Phylogeography of western Mediterranean Cymbalaria (Plantaginaceae) reveals two independent long-distance dispersals and entails new taxonomic circumscriptions

Pau Carnicero 1,2, Peter Schönswetter 2, Pere Fraga Arguimbau 3, Núria Garcia-Jacas 4, Llorenç Sáez 5,6 & Mercè Galbany-Casals 1

The Balearic Islands, Corsica and Sardinia (BCS) constitute biodiversity hotspots in the western Mediterranean Basin. Oligocene connections and long distance dispersal events have been suggested to cause presence of BCS shared endemic species. One of them is Cymbalaria aequitriloba, which, together with three additional species, constitute a polyploid clade endemic to BCS. Combining amplified fragment length polymorphism (AFLP) fingerprinting, plastid DNA sequences and morphometrics, we inferred the phylogeography of the group and evaluated the species' current taxonomic circumscriptions. Based on morphometric and AFLP data we propose a new circumscription for C. fragilis to additionally comprise a group of populations with intermediate morphological characters previously included in C. aequitriloba. Consequently, we suggest to change the IUCN category of C. fragilis from critically endangered (CR) to near threatened (NT). Both morphology and AFLP data support the current taxonomy of the single island endemics C. hepaticifolia and C. muelleri. The four species had a common origin in Corsica-Sardinia, and two long-distance dispersal events to the Balearic Islands were inferred. Finally, plastid DNA data suggest that interspecific gene flow took place where two species co-occur.

Islands in the Mediterranean Basin harbour both high species diversity and endemism. For instance, from the around 5000 islands scattered in the Mediterranean Sea 1 all the largest ones (i.e. Balearic Islands, Corsica, Crete, Cyprus, Sardinia and Sicily) constitute Mediterranean biodiversity hotspots 2. Most of the Mediterranean islands are of continental origin and have been separated from the mainland and from each other through progressive geomorphological processes, often with posterior re-connections due to sea level variation 3. Such processes are thought to have played an important role in the evolution and diversification of biota 4.

The Balearic Islands, Corsica and Sardinia (hereafter termed BCS) originated from progressive splitting and rotation of the Hercynian belt, an Oligocene mountain range linking the Alps with the Iberian Peninsula that started to disintegrate around 30 Million years ago (Ma) 5. The split between Corsica-Sardinia and the Balearic Islands occurred around 20 Ma 6. Endemic plant species shared by Corsica, Sardinia and the eastern Balearic Islands (Mallorca and Menorca) have traditionally been considered relict palaeoendemic remnants of
Citing features such as white to pale violet flowers, fleshy leaves and fragile stems, dense hairiness and finely
textured petals, the Mediterranean genus comprising ca. 17 taxa, four of which are endemic to BCS 20 (Fig. 1). The BCS endem-
ics form an independent lineage of presumably polyploid origin that diversified during the Pliocene and the Pleistocene, as revealed by a nuclear ribosomal DNA (nrDNA) phylogeny12. However, interspecific phylogenetic relationships were poorly resolved, and therefore using more variable molecular markers such as AFLPs would be desirable. Chromosome counts suggest that the BCS endemics are hexa- to octoploids (2n = 42, 56), whereas all other Cymbalaria species are diploids or tetraploids (2n = 14, 2820 and references therein). While C. aequitriloba is widely distributed throughout BCS and occurs in all main islands except for the western Balearic Islands Eibissa and Formentera, the other three species are single island endemics according to the current taxonomic treatments26,27. This lineage is therefore an excellent example to understand the mechanisms underlying the endemic patterns in BCS.

Pleistocene climatic oscillations caused massive changes of the sea level that led to repeated appearance of land bridges between some islands during cold stages, and to transgressions reducing the area of islands during interglacials16. During glaciation periods Corsica-Sardinia and the eastern Balearic Islands (Mallorca, Menorca and Cabrera), respectively, formed single landmasses. The resultant connections have often been used to explain phylogeographic patterns and shared endemics within each archipelago (e.g. Mansi et al.17; Salvi et al.18; Mayol et al.19).

One of the genera with highest number of endemics in BCS is Cymbalaria Hill (Plantaginaceae), a Mediterranean genus comprising ca. 17 taxa, four of which are endemic to BCS20 (Fig. 1). The BCS endem-
ics form an independent lineage of presumably polyploid origin that diversified during the Pliocene and the Pleistocene, as revealed by a nuclear ribosomal DNA (nrDNA) phylogeny12. However, interspecific phylogenetic relationships were poorly resolved, and therefore using more variable molecular markers such as AFLPs would be desirable. Chromosome counts suggest that the BCS endemics are hexa- to octoploids (2n = 42, 56), whereas all other Cymbalaria species are diploids or tetraploids (2n = 14, 2820 and references therein). While C. aequitriloba is widely distributed throughout BCS and occurs in all main islands except for the western Balearic Islands Eibissa and Formentera, the other three species are single island endemics according to the current taxonomic treatments26,27. This lineage is therefore an excellent example to understand the mechanisms underlying the endemic patterns in BCS.

The two single island endemics from Corsica and Sardinia have been traditionally considered as clearly dis-
tinct species. Cymbalaria hepaticifolia Wettst. (2n = 5630) occurs in the mountains of Corsica, from subalpine hab-
itats to fresh and humid forests22,23. Cymbalaria muelleri (Moris) A.Chev., comprising ca. ten known populations, is a hexaploid (2n = 4230) endemic of the mountains of central Sardinia35.

Cymbalaria fragilis (J.J.Rodr.) A.Chev. is a critically endangered (CR) octoploid (2n = 5638) species endemic to Menorca26. It has been treated either as a subspecies of C. aequitriloba26,27, or as a separate species37. Furthermore, the circumscription of this taxon is uncertain. According to Güemes31, C. fragilis is characterized by several dis-
tinctive features such as white to pale violet flowers, fleshy leaves and fragile stems, dense hairiness and finely
alveolate to nearly smooth seeds. However, some populations in Menorca, formerly identified as C. fragilis37, and Cabrera exhibit all distinctive features but have deeply alveolate seeds similar to those of C. aequitriloba21,26. Seed ornamentation has been regarded as highly relevant in the taxonomy of the tribe Antirrhineae30–32 and this character was later considered sufficient by Güemes26 to attribute these populations to C. aequitriloba. The latter is the most widespread species among the four BCS endemics, occurring in the eastern Balearic Islands, Corsica, Sardinia and the Tuscan Archipelago. Octoploid counts (2n = 56) have been reported from Corsica and the Balearic Islands33–35; older hexaploid counts (2n = 4238) were considered erroneous by Verlaque et al.35. Indeed, the reference cited by Sutton20 for the hexaploid counts36 does not report any chromosome number for C. aequitriloba, and we were unable to trace any other hexaploid count in the literature. Cymbalaria aequitriloba
exhibits high morphological variability, which is the reason why sparsely hairy forms from Mallorca have sometimes been confused with C. hepaticifolia. Some morphological extremes were described as separate taxa, but they were later synonymised with C. aequitriloba. Nuclear ribosomal and plastid DNA sequences, however, did not support the species’ monophyly, and it is thus doubtful whether C. aequitriloba is indeed a single taxon. Some authors, based on chromosome number information, considered C. aequitriloba an example of a palaeoendemism, although some of its aforementioned features, such as its high morphological variability, do not fit with the classical definition of the term. Most importantly, its recent origin inferred from a dated phylogeny rejected that hypothesis.

Here, we focus on the phylogeography and systematics of the BCS endemics of the genus Cymbalaria. We use a combination of AFLPs, plastid DNA sequences and morphometrics to tackle the following questions: (1) Are genetic and morphological data congruent and do they support the current taxonomy? (2) Specifically, is C. aequitriloba monophyletic based on AFLP data? (3) What is the appropriate taxonomic rank and circumscription of C. fragilis? (4) What was the pattern of colonization of the BCS endemics among Sardinia, Corsica and the Balearic Islands?

Results

AFLPs. In total, 834 AFLP fragments were scored for 304 individuals (Supplementary Table S1), of which 779 polymorphic, high-quality, reproducible AFLP-fingerprints were obtained. Fifty-five fragments with singular presences or absences were excluded. The initial error rate before the exclusion of non-reproducible fragments was 3.17%. The final dataset for C. aequitriloba comprised 175 individuals and 557 AFLP fragments.

The NJ analysis revealed that the four taxa analysed constituted a monophyletic group with high bootstrap support (BS 100%, Fig. 2). The main split in the NJ tree separated C. aequitriloba and C. fragilis from C. hepaticifolia and C. muelleri. Populations of C. aequitriloba formed a clade (BS 62%) excluding all populations with distinctive features of C. fragilis, but bearing alveolate seeds (f1, f3, f4, f5, f8 and f9), which instead grouped with C. fragilis (BS 100%). One well supported clade each grouped populations of C. hepaticifolia (BS 100%) and C. muelleri (BS 96%). Individuals from the same population – with the exception of populations a14, a23 and h4 – formed monophyletic groups. The C. fragilis clade showed a main split between the populations from Menorca (f1 to f7, BS 79%) and those from Cabrera traditionally included in C. aequitriloba (f8 and f9, BS 100%).

The most frequent species tree topology (61%) obtained by the SNAPP analysis coincided with the NJ analysis (Supplementary Fig. S1). However, a second most frequent topology (30%) indicates different relationships among the species: C. hepaticifolia diverged first, followed by the differentiation of C. muelleri, which is sister to the C. aequitriloba-C. fragilis clade. Both topologies showed how the diversification of the four western Cymbalaria species occurred roughly simultaneously. Three alternative topologies represented less than 10% of the total number of trees.

The results obtained by the Bayesian analyses with BEAST were congruent with the NJ analysis, minor differences concerned nodes with low support (Supplementary Fig. S2). The origin of diversification was estimated to be in northern Sardinia (Fig. 3), from where dispersal events to southern Sardinia, Corsica and the Balearic Islands occurred. Later, further dispersals from Sardinia to Corsica and the Balearic Islands (Mallorca) took place.

A hierarchical AMOVA (Supplementary Table S2) assigned 28.7% of the entire variation to the among-taxa component. In C. aequitriloba the highest variation among populations within an island was found in Sardinia (44.7%), and the lowest in Corsica (27.7%). Gene diversity expressed as π was significantly lower in the Balearic Islands than in Corsica and Sardinia. In the separate NJ analysis for C. aequitriloba (BS 100%, Fig. 4) the populations from the Balearic Islands formed a clade (BS 100%), which contained a subclade comprising populations from Menorca (a15–a20, BS 91%). Populations from Corsica (a9–a14) formed a supported clade (BS 99%) whereas populations from Sardinia were distributed in three main clades (a1, BS 100%; a2, a7 and a8, BS 70%; a3–a6, BS 97%). Nonhierarchical K-means clustering resulted in an optimal separation of the dataset into four groups (Fig. 4) composed by 1) the Sardinian populations a3–a6; 2) the Sardinian populations a1, a2, a7 and a8; 3) all Corsican populations; and 4) all Balearic populations. At K = 6 populations a1–a8 from Sardinia were split into four clusters, whereas the other two groups included all Corsican and all Balearic populations, respectively (Fig. 4).

Plastid DNA Sequences. The concatenated ndhF and rpl32-trnL(UAG) sequences consisted of 2776 aligned positions. We found 26 plastid haplotypes that differed by one to 14 substitutions including codified indels (Fig. 5). There was no clear correlation with taxonomy; instead, some geographically close populations of different taxa shared haplotypes. For instance, haplotype 4 was found in almost all populations of C. aequitriloba and C. fragilis from the Balearic Islands, haplotype 9 was found in two populations of C. aequitriloba and one population of C. hepaticifolia from Corsica, and haplotype 22 was found in two populations of C. aequitriloba and one population of C. muelleri from southern Sardinia (Fig. 5). The Balearic Islands clearly showed low haplotype diversity in comparison to Corsica and Sardinia. Relationships inferred by MP and Bayesian analyses were congruent and consequently we only show the topology from the Bayesian analysis (Fig. 5). Two main clades were found in the tree: the first contains the strongly divergent haplotype 26 from Sardinia and a haplotype of C. muralis (SP 1; BS 100%), and it is sister to all other haplotypes found in BCS (PP 0.89). Phylogenetic relationships among haplotype groups (colour-coded in Fig. 5) were in general statistically weakly supported, whereas relationships among closely related haplotypes reflected the pattern of the parsimony network.

Morphometric analyses. Correlation coefficients did not exceed 0.95 for any pair of characters; therefore, all characters were retained for further analyses (Supplementary Table S4).
Dataset 1. In the PCA, the first axis accounted for 34% of the variation and the second axis for 28%. The ordination diagram (Fig. 6a) suggested three clusters, two corresponded to *C. aequitriloba* and *C. hepaticifolia*, respectively, and a third one grouped *C. fragilis* and *C. muelleri*, including the controversial *C. aequitriloba* populations f1, f3, f4, f5, f8 and f9 from Cabrera and Menorca. The characters with most weight in the separation of the three groups were those related with indumentum density. The CDA (Fig. 6b–e) supported morphological differentiation of the four main AFLPs clades. The first two axes (41% and 35% of the explained variation) supported the three groups observed in the PCA, with some overlap between *C. aequitriloba* and *C. fragilis* plus *C. muelleri*. The characters with the highest contribution to these axes were derived from the leaf indumentum. The third axis (24%) supported the differentiation between *C. fragilis* and *C. muelleri*. The characters with the highest contribution to the third axis were spur length and width.

Dataset 2. The first PCA axis accounted for 44% of the variation and the second axis for 14%. The ordination diagram showed a structure similar to that retrieved by the analysis of dataset 1, albeit with stronger overlap of *C. aequitriloba* and the *C. fragilis* – *C. muelleri* group (Supplementary Fig. S3).
Discussion

The combination of molecular and morphological data has been widely used to reevaluate systematics at different taxonomic levels (e.g.32,40–42). Specifically, many studies have used AFLP data and morphometrics to address taxonomic conflicts in groups of closely related plant taxa (e.g.43–47). In the case of *Cymbalaria*, AFLP (Fig. 2) and morphometric analyses (Fig. 6) allow for the recognition of four distinct entities within a previously identified monophyletic group of *Cymbalaria* species endemic to the Balearic Islands, Corsica and Sardinia12. These entities coincide with the current taxonomic circumscription of *C. hepaticifolia* and *C. muelleri*, but require novel circumscriptions of *C. aequitriloba* and *C. fragilis* (Figs 2 and 6).

A novel circumscription of *C. fragilis* and *C. aequitriloba*. Güemes21 suggested that *C. fragilis* is restricted to three populations in southwestern Menorca, and assigned reports from the North and East of the island (populations f1 and f4) as well as those from Cabrera (populations f8 and f9) to *C. aequitriloba*, based on their alveolate seeds. However, these populations exhibit all distinctive morphological features of *C. fragilis* such as (1) villous indumentum throughout the vegetative parts, (2) highly fragile stems, petioles and pedicels, (3) fleshy leaves, (4) white to pale violet corollas, (5) spur proportionally shorter and wider than in *C. aequitriloba* (ratio spur length/spur width 0.9–2.8), (6) seeds, which are often smaller than those of *C. aequitriloba* (0.6–0.9 mm long, vs. 0.8–1.2 mm long in *C. aequitriloba*, see identification key), as well as (7) a late flowering time compared to *C. aequitriloba* (June as compared to early spring in *C. aequitriloba* from the same altitude). Interestingly, in population f6 we observed both types of seed ornamentation. In the AFLP (Figs 2 and 4) as well as in the morphometric analyses (Fig. 6), these populations clustered with *C. fragilis*. Therefore, *C. fragilis* must be redefined to include populations with the before-mentioned distinctive morphological characters from Menorca and Cabrera in spite of their aberrant seeds. Seed ornamentation – in this particular case – must thus be considered a variable character without taxonomic value. The strongly supported monophyly in the AFLP analyses (Fig. 2) together with the clear morphological differentiation (Fig. 6) and the overlapping distributions of *C. fragilis* and *C. aequitriloba* in Menorca do not match the definition of subspecies48–50. In contrast, the available evidence strongly suggests treating *C. fragilis* as a separate species in accordance with the original description51, rather than as a subspecies of *C. aequitriloba*52. The proposed change of circumscription brings about a change in the species’ threat category according to IUCN53 criteria. We sampled all known populations of *C. fragilis* to date, updating the number of populations from three to nine, and therefore suggest that Near Threatened (NT) would be a more adequate category than the formerly applied Critically Endangered (CR)27.

Figure 3. Snapshots of estimated ancestral node areas in the Maximum Clade Credibility tree obtained from the BEAST continuous phylogeographical analysis of AFLP data at different time horizons as visualized using the software SPREAD. The 80% highest posterior density areas for nodes are indicated as grey polygons. The star in the upper left panel indicates the starting point of diversification and the relative time scale of diversification is indicated by numbers (1 = origin, 0 = present). Numbers indicate LDD events to the Balearic Islands leading to (1) the origin of *C. fragilis* and (2) a range expansion of *C. aequitriloba*. Lines indicating the two LDD events were modified to prevent them from crossing. Maps generated from the altitude layer obtained from http://www.worldclim.org132, and modified using QGIS v2.18.2133.
Figure 4. Neighbor-Joining tree based on AFLP data generated for *C. aequitriloba* following the taxonomic concept adopted in the present study. Distributions of clusters identified by nonhierarchical K-means clustering at $K = 4$ and $K = 6$ are shown to the right. Circles next to population names indicate the island of origin.

**A Corso-Sardinian common ancestor for the BCS endemics.** Spatial reconstruction of evolutionary dynamics using relaxed random walks (Fig. 3) suggests a Sardinian origin of the BCS endemic species of *Cymbalaria*. The low haplotype diversity found in the Balearic Islands (Fig. 5) is congruent with a Sardinian origin, but does not allow to reject the alternative hypothesis of a Corsican origin. An east (Corsica-Sardinia) to west (Balearic Islands) pattern of colonization has been also suggested for the BCS endemics *Arum pictum* and *Thymus herba-barona*.
AFLP and morphological data confirmed *C. hepaticifolia* and *C. muelleri* as two well differentiated species (Figs 2 and 6). They diverged from a common ancestor probably via an allopatric speciation event, as suggested by their current distribution areas (Fig. 1, Supplementary Fig. S1). However, the SNAPP analysis revealed an alternative topology supported by 30% of trees, which resolved *C. hepaticifolia* but previously assigned to *C. aequitriloba* owing to the presence of alveolate seeds. This uncertainty could be caused by rapid simultaneous diversification of the BCS endemics, as suggested for the initial diversification of *Cymbalaria* 12. Indeed, rapid diversification rates after colonisation of new areas by polyploid lineages have been previously suggested 54,55. Alternatively, the closer relationship of *C. muelleri* with the *C. aequitriloba-C. fragilis* clade could have been caused by gene flow between *C. muelleri* and *C. aequitriloba* or its ancestor in Sardinia. Plastid sequences revealed, indeed, close relationships among the haplotypes found in *C. hepaticifolia* and *C. muelleri* and those of sympatric *C. fragilis* populations (Fig. 5).

**Figure 5.** Sampled populations and patterns of plastid DNA (*ndhF* and *rpl32-trnL(UAG)*) variation. Colours correspond to groups of closely related haplotypes. Taxonomy follows the taxonomic concept adopted in the present article (a: *C. aequitriloba*, b: *C. fragilis*, h: *C. hepaticifolia*, m: *C. muelleri*). Maps generated from the altitude layer obtained from http://www.worldclim.org132, and modified using QGIS v2.18.2133. (a,b) Pie charts indicate the proportions of haplotype groups sampled in each population. Asterisks indicate populations exhibiting mostly distinctive features of *C. fragilis* but previously assigned to *C. aequitriloba* owing to the presence of alveolate seeds. (c) Phylogram from the Bayesian Inference analysis. Numbers above branches indicate Bayesian posterior probabilities/Maximum parsimony bootstrap support values. Haplotypes are given for each lineage next to the tips, numbering corresponds to Table S1. (d) Statistical parsimony haplotype network. Black dots indicate hypothetical unsampled haplotypes, numbering corresponds to Table S1.
Both the strongly reduced haplotype diversity of *C. fragilis* in comparison with Corsican and Sardinian populations of the three other studied species (Fig. 5) and the continuous phylogeographic analysis (Fig. 3) support the hypothesis that a LDD event to the Balearic Islands led to the origin of *C. fragilis*. Genetic diversity of *C. fragilis* was not significantly lower than that of *C. aequitriloba* (Supplementary Table S3), suggesting that the LDD event was ancient enough to allow for a recovery of genetic diversity56. *Cymbalaria aequitriloba* has its centre of genetic diversity in Sardinia, as shown by the highest among-population variation in the AMOV As (Supplementary Table S2), the significantly higher genetic diversity observed in Sardinian populations as compared to populations from the other islands (Supplementary Table S3), the high haplotype diversity (Fig. 5) and the high number of separable AFLP gene pools (Fig. 4).

**Phylogeography of *C. aequitriloba* and *C. fragilis***. *Cymbalaria aequitriloba* exhibits a higher genetic diversity in Sardinia than in Corsica, as shown by the higher number of plastid haplotypes and AFLP clusters (Figs 4 and 5). This is most probably explained by Pleistocene glaciation periods, which had a stronger effect in Corsica than in Sardinia as illustrated by the large glaciers documented in Corsica during the Last Glacial Maximum57 (18,000 years ago), and could have caused local extinctions58,59. Later range expansion might have occurred from either lowland Corsican or Sardinian refuge areas, via land bridges connecting both islands multiple times during the Pleistocene60,61 or across the narrow Bonifacio strait separating Corsica and Sardinia62. Populations from northern and western Sardinia showed a high genetic similarity with Corsican populations (Figs 4 and 5), supporting contacts between the populations of the two islands as shown for *Cistus creticus* L.62.

**Figure 6.** Ordination diagrams of morphometric analyses of 18 floral and vegetative characters obtained for 129 specimens (dataset 1) of *C. aequitriloba* (empty circles), *C. fragilis* (grey dots), *C. hepaticifolia* (black dots) and *C. muelleri* (crosses). Grey dots with thicker outline indicate *C. fragilis* specimens previously considered *C. aequitriloba*. (a,b) Principal component analysis. (a) Scatter plot of principal component scores for the first two components of morphological variation. (b) Character relationships and contribution to the first two axes. (c–f) Canonical discriminant analysis. (c,e) Scatter plots of discriminant scores for two combinations of the three discriminant axes. (d,f) Character relationships and contribution in the same ordination space as (c,e). Characters with high contributions in each combination are identified. (ind. = indumentum).
Cymbalaria reached the Balearic Islands via two independent LDD events, as supported by the NJ tree (Fig. 2) and the continuous phylogeographic analysis (Fig. 3), combined with evidence of a relatively young origin of Cymbalaria BCS endemics, ca. 15–20 myr later than the last land connection between Corsica-Sardinia and the Balearic Islands12. The first LDD events ultimately led to the origin of C. fragilis and the second caused a range expansion of C. aequitriloba. These two events were also supported by the shorter terminal splits depicted in the NJ tree for all Balearic populations indicating strong similarity among individuals (Figs 2 and 4). Significantly lower average genetic diversity of Balearic populations of C. aequitriloba as compared to those from Corsica and Sardinia also support the hypothesis of a recent LDD event (Supplementary Table S4). In addition, the strongly reduced haplotype diversity in the Balearic Islands (Fig. 5) may evidence a strong bottleneck resulting from a founder event (e.g., Schonswetter et al.63, Burnier et al.64). With the exception of haplotype H11, the close phylogenetic relationships among haplotypes from the Balearic Islands (Fig. 5c) suggest a common ancestral haplotype for C. aequitriloba and C. fragilis. This common ancestor might have occurred in western and northern Sardinia as shown by the BEAST continuous phylogeographic analysis and the close phylogenetic relationship between haplotypes H1–H4 from the Balearic Islands and H5–H10 from Corsica and Sardinia (Figs 3 and 5). In the same line, the geographic setting supports a dispersal from western Sardinia, for instance from the surroundings of population a2. It is the closest to the Balearic Islands and thrives on the summit of the Monte Ferru massif, at 1050 m altitude only 12 km from the coast. Alternatively, or in addition, at a later stage hybridization with C. fragilis could have led to the sharing of haplotype H4 across species boundaries65.

The history of the Balearic populations of both species was likely influenced by sea level variation during the Pleistocene. Land connections during glaciation periods18 might have aided C. aequitriloba and C. fragilis to disperse among the Eastern Balearic Islands until both species achieved their current distributions. However, during early Pleistocene interglacials (1.6–0.7 Ma) current level, leading to a considerable reduction of the surface of Menorca, division of Mallorca into two islands and transgression of most of Cabrera19,66. This was likely a major obstacle for the expansion of C. fragilis to Cabrera, since only a small area of that island protruded from the sea, suggesting that the colonization most probably occurred after this period. In any case, presence of well-supported clades in Menorca and Cabrera (Fig. 2) supports the current isolation.

In contrast to other members of the tribe Antirrhineae exhibiting winged seeds or adaptation to marine dispersal (e.g. Linaria Mill., Cymbalaria longipes (Boiss. & Heldr.) A.Chev.)50 neither C. aequitriloba nor C. fragilis show any apparent adaptation for LDD. Moreover, both species are at least partly achorous and actively prevent seed dispersal by positioning mature capsules in rock crevices50. Notably, very similar features were used as an argument against LDD for the BCS endemic Arenaria balearica50. However, the importance of non-standard vectors has been recognised67,68 and rare stochastic events such as LDDs have been highlighted as important drivers of organisms’ histories69,70. In fact, LDD has been invoked in several cases in spite of the apparent lack of particular adaptations (e.g. Guzmán and Vargas71; Dixon et al.72, Escudero et al.73, Santos-Gally et al.74, Piñeiro et al.75). In the case of Cymbalaria, the small size of seeds could have favoured LDD as a consequence of rare stochastic events, as have been invoked for Androsace69 or suggested as an a priori possibility for A. balearica76. The success in colonization after a LDD is thought to be governed by the availability of adequate niches rather than by dispersal abilities5. In addition, polyploids are better buffered against genetic bottlenecks and subsequent inbreeding resulting from founder effects76, a feature that could have helped C. aequitriloba and C. fragilis to successfully establish in the Balearic Islands.

Finally, our previous results12 strongly suggested that C. aequitriloba is not a palaeoendemic that originated and expanded its range when Corsica-Sardinia and the Balearic Islands were connected8 (~20 Ma), but rather colonized the Balearic Islands much later, 1–4 Ma. The results presented here support this statement and furthermore show that C. aequitriloba does not meet the criteria attributed to relict taxa, i.e., low levels of infraspecific morphological and genetic variation as a result of a long process of adaptation in refugial isolation, and taxonomic isolation due to extinction of close relatives83,87. In contrast, C. aequitriloba is highly morphologically variable20,21 (Fig. 6), genetically equally diverse as its closest relatives (Supplementary Table S3) and not taxonomically isolated. Obviously, knowledge of the temporal framework is crucial for any study of palaeoendemism, since it allows for evaluating the congruence between a biogeographic hypothesis and, for instance, the geomorphological distribution of landmasses. In at least two cases focussing on BCS, dating analyses have rejected the hypothesis of a continuous palearctic distribution in the Oligocene landmass followed by fragmentation (A. pictum, T. herba-barona). To date, molecular dating has resulted in a temporal framework congruent with a continuous palearctic distribution only in the case of H. muscivorus. Although Bobo-Pinilla et al.32 suggested A. balearica to be a palaeoendemic, the lack of absolute time estimates prevented them from firmly rejecting a more recent origin. Often, dating analyses are revealing more recent divergent times than formerly thought (e.g. Salvo et al.13, Fernández-Mazuecos et al.14) and are thus increasingly shifting the biogeographic paradigm for the Mediterranean area.

Incongruence between AFLP and sequence data. The monophyly of each of the four species in the AFLP analyses (Fig. 2) suggests that reproductive isolation is effective at present. However, interspecific gene flow appears possible due to the close proximity between populations of C. aequitriloba and C. hepaticifolia in Corsica, and between C. aequitriloba and C. muelleri in Sardinia (P.C. & M.G., personal field observations). A few incongruities between mostly nuclear-derived AFLPs77 and maternally inherited plastid DNA sequences78 could indicate past reticulation events. Although incomplete lineage sorting (ILS) of ancient polymorphisms cannot be ruled out as cause for these incongruities79–83 hybridization appears more likely due to the grouping of haplotypes retrieved from geographically close populations of different taxa84,85. The most obvious case is the occurrence of haplotype 26 otherwise characteristic of the phylogenetically distant (cf. Carnicero et al.12) C. muralis in populations a4 and a6 of C. aequitriloba from Sardinia (Fig. 4). Cymbalaria muralis, a diploid species native to the Apennine and northern Balkan Peninsulas, is naturalized almost worldwide in temperate areas86. The population
of *C. muralis* used here is from the same area in Sardinia where haplotype 26 was found; hybridization resulting in chloroplast capture of a *C. muralis* haplotype is, therefore, likely.

Two statistically supported incongruities were found between the AFLP dataset and the previous phylogeny. The latter retrieved a clade formed by *C. aequitriloba* and *C. muralis*, and inferred a common ancestor for *C. fragilis* and *C. muelleri*. These two clades did not appear in the AFLP analysis (Fig. 2). Although nrDNA and AFLPs have been shown to yield mostly congruent signals, incongruities have also been reported. Here, we suggest that hybridization followed by concerted evolution of nrDNA towards one parental copy is a probable cause for incongruences affecting reproductive barriers (P.C. & M.G., personal field observations). In contrast, ILS or insufficient variation among nrDNA sequences seem the most probable causes for incongruence affecting closely related species, especially when dealing with complex evolutionary processes such as polyploidy and hybridization.

**Identification Key.** Here we provide a dichotomous key to all *Cymbalaria* species occurring in BCS.

1. Leaves and stems glabrous or only with occasional hairs on young organs .................................................. 2
2. Leaves entire to 5-lobed. Corolla 9.7–12.6(13.9) mm long .................................................. *C. hepatica*

3. Leaves mostly 5- to 9-lobed. Corolla 8–11 mm long .................................................. *C. muralis* subsp. *muralis*
3. Lower lip of the corolla 8.2–16.6(17.7) mm wide. 9–12 seeds per capsule. Seeds 1.1–1.3 mm long, alveolate to tuberculate .................................................. *C. muelleri*
4. Stems, petioles and pedicels highly fragile when fresh, villous. Corolla white to pale violet, ratio spur length/spur width 0.9–2.8. Seeds 0.6–0.9 mm long, smooth to finely alveolate .......................... *C. fragilis*

4. Stems, petioles and pedicels not fragile, sub-glabrous to villous. Corolla pale violet to violet or pink, ratio spur length/spur width 1.6–5.6. Seeds 0.8–1.1 (1.2) mm long, alveolate .......................... *C. aequitriloba*

**Methods**

**Plant material and DNA extraction.** Leaf material for molecular analyses was collected in the field in 2012 and 2013, dried and stored in silica gel. The sampling localities are shown in Fig. 1 and voucher details are provided in Supplementary Table S1. Total genomic DNA was extracted from ca. 10–30 mg of dried leaf material following the CTAB protocol with some modifications. When less than 10 mg of dried material was available, no sorbitol washing was applied.

**AFLP fingerprinting.** AFLP fingerprinting was performed with usually eight to ten individuals per population (Supplementary Table S1), using *C. microcalyx* (Boiss.) Wettst. and *C. pubescens* (C.Presl) Cuifod. as outgroup taxa. AFLP profiles were generated following established protocols with modifications. Three blanks (DNA replaced by water) were included to test for contamination, and 32 samples (10.5%) were used as replicates between PCR batches to test the reproducibility of AFLP fingerprinting. Based on an initial primer trial the following three selective primer combinations were chosen for selective PCR (fluorescent dye in brackets): EcoR I (6-FAM)ACA/MseI-CAT, EcoR I (VIC)AGG/MseI-CAG, and EcoR I (NED)AAC/MseI-CAG (6-FAM labelled primers: Sigma-Aldrich; NED and VIC labelled primers: Applied Biosystems). Selective PCR products were purified using Sephadex G-50 Fine (GE Healthcare Bio-Sciences, Uppsala, Sweden) applied to a MultiScreen-HV plate for 1, 1 and 5 min, respectively. Then 0.8 μl of the elution product was mixed with 10 μl formamide (Applied Biosystems) and 0.125 μl GeneScan 500 ROX (Applied Biosystems) and run on an ABI 3130 automated capillary sequencer at the Department of Botany & Biodiversity Research of the University of Vienna, Austria.

**Analyses of AFLP data.** Electropherograms were analysed with Peak Scanner version 1.0 (Applied Biosystems) using default peak parameters except employing light peak smoothing. The minimum fluorescent threshold was set to 50 relative fluorescence units. Automated binning and scoring of the AFLP fragments were performed using RawGeno 2.0-1.0® for R 2.15.0® with the following settings: scoring range 140–500 bp, minimum intensity 50 relative fluorescence units, minimum bin width 1 bp, and maximum bin width 1.5 bp. We chose not to score AFLP fragments shorter than 140 bp as short fragments are more prone to homoplasmy than long ones. Fragments with a reproducibility lower than 80% based on sample-replicate comparisons were eliminated. The error rate was calculated as the ratio of mismatches (scoring 1 versus 0) over phenotypic comparisons in AFLP profiles of replicated individuals. Fragments present/absent in only one individual were excluded. A Neighbour-Joining (NJ) analysis based on Nei-Li genetic distances was conducted and bootstrapped (2000 pseudo-replicates) with TREECON v.1.3b.

As an alternative approach to infer the relationships among the studied species, we conducted a species tree analysis using SNAPP. SNAPP is a Bayesian MCMC sampler, which interfaces with the BEAST package to

---

**Identification Key.** Here we provide a dichotomous key to all *Cymbalaria* species occurring in BCS.

1. Leaves and stems glabrous or only with occasional hairs on young organs .................................................. 2
2. Leaves entire to 5-lobed. Corolla 9.7–12.6(13.9) mm long .................................................. *C. hepatica*

3. Leaves mostly 5- to 9-lobed. Corolla 8–11 mm long .................................................. *C. muralis* subsp. *muralis*
3. Lower lip of the corolla 8.2–16.6(17.7) mm wide. 9–12 seeds per capsule. Seeds 1.1–1.3 mm long, alveolate to tuberculate .................................................. *C. muelleri*
4. Stems, petioles and pedicels highly fragile when fresh, villous. Corolla white to pale violet, ratio spur length/spur width 0.9–2.8. Seeds 0.6–0.9 mm long, smooth to finely alveolate .......................... *C. fragilis*

4. Stems, petioles and pedicels not fragile, sub-glabrous to villous. Corolla pale violet to violet or pink, ratio spur length/spur width 1.6–5.6. Seeds 0.8–1.1 (1.2) mm long, alveolate .......................... *C. aequitriloba*
estimate species trees using a coalescent multispecies approach from biallelic input data. Due to the long run times, we reduced the sampling to two individuals per population. Three MCMCs were initially intended to run for 10 million generations, sampling trees every 1000 generations. However, runs were stopped before reaching the final generation number after checking the convergence of most parameters in Tracer 1.6.0103 due to the long run times. Therefore, the first MCMC was run for 153000 generations, the second for 126000 generations, and the third for 392000 generations. After discarding the first 10% of trees of each run, the resulting 980 trees were combined with LogCombiner v.1.8.2, and the topologies were visualized in DensiTree v2.2.5104.

A Bayesian phylogeographical diffusion model in continuous space105 was performed on the AFLP dataset using BEAST v1.8.2106 to infer the main pattern of colonization between the Balearic Islands and Corsica-Sardinia. Geographic input data were coordinates of each population (Supplementary Table S1). We used a simple substitution model with estimated state frequencies for phylogeny inference and the Bayesian skyline coalescent prior107 and a piecewise-linear skyline model was set to model population growth. A strict molecular clock was used due to the simple presence-absence structure of AFLP data. The diffusion process was modelled by a lognormal relaxed random walk process, which is an extension of a phylogenetic Brownian motion process that rescales the precision matrix by a branch-specific scalar that is drawn independently from an identical distribution108, in our case a lognormal distribution centred on 1. We specified a prior exponential distribution on the standard deviation of the lognormal distribution with a mean of 5. We added random jitter with a window size 1.0 to the tips, as more individuals were sampled from the same location. The analysis of the diffusion inference was run for 200 million generations, logging parameters every 5,000 generations. The performance of the analysis as well as mixing and effective sample size values for all parameters were checked in Tracer 1.6.0108. A maximum clade credibility tree (MCC) was produced and annotated with Tree Annotator (part of the BEAST package) after removing 25% of trees as burnin and visualized with FigTree 1.4.2. The diffused MCC tree with annotated diffusion estimates was visualized in SPREAD v1.0.6110 and projected together with polygons representing ancestral areas using ArcGIS 10.3.

Analyses of molecular variance (AMOVAs) and genetic diversity were computed with ARLEQUIN 3.5.2. Groups were defined according to taxa and – for C. aequitriloba – to geographic distribution (islands and archipelagos). Average gene diversity over loci ($\pi$) was obtained for populations and groups after removing populations with fewer than four sampled individuals. Comparison of gene diversity among groups was performed with Student's t-tests after testing for normality with Shapiro-Wilk normality tests.

A separate dataset for C. aequitriloba was prepared and, after independent automated binning and scoring, analysed in the same way. It comprised all populations of the C. aequitriloba clade (see Results) plus C. fragilis populations f8 and f9 as outgroup, based on the results of the NJ analysis for the entire dataset. In addition, we performed nonhierarchical K-means clustering111, a model-free clustering approach. We used a script of Arrigo et al.112 in R107 to identify genetically homogeneous groups in a dataset pruned to C. aequitriloba. We performed 100,000 independent runs (i.e., starting from random points) for each assumed value of K (i.e., the number of groups) ranging from 2 to 10.

**Amplification and sequencing of plastid DNA markers.** We amplified and sequenced the plastid ndhF region and the rpl32-trnL(UAG) spacer for three individuals per population, with the exceptions of populations a20 and m1 – for which we could only obtain sequences from two individuals –, and the outgroup (C. muralis G.Gaertn., B.Mey. & Scherb. and C. pubescens) – with a single individual per population. We used the primers 3'-F113, 3'-F114 and the internal specific primers ndhF CymbF and ndhF CymbR12 for the ndhF region and rpl32F and trnL(UAG)115 for the rpl32-trnL spacer. All reactions were carried out in a MasterCycler Gradient thermocycler (Eppendorf).

For both plastid regions, the reaction mix (total volume 25 μL) contained 9 μL of RedTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich), 1.05 μL bovine serum albumin (BSA; 1 mg mL$^{-1}$; Promega), 1.05 μL of each primer (10 μM) and 1.5 μL template DNA. We followed the PCR profiles described in Galbany-Casals et al.116 for the ndhF and Magauer et al.45 for the rpl32-trnL(UAG) spacer.

The quality of the PCR products was checked on 1% TBE agarose gels. Subsequently, the amplification products were purified enzymatically using Exonuclease I and Shrimp Alkaline Phosphatase (Fermentas) according to the manufacturer’s instructions. Sanger sequencing was conducted by a commercial sequencing facility (Eurofins MWG Operon, Munich, Germany) using the primers 3'-F, 3'-F1, rpl32F and trnL(UAG). Sequences were examined and aligned by hand using Chromas Lite 2.0 (Technelysium Pty Ltd, Tewantin, Australia) and Mega 6.06117. GenBank accession numbers are given in Supplementary Table S1.

**Analyses of sequence data.** The alignment of the concatenated plastid markers was analysed using statistical parsimony as implemented in TCS 1.2118 with the connection limit set to 95%. For this analysis, an indel longer than 1 bp was reduced to a single base pair column allowing this structural mutation to be counted as single base pair mutation only. Two overlapping, but clearly non-homologous indels were detected and considered different. Differences on the indel length (1 bp) caused by different lengths of poly-T regions were not considered due to the high likelihood of homoplastic evolution119. Specimens with missing data in polymorphic positions were excluded to avoid ambiguous haplotype assignments. After these modifications, gaps were treated as fifth character state in this analysis.

Maximum parsimony (MP) analyses as well as MP bootstrap analyses of both data sets were performed using PAUP 4.0b1020. Indels were coded as binary characters using the simple indel coding method121 using SeqState 1.4.1122. The most parsimonious trees were searched heuristically with 1000 replicates of random sequence addition, TBR swapping, and MulTrees on. The swapping was performed on a maximum of 1000 trees (nchuck = 1000). All characters were equally weighted and unordered. The data set was bootstrapped using full
heuristics, 1000 replicates, TBR branch swapping, MulTrees option off, and random addition sequence with five replicates. *Cymbalaria pubescens* was used as outgroup based on a previous study. Bayesian Inference (BI) analysis was performed employing MrBayes v.3.2[12] applying the GTR + G substitution model proposed by the Akaike information criterion implemented in JModelTest 0.1.12[14]. Indels were coded as binary characters using the simple indel coding method[12,19]. We generated 10,000 trees running MrBayes for 5,000,000 generations and sampling one of every 500 generations. After ensuring that the Markov chain Monte Carlo (MCMC) reached stationarity, we discarded the first 2500 trees as burn-in.

**Morphometric analyses.** Twenty-seven characters were scored on the basis of previous studies on the tribe Antirrhineae[20-125], of which 18 were quantitative and seven were ordinal (Supplementary Table S4). Given that ordination analyses perform better with quantitative characters[126], qualitative characters such as stem fragility and leave turgescence, which are diagnostic characters in the group, have not been included in these analyses. In *Cymbalaria*, indumentum and seed ornamentation, treated here as ordinal characters, have been considered of main diagnostic importance[21,23,127,128]. Terminology used for indumentum characters follows Payne[126] and that for seed characters follows Sutton[20]. Seven characters corresponded to vegetative features, 11 to floral features, and seven to fruit and seed features. For the vegetative characters, three measurements per specimen were averaged when possible. As it was not possible to score both floral and fruit or seed characters for all populations, we analysed two datasets. Dataset 1 included 11 floral and seven vegetative characters and dataset 2 included seven fruit and seven vegetative characters. Dataset 1 comprised 129 specimens from 31 populations and dataset 2 comprised 85 specimens from 33 populations (Supplementary Table S1). All specimens used in these analyses were also included in the molecular analyses, with the exception of *C. muelleri* and some individuals from population a5 of *C. aequitriloba* included in dataset 1, which were added later in order to provide floral measurements. Floral measurements were performed in the field with a caliper. Features of indumentum, calyx, fruits and seeds were examined under a binocular stereoscopic microscope. The other characters were measured on scanned specimens using ImageJ[130]. Analyses were conducted using a set of R functions contained in MorphoTools ver. 1.01[131]. Pearson and Spearman correlation coefficients were computed to reveal correlation structure among the characters and to ensure that no strong correlations (>0.95) were present. After standardization to zero mean and unit variance, principal component analysis (zero-centred PCA based on a covariance matrix) was applied to datasets 1 and 2 to display the overall variation pattern along the first two components. A canonical discriminant analysis (CDA) was performed on dataset 1 to assess the morphological differentiation among the four taxa. Morphological intermediate specimens between *C. fragilis* and *C. aequitriloba* from populations f1, f3, f4, f5, f8 and f9 were included in *C. fragilis* according to the results of the molecular analyses and the PCA.

**Data Accessibility**

DNA sequences have been deposited in GenBank (see Supplementary Table S1). AFLP data matrix, alignments of concatenated plastid sequences and the xml files with details of the BEAST analyses are available at Dryad: https://doi.org/10.5061/dryad.0cr574s.

**References**

1. Vogiatzakis, I. N., Pungetti, G. & Mannion, A. M. Mediterranean island landscapes: natural and cultural approaches (Springer Publishing, New York, 2008).

2. Médail, F. & Quézel, P. Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. *Conserv. Biol.* **13**, 1510–1513 (1999).

3. Thompson, J. D. Plant evolution in the Mediterranean. (Oxford University Press, Oxford, 2005).

4. Nieto Feliner, G. Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspect. Plant Ecol. Evol. Syst.* **16**, 265–278 (2014).

5. Rosenbaum, G., Lister, G. S. & Duboz, C. Reconstruction of the tectonic evolution of the western Mediterranean since the Oligocene. *J. Vrtr. Ex.* **8**, 107–130 (2002).

6. Speranza, F. et al. Age of the Corsica–Sardinia rotation and Liguro–Provencal Basin spreading: new palaeomagnetic and Ar/Ar evidence. *Tectonophysics* **347**, 231–251 (2002).

7. Contandriopoulos, J. & Cardona, M. A. Caractère original de la flore endémique des Baléares. *Bot. Helv.* **94**, 101–132 (1984).

8. Contandriopoulos, J. Spécificité de l’endémisme Corse. *Atti Dei Convegni Lincei* **85**, 394–416 (1990).

9. Mansion, G. et al. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean Basin by the angiosperm family Araceae. *Syst. Biol.* **57**, 869–885 (2008).

10. Bobo-Pinilla, J. et al. Phylogeography of *Arenaria balareia* L. (Caryophyllaceae): evolutionary history of a disjunct endemic from the Western Mediterranean continental islands. *Peer. eLife* **4**, e2618 (2016).

11. Molins, A., Bacchetta, G., Rosato, M., Rossello, J. A. & Mayol, M. Molecular phylogeography of *Thymus herba-barona* (Lamiaceae): insight into the evolutionary history of the flora of the western Mediterranean islands. *Taxon* **60**, 1295–1305 (2011).

12. Carnicerio, P., Sáez, L., Garcia-Jacas, N. & Galbany-Casals, M. Different speciation types meet in a Mediterranean genus: the biogeographic history of *Cymbalaria* (Plantaginaceae). *Taxon* **66**, 393–407 (2017).

13. Salvo, G., Ho, S. Y. W., Rosenbaum, G., Ree, R. & Conti, E. Tracing the temporal and spatial origins of island endemics in the Mediterranean region: a case study from the citrus family (*Ruta* L., Rutaceae). *Syst. Biol.* **59**, 705–22 (2010).

14. Fernández-Mazuecos, M., Jiménez-Mejías, P., Rotillán-Puig, X. & Vargas, P. Narrow endemics to Mediterranean islands: moderate genetic diversity but narrow climatic niche of the ancient, critically endangered *Nanphyra* (Apiaceae). *Perspect. Plant Ecol. Evol. Syst.* **16**, 190–202 (2014).

15. Magri, D. et al. The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Mol. Ecol.* **16**, 5259–5266 (2007).

16. Veiga, P. L. et al. Late Pleistocene Paleoclimates and sea-level change in the Mediterranean as inferred from stable isotope and U-series studies of overgrowths on speleothems, Mallorca, Spain. *Quat. Sci. Rev.* **19**, 865–879 (2000).

17. Mansion, G., Selvi, F., Guggisberg, A. & Conti, E. Origin of Mediterranean insular endemics in the Boraginaceae: integrative evidence from molecular dating and ancestral area reconstruction. *J. Biogeogr.* **36**, 1282–1296 (2009).

18. Salvi, D., Harris, D. J., Bombi, P., Carretterio, M. A. & Bologna, M. A. Mitochondrial phylogeography of the Bedriagàs rock lizard, *Arachnoidea aedosa* (Reptilia: Lacertidae) endemic to Corsica and Sardinia. *Mol. Phylogenet. Evol.* **56**, 690–697 (2010).
113. Eldenäs, P., Källersjö, M. & Anderberg, A. Phylogenetic placement and circumscription of tribes Inuleae s. str. and Plucheeae (Asteraceae): evidence from sequences of chloroplast gene ndhF. *Mol. Phylogenet. Evol.* **13**, 50–58 (1999).

114. Kim, K. & Jansen, R. ndhF sequence evolution and the major clades in the sunflower family. *Proc. Natl. Acad. Sci.* **92**, 10379–10383 (1995).

115. Shaw, J., Lickey, E., Schilling, E. E. & Small, R. L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* **94**, 275–288 (2007).

116. Galban-Casals, M., Carnicerio-Campmany, P., Blanco-Moreno, J. M. & Smisken, R. D. Morphological and genetic evidence of contemporary interregional hybridisation in Mediterranean *Helichrysum* (Asteraceae, Gnaphalieae). *Plant Biol.* **14**, 789–800 (2012).

117. Tamura, K., Stecher, G. & Peterson, D. MEGAv6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

118. Clement, M., Posada, D. & Crandall, Ka TCS: a computer program to estimate gene genealogies. *Mol. Evol.* **9**, 1657–9 (2000).

119. Jolliffe, I. T. Principal Component Analysis, 2nd edn. (Springer Series in Statistics, New York, 2002).

120. Tamura, K., Stecher, G. & Peterson, D. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

121. Simmons, M. P. & Ochoterena, H. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**, 369–381 (2000).

122. Müller, K. SeqState. *Appl. Bioinformatics* **4**, 65–69 (2005).

123. Müller, K. SeqState. *Appl. Bioinformatics* **4**, 65–69 (2005).

124. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

125. Sáez, L. & Crespo, M. B. A taxonomic revision of the *Flora d'Italia* (Edagricole, Bologna).

126. Jolliffe, I. T. Principal Component Analysis, 2nd edn. (Springer Series in Statistics, New York, 2002).

127. Webb, D. A. 1972. *Flora Europaea*, Vol. 3, Cymbalaria Hill. (eds Tutin, T.G., Heywood, V. H., Burges, N. A., Moore, D. M., Valentine, D. H., Walters, S. M. & Webb, D. A.) 236–238 (Cambridge University Press, Cambridge, 1972).

128. Pigott, S. 1982. *Flora Iberica* (Edagricole, Bologna).

129. Payne, W. Glossary on plant hair terminology. *Brittonia* **30**, 239–255 (1978).

130. Atramoff, M. D., Magalhães, P. I. & Ram, S. J. Image processing with imageJ. *Biophotonics Int.* **11**, 36–41 (2004).

131. Hufnagl, P., Mithofer, A., Karrasch, S. & Steinkellner, M. Characterization of the *Flora verticillata* (Liliaceae, Liliales) project, CGL2011-28613-C03-01) as well as the Catalan government (2014-SGR 514). Pau Carnicer benefited from the support of a PIF PhD student fellowship from the Universitat Autònoma de Barcelona and an Erasmus+ traineeship mobility fellowship, which supported a research stay at the Department of Botany of the University of Innsbruck.

**Author Contributions**

P.C., M.G. and P.S. designed and conducted the analyses. All authors contributed to the interpretation of results. P.C., M.G. and P.S. drafted the manuscript. All authors revised the manuscript.

**Additional Information**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

© The Author(s) 2018