfourii* (Socotra). *Molecular Phylogenetics and Evolution*, **69,** 828–836.

Avise, J.C. (2004) Molecular markers, natural history, and evolution. 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Bandelt, H.J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*, **27,** 2910–2912.

Bintanja, R., van de Wal, R.S. & Oerlemans, J. (2005) Modelled atmospheric temperatures and global sea levels over the past million years. *Nature*, **437**, 125–128.

Blanco-Pastor, J.L., Vargas, P. & Pfeil, B.E. (2012) Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *PLoS ONE,* **7,** e39089.

Bouckaert, R.R. (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, **26**, 1372–1373.

Clement M., Posada D. & Crandall K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology,* **9,** 1657–1659.

Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C.S., Carton, J.A. & Smith, R.D. (2006) The community climate system model version 3 (CCSM3). *Journal of Climate*, **19,** 2122–2143.

Désamoré, A., Laenen, B., Devos, N., Popp, M., González- Mancebo, J.M., Carine, M.A. & Vanderpoorten, A. (2011) Out of Africa: north- westwards Pleistocene expansions of the heather *Erica arborea*. *Journal of Biogeography*, **38**, 164–176.

Dorman, C.E., Beardsley, R.C. & Limeburner, R. (1995) Winds in the Strait of Gibraltar. *Quarterly Journal of the Royal Meteorological Society*, **121,** 1903–1922.

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

453 Duggen, S., Hoernle, K., Van Den Bogaard, P., Rüpke, L., & Morgan, J. P. (2003) Deep roots of the
454 Messinian salinity crisis. *Nature*, **422,** 602–606.
455 Eckert C.G., Samis K.E. & Lougheed S.C. (2008) Genetic variation across species' geographical ranges: the
456 central-marginal hypothesis and beyond. *Molecular Ecology*, **17,** 1170–1188.
457 Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software
458 STRUCTURE: a simulation study. *Molecular Ecology*, **14,** 2611–2620.
459 Excoffier L., Laval G. & Schneider S. (2005) ARLEQUIN ver. 3.0: an integrated software package for
460 population genetics data analysis. *Evolution and Bioinformatics Online*, **1,** 47–50.
461 Farris, J.D., Kallersjo, M.A., Kluge,G. & Bult., C. (1995) Constructing a significance test for incongruence.
462 *Systematic Biology*. **44**, 570– 572.
463 Fernández-Mazuecos, M., & Vargas, P. (2011). Genetically depauperate in the continent but rich in oceanic
464 islands: *Cistus monspeliensis* (Cistaceae) in the Canary Islands. *PLoS One*, **6**, e17172.
465 Fiz-Palacios, O. & Valcárcel, V. (2013) From Messinian crisis to Mediterranean climate: a temporal gap of
466 diversification recovered from multiple plant phylogenies. *Perspectives in Plant Ecology, Evolution and*
467 *Systematics*, **15**, 130–137.
468 García-Verdugo, C., Sajeva, M., La Mantia, T., Harrouni, C., Msanda, F. & Caujapé-Castells, J. (2015) Do
469 island plant populations really have lower genetic variation than mainland populations? Effects of
470 selection and distribution range on genetic diversity estimates. *Molecular Ecology*, **24,** 726–741.
471 Gómez, A. & Lunt D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the
472 Iberian Peninsula. (eds Weiss S. & Ferrand N.), pp. 155–188. Springer, Amsterdam.
473 Hasumi, H. & Emori, S. (2004) K-1 coupled gcm (miroc) description. *Center for Climate System Research,*
474 *University of Tokyo, Tokyo*.
475 Heled, J. & Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Molecular*
476 *Biology and Evolution*, **27,** 570–580.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature,* **405,** 907–913.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* **25,** 1965–1978.

Husemann, M., Schmitt, T., Zachos, F.E., Ulrich, W. & Habel, J.C. (2014) Palaearctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaearctic biota. *Journal of Biogeography*, **41,** 81–94.

Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Ecology and Evolution,* **23,** 254–267.

Huson, D.H. & Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, **61,** 1061–1067.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research,* **33,** 511–518.

Lambeck, K., Yokoyama, Y. & Purcell, T. (2002) Into and out of the Last Glacial Maximum: Sea-level change during oxygen isotope stages 3 and 2. *Quaternary Science Reviews,* **21,** 343–360.

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

Lemey, P., Rambaut, A., Drummond, A.J. & Suchard, M.A. (2009) Bayesian phylogeography finds its roots. *PLoS Computational Biology*, **5,** e1000520.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* **25,** 1451–1452.

Mairal, M., Sanmartín, I., Aldasoro, J.J., Culshaw, V., Manolopoulou, I. & Alarcón, M. (2015) Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canarina canariensis* (Campanulaceae). *Molecular Ecology,* **24**, 3944–3963.

500 Manni, F. & Guérard, E. (2004) *BARRIER, version 2.2 Manual*. Population Genetics team, Museum of
501 Mankind, Paris.
502 Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a
503 crucial process in island clades. *Systematic Biology*, syu056.
504 Médail, F. & Diadema, K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean
505 Basin. *Journal of Biogeography*, **36,** 1333–1345.
506 Molina-Venegas, R., Aparicio, A., Pina, F.J., Valdés, B. & Arroyo, J. (2013) Disentangling environmental
507 correlates of vascular plant biodiversity in a Mediterranean hotspot. *Ecology and Evolution*, **3**, 3879–
508 3894.
509 Myers, P.G. & Haines, K. (2000) Seasonal and interannual variability in a model of the Mediterranean under
510 derived flux forcing. *Journal of Physical Oceanography,* **30,** 1069–1082.
511 Nei, M. & Li W.H. (1979) Mathematical model for studying genetic variation in term of restriction
512 endonucleases. *Proceedings of the National Academy of Sciences of the United States of America,* **76,**
513 5269–5273.
514 Nyman, Y.Y. (1991) Crossing experiments within the *Campanula dichotoma* group (Campanulaceae). *Plant*
515 *Systematics and Evolution*, **177,** 185–192.
516 Olesen, J.M., Alarcón, M., Ehlers, B.K., Aldasoro, J.J. & Roquet, C. (2012) Pollination, biogeography and
517 phylogeny of oceanic Island bellflowers (Campanulaceae). *Perspectives in Plant Ecology and Evolution*
518 *Systematics,* **14,** 169–182.
519 Ortiz, M.A., Tremetsberger, K., Stuessy, T.F., Terrab, A., García- Castaño, J.L. & Talavera, S. (2009)
520 Phylogeographic patterns in *Hypochaeris* section *Hypochaeris* (Asteraceae, Lactuceae) of the western
521 Mediterranean. *Journal of Biogeography*, **36,** 1384–1397.
522 Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H. & Hu, A. (2006) Simulating Arctic climate
523 warmth and icefield retreat in the last interglaciation. *Science*, **311,** 1751–1753.

524 Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic
525 distributions. *Ecological Modelling*, **190,** 231–259.
526 Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25,** 1253–
527 1256.
528 Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype
529 data. *Genetics*, **155,** 945–959.
530 Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P. & Arroyo, J. (2008) The Strait of Gibraltar as
531 a melting pot for plant biodiversity. *Quaternary Science Reviews*, **27,** 2100–2117.
532 Rohlf, F.J. (1998) *NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System, v.1.80*. Exeter
533 Software, New York.
534 Roquet, C., Sanmartín, I., García-Jacas, N., Sáez, L., Susanna, A., Wikstrom, N. & Aldasoro, J.J. (2009)
535 Reconstructing the history of Campanulaceae with a Bayesian approach to molecular dating and
536 dispersal-vicariance analyses. *Molecular Phylogenetics and Evolution*, **52,** 575–587.
537 Sanmartín, I., Anderson, C.L., Alarcón, M., Ronquist, F. & Aldasoro, J.J. (2010) Bayesian island
538 biogeography in a continental setting: The Rand Flora case. *Biology Letters*, **6,** 703–707.
539 Stuessy, T.F., Takayama, K., López-Sepúlveda, P. & Crawford, D.J. (2014) Interpretation of patterns of
540 genetic variation in endemic plant species of oceanic islands. *Botanical Journal of the Linnean*
541 *Society*, **174**, 276–288.
542 Valtueña, F. J., López, J., Álvarez, J., Rodríguez- Riaño, T. & Ortega- Olivencia, A. (2016) *Scrophularia*
543 *arguta*, a widespread annual plant in the Canary Islands: a single recent colonization event or a more
544 complex phylogeographic pattern? *Ecology and Evolution*. **6**, 4258–4273
545 Vekemans, X. (2002) *AFLP-SURV version 1.0*. Laboratoire de Génétique et Ecologie Végétale, Université
546 Libre de Bruxelles, Brussels.

Vos, P., Hogers, R. & Bleeker, R. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research,* **23,** 4407–4414.

Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R.K. & Burke, T. (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, **8,** 725–735.

Wolfe, K.H., Li, W.H. & Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA*, **84,** 9054–9058.

Zhang, Z., Ramstein, G., Schuster, M., Li, C., Contoux, C. & Yan, Q. (2014) Aridification of the Sahara Desert caused by Tethys Sea shrinkage during the Late Miocene. *Nature*, **513,** 401–404.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Supplementary Tables.

Appendix S2 Supplementary Figures.

BIOSKETCH

Sara García Aloy is currently working on her PhD project in the Biodiversity and Evolution Department at Institut Botànic de Barcelona. All the authors are interested in biogeography and evolution of African plants, with the specific focus on macro- and microevolutionary processes in Campanulaceae and Geraniaceae.

Author contributions: all the members of the research contributed to design the study; S.G.A. and M.A ran the molecular data analyses. S.G.A and D.V. performed the species distribution analyses. S.G.A. led the writing with substantial contributions from all co-authors. All authors approved the final manuscript.

Editor: Alain Vanderpoorten

586 **TABLES**

587 **Table 1** Descriptors of within population genetic diversity in plastid haplotypes and AFLPs for each *C.*
588 *kremeri* and *C. occidentalis* population. Abbreviations refer to: number of samples (*n*), haplotype diversity
589 [H(d)], nucleotide diversity (π), nucleotide heterozygosity (θ), number of polymorphic fragments (PF),
590 percentage of polymorphic fragments (%PF), Nei's gene diversity and standard error [Hj (se)] and number of
591 private fragments (Np).

592

Locality, population and code		**Plastid haplotypes**					**AFLPs**				
		n	**Haplotypes**	**H (d)**	**π**	**θ**	***n***	**PF**	**%PF**	**Hj (se)**	**Np**
Iberia, Istán	IP1	10	H8	0	0	0	11	362	45.5	0.1470 (0.006)	1
Iberia, Benarrabá	IP2	12	H7, H8	0,167	0,00012	0,00023	11	236	29.6	0.1130 (0.005)	0
Iberia, S. Pedro	IP3	12	H8	0	0	0	11	356	44.7	0.1385 (0.006)	0
***C. kremeri* Iberia**		34		0,451	0,00032	0,00017	33	281	35.3	0.1275	
W Rif, J. Tissouka	WR1	5	H9, H10	0.400	0.00028	0.00034	11	466	58.5	0.1859 (0.006)	1
W Rif, Talassemtane	WR2	10	H8	0	0	0	10	352	44.2	0.1582 (0.006)	4
***C. kremeri* W Rif**		15		0.514	0,00043	0,00043	21	458	57.5	0.1867	
E Rif, Al Hoceima	ER1	4	H1	0	0	0	7	426	53.5	0.2029 (0.006)	12
M. Atlas, Taza	AT1	5	H1, H2	0.600	0.00049	0.00039	6	387	48.6	0.1912 (0.006)	5
C. Morocco, Mulay Idris	AT2	3	H3, H6	0.667	0.00109	0.00109	3	213	26.8	0.1116 (0.006)	3
H. Atlas, Afourer	AT3	7	H3, H5, H6	0.667	0.00102	0.00102	3	282	35.4	0.1838 (0.007)	0
A. Atlas, Chichaoua	AT4	7	H3, H4	0.286	0.00023	0.00033	7	341	42.8	0.1625 (0.006)	4
A. Atlas, Oued Mrabet	AT5	9	H3	0	0	0	9	350	44.0	0.1479 (0.006)	7
***C. kremeri* E Rif+Atlas**		35		0.640	0.00283	0.00239	35	464	58.3	0.1959	
Lanzarote, 7 Leguas	CI1	10	H11	0	0	0	10	291	36.6	0.0921 (0.005)	0
Fuerteventura, Pájara	CI2	11	H11	0	0	0	12	230	28.9	0.0979 (0.005)	1
Tenerife, Anaga	CI3	9	H11	0	0	0	10	294	36.9	0. 0959 (0.005)	0
Tenerife, Guimar	CI4	10	H11	0	0	0	10	328	41.2	0. 1107 (0.005)	1
Tenerife, Masca	CI5	10	H11	0	0	0	10	322	40.5	0.1161 (0.005)	0
***C. occidentalis* Canary Is.**		50		0	0	0	52	270	33.9	0.0996	

593
594

595 **FIGURE LEGENDS**
596

597 **Figure 1** Consensus tree of *Azorina*-group of species. Maximum-clade-credibility (MCC) time-calibrated tree
598 inferred and dated with a Bayesian analysis implemented in BEAST. Numbers above branches indicate
599 posterior probabilities and blue bars represent the posterior distribution of divergence-time estimates.
600 Biogeographic reconstruction was inferred with the DEC+J model as implemented in BioGeoBEARS.
601 Squares in the left of the taxon names indicate distribution areas (see legend). Red and green stripes show the
602 periods of Messinian Salinity Crisis and the refilling of the Mediterranean respectively.
603

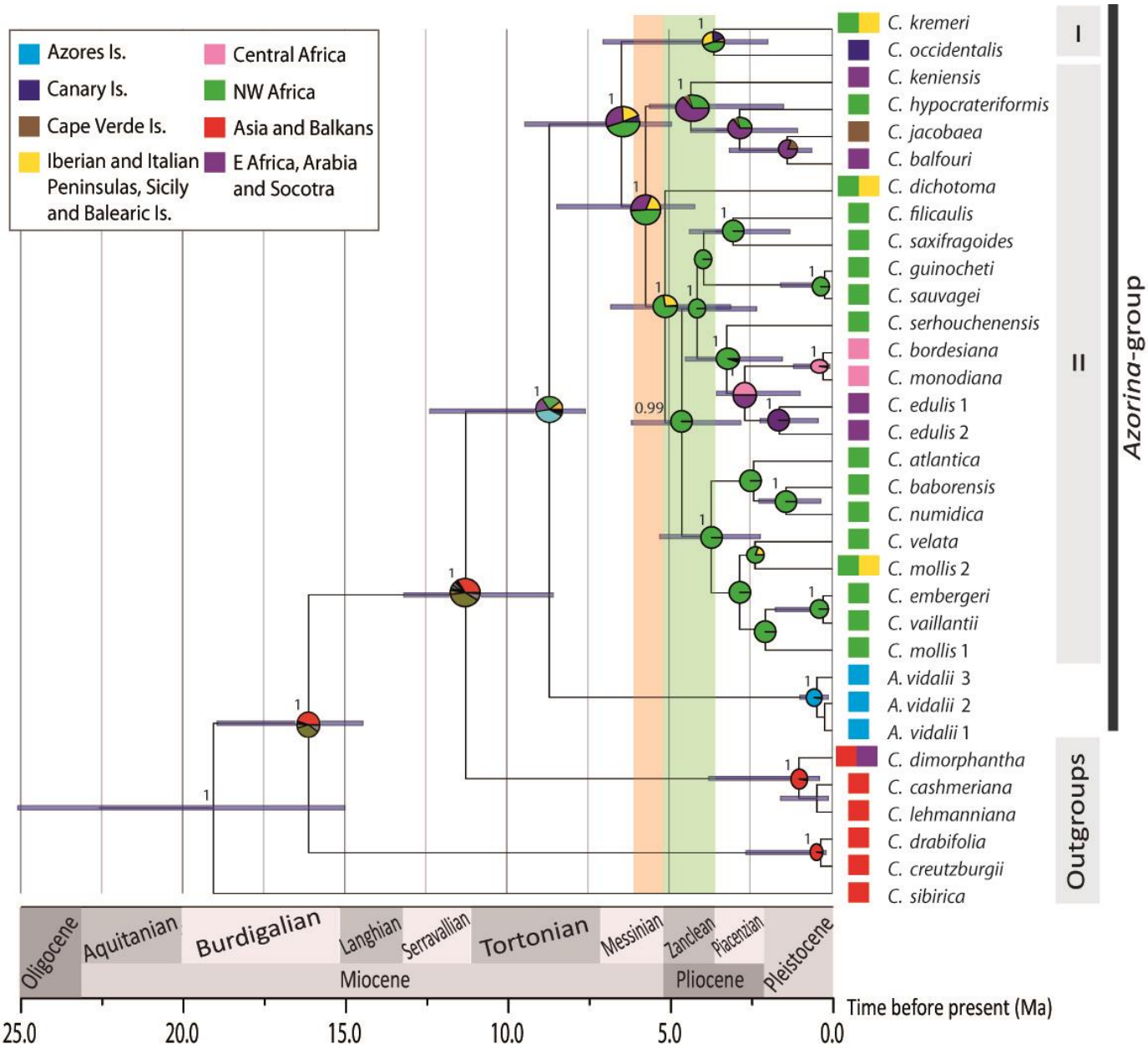
604 **Figure 2** (a) Consensus tree of *C. kremeri*-*C. occidentalis* populations. MCC time-calibrated tree based on
605 plastid and nuclear markers and obtained with Bayesian analysis implemented in BEAST. Plastid markers are
606 *trn*L-*trn*F, *trn*S–*trn*G, *pet*B–*pet*D and *rpl*32-*trn*L; nuclear ones are ITS and PPR11. Numbers above branches
607 indicate posterior probabilities and blue bars represent the posterior distribution of divergence-time estimates.
608 The squares represent the three main lineages: red, Atlas and eastern Rif; green, western Rif and Iberian
609 Peninsula; and blue, Canary Islands. Areas are coded: IP- Iberian Peninsula; WR- Western Rif; ER- Eastern
610 Rif; AT- Atlas; and CI- Canary Islands.
611 Rooted hybridization networks generated from the MCC trees of the plastid and nuclear data implemented in
612 BEAST: (b) network built using the level-k 47.5 algorithm implemented in DENDROSCOPE 3. The squares
613 represent the three main lineages as in Fig. 2a; and (c) multilabelled MCC tree obtained in the *BEAST species
614 tree analysis. Lineages inferred to be of hybrid origin are labelled "*C. kremeri* (A) N" (progenitor lineage of
615 Atlas populations that carry the nuclear compartment) and "*C. kremeri* (A) P" (progenitor lineage of Atlas
616 populations that carry the plastidial compartment).
617

618 **Figure 3** (a) Map representing the distribution of plastid haplotypes and BSSVS analysis of *Campanula*
619 *kremeri* and *C. occidentalis,* showing migration events with a BF support >3. The sampled localities are

showed as red spots, and the herbaria records as small white spots. Inside the square a Coalescent MCC tree showing the results of the plastid BSSVS analysis. Branch colour indicates the ancestral range with the highest posterior probability for each lineage. Numbers above branches indicate Bayesian *PP*. (b) TCS network of plastid markers (*rpl32*-*trn*L and *trn*S-*trn*G), each haplotype has a colour. Black stripes represent nucleotide changes, and the circle size is proportional to the number of individuals for each haplotype (Table 1). (c) Median network of nuclear markers (ITS and PPR11) implemented in NETWORK 4.2.0.143 using the same accessions as in Figure 2.

Figure 4 Results from the analysis of AFLP markers using K=3. Histograms show the Bayesian clustering of individuals within populations (STRUCTURE), colours represent the individual membership to each inferred Bayesian group. Dotted lines indicate barriers to gene flow and their percentage, as inferred by BARRIER. Inside the square is represented the Neighbor-Net analysis inferred for individuals and populations using SPLITSTREE 4.10.

634 **Figure 1**



635

636

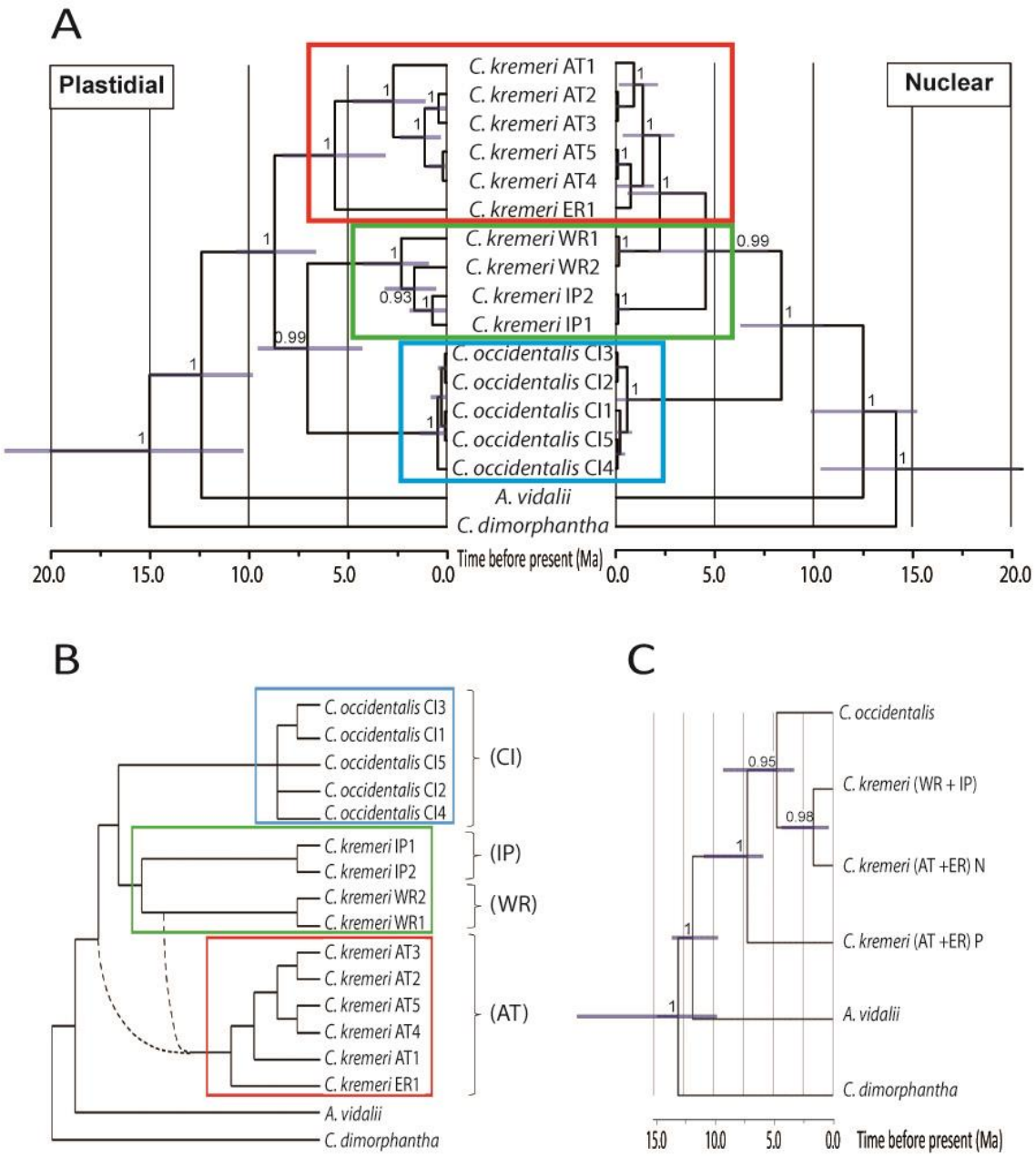
637

638

639

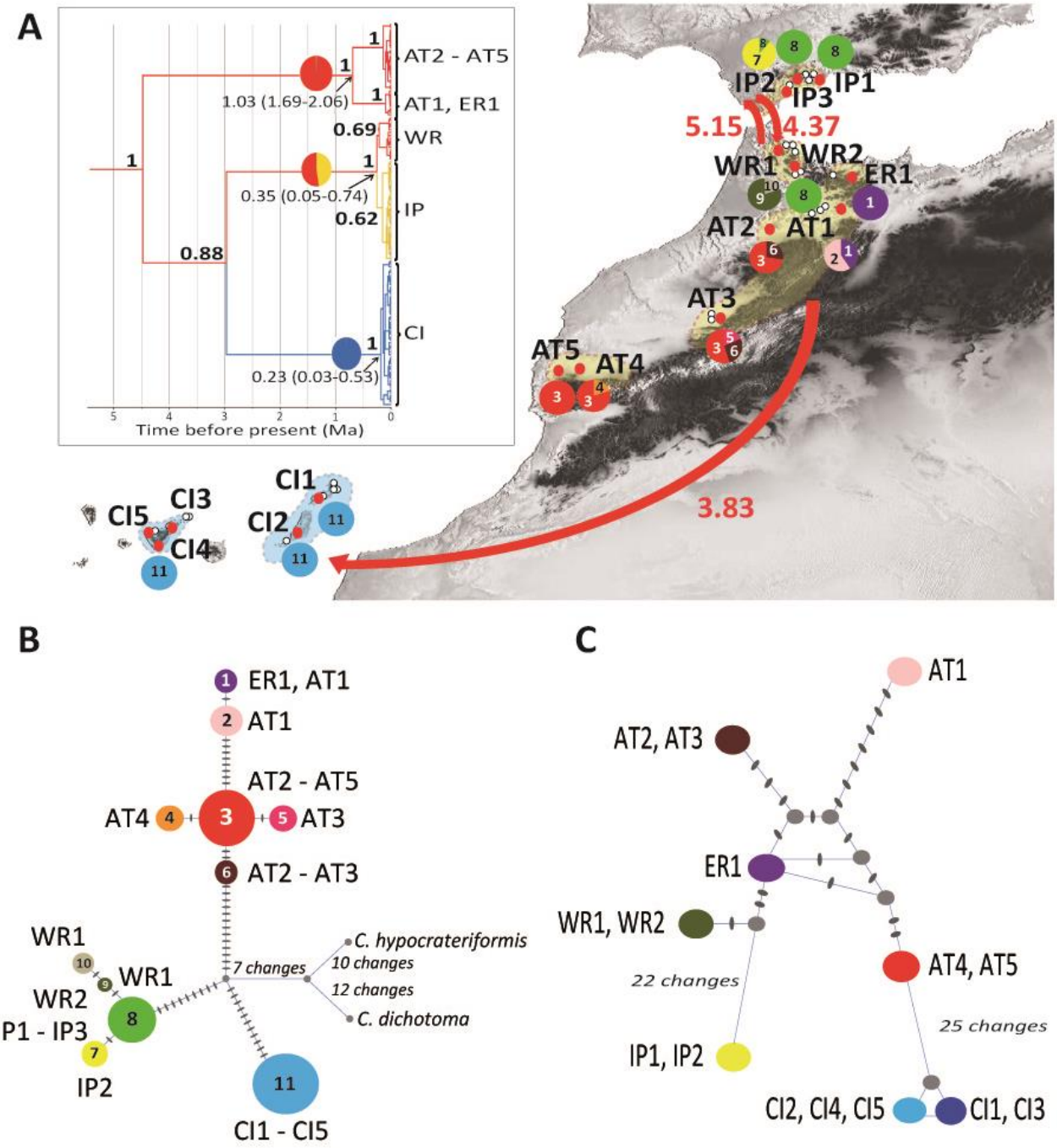
640

641 **Figure 2**



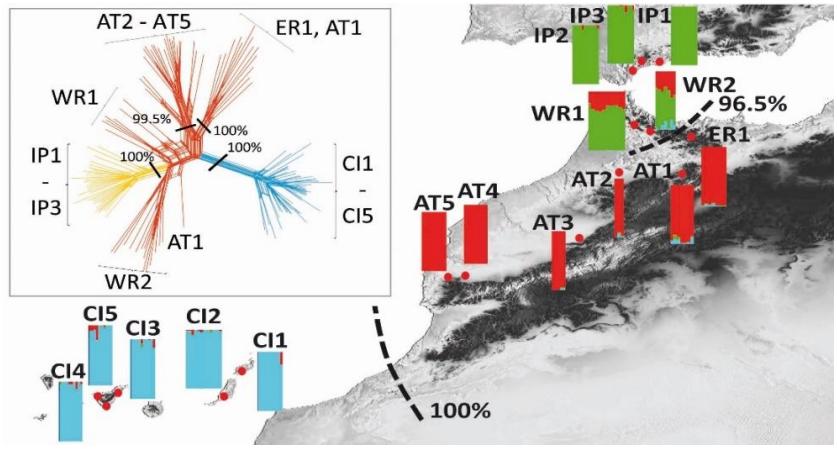
642
643

644 **Figure 3**



645

646 **Figure 4**



647