Phylogeny and genetic structure of *Erophaca* (Leguminosae), a East–West Mediterranean disjunct genus from the Tertiary

Ramón Casimiro-Soriguer a,*, María Talavera a, Francisco Balao a, Anass Terrab a,b, Javier Herrera a, Salvador Talavera a

a Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Apdo. 1095, E-41080 Sevilla, Spain

b Department of Systematic and Evolutionary Botany, University of Vienna, Rennweg 14, A-1030 Vienna, Austria

**A R T I C L E   I N F O**

Article history:
Received 24 November 2009
Accepted 19 February 2010
Available online 26 February 2010

**Keywords:**
AFLP
*Astragalus lusitanicus*
*Erophaca baetica*
Genetic diversity
Gibraltar Strait
Leguminosae
nrDNA ITS
Mediterranean
Phylogeography
Phylogeny

**A B S T R A C T**

The genus *Erophaca* comprises a single herbaceous perennial species with two subspecies distributed at opposite ends of the Mediterranean region. We used nrDNA ITS to investigate the phylogeny of the genus, and AFLP markers (9 primers, 20 populations) to establish the genetic relationship between subspecies, and among populations at each side of the Gibraltar Strait. According to nrDNA ITS, *Erophaca* is monophyletic, old (Miocene), and sister to the Astragalean clade. Life form attributes and molecular clock estimates suggest that *Erophaca* is one of the many Tertiary relicts that form part of the present Mediterranean flora. Within the occidental subspecies, European plants are clearly derived from North-African populations (Morocco) which, despite being rare on a regional scale, present the highest genetic diversity (as estimated by private and rare fragment numbers). In general, genetic diversity decreased with increasing distance from Morocco. AFLP and nrDNA ITS markers evidenced that the Eastern and the Western subspecies are genetically distinct. Possible causes for their disjunct distribution are discussed.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The Mediterranean region is considered one of the World’s biodiversity “hotspots” (Myers et al., 2000), but actually it comprises two separate areas of very high diversity, to the East (specially the Balkans and Turkey) and the West (Iberian Peninsula and NW Africa) (Medail and Quezel, 1999; Schmitt, 2007). Many species that now inhabit these areas experienced intermittent shifts in their altitudinal/latitudinal ranges as a result of the climatic oscillations of the Pleistocene (Hewitt, 1996, 2000; Taberlet et al., 1998; Thompson, 2005; Rodríguez-Sánchez et al., 2008), and such readjustments have resulted in a number of interesting biogeographical situations, including a high incidence of endemism and many disjunct distributions of plants and animals (e.g., Davis and Hedge, 1971; Rosselló et al., 2007). In the West Mediterranean, for example, lowering of sea levels during Pliocene and Pleistocene glacial maxima could have permitted migrations to occur across the Strait of Gibraltar (Collina-Girard, 2001; Ortiz et al., 2007), while warmer periods allowed some species to expand and others to become restricted or extinct (Taberlet et al., 1998). Most often the Strait is considered a formidable obstacle to plant expansion (e.g., Terrab et al., 2007; Cano-Maqueda et al., 2008), but there are numerous examples of species that have successfully crossed this barrier, both from the North to the South (e.g., Caujape-Castells and Jansen, 2003; Escudero et al., 2008) and from South to North (e.g., Carranza et al., 2006; Ortiz et al., 2009).

The Legume family is one of the most diverse groups of flowering plants with about 19,000 species (Culham, 2007). Many are major constituents of Mediterranean scrub that covers vast areas in Morocco and the Iberian Peninsula, and they are, therefore, of great ecological importance in the region. However, few molecular studies have been undertaken in this area on legume taxa. Exceptions are the use of chloroplast markers in *Ulex* (Cubas et al., 2005) and *Stauracanthus* (Pardo et al., 2008), and allozymes in *Calicotome* (Arroyo et al., 2008). These studies have shown that the Strait of Gibraltar can be an effective barrier to gene flow in some cases (e.g., *Ulex, Stauracanthus*) but not in others (*Calicotome*). The AFLP (Amplified Fragment Length Polymorphism) technique is appropriate for phylogeographic studies because it allows genotyping of large numbers of loci in many different individuals, and hence the elucidation of genetic population structures and recent phylogenies (Weising et al., 2005). Phylogeographic studies using AFLP have been successfully carried out of legume taxa in Europe and...
America (e.g., *Anthyllis montana*; Kropf et al., 2002; *Astragalus ampullarioides*; Breinholt et al., 2009; *Astragalus cremnophylox*; Travis et al., 1996; *Oxytropis campestris*; Chung et al., 2004), as well as in Mediterranean species of diverse plant groups (Ortiz et al., 2008; Piñeiro et al., 2007; Terrab et al., 2008).

The monotypic genus *Erophaca* Boiss. is endemic to the Mediterranean and its sole species, *E. baetica* (L. Boiss.), was long included in *Astragalus* (as *A. lusitanicus* Lam.) until transferred to *Erophaca* by Podlech (1993), who assigned the genus to the tribe Sophoreae on the basis of morphology. Also on morphological data, Talavera (1999) retained *Erophaca* in the tribe Astragaleae but proposed a new subtribe for it, the Erophacineae. Molecular data (Liston, 1995; Wojciechowski et al., 2000) point to a phylogenetic relationship between *Erophaca* and the Astragalae clade, but as of yet the phylogeny of the genus is not fully understood. Furthermore, *E. baetica* has two subspecies distributed at opposite ends of the Mediterranean area, and is thus a good subject for studying evolutionary changes below the species level over a relatively large geographical range. Lastly, the Western subspecies lives at both sides of the Strait of Gibraltar, and this gives an opportunity to check the effects of a major geographic barrier on the genetic differentiation of populations. As part of a larger study on the ecology, distribution and evolution of *E. baetica* (Casimiro-Soriguer, unpublished), here we used nuclear ribosomal Internal Transcribed Spacers DNA sequencing (nrDNA ITS) and AFLP analyses to: (1) contribute to clarify the phylogenetic status of *Erophaca* within the Leguminosae; (2) assess the genetic differences among its subspecies; and (3) provide insights into the phylogeographic relationships between North African and S. Iberian populations.

### 2. Materials and methods

#### 2.1. Study plant, populations sampled and DNA isolation

Subtribal assignments used in the present study follow Polhill (1981), and Talavera (1999), and the tribe to which *Erophaca* belongs will be referred to as Galegaeae in accordance with most modern systematic works (e.g., Lewis et al., 2005). As for the subspecies of *E. baetica* that populate the West and the East Mediterranean, they are, respectively, the subs. *baetica* (S. Iberian Peninsula, Morocco and Algeria), and the subs. *orientalis* (Chater & Meikle) Podlech (Greece, Turkey, Cyprus and Lebanon). These subspecies can be distinguished morphologically on the basis of leaflet hairiness (appressed hairs present in subsp. *orientalis* vs. glabrous leaflets in subsp. *baetica*); the teeth of the calyx (longer than half calyx in the Eastern subspecies vs. much shorter in the Western one); the color of their ‘stipelles’ (i.e., the stipule-like outgrowths on the base of the leaflets) which are either purple (Eastern) or yellowish (Western subspecies); and the fruits (usually smaller than 70 × 25 cm in subsp. *orientalis*, and larger than that in subsp. *baetica*).

Our study focused primarily on *E. baetica* subsp. *baetica*, endemic to the Iberian Peninsula, Morocco and Algeria (Fig. 1, inset). This is a long-lived, perennial plant with a woody tuber-like organ (xylopodium) and numerous, fast growing herbaceous stems that start to sprout at the beginning of winter (December), in synchrony with the main rainy season under the Mediterranean climate. Plants up to 2 m in height can be found on shady places, but more often they are around 1 m tall. The stems converge at ground level on the xylopodium, so that individuals can be easily differentiated in the field. Vegetative reproduction (e.g., by stolons) has never been observed but, in order to minimize the risk of genotype resampling, only plants more than 4 m apart were used in the present study. From December to March the stems bear clusters of showy white, nectariferous flowers that are visited mainly by bees (Casimiro-Soriguer, unpublished). Ovaries develop into inflated legumes with up to nine, relatively large (7–10 mm) seeds and explosive dehiscence. Biotic seed dispersal agents are unknown. By mid spring (May) the plants shed their leaves, all above ground plant parts senesce and become dry, and the buried xylopodia become dormant until next winter. Throughout its distribution range populations occur as isolated patches, usually in cork oak (*Quercus suber*) woodlands from sea level up to 1300 m. The foliage is very toxic to sheep, goats and cattle (Bel-Kassoua et al., 2008) and for this reason plants are sometimes cut down by shepherds and farmers, presumably to prevent them from spreading (Casimiro-Soriguer, personal observation). Consequently, in parts of its range (particularly Morocco) humans may have decimated populations for centuries. Developing seeds can be heavily preyed upon by Lycaenid butterfly larvae (Jordano et al., 1990).

We studied 20 populations covering most of the distribution range (Fig. 1). Due to the rarity of *E. baetica* in North Africa, and despite considerable sampling effort, only five Moroccan populations could be analyzed. Leaves were picked at random from around 10 individuals per population, stored in silica gel until processed for DNA isolation, and vouchers deposited in the herbarium of the University of Seville, Spain (SEV). Furthermore, in order to investigate the relationship among subspecies, 11 individuals of the Eastern subspecies were sampled from a Greek population. Both for the nrDNA ITS and AFLP analyses, genomic DNA was extracted following the CTAB protocol (Doyle and Doyle, 1987) with modifications (Terrab et al., 2008). Extracts were checked for the presence and amount of DNA on 1% TAE agarose gels.

#### 2.2. nrDNA ITS

Two randomly chosen plants per population, from nine populations of *E. baetica* subsp. *baetica*, and one population of subsp. *orientalis* were analyzed (Genbank No. FNS8233 through FNS8252). Using universal primers ITS4 and ITS5 of White et al. (1990), the nuclear ribosomal Internal Transcribed Spacers regions ITS1, 5.8S and ITS2 were amplified under the following conditions of the polymerase chain reaction (PCR): 18 μl of PCR Reddy mix at 1.4× (supplied by Abgene, Vienna, Austria), 0.32 μl (20 pm) of each primer (forward or reverse), 0.4 μl DMSO, and 1 μl (2–8 ng/μl) of DNA template. Amplified products were purified adding 0.5 μl Exol and 2 μl CIAP (Fermentas, St. Leon-Rot, Germany), then heated for 45 min at 37 °C, and 15 min at 85 °C. The polymerase chain reaction (PCR) sequence profile was one cycle of 1 min at 96 °C, followed by 35 amplification cycles. Each cycle consisted of 10 s denaturation at 96 °C, 5 s annealing at 50 °C, and 3 min elongation at 60 °C. Lastly, the sample was extended at 72 °C for 8 min. The purified fragments were directly sequenced using dye terminator chemistry and the following cycling conditions: 38 cycles of 10 s at 96 °C, 5 s annealing at 50 °C, and 3 min elongation at 60 °C. Lastly, the sample was extended at 72 °C for 8 min. The purified fragments were directly sequenced using dye terminator chemistry and the following cycling conditions: 38 cycles of 10 s at 96 °C, 5 s annealing at 50 °C, and 3 min elongation at 60 °C. The samples were run on a 3130xl Genetic Analyzer capillary sequencer (Applied Biosystems). Forward and reverse sequences were assembled with AutoAssembler ver. 1.4.0 (Applied Biosystems), and alignment of the complete ITS1, 5.8S and ITS2 sequences were performed manually using MEGA version 4 (Tamura et al., 2007). To select the nucleotide substitution model that best fitted the data, Modeltest ver. 3.06 (Posada and Crandall, 1998) was used. According to Akaike Information Criterion the optimal model was GTR+Γ.

The sequences obtained, together with GenBank sequences of a number of closely or distantly related taxa, were used to construct a phylogeny based on heuristic parsimony searches, based on 21 taxa and 779 characters. Computations were performed in PAUP* ver. 4.0 beta 10 (Swofford, 2002), with bootstrapping values calculated using the full heuristic search option (10,000 replicates). GenBank provided sequences for the following taxa (accession numbers): *Anagyris foetida* (AF330637); *Astragalus boeticus* (AF12...
A. membranaceus (AF121675); Caragana hololeuca (AB262524); C. tragacanthoides (AB262536); Chesneya deshungrica (U50350); E. baetica subsp. orientalis (EU070920); Galega orientalis (AY553709); Glycyrrhiza glabra (AY065623); Hedysarum coronarium (AY772225); Oxytropis multiceps (AF121760); O. splendens (AF121761); Sophora flavescens (AF123452); Swainsona purpurea (AF113864); and S. stenodonta (AF113865).

Divergence times (either within E. baetica, or from its ancestors) were estimated from phylogeny using the Bayesian relaxed molecular clock method in BEAST 1.4.7 (Drummond and Rambaut, 2007). BEAST estimates divergence times (branch lengths) using the topology it generates, and a calibration time (uncertainty) that is provided as input. Because the fossil record is poor for the clade to which Erophaca is presumed to belong (i.e., the one that lacks the inverted repeat, or IRLC clade of Wojciechowski et al. (1999)) the clock was calibrated using the estimate by Lavin et al. (2005) of 14.8 ± 2 mya for the divergence of Colutea arborescens and Astragalus americanus (as representative case examples of the subtribe Coluteinae and Astragalinae). Furthermore (and as suggested by the topology of the tree) the genera Swainsona, Astragalus and Oxytropis were chosen to represent the sister-group of Erophaca, while the genus Anagyris was used to root the tree. Other assumptions of the model were: (1) uncorrelated, log-normal prior model of rate change; (2) Yule model for speciation; and (3) automatic tuning of operators. Analyses were run for 100 million generations (sampling every 1000th), with a burn-in of one million generations.

2.3. AFLP

Overall, 200 plants from N. Africa, the Iberian Peninsula, and the E. Mediterranean were studied (Table 1). Following the protocols established by Vos et al. (1995), genomic DNA (0.5 µg on average) was digested with two restriction endonucleases (EcoRI and MseI), fragments ligated to double-stranded adaptors (EcoRI and MseI) at 37 °C for 2 h., then diluted 20-fold with TE0.1 buffer. Fragments with the matching nucleotides were amplified (downstream of the restriction sites) using preselective primers based on EcoRI and MseI adaptors. This resulted in a ~16-fold decrease in the number of fragments eventually amplified. Preselective and selective amplifications were performed in a thermal cycler (Gene

Fig. 1. Distribution and genetic structure in 19 populations of E. baetica subsp. baetica, and geographic range of the taxon (gray area). Populations are identified by their numbers in Table 1, and colors indicate genetic groups inferred from Bayesian clustering. The known distribution of the genus Erophaca is shown as an insert.
Table 1
Population-based estimates of genetic diversity in *E. baetica*. Averages for major geographic areas in bold.

| Area                  | Sub-area       | Pop. | Localitiesa | Elevation (m) | Coordinatesb | Collector No. | N  | AFLP fragments | % Polymorphic | Hc\(^c\) | Fragspc | DWp | Fragsqd | DWqd |
|-----------------------|----------------|------|-------------|---------------|--------------|---------------|----|----------------|---------------|---------|---------|-----|---------|------|-------|
|                       |                |      |             |               |              |               |    |                |               |         |         |     |         |      |       |
| Iberian Peninsula     | Iberian Massif | 1    | Miranda     | 553           | 40°29′55″57″  | RCS et al. 315/06 | 11 | 246            | 73.6          | 0.10 (0.052) | 1        | 3   | 1.8   | 4.4  |        |      |       |
|                       |                | 2    | Guadalupe   | 670           | 39°22′5″20″   | RCS et al. 321/06 | 10 | 252            | 74.2          | 0.11 (0.060) | 2        | 2.0  |        |      |       |      |       |
|                       |                | 3    | Niefa       | 806           | 38°31′4″22″   | RCS et al. 204/06 | 9  | 281            | 83.3          | 0.13 (0.070) | 4        | 2.5  |        |      |       |      |       |
|                       |                | 4    | Cerro Hierro| 658           | 37°57′5″37″   | RCS et al. 2/07 | 10 | 251            | 74.9          | 0.11 (0.059) | 2        | 4    | 1.8   | 4.2  |        |      |       |
|                       |                | 5    | Almonaster  | 689           | 37°52′6″46″   | RCS et al. 323/06 | 9  | 270            | 77.4          | 0.11 (0.059) | 0        | 2.0  |        |      |       |      |       |
|                       |                | 6    | Ronquillo   | 353           | 37°43′6″11″   | RCS 194/06     | 10 | 279            | 78.5          | 0.13 (0.068) | 6        | 2.9  |        |      |       |      |       |
|                       |                | 7    | Aljezur     | 40            | 37°20′8″50″   | RCS et al. 233/06 | 6  | 244            | 61.9          | 0.10 (0.059) | 5        | 2.1  |        |      |       |      |       |
|                       |                | 8    | Hinojos     | 90            | 37°17′6″25″   | RCS 192/07     | 11 | 265            | 76.9          | 0.11 (0.058) | 0        | 1.8  |        |      |       |      |       |
|                       |                | 9    | Loué        | 280           | 37°12′7″57″   | RCS et al. 218/06 | 12 | 277            | 78.3          | 0.11 (0.058) | 6        | 10   | 2.4   | 5.3  |        |      |       |
|                       |                | 10   | Donhana     | 22            | 37°00′6″30″   | RCS 193/07     | 10 | 255            | 72.2          | 0.10 (0.051) | 1        | 1.9  |        |      |       |      |       |
|                       |                | 11   | Puerto Real | 69            | 36°31′6″04″   | FB 18/07       | 10 | 280            | 77.9          | 0.12 (0.064) | 2        | 2.1  |        |      |       |      |       |
| Betic Ranges          |                | 12   | Cartagena   | 277           | 37°38′1″11″   | RCS & MT 17/07 | 11 | 241            | 78.3          | 0.11 (0.055) | 6        | 13   | 2.7   | 5.2  |        |      |       |
|                       |                | 13   | Bédar       | 563           | 37°12′2″00″   | RCS & MT 16/07 | 12 | 235            | 73.6          | 0.09 (0.047) | 1        | 1    | 1.8   | 3.4  |        |      |       |
|                       |                | 14   | Cortes de la Frontera | 510 | 36°40′5″16″ | RCS 1/07 | 12 | 270 | 77.4 | 0.12 (0.061) | 1 | 7 | 2.1 | 4.5 |        |      |       |
|                       |                | 15   | Gourougou   | 603           | 35°14′2″58″   | RCS et al. 116/07 | 7  | 275            | 76.7          | 0.13 (0.075) | 2        | 4.2  |        |      |       |      |       |
|                       |                | 16   | Asifane     | 900           | 35°08′4″57″   | RCS et al. 57/07 | 9  | 312            | 79.2          | 0.15 (0.078) | 8        | 4.6  |        |      |       |      |       |
|                       |                | 17   | TarguiSA     | 1300          | 34°57′4″19″   | RCS et al. 58/07 | 10 | 319            | 80.6          | 0.16 (0.082) | 13       | 35   | 5.0   | 9.1  |        |      |       |
|                       |                | 18   | Romannia    | 111           | 33°38′7″57″   | RCS et al. 170/06 | 9  | 329            | 79.9          | 0.15 (0.081) | 11       | 20   | 4.7   | 8.7  |        |      |       |
|                       |                | 19   | At Atab      | 850           | 32°09′6″46″   | FB et al. 205/07 | 11 | 312            | 82.4          | 0.15 (0.077) | 9        | 12   | 4.2   | 7.5  |        |      |       |
|                       |                | 20   | Olympia     | 203           | 37°38′21″44′E | RS 19/07      | 11 | 182            | 47.8          | 0.05 (0.026) | 42       | 7.9  |        |      |       |      |       |
|                       |                |      |             |               |              |               |    |                |               |         |         |     |         |      |       |
|                       |                |       |             |               |              |               |    |                |               |         |         |     |         |      |       |

- Asterisks sequenced for ITS variation.
- All N/W except otherwise indicated.
- Mean genetic diversity.
- a, based on all studied plants; b, sub-area based, with nine plants per population and three (randomly chosen) populations per group.
Amp® PCR system 9700, PE Applied Biosystems). In order to assess the reliability of the method a random fraction (N = 23, i.e., 12%) of the samples was analyzed twice. It was found that duplicate analyses were largely indistinguishable, with a repeatability of 98.8%.

A screening of selective primers (consisting of a battery of 72) was run on eight individuals from eight populations, and the nine primers that generated the most polymorphic fragments were selected. These were: EcoRI (Fam)-ACT/MseI-CAT; EcoRI (Fam)-ATC/MseI-CAG; EcoRI (Fam)-AGC/MseI CAG; EcoRI (Fam)-AGC/MseI CAA; EcoRI (Ned)-AAC/MseI-CAT; EcoRI (Ned)-ACC/MseI-CTA; EcoRI (Ned)-ACA/MseI-CAT; EcoRI (Fam)-ATG/MseI-CAG; and EcoRI (Fam)-ACA/MseI-CAA. Fluorescence-labeled selective amplification products were separated on a 3130xl Genetic Analyzers (Applied Biosystems) capillary sequencer with an internal size standard (GeneScan-500 ROX, PE Applied Biosystems). Alignment of the raw data with the size standard was performed with ABI Prism GeneScan software 2.1 (PE Applied Biosystems). AFLP fragments were scored for bands 80–500 bp in length with software developed in the Montana State University (Genographer ver. 1.6.0, available at http://hordeum.oscs.montana.edu/genographer) using the ‘thumbnail’ option. Results were summarized as a matrix of presence/absence.

2.3.1. AFLP-based genetic diversity and population differentiation

Total and percent polymorphic fragments were determined for each population using FAMD version 1.08 (Schlüter and Harris, 2006). The numbers of private fragments (Fragpriv) and of private fragments shared by any pair of populations (Fragpa) were also calculated. Mean genetic diversity (Hs) was computed with Arlequin version 3.01 (Excoffier et al., 2005), and the Rarity 1 Index (equivalent to the frequency of down-weighted marker values; i.e., DW) determined with R Statistical Software ver. 2.8.1 (R Development Core Team) and AFLPdat (Ehrich, 2006). As both the number of private fragments and the rarity index are highly dependent on the spatial distances that separate the populations and the number of plants analyzed (Tremetsberger et al., 2009), for simplicity we calculated these parameters using three populations chosen at random per geographic group, and with nine randomly chosen individuals per population (Schönswetter and Tribisch, 2005). General Linear Models (module GLM in SPSS) were used to study the effect of geographic origin on Hs, Fragpol, Fragp and DW.

The extent of genetic differentiation (measured as FST, the fixation index) was determined both for populations and for geographically-based sets of populations (Arlequin 3.01; Excoffier et al., 2005), and affinities between individuals and populations checked using the random reallocation method implemented in AFLPOP v 1.1 (Duchesne and Bernatchez, 2002). The relationship between pairwise FST values and geographical distances was investigated with Mantel tests based on Spearman rank correlations with 100,000 permutations (package ade4 in R Statistical Software; Dray and Dufour, 2007). Lastly, the relationship between the two subspecies of E. baetica was depicted in a dendrogram built (on the basis of among-population pairwise FST distances) in SPLITSTREE 4.6 (Huson and Bryant, 2006), with neighbor-joining clustering and 10,000 replicates.

2.3.2. Population and plant clustering

Among-plant genetic distances (Nei and Li, 1979) were calculated from the matrix of AFLP scores and represented graphically in a consensus tree built with PAUP® ver. 4.0 beta 10 (Swofford, 2002; neighbor-joining clustering, 10,000 replicates). Furthermore, the partitioning (within and among populations) of genetic variation was determined with an Analysis of Molecular Variance (AMOVA; Excoffier et al., 2005) performed over all the studied samples. The effect on genetic variation of evident geographical barriers (i.e., Strait of Gibraltar, Guadalquivir River Valley, or both) was also studied with AMOVA.

Genetic relationships among populations were further explored with a Bayesian assignment technique implemented in STRUCTURE 2.2.3 (Falush et al., 2003; Pritchard et al., 2000) and carried out at the University of Oslo (http://www.bioportal.uio.no/). This approach uses individual multilocus genotypes to construct clusters of genetically similar plants, in a way that maximizes Hardy–Weinberg equilibrium and minimizes linkage disequilibrium within clusters. Following Evanno et al. (2005) and Falush et al. (2003) we used the ‘admixture’ and the ‘correlated allele frequency' models, a combination judged appropriate when the aim is to infer subtle population structures, as expected in our dataset. To ensure convergence of the Monte Carlo Markov Chain (MCMC) we used 50,000 burn-ins followed by 500,000 iterations. The true number of genetically distinct sets of populations (i.e., ‘archipelagos', sensu Evanno et al., 2005) was determined from analyses that had varying values of K, from 1 to a maximum of 19 (i.e., the total number of populations). Ten independent runs were performed and averaged for each value of K. The best estimate of the true number of archipelagos (Delta-K, as described by Evanno et al. (2005)) was computed with Structure-sum 2.2 (Ehrich, 2006).

3. Results

3.1. nrDNA ITS phylogeny

Fig. 2 presents one of two equally-parsimonious trees based on nrDNA ITS. The tree has 30 steps, a Consistency Index of 0.804, and a Retention Index of 0.746. As regards their nrDNA ITS sequences, the studied populations of E. baetica subsp. baetica were undistinguishable. Nevertheless, they could be differentiated from subsp. orientalis on the basis of four parsimoniously informative changes in region ITS1. Furthermore, the genus Erophaca itself appeared as monophyletic and separated from the other genera with a bootstrap value of 100. The tree also showed that Erophaca was only distantly related to Sophora, but should be considered sister to Oxytropis, Astragalus, and Swainssona.

According to the relaxed molecular clock, the most recent common ancestor of the two subspecies of E. baetica lived 11.9 mya (Fig. 2), although this estimate had considerable uncertainty (95% Highest Posterior Density confidence interval, HPD = 3.3–19.1 my, i.e., from the Late Miocene to the Early Pliocene). As for the genus Erophaca itself, it would have split away from its sister clade 16.6 mya (HPD = 8.5–20.5 my), during the Mid Miocene.

3.2. AFLP-based genetic diversity and population differentiation

A total of 561 unambiguously scoreable DNA fragments were detected in E. baetica subspecies baetica, of which 527 (94%) were polymorphic. Population-based estimates in Table 1 indicate a maximum of 329 fragments per population, and a minimum of 235 (61–82% polymorphic). Mean genetic diversity (Hs) was markedly dependent on the geographic location of populations (i.e. Iberian Massif, Betic Ranges, or North Africa; F2,16 = 19.31, p < 0.001; General Linear Model ANOVA). On average, Hs was significantly higher for Moroccan populations than for Iberian counterparts (p < 0.001; Tukey’s pairwise comparisons). Values of the Rarity Index (i.e., the sub-area based DW in Table 1) were also dependent on geographic location (F2,16 = 25.36, p = 0.001), and again significantly higher for North-African populations than for the rest (8.4 on average; p < 0.001, Tukey’s test). Populations from Morocco were more diverse than those from Iberia also as regards polymorphic and private fragment numbers.
Table 2 reports genetic distances and numbers of shared private fragments among populations of *Erophaca baetica*. The 14 populations from the Iberian Peninsula had few (at most two) or no private fragments in common, while the five Moroccan shared up to 11. According to Mantel tests, the genetic dissimilarity among any two populations (as $F_{ST}$) increased with increasing spatial distance. This proved significant for the Iberian subset of populations ($r = 0.561$, $N = 14$, $p < 0.001$); for the Moroccan subset ($r = 0.721$, $N = 5$, $p < 0.05$); and for the 19 West Mediterranean populations altogether ($r = 0.633$, $N = 19$, $p < 0.001$). In addition, following random reallocation in AFLPOP, plants were most often correctly assigned to their populations of origin (with the implication that populations had genetic profiles definite enough to become predictable). Specifically, all Moroccan plants except one (i.e., 98%) were allocated to their true source, all but two from the Betic Ranges (94%), and all but 12 from the Iberian Massif (89%).

### 3.3. Genetic structure and geographic areas

The Bayesian analysis of individual genotypes showed that the most informative representation of overall genetic structure was achieved with $K = 3$ geographical groups, namely Morocco, the Betic Ranges, and the Iberian Massif. The same can be concluded from Fig. 3 (left), depicting neighbor-joined individual plants based on AFLP data: two well-supported (98% bootstrap) clusters existed on each side of the Strait of Gibraltar, that corresponded to Morocco and the Iberian Peninsula. Within Morocco, two sub-clusters could be distinguished, one including Middle Atlas plants and another with plants from the Rif Mountains (populations 15, 16, and 17). Within the Iberian Peninsula, population 12 appeared distinctly separated from the rest at 100% bootstrap, while the rest encompassed plants from the Betic Ranges on the one hand (populations 13 and 14), and plants from the Iberian Massif on the other (i.e., populations 1 through 11).

The population-based graph in Fig. 3 (right) reiterated the genetic relationships described above and, additionally, suggested that subsp. *orientalis* could relate more to Moroccan plants (of subsp. *baetica*) than to Iberian plants. Moreover, Moroccan populations also shared a few private fragments with the (single) Greek population (see Table 2).

Estimates of genetic diversity and differentiation calculated specifically for geographic sub-areas are shown in Table 3. As expected, fewer fragments were shared across continents (from 1 to 11) than within the Iberian Peninsula (31) or North Africa (36). Sub-areal dissimilarity (estimated from fixation indices) showed an equivalent trend: most dissimilar were plants from Middle Atlas and the Iberian Massif ($F_{ST} = 0.28$), while those from...
North African sub-areas (e.g., Rif Mountain and Middle Atlas) were comparatively alike ($F_{ST} = 0.13$).

Table 4 summarizes results of several AMOVA models analyzing the effects of major geographic barriers on the partitioning of molecular variance. In general, more than half of the genetic variation occurred among plants within populations, but this was particularly evident if no geographic barrier was included in the models. Including the Strait of Gibraltar explained 18% of variation, whereas the Guadalquivir River Valley accounted for around 15%. The model that explained a larger fraction of variation (19%) on the basis of geographical barriers was the one considering both continental (i.e., Strait of Gibraltar) and regional (Guadalquivir River) differentiation.

### Discussion

#### 4.1. Evolution, phylogeny and systematics of the genus Erophaca

Our relaxed molecular clock estimate for the origin of Erophaca (16.6 my) had considerable uncertainty, and yet it matched fairly well previous estimates for the Astragalean clade as a whole (16.1 my; Wojciechowski, 2005). Interestingly, this predates the onset of Mediterranean climate which occurred approximately 3 mya (Suc, 1984), and is compatible with estimates for the evolution of coexisting trees (e.g., Quercus suber; Magri et al., 2007). Therefore, the genus Erophaca could be reasonably included in the group of (mostly woody) Tertiary relicts that form part of the...
Table 3
Genetic affinities among geographic sub-areas inhabited by E. baetica subsp. baetica. Pairwise fixation indices \( F_{st} \), above appear below, and numbers of shared private fragments \( \text{FragPriv} \), above the diagonal. The sub-area-specific numbers of private fragments \( \text{FragPriv} \) are given in bold.

<table>
<thead>
<tr>
<th>Sub-areas</th>
<th>FragPriv</th>
<th>Iberian Massif</th>
<th>Betic Ranges</th>
<th>Rif Mountain</th>
<th>Middle Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iberian Massif</td>
<td>–</td>
<td>31</td>
<td>5</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>Betic Ranges</td>
<td>0.19</td>
<td>0.19</td>
<td>0.27</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>Rif Mountain</td>
<td>0.27</td>
<td>17</td>
<td>1</td>
<td>1</td>
<td>0.22</td>
</tr>
<tr>
<td>Middle Atlas</td>
<td>0.28</td>
<td>0.28</td>
<td>0.27</td>
<td>0.13</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Mediterranean flora at present (e.g., Chamaerops, Arbutus, Olea, Piscata, or Phillyrea; see Palamarev (1989) for a review), and the fact that it presents some attributes relatively unusual in typical Mediterranean plants (e.g., blooming in winter, and a woody xylepode that acts as the underground storage organ) is hardly surprising. This relictic, pre-Mediterranean status has also been claimed for other legume genera considerably younger than Erophaca (e.g., Anagyris, Ortega-Olencia and Catalán, 2009).

Erophaca was long included in Astragalus (as A. lasiaticus Lam.) until Podlech (1993) placed it (on the basis of morphology) in a different genus. This is strongly supported by molecular data in the present study (nrDNA ITS, Fig. 2) which, in addition, identified Astragalus, Oxytropis and Swainsona as the closest relatives of Erophaca. Furthermore, on the basis of nrDNA ITS the genus should not be considered close to Cheneseya (but see Wojciechowski et al. (2000)). Nevertheless, nrDNA ITS sequences for Gueldenstaedtica and other genera are still unavailable, so more data are needed before the phylogenetic relationships among these genera can be clearly established. A phylogenetic tree based on matK sequences (Wojciechowski, personal communication) also distances Erophaca from Cheneseya and Gueldenstaedtica.

In the current study, 41 private fragments (Table 3) were identified between the Iberian Massif and all other regions, the Rif Mountain region, and the Middle Atlas region. These private fragments might be contributed as part of the phylogenetic differentiation process that has resulted in the current distribution of E. baetica subsp. baetica.

AFLP data in the present study identified Morocco, the Betic Sierras, and the Iberian Massif as sub-areas that shelter genetically identifiable variants of this subspecies. The Strait of Gibraltar undoubtedly represents an important geographic barrier to gene flow, and the long-lasting separation between African and Iberian populations (the opening of the Gibraltar Strait is dated around 5.3 mya; Duggen et al., 2003) has allowed genetic differentiation to accumulate, just as in many other herbaceous taxa (Escudero et al., 2008), woody legumes (Pardo et al., 2008), and even tree species (Terrab et al., 2008). It must also be noted that African populations are much more diverse genetically than European ones, despite the plant being (currently) relatively rare in N. Africa. Our data for Erophaca support the view that Morocco is an important reservoir of genetic diversity in plants (Medail and Diadema, 2009). This diversity can be due to the ancestral nature of populations (in the case of Erophaca, they present the largest numbers of private and rare fragments which tend to accumulate over time; Stehlik et al., 2002; Schönswetter and Tribsch, 2005) or, alternatively, be a by-product of larger population sizes during Pleistocene climatic fluctuations.

As regards regional genetic differentiation at each side of the Guadalquivir River Valley, even though this has been demonstrated in other taxa (e.g., Hypochaeris radicata; Ortiz et al., 2008), in the case of E. baetica the role of this barrier impeding gene flow should be considered less important (compared to the Strait of Gibraltar). According to Mantel tests, southernmost Iberian populations (i.e., on the Betic Sierras) were genetically similar to the African coastal pool, whereas those from the Iberian Massif appeared relatively impoverished and poorly structured (suggesting they are relatively recent). In general, however, genetic structuring was modest within the Iberian Peninsula and more than half of genetic variance occurred among plants within populations, as could be expected in a long-lived, obligate outcrosser with low seed dispersibility (Hamrick et al., 1992; Nybom, 2004).

4.3. Subspecies evolution and extant distribution

According to data in the present study, the most recent ancestor shared by the two disjunct subspecies of E. baetica would be around 11.9 mya old, so they most likely evolved during the cool-dry cycles of the Miocene that led to disappearance of many tropical elements of the flora (Quezel, 1978; Thompson, 2005; Postigo Mijarra et al., 2009). It may seem at first sight that the two subspecies have accumulated very few differences after being separated for such a long period of time (just four nrDNA ITS substitutions in nearly 12 my), but this is not unexpected in long-lived herbaceous perennials (Kay et al., 2006).

E–W disjunctions abound in the Mediterranean flora, and are often explained on the basis of contraction of once more continuous areas (i.e., vicariance: Davis and Hedge, 1971; Rosselló et al., 2007), mainly brought about by an accentuation of aridity (Hsu et al., 1973; Griffin, 2002; Thompson, 2005). This process can provide a likely explanation for the present distribution of our two
subspecies. On the other hand, molecular data could be interpreted to provide an alternative scenario, namely that the subspp. orientalis derives from the Western subspecies. If this was the case, and judging from the number of shared private fragments (Table 2), candidates to represent the ancestral pool would be the N. African populations. It is often assumed in the biogeography of the Mediterranean flora that ‘Eastern’ is equivalent to ancestral, so the reverse scenario just described would be quite exceptional. To our knowledge, the only example of a similar (but non-Mediterranean) biogeographical pattern is that of the montane subspecies of Anthyllis montana, in which the western ones have been demonstrated to be ancestral, whereas Eastern subspecies are derived (Kropf et al., 2002). Palaeogeographical reconstructions for Mediterranean animal groups have often proposed the North of Africa as an ancestral area (Oosterbroek and Arntzen, 1992; Sanmartin, 2003; Carranza et al., 2006).

Acknowledgments

We are grateful to Prof. P.E. Gibbs for a critical reading of an early version of the manuscript, and to Prof. M. Wojciechowski for valuable suggestions on the phylogeny of the Atragelinae clade. Prof. T.F. Stuessy allowed access to laboratory facilities at the University of Vienna, and staff of the Biology Laboratory Service (CITIUS, Seville University) assisted during DNA extractions. R. Santos kindly supplied leaves of E. baetica subsp. orientalis. This study was financially supported by a predoctoral grant of the Junta de Andalucía; and by Grant CGL2006-2005/RNM848 of the Junta de Andalucía; and by Grant CGL2006-00817 of the Spanish Ministerio de Educación y Ciencia to S. Tala vera. The Doñana National Park (Project 33/2004) provided permission to access sampling sites.

References


Pardo, C., Cubas, P., Tahiri, H., 2008. Genetic variation and phylogeography of 
Stearanthus (Fabaceae, Genistaeae) from the Iberian Peninsula and northern 
Morocco assessed by chloroplast microsatellite (cpSSR) markers. Am. J. Bot. 95, 
98–109.
Mediterranean disjunction in the sand-dune shrub Armeria pungens 
Sendtner 1, 267–272.
Bioinformatics 14, 817–818.
changes in the Iberian Peninsula and Balearic Islands (south-west Europe) 
during the Cenozoic. J. Biogeogr. 36, 2025–2043.
using multilocus genotype data. Genetics 155, 945–959.
Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P., Arroyo, J., 2008. The 
Strait of Gibraltar as a melting pot for plant biodiversity. Q. Sci. Rev. 27, 2100– 
2117.
Buxus balearica (Buxaceae) as evidenced by nuclear ribosomal markers: when ITS 
biogeography of the Paleartic Pachydeminae (Coleoptera, Scarabaeoidea). J. 
Biogeogr. 30, 1883–1897.
Press, USA.
Travis, S.E., Maschinski, J., Keim, P., 1996. An analysis of genetic variation in 
Astragalus cremnophylax var. cremnophylax, a critically endangered plant, using 
Tremetsberger, K., Urtubey, E., Terrab, A., Baeza, C.M., Ortiz, M.A., Talavera, M., 
refugia and polytopic replacement of diploids by tetraploids in the Patagonian 
18, 3668–3682.
Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, 
J., Peléman, J., Kuiper, M., Zabeau, M., 1995. AFLP: a new concept for DNA 
Weising, K., Nybom, H., Wolff, K., Kahl, G., 2005. DNA Fingerprinting in Plants: 
Principles, Methods, and Applications, 2nd ed. CRC Press, Boca Raton.
White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of 
fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., 
Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and 
Wojciechowski, M.F., 2005. Astragalus (Fabaceae): a molecular phylogenetic 
perspective. Brittonia 57, 382–396.
Astragalus Fabaceae) and its major subgroups based on nuclear ribosomal DNA 