INTRODUCTION

Islands constitute major plant biodiversity centers in the Mediterranean basin (Médail & Quezel, 1999). Mediterranean islands and islets are not only singular for their species richness, but also for the narrow distribution of most of their flora. Typically, high endemicity rates characterize the flora of many Mediterranean islands. Thus, the endemic element of the Balearic Islands (Médail & Verlaque, 1997), Corsica (Gamisans & al., 1985), Sardinia (Médail & Verlaque, 1997), Sicily (Médail & Verlaque, 1997), Crete (Turland & al., 1993), and Cyprus (Alziar, 1995) ranges from 5.3% to 12% of their floras. Endemic species may even characterize single small islands and islets of much reduced extension throughout the Mediterranean basin.

One of the salient features of the endemic flora of the Mediterranean islands is that very few species are shared between two or more distant archipelagos. In fact, only two single endemic taxa, Helichrysum microphyllum Willd. (Asteraceae) and Cephalaria squamiflora (Sieber) Greuter (Dipsacaceae), have been reported to be restricted to both Western and Eastern Mediterranean islands (Thompson, 2005; Galbany-Casals & al., 2006). Cephalaria squamiflora, together with C. leucantha (L.) Roem. & Schult., C. boetica Boiss. and C. lineari-folia Lange, constitutes the subgenus Fimbriatocarpus Szabó, a small assemblage of perennial species with a narrow distribution centered in the Western Mediterranean basin (Verlaque, 1985; Devesa, 2007). It was hypothesized that subgenus Fimbriatocarpus is primitive within Cephalaria and probably derived from ancestors belonging to the South African subgenus Lophocarpus (Verlaque, 1985). A cladistic study of the Dipsacaceae using morphological and palynological characters showed a monophyletic Dipsacus clade, including Dipsacus and Cephalaria, characterized by a synapomorphic unilobate stigma (Caputo & Cozzolino, 1994). Dipsacus appeared paraphyletic and basal to Cephalaria; Cephalaria subgg. Lophocarpus and Fimbriatocarpus formed the most basal grade in Cephalaria, but were not sister taxa (Caputo & Cozzolino, 1994). Phylogenetic analysis of Dipsacaceae

Cephalaria squamiflora is a chamaephyte restricted to rupicolous habitats in islands of the Western (Balearic Islands, Sardinia) and Eastern Mediterranean (Crete and few Aegean islands). Four narrowly distributed races (subsp. squamiflora, mediterranea, ebusitana, balearica) have been described to encompass the morphological variation within the species. We have used nuclear ribosomal ITS and cpDNA sequences to assess how the patterns of molecular differentiation are related to taxonomic and geographic boundaries. Extensive intragenomic ITS variation was detected in samples from all territories, the average sequence divergence among cloned ribotypes was 1.339%. The parsimony network of cloned ITS sequences suggests a split between Eastern and Western Mediterranean accessions. Chloroplast DNA sequences showed five distinct haplotypes, only one of which was shared between islands (Majorca and Sardinia). Both nuclear and cpDNA markers supported the monophyly of the C. squamiflora complex and identified a highly structured pattern of molecular variation composed by sister monophyletic lineages that mirror major biogeographic units (Western and Eastern Mediterranean). The molecular evidence supports the hypotheses that vicariance events linked to the geological history of the region or dispersal across the Mediterranean may explain the distribution of the complex.

KEYWORDS: continental islands, Dipsacaceae, ITS paralogs, trnL-trnF spacer
using nuclear ribosomal and chloroplast DNA sequences supported the Dipscacus clade, but further resolution within this clade was very limited due to the low sampling (Caputo & al., 2004).

Cephalaria squamiflora is a long-lived rupicolous chamaephyte that lives in calcareous rocky places (crevices of vertical cliffs and overhangs), forming scattered populations from sea level to about 1,500 m altitude. The species shows a distribution restricted to the Balearic Islands (Majorca, Ibiza) and Sardinia in the Western Mediterranean, and Crete and several Aegean islands (Karpathos, Ikaria, Chios, Alonissos, Kyra Panagia, Gioura, Skyros, Sifnos) in the Eastern Mediterranean (Greuter, 1967; Arrigoni, 1978, Mus & al., 1990; Fielding & Turland, 2005; Bacchetta & al., 2008).

The populations of C. squamiflora show an overall morphological similarity. However, subtle variation in leaf shape and hairiness, shape of floral bracts, and epicalyx ornamentation has been reported to be correlated with geography (Illario, 1938; Greuter, 1967; Pignatti, 1977; Arrigoni, 1978), allowing the recognition of up to five very closely related races. With the exception of Bacchetta & al. (2008, who adopt a very narrow species concept, all authors dealing with the C. squamiflora group (Greuter, 1967; Pignatti, 1977; Arrigoni, 1978, Mus & al., 1990; Devesa, 2007) have treated the major geographic variants as subspecies of C. squamiflora: subsp. squamiflora (Crete, Aegean islands), subsp. mediterranea (Viv.) Pignatti (Sardinia, Ibiza), and subsp. balearica (Coss. ex Willk.) Greuter (Majorca). However, no agreement has been reached concerning the number and delimitation of taxonomic entities from the Western Mediterranean. Thus, Greuter (1967) pointed out that the recognition of subsp. mediterranea from subsp. balearica on morphological grounds was not justified and suggested that both subspecies should be merged within a single taxon. In addition, authors have suggested that subsp. mediterranea should include all populations from Sardinia and Ibiza (Mus & al., 1990; Devesa, 2007). But this view it is not followed by others claiming that the name subsp. mediterranea should be restricted to Sardinian individuals and the populations from Ibiza referred to a distinct entity either at the subspecific, C. squamiflora subsp. ebusitana (O. Bolòs & Vigo) Romo (Bolòs, 1991; Romo, 1994) or specific level, C. ebusitana (O. Bolòs & Vigo) Bacch., Brullo & Giusso (Bacchetta & al., 2008). Recently, Bacchetta & al. (2008) have described a new species (C. bigazzii Bacch., Brullo & Giusso) from a single population of SW Sardinia previously identified as C. squamiflora subsp. mediterranea (Ballero & al., 2000; Maxia & Usai, 2004).

Cephalaria squamiflora belongs to the large group of endemic insular species growing exclusively in rupicolous habitats. These environments are usually characterized by singular plant communities of very limited and fragmented distribution, including hotspots of endemism and refuge centers for relict taxa (e.g., Simmons & al., 1998; Bacchetta & al., 2008). Morphological variability within populations is usually low in rupicolous species, but discontinuities are usually present among allopatric populations. This geographical variability can lead to unsatisfactory taxonomic treatments, since the peripheral populations of allopatric species may sometimes be linked by a mosaic of morphologically intermediate, but geographically still discontinuous, non-interbreeding and long-isolated populations. A further question to be addressed in chasmophytes groups is to what extent the groups that taxonomists have defined as cohesive conspecific populations may be the products of recurrent evolution from separate lineages (Levin, 2001). This is a possibility given the similar biological features associated with saxicolous environments which have been favoured or retained by similar selective pressures in unrelated taxonomic groups (Davis, 1951).

Due to its exclusively chasmophyte habitat, interpopulation differentiation, and restricted and discontinuous distribution, C. squamiflora is a suitable taxon (1) to assess whether molecular discontinuities exist within the species, and if these are linked to island or archipelago boundaries, (2) to explore how many evolutionary lineages may be detected within the species, (3) to assess the patterns of molecular differentiation in narrowly distributed insular endemics using molecular markers. In this paper we report the patterns of molecular variation within C. squamiflora using sequences obtained from the nuclear and cpDNA genomes. The nuclear ribosomal ITS region (ITS1-5.8S-ITS2), the trnL (UAA) intron and the cpDNA intergenic spacer trnP (GAA) have been previously used to infer evolutionary relationships at various taxonomic levels in Dipsacaceae (Zhang & al., 2003; Caputo & al., 2004) and could constitute molecular targets to assess patterns of molecular variation within Cephalaria species.

MATERIALS AND METHODS

Plant material and DNA extraction. — Populations of the C. squamiflora group were sampled throughout their distribution range (Appendix; Fig. 1). The report of the species in Corsica (Viviani, 1825) is likely due to a herbarium labeling error (Arrigoni, 1978) and its presence has not been reconfirmed in the island in recent times (Gamisans & Jeanmonod, 1993; Gamisans & Marzocchi, 1996). For the purporses of this paper C. ebusitana and C. bigazzii have been included under C. squamiflora subsp. mediterranea, and C. balearica has been treated at the subspecific level under C. squamiflora. Total DNA was extracted from fresh, silica gel–dried leaves or
herbarium specimens using the CTAB protocol of Doyle & Doyle (1987), scaled down to perform the process in 1.5 mL microfuge tubes.

**Nuclear ribosomal ITS sequences.** — The region including ITS-1, 5.8S and ITS-2 was amplified using the primer pair ITS-5/ITS-4 (White & al., 1990). In some instances, direct sequences were unreadable from the electrophoreograms due to the presence of putative paralogous products of different length. In these cases, the primers ITS-2 and ITS-3 (White & al., 1990) were also used to determine the bases of the remaining sequence. PCR reactions were performed in 50 μl, containing 75 mM Tris-HCl (pH 9.0 at 25°C), 5 mM KCl, 20 mM (NH₄)₂SO₄, 0.001% BSA, 1.5 mM MgCl₂, 5% DMSO, 200 μM of each dATP, dCTP, dGTP and dTTP, 0.2 μM of each primer, approximately 50–100 ng of genomic DNA and 0.75 units of Taq polymerase. Amplifications comprised an initial incubation step at 94°C for 3 min and 40 cycles of 15 s at 94°C, 15 s at 56°C, 30 s at 72°C. A final extension step at 72°C for 5 min was included. The PCR products were separated on 1.0% agarose gels and purified using the High Pure PCR Product Purification Kit (Roche Diagnostics). PCR products from eight accessions encompassing all major biogeographic territories (Western and Eastern Balearics, Sardinia, Aegean islands) were gel-purified and ligated into the vector provided with the p-GEM-T (Promega) cloning kit. Plasmid DNA from individual recombinant colonies was isolated according to a miniprep protocol (High Pure Plasmid Isolation Kit, Roche Diagnostics). For sequencing, purified PCR or cloned ITS products were reacted with BigDye Terminator Cycle Sequencing Ready Reaction (Perkin-Elmer, Applied Biosystems) using the amplification primers. GenBank accession numbers are provided in the Appendix.

**Integrity and secondary structure stability of the ribosomal sequences.** — The boundaries of the ITS sequences and ribosomal coding regions were determined by comparison with published sequences of Dipsacaceae.

![Fig. 1. Distribution and sampling locations of Cephalaria squamiflora. Inset 1 (Balearic Islands), C. squamiflora subsp. balearica (squares) and C. squamiflora subsp. mediterranea (circles); inset 2 (Sardinia), C. squamiflora subsp. mediterranea (circles); inset 3 (Crete, Aegean islands), C. squamiflora subsp. squamiflora (polygons). Details of accession codes are indicated in the Appendix.](image-url)
available in GenBank (Caputo & al., 2004). Sequences were compared for length variation and guanine plus cytosine content (G+C%, hereafter). We explored the DNA sequences for the presence of several structural motifs. Thus, at the ITS-1 region we searched for the presence and length of the conserved angiosperm motif GGCRY-(4 to 7)n-GYYGYCAAGGAA (Liu & Schardl, 1994). This core sequence is most often associated with part of a hairpin structure, whereas the AAGGAA motif is predicted to stand as a non base-paired sequence in an internal loop structure. It has been hypothesized that the whole or partial conserved motif serves as a critical processing signal in the enzymatic processing of the ribosomal RNA (Liu & Schardl, 1994). Lastly, secondary structures of the ITS-2 RNA transcripts were examined for the presence of the tetranucleotide UGGU, which is preserved across the flowering plants. This motif is located within the C4 domain and is contained near the apex of helix III on the 5’ side (Mai & Coleman, 1997). Predicted secondary structures of the ITS-1 and ITS-2 RNA transcripts and associated free energy values were evaluated using the minimum-free energy (MFE) algorithm (Zuker, 1989) and the latest free energy rules (Mathews & al., 1999; Zuker & al., 1999). Fold predictions were made at the M. Zuker web server (Zuker, 2003; http://www.bioinfo. rpi.edu/~zukerm/rna/) using the MFOLD program (version 3.1). Foldings were conducted at 37°C using a search within 5% of the thermodynamic optimality set. Individual bases within foldings were annotated with the descriptor P-num (Jaeger & al., 1989, 1990), which indicates the propensity of individual nucleotides to participate in base pairs and consequently whether or not a predicted base pair is well determined. RNA structures were displayed with the RnaViz package (De Rijk & De Wachter, 1997).

**cpDNA sequences.** — The region including the trnL intron and the intergenic spacer between the trnL (3’) and trnF (5’) coding sequences was amplified using the universal primers e and f described in Taberlet & al. (1991). PCR reactions, purifications and direct sequencing were performed as above.

**Phylogenetic analyses.** — Cephalaria sequences were of similar length and were manually aligned and edited with DAMBE (Xia & Xie, 2001).

Network methods are designed for DNA haplotypes and are believed to work better with infraspecific phylogenies than standard parsimony algorithms (Posada & Crandall, 2001) since it makes fewer assumptions about the direction of evolution, the extent of sexual isolation, and the pattern of ancestry and descent (Allaby & Brown, 2001). The optimal use of networks when analyzing ribosomal multigene families subjected to homogenization processes, such as unequal crossing over and biased gene conversion has been reported (Allaby & Brown, 2001). Poor resolution using multigene families when conventional tree building methodologies are used is likely due on the facts that ribosomal repeat units with multiple ancestry are related by a multifurcating rather than dichotomous branching pattern and the true relationship between groups of sequences is reticulate rather than linear (Allaby & Brown, 2001).

Parsimony network was implemented with the Network version 4.5 software (available at the website http:// www.fluxus-engineering.com) using the Reduced Median (RM) network method (Bandelt & al., 1995). In order to compare the parsimony network with other tree building methods, the datasets were analyzed with conventional parsimony and distance-based methods.

Phylogenetic analyses using standard parsimony were conducted using Fitch parsimony with equal weighting of all characters and of transitions/transversions as implemented in PAUP*4.0b10 (Swofford, 2002). Heuristic searches were replicated 100 times with random taxon-addition sequences, Tree Bisection-Reconnection (TBR) branch swapping, and with the options Multrees and Steepest Descent in effect. Support for monophyletic groups was assessed by both “fast” bootstrapping (10,000 resamplings of the data) and “full” bootstrapping (100 resamplings) using the heuristic search strategy as indicated above (see Mort & al., 2000 for discussion). In addition, phylogenetic reconstructions of ITS sequences were also performed by using distance-based methods using PAUP* 4.0b10, the Kimura 2-parameter distance model (Kimura, 1980), and the Neighbor-Joining method (Saitou & Nei, 1987). For parsimony analysis, inferred indels from the aligned sequences (except a variable poly-A satellite region) were coded as an additional binary data and added to the molecular matrix. Available GenBank sequences from Cephalaria (C. leucantha: ITS-1, AJ426523; ITS-2, AJ426523; trnL intron, AJ 427376; C. syriaca: ITS-1, AJ426525; ITS-2, AJ426526; trnL intron, AJ427377) and Dipsacus species (that are sister to Cephalaria; Caputo & al., 2004) were downloaded for comparative purposes.

**RESULTS**

**Ribosomal ITS sequences: direct sequencing.** — Direct ITS sequences from Cephalaria accessions showed polymorphic sites at 33 positions, of which 14 were located in the ITS-1 and 19 in the ITS-2 region. Unreadable electrophoreograms were found in direct sequences at the ITS-1 region in all the Western Mediterranean accessions analyzed. Particularly, the sequences could not be read after a poly-A tract located at the 5’ end. The three accessions from the Aegean islands showed unreadable sequences near the middle of the ITS-2 region in forward and reverse sequencing electrophoreograms. The displacement observed between peaks in both
spacers suggests that intragenomic ITS variants differing in length could be present within the PCR-amplified ribosomal pool. In addition, most accessions showed double peaks in different sites at the ITS-1 and ITS-2 regions. Overall, we inferred 8 intragenomic polymorphisms in ITS-1 and 18 in ITS-2. Given such sequence ambiguities it was necessary to clone the PCR products in order to assess the number of intragenomic ribotypes present in selected accessions.

**Cloned sequences.** — We cloned the ribosomal products amplified from eight individuals coming from separate geographic regions for which direct ITS sequences were previously generated: one from the Aegean islands (Karpathos), four from Sardinia (Lisandrus, Mt. Irveri, Mt. Fumai, Janna Nurai), two from Ibiza (Cap Jueu, Ses Roques Altes), and one from Majorca (Son Torrella). Altogether, we sequenced 50 clones, 5 from subsp. *squamiflora*, 8 from subsp. *balearica*, and 37 from subsp. *mediterranea* (25 from Sardinia and 12 from Ibiza accessions). The number of analyzed clones per individual ranged from three to ten. The G+C% was similar in all clones, ranging from 63.8 to 65.2, and averaged 64.5%. The length of ITS-1 varied from 246 to 250 bp. Most of the length variation was due to the presence of a poly-A stretch (showing between six and ten adenines) located at the 3′ end of the ITS-1 spacer. In addition, clones from subsp. *squamiflora* showed slightly shorter ITS-1 sequences than those from subspp. *mediterranea* and *balearica* due to an indel 1 bp long. The number of variable sites (excluding length variants) in this spacer was twelve. ITS-2 showed a uniform length of 248 bp, and 17 polymorphic sites were present across the whole dataset. Overall, 32 distinct clones were detected and the average sequence divergence among ribotypes (Kimura 2-parameter model) was 1.339% ± 0.002. The most divergent ribotypes were from subsp. *squamiflora* (Mt. Kali Limni, KA) and subsp. *mediterranea* (Lisandrus, BU), differing by 3.311% ± 0.007. Sequence divergence values within and between ribotypes retrieved from each subspecies, territories and accessions are depicted in Table 1. Within single accessions, intragenomic divergence among ribotypes ranged between 0.000 ± 0.000 (C. *squamiflora* subsp. *mediterranea* from Fumai, FU), but only three clones sequenced) to 1.328% ± 0.003 (C. *squamiflora* subsp. *mediterranea* from Lisandrus, BU). Within territories, ribosomal sequences from Sardinian *C. squamiflora* subsp. *mediterranea* showed the highest sequence divergence values (2.265% ± 0.006), whereas subsp. *squamiflora* and Ibiza subsp. *mediterranea* samples showed the lower values (1.221 ± 0.004 and 1.227 ± 0.004), respectively.

**Comparison of direct and cloned ITS sequences.** — Usually, the polymorphic sites found in the cloned sequences matched the diversity found in the electrophoretograms of the direct sequences. However, in some accessions cloning revealed the presence of bases that were not apparent in the direct sequences. These intragenomic polymorphisms could be either related to cloning or sequencing artifacts. In addition, due to a non-exhaustive cloning screening effort, not all the gene pool of ribotypes that should be present within individual genomes, as inferred by direct sequencing, was retrieved. A close inspection of both direct and cloned sequences

### Table 1. Number of sequenced clones, number of ribotypes, and sequence divergence (Kimura 2-parameter) within accessions and among ribosomal ITS clones retrieved from *C. squamiflora*.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Accession</th>
<th>Sequenced clones (ribotypes)</th>
<th>Sequence divergence [%] ± standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Min.</td>
</tr>
<tr>
<td>Cephalaria squamiflora</td>
<td>subsp. <em>squamiflora</em></td>
<td>Karpathos KA (5 (4))</td>
<td>0.000 ± 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subsp. <em>mediterranea</em></td>
<td>Batlló BU (10 (8))</td>
<td>10.000 ± 0.000</td>
<td>1.328 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>Janna Nurai IR (6 (4))</td>
<td>0.202 ± 0.002</td>
<td>0.814 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>Cap Jueu CJ (6 (4))</td>
<td>1.000 ± 0.000</td>
<td>1.152 ± 0.002</td>
</tr>
<tr>
<td></td>
<td>Ses Roques Altes RA (6 (4))</td>
<td>0.000 ± 0.000</td>
<td>0.612 ± 0.002</td>
</tr>
<tr>
<td></td>
<td>Mt. Irveri IR (6 (4))</td>
<td>0.202 ± 0.002</td>
<td>0.814 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>Lisandrus BU (10 (8))</td>
<td>10.000 ± 0.000</td>
<td>1.328 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>Son Torrella JN (6 (4))</td>
<td>0.000 ± 0.000</td>
<td>0.612 ± 0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subsp. <em>balearica</em></td>
<td>Majorca ST (8 (4))</td>
<td>0.000 ± 0.002</td>
<td>1.055 ± 0.003</td>
</tr>
</tbody>
</table>

5
revealed (1) conspicuous fixed differences (seven mutations) between Western and Eastern Mediterranean samples; (2) shared ribotypes between accessions from different islands were extremely rare, and in fact only a single ribotype was shared between subsp. *mediterranea* (Ibiza) and subsp. *balearica*; (3) no diagnostic fixed ITS markers differentiated subsp. *balearica* and subsp. *mediterranea* accessions.

**ITS sequences in other Cephalaria and Dipsacaceae species.** — Inspection of the aligned sequences in the *C. squamiflora* group with those available of the *Cephalaria-Dipsacus* clade showed (1) that intragenomic polymorphisms (as inferred from base ambiguities reported from direct readings) were apparently rare (C. *leucantha*: three indeterminations, AJ426523, AJ426524) or absent (C. *syriaca*: AJ426525, AJ426526; *D. mitis*: AY236187; *D. pilosus*: AY290016; *D. sylvestris*: AJ426527, AJ426528); (2) that the intragenomic polymorphic sites present in the *C. squamiflora* group were monomorphic in the other *Cephalaria* species; and (3) intragenomic polymorphisms in *C. leucantha* were monomorphic in the other congeneric taxa.

**Fig. 2. Parsimony network of cloned ITS sequences from *C. squamiflora*.** Sequences from *C. leucantha* (leu) and *C. syriaca* (syr) have been incorporated for comparative purposes. Open symbols, inferred ribotypes not sampled; stars, *C. squamiflora* subsp. *squamiflora*; black circles, *C. squamiflora* subsp. *balearica*; grey circles, *C. squamiflora* subsp. *mediterranea* (ribotypes from Ibiza show a black ring). Size of the symbols is proportional to the number of analyzed ribotypes. Numbers close to shaded ribotypes indicate the bootstrap values of the associated clades assessed by maximum parsimony.

**Structural integrity and RNA secondary structure.** — Cloned sequences from *C. squamiflora* showed structural landmarks typical of angiosperm ribosomal ITS spacers (Mayol & Rosselló, 2001; Nieto Feliner & Rosselló, 2007): (1) the presence in the ITS-1 region of the universal motif GGCRY-(5n)-GYGYCAAGGAA [GGCGCGATCTGCGCCAAGGA] that is associated with part of a hairpin structure and a non-base-paired sequence in an internal loop structure; (2) the universally conserved pyrimidine-pyrimidine mismatch in helix II of the predicted structure of the ITS-2 RNA transcript (C U, U U); (3) the predicted secondary models of the ITS-2 RNA transcripts conform well with a cruciform (four helix) structure; in addition, helix III is the longest stem-loop, as early reported for flowering plants; (4) thermodynamically, the secondary structures of the RNA transcripts of all sequences were similar and stable, as inferred from the free energy values (near 98 kcal/mole) of the folded sequences; (5) we noted the presence of the tetranucleotide UGGU, contained near the apex of helix III of the ITS-2 RNA transcript on the 5′ side.

**Phylogenetic analysis.** — The parsimony network of cloned ITS sequences is depicted in Fig. 2. Complex patterns were present in the network, including several unresolved loops that are the result of homoplasious sites. With the single exception of clones from subsp. *squamiflora* (that were grouped together and differ by the remaining *Cephalaria* accessions by a minimum of eight mutational events), there were no major groupings related to geographic or taxonomic boundaries. Clones from both Majorcan and Sardinian accessions were located on opposite terminals of the network, whereas Ibiza sequences were preferentially located between them. The direct ITS sequence of *Cephalaria leucantha*, the other species of subgenus *Fimbriatocarpus*, was linked closely to the Ibiza clones, differing by a minimum of only three mutations. By contrast, *C. syriaca* was clearly differentiated in the network and differed from other *Cephalaria* taxa by a minimum of 26 mutations. The joint analysis of *Cephalaria* ITS sequences using other phylogenetic methods (data not shown) always supported a monophyletic Eastern Mediterranean clade. However, they do not improve relationships between haplotypes from Western Mediterranean.

**cpDNA sequences.** — The *trnL* intron and coding boundaries ranged from 498 to 505 bp across accessions whereas the *trnL*-*trnF* region showed a uniform length of 438 bp. The concatenated cpDNA region yielded a 949-bp aligned sequence. Only six positions were polymorphic within the ingroup (*trnL* intron: 5; *trnL*-*trnF* spacer: 1), resulting in five distinct haplotypes (Table 2). Haplotype 1 was exclusive of Aegean subsp. *squamiflora* accessions; haplotype 2 was present in samples of subsp. *mediterranea* from Ibiza; haplotypes 3 and 5 were restricted
to subsp. *mediterranea* (Sardinia), and haplotype 4 was shared between samples of subsp. *balearica* and some accessions of subsp. *mediterranea* from Sardinia. The haplotypes of *C. leucantha* and *C. syriaca* were divergent from the ingroup. A parsimony network depicting haplotype relationships is shown in Fig. 3. Haplotype 4, the most frequent in the accessions studied, showed an internal position in the network, from which the haplotypes 1, 2, and 5 derived. Phylogenetic analyses using parsimony and distance methods of an expanded dataset including *C. leucantha*, *C. syriaca*, *Dipsacus sylvestris*, *D. pilosus*, and *D. muticus* strongly support (BS > 90%) a monophyletic *C. squamiflora* clade (data not shown), but further resolution of the group was not obtained.

Fig. 3. Parsimony network and geographical distribution (A, Balearic Islands; B, Sardinia; C, Aegean islands) of cp-DNA haplotypes (*trnL-trnF* region) from *C. squamiflora*. Sequences from *C. leucantha* (leu) and *C. syriaca* (syr) have been incorporated for comparative purposes. Size of the symbols is proportional to the number of analyzed haplotypes. Numbers refer to haplotypes from Table 2.

### Table 2. Polymorphic sites in the *trnL-trnF* region in *C. squamiflora* and congeneric taxa.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Accession</th>
<th>Haplo-</th>
<th><em>trnL</em> intron</th>
<th><em>trnL-trnF</em> spacer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>type</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>89</td>
<td>128</td>
</tr>
<tr>
<td><em>Cephalaria squamiflora</em> subsp. <em>squamiflora</em></td>
<td></td>
<td></td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>KA</td>
<td>1</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>KA-2</td>
<td>1</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>IK</td>
<td>1</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td><em>Cephalaria squamiflora</em> subsp. <em>mediterranea</em></td>
<td></td>
<td></td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>CJ</td>
<td>2</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>RA</td>
<td>2</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>BU</td>
<td>3</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>TU</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>IR</td>
<td>5</td>
<td>A</td>
<td>C</td>
<td>TAAAC</td>
</tr>
<tr>
<td>FU</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>JN</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>UR</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td><em>Cephalaria squamiflora</em> subsp. <em>balearica</em></td>
<td></td>
<td></td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>PM</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>ST</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>CB</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>PA</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>FO</td>
<td>4</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td><em>C. leucantha</em></td>
<td>AJ427376</td>
<td>A</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td><em>C. syriaca</em></td>
<td>AJ427377</td>
<td>T</td>
<td>A</td>
<td>-</td>
</tr>
</tbody>
</table>
DISCUSSION

Ribosomal ITS paralogs in *C. squamiflora*. — The ribosomal ITS region is the most frequently used nuclear region in phylogenetic reconstructions (cf. Baldwin & al., 1995). Although the ITS region forms part of a multigene family with thousands of tandem copies clustering in arrays, and usually several ribosomal loci are present within plant genomes, only a single consensus ITS sequence is usually retrieved after sequencing PCR products from single individuals. Some authors have cautioned about the use of rDNA nucleotide sequences in phylogenetic reconstructions after the finding that major ribosomal loci may move within and among chromosomes and that their movement may potentially occur via magnification of minor loci consisting of a few rDNA copies (Dubcovsky & Dvorak, 1995). However, this violation of the assumed orthology is usually neglected in standard phylogenetic practices. In addition, if only direct, not cloned, sequences are routinely generated from an organism it is highly likely that much of the intragenomic ribosomal diversity present within organisms is overlooked. This is due to the fact that molecular homogenizing mechanisms of the ribosomal multigene family could be relaxed, and from two to several ribotypes can be present within organisms if (1) the rate of mutation among copies is faster than that driving the concerted evolution of the array, (2) duplicate ribosomal loci evolve without selective constraints and accumulate mutations, and (3) homeologous loci from other species are incorporated into a single genome through hybridization-mediated processes. In this study, cloning procedures have revealed that significant intragenomic variation in ribosomal sequences is present in a narrowly restricted species with a low population size. Levels of intraindividual sequence divergence (up to 2.1%, Table 1) strongly suggest that the intragenomic ITS variability found should not only be explained by PCR or sequencing artifacts. The presence of paralogous ribosomal loci within the nuclear genome or, alternatively, the occurrence of heterogeneous ribosomal copies within arrays should be postulated to occur in *C. squamiflora*. The length and G plus C content of the spacers and the coding 5.8S region, the presence of all structural landmarks in the primary DNA sequence, the null rate of substitutions and deletions in the 5.8S region and in the highly conserved motifs, the high thermodynamic stability and the conserved cruciform foldings of the secondary RNA structures suggest that the recovered ITS sequences they are not pseudogenes.

*Cephalaria squamiflora* is diploid (*2n = 18*) and shows two hypothesized plesiomorphic features, like a *x* = 9 basic chromosome number and an asymmetric karyotype (Verlaque, 1985). Since *C. squamiflora* is predominantly allopatric respect to other species of the genus that usually grow in distinct non-rupicolous environments, it is unlikely that the ITS polymorphism could be generated by hybridization events with other extant species. In fact, no other *Cephalaria* species is known to occur on Majorca, Ibiza, Crete, Karpathos and Ikaria territories, and only *C. leucantha* co-occurs with *C. squamiflora* in Sardinia. The inspection of ITS sequences reported for *C. leucantha* does not support the view that this species has contributed to the ITS heterogeneity found in *C. squamiflora*. The haploid chromosome complement of *C. squamiflora* shows secondary constrictions that could be sites of ribosomal loci in five out of nine chromosomes (Verlaque, 1985). This relative high number of secondary constrictions is not unique to *C. squamiflora*, but it has also been reported in other species of *Cephalaria* and in most genera of Dipsacaceae (Verlaque, 1985, 1986a, b). Karyotype rearrangements resulting in gene duplications have been hypothesized to have occurred before the divergence of several genera of the family (*Sca-biosa, Cephalaria, Knautia*, and *Succisa*; Van Treuren & Bijlsma, 1992), and could be also responsible for the expansion of ribosomal loci in Dipsacaceae. Nevertheless, significant intragenomic ribosomal polymorphisms have not been apparently detected in other representatives of Dipsacaceae so far analyzed (Bell, 2004; Caputo & al., 2004). However, this does not necessarily suggest its absence since non-cloned, consensus sequences were generated in both studies. Independently of their origin, the persistence of ribosomal paralogs in *C. squamiflora* implies that the molecular forces driving concerted evolution of this multigene family are not fully operating in this narrowly-distributed species. If, as it is here speculated, the divergent ribosomal families are on loci located in separate chromosomes, this could prevent concerted evolution since it has been suggested that the chromosomal location of rDNA loci could have a more substantial impact than the number of loci on the tempo of concerted evolution through either unequal crossing-over or gene conversion (Zhang & Sang, 1999).

Geographic patterns of molecular variation in *C. squamiflora*. — The distribution of *C. squamiflora*, restricted to three Western and seven Eastern Mediterranean islands, is unique among the biogeographic patterns shown by the approximately 25,000 plant vascular species native to the Mediterranean region (Contandriopoulos & Cardona, 1984; Thompson, 2005); a further entity, *Helichrysum microphyllum*, is restricted to a different combination of Western and Eastern Mediterranean islands. Four major hypotheses can be proposed to explain this pattern of distribution. First, insular populations could have diverged from a common ancestor by the splitting of a continuous range through successive vicarious events. The scattered populations are thus relics of a formerly widespread species that went to extinction throughout
most of its area, excepting a few islands that acted as refuge. Second, during the late Miocene (Messinian period) closing of the Strait of Gibraltar and the concomitant negative hydric balance caused the desiccation of the Mediterranean. Connections across land corridors could have allowed biotic interchanges between circum-Mediterranean territories through dispersal. Third, some insular populations could have been originated by stochastic events involving long range dispersal (either by water, wind or birds) of fruits coming from a distant Eastern or Western gene pool. Lastly, the species could not be a monophyletic lineage and the shared morphological similarities between races could be obtained by convergence, since plants from all populations share the same rupicolous habitat. Under the first two hypotheses we would expect molecular markers to be geographically structured, and molecular footprints should show sister monophyletic lineages. In contrast, under the third hypothesis it would be expected to find an incomplete differentiation among major biogeographic units, and paraphyletic lineages would be detected. The fourth hypothesis would be supported if polyphyletic evolutionary units were depicted within a phylogenetic framework including other congeneric species.

In this work, both cpDNA and ribosomal sequences have shown a highly structured pattern of molecular variation composed by sister monophyletic lineages that mirror major biogeographic units (Western and Eastern Mediterranean). This East-West vicariance is common in the Mediterranean basin and several climatic and historical factors have been postulated to contribute to East-West floristic and species differentiation (Thompson, 2005). Thus, available molecular evidence supports the hypotheses that (1) vicariance events linked to the geological formation of the present geography, or alternatively, (2) dispersal across an almost dry Mediterranean could explain the distribution of the complex. *Cephalaria squamiflora* fruits, although having developed a short fimbriated or scarious involucel-corona (Verlaque, 1984, 1985), are fully covered by the stiff receptacular bracts at ripeness (Caputo & Cozzolino, 1995). This favours dispersal at short distances due to projecting stems and a catapult mechanism (Verlaque, 1984). Since efficient adaptations for dispersion over great distances by wind dispersal and epizoochory are lacking, it is unlikely that long-range dispersal has played a significant role in shaping its distribution and genetic structure across the Mediterranean.

The Messinian salinity crisis has been repeatedly invoked to explain the current distribution of many plant and animal taxa in the Mediterranean basin (Bocquet & al., 1978; Kiefer & Bocquet, 1979). However, dispersal across a dry Mediterranean basin that was likely a saline hot desert would be a prowess beyond the abilities of most species (Altaba, 1998). In view of the similar ecological requirements of *C. squamiflora* throughout its range, it is hard to envisage how a non-halophyte and strict rupicolous species could successfully spread in the unfavourable saline environment. Further, it is difficult to explain how, if a Messinian dispersal is taken as a general explanation for the shaping of its area, this has not resulted in a more widespread, circum-Mediterranean distribution of the species.

The co-occurrence of shared endemic species, or closely related species pairs, restricted to the Balearic archipelago and Corsica and Sardinia islands is a relevant biogeographic feature that has been linked to the paleogeographic history of the Western Mediterranean basin and has been repeatedly invoked as a paradigm for vicariance and gradual plant speciation (Cardona & Contandriopoulos, 1979; Contandriopoulos & Cardona, 1984). Usually, these floristic connections between Corsica, Sardinia and the Balearic Islands involve endemic species occurring in the Eastern (Majorca or Minorca) rather than from the Western (Ibiza and Formentera) Balearic Islands. The single puzzling exception to this pattern is *C. squamiflora*, since subsp. *mediterranea* is shared between Ibiza and Sardinia whereas subsp. *balearica* is endemic to Majorca. Post-Messinian terrestrial connections between the Balearic Islands (including Ibiza) and the Corso-Sardinian microplate have not been documented. Accordingly, unless a long–dispersal event should be invoked, the shared presence of *C. squamiflora* subsp. *mediterranea* in both islands could be a likely remnant of the Oligocene land connections between Western Mediterranean territories (Thompson, 2005).

The Mediterranean basin has a complex geological history, and a different past configuration and situation of the Western islands has been postulated (Thompson, 2005, and references therein). The Proto-Ligurian Massif was an Oligocene territory formed by emerged lands now belonging to the Balearic Islands, the Corso-Sardinian archipelago, Southern France, the Kabylies, Calabria and NE Spain (Alvarez, 1972, 1976; Alvarez & al., 1974; Van der Voo, 1993; Westphal & al., 1973, 1976; Speranza & al., 2002). Successive splitting of the Proto-Ligurian Massif in the Late Oligocene and posterior tectonic movements of microplates led to the segregation of the Corso-Sardinian microplate and the configuration of Balearic land masses. The splitting of the Western and Eastern Balearic islands took place after the Messinian transgression (about 5.3 million years ago; Gautier & al., 1994) and allowed their isolation from other Western Mediterranean continental and insular lands until the present.

 Parsimony network reconstruction of *C. squamiflora* genotypes from Western Mediterranean islands is to a high extent concordant with the postulated paleobiogeographic relationships based on microplates movements.
The extensive ribosomal paralogy found in Western Mediterranean samples of *C. squamiflora* complicates the use of ITS sequences in postulating a sound biogeographic scenario for this area. However, some results are relevant for the relationships of subsp. *mediterranea* accessions from Ibiza and Sardinia. First, contrary to expectations, the unique shared ITS haplotype between islands was from Ibiza (subsp. *mediterranea*) and Majorcan (subsp. *balearica*) samples. Second, ITS sequences from subsp. *mediterranea* from Ibiza are on average more closely related to Majorcan accessions of subsp. *balearica* (Kimura 2-parameter distances: 0.95% ± 0.44) than to Sardinian sequences of subsp. *mediterranea* (1.20% ± 0.53). In addition, results from the chloroplast genome indicated that the haplotype present in Ibiza is derived and originated from the ancestral haplotype 4, which is present in both Majorca and Sardinia islands. Thus, both kinds of markers are consistent with the hypothesis that populations from Ibiza and Majorca derived from a recent ancestor not shared by Sardinian populations.

Molecular markers agree with a biogeographic scenario for the Western Mediterranean involving two vicarious events. The splitting of the ancestral populations of *C. squamiflora* present in the Proto-Ligurian Massif, leading to the Corso-Sardinian plate and its rifting eastward with a counterclockwise rotation (Cherchi & Montadert, 1982; Deino & al., 2001; Rosenbaum & al., 2002; Speranza & al., 2002) is consistent with a first pattern of West-East vicariance (Sardinian and Balearic evolutionary units). The shared presence of the putative ancestral cpDNA haplotype 4 in the populations of the Corso-Sardinian plate and the remaining proto-Balearic Islands suggests a Proto-Ligurian origin for the Western Mediterranean populations of *C. squamiflora*. This Balearic-Sardinian split would explain why populations located in Majorca and Ibiza shared ribosomal paralogs not found in Sardinia. The two private cpDNA haplotypes found in Sardinia could have evolved after the fragmentation of the Proto-Ligurian microplate under insularity environments. A later vicarious event involved the SW-NE fragmentation of the Proto-Balearic islands leading to Western and Eastern subarchipelagos, leading to the persistence of the ancestral cpDNA haplotype in Majorca and the differentiation of the derived haplotype present in Ibiza. If this scenario is correct, the taxonomic adscription of the populations from Ibiza to subsp. *mediterranea* should be revisited, since the overall morphology shared between them and Sardinia samples is not due to recent ancestry but to convergence. Molecular markers are not conclusive to support the splitting of *C. squamiflora* in five species, as recently suggested (Bacchetta & al., 2008). Other fast-evolving regions should be analyzed to assess in more detail the phylogeographic and taxonomic patterns of the Western Mediterranean populations.

We thank M. Mus, N. Torres and M.A. Conesa who helped in locating Balearic populations of *C. squamiflora* and efficiently assisted with field work. This work was partially funded by the project CGL2004-00223/BOS.

**LITERATURE CITED**


Appendix. List of investigated accessions of Cephalaria squamiflora and related taxa, including GenBank codes.

Taxon, location (abbreviation), date, collector (herbarium), GenBank accession numbers