

---

This is the **accepted version** of the journal article:

Calleja, Juan Antonio; Garcia-Jacas, Núria; Roquet, Cristina; [et al.]. «Beyond the rand flora pattern : Phylogeny and biogeographical history of Volutaria (compositae)». *Taxon*, Vol. 65, Issue 2 (May 2016), p. 315-332. DOI 10.12705/652.8

---

This version is available at <https://ddd.uab.cat/record/288610>

under the terms of the  **IN**  
COPYRIGHT license

Original Article

# **Beyond the Rand Flora pattern: phylogeny and biogeographical history of *Volutaria* Cass. (Compositae)**

**Juan Antonio Calleja\*<sup>1,2</sup>, Núria Garcia-Jacas<sup>3</sup>, Cristina Roquet<sup>4</sup> & Alfonso Susanna<sup>3</sup>**

<sup>1</sup> *Botánica (Dpto. Biología Animal, Vegetal y Ecología), Universidad Autónoma de Barcelona, ES-08193 Bellaterra, Spain.*

<sup>2</sup> *CREAF, Cerdanyola del Vallés 08193, Spain*

<sup>3</sup> *Institut Botànic de Barcelona (IBB – CSIC – ICUB), Pg. del Migdia s.n., ES-08038 Barcelona, Spain.*

<sup>4</sup> *Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Grenoble Alpes, 2233 Rue de la Piscine, FR-38041 Grenoble Cedex 9, France*

\*Author for correspondence: *Juan Antonio Calleja, [juan.calleja@gmail.com](mailto:juan.calleja@gmail.com)*

**Abstract** The Rand Flora describes the disjunct phytogeographical pattern of a wide range of taxa distributed at the periphery of Africa and adjacent islands, as well as the Arabian Peninsula. We focused on *Volutaria* (Compositae, Cardueae-Centaurineae), a genus of ca. 18 species that conforms to the Rand Flora pattern; yet its phylogeny, interspecific relationships, and biogeographical history remain poorly known. We aim to construct a robust phylogeny that will allow us to interpret the biogeography and the diversification of this genus, together with its present distribution pattern, and to ascertain whether the latter arose by single or multiple gradual expansion processes, vicariance, or long-distance dispersal events. We sampled all extant *Volutaria* species, covering its entire geographical range, and generated sequences of nuclear-ribosomal DNA and three plastid regions, which were analyzed by Bayesian inference and maximum parsimony. Samples yielding more than one sequence in direct sequencing were cloned. Historical biogeographical analysis was performed using BioGeoBEARS based on a phylogeny dated using a relaxed molecular clock calibrated with a previous dating undertaken for the tribe Cardueae. *Volutaria* is a monophyletic taxon, having an Asian ancestor, and its present diversity is represented by four main clades that emerged in the Miocene. The earliest taxa of *Volutaria* to diverge are preserved in North Africa, whereas clades that diverged more recently have done so on both sides of the Sahara. This process involved both ancient and recent interspecific introgression and hybridization events, as indicated by incongruities between plastid and nuclear results, and by cloning of the ITS region. The distribution of *Volutaria* around two diversity poles conforms to the Rand Flora pattern, but this did not arise by a single event or process. Instead, the historical biogeography of *Volutaria* involves at least one migration wave from east to west, followed by the extinction of ancestral taxa and subsequent expansion and retraction events, together with speciation processes on both

sides of the Sahara. The intense tectonic and climatic changes that occurred in North Africa and western Asia throughout the Neogene and Pleistocene periods might explain the present diversity and distribution pattern of the genus.

**Keywords** Africa; biogeography; Cardueae-Centaurineae; hybridization; Macaronesia; Rand Flora

**Running head** Phylogeny and biogeography of *Volutaria*

**Supplementary Material** Electronic Supplement (Appendix S1, Figs. S1--S3 and Tables S1--S2) are available in the Supplementary Data section of the online version of this article (<http://www.ingentaconnect.com/content/iapt/tax>).

## INTRODUCTION

The name of the phytogeographical pattern known as Rand Flora, which occurs on a continental scale, was coined for a group of closely related plants belonging to very different and unrelated families, scattered at the periphery of the African continent and the adjacent islands of Macaronesia, Madagascar and Socotra (Christ, 1892; Quézel, 1978; Andrus & al., 2004; Sanmartín & al., 2010; Pokorny & al., 2015). The main thrust of the hypothesis for the origin of the Rand Flora proposes eastern-western and eastern-southern vicariations, resulting from the fragmentation of a continental macroflora that was originally widespread across Africa, as a result of the development of environmental barriers (e. g. the Sahara Desert) throughout the Neogene (Monod 1971, Axelrod & Raven, 1978; Quézel, 1978; Sanmartín & al., 2010; Pokorny & al., 2015). Conversely, a second, but non-exclusive hypothesis, proposes long-distance or medium-range dispersal events or migration tracks (Quézel, 1978), either from southern and eastern Africa (Quézel, 1978; Galbany & al., 2009; Sanmartín & al., 2010), or from western Asia via the Arabian Peninsula (Davis & Hedge, 1971; Roquet & al., 2009; Sanmartín & al., 2010; Barres & al., 2013; Mairal & al., 2015). Thus, the present geographical pattern of the Rand Flora might be a consequence of multiple dispersal and fragmentation events that occurred at different times (Axelrod & Raven, 1978; Sanmartín & al., 2010; Pokorny & al., 2015). At genus level, this pattern was found to be due either to a single (Thiv & al., 2010; Mairal & al., 2015) or repeated events (Barres & al., 2013; Pokorny & al., 2015 and references therein).

The Rand Flora comprises more than 20 vascular plant families, of which Compositae is remarkable for having the largest number of genera that conform to this biogeographical pattern (Engler, 1879; Christ, 1892; Monod 1971; Bramwell, 1985; Andrus & al., 2004). Within tribe Cardueae (Compositae), the group *Volutaria* (Garcia-

Jacas & al., 2001; Barres & al., 2013) comprises seven genera, namely *Amberboa* (Pers.) Less., *Goniocaulon* Cass., *Plagiobasis* Schrenk, *Russowia* C. Winkl., *Tricholepis* DC., *Schischkinia* Iljin and *Volutaria* Cass., all of which are distributed in Asia and the Mediterranean region (Susanna & al., 2011). *Volutaria* is amongst the most species-rich of the group, and comprises 18 species (Susanna & Garcia-Jacas, 2007) unevenly distributed throughout western Asia, Socotra, the Horn of Africa, northwestern Africa, Macaronesia and southwestern Europe. Most of these colonize open habitats in steppic, Mediterranean and desert climatic regimes (Hellwig, 2004), sometimes with monsoonal or oceanic influences (Murbeck, 1897; Scholte & De Geest, 2010). *Volutaria* shows two diversity poles, the first located in northwestern Africa, and the other located on the Horn of Africa and the Arabian Peninsula, with endemic species found on both sides of the Sahara Desert (Wagenitz, 1991).

The true diversity and phylogeny of the genus are currently unknown, since only incomplete taxonomic studies of *Volutaria* (Wagenitz, 1989, 1991) have so far been undertaken. Karyological studies are also only partially complete (Hellwig, 1994; Wagenitz & Hellwig, 1996a), but nonetheless, have indicated the presence of polyploidy and possible reticulation events. Basic chromosome numbers are high, suggesting ancient polyploidy. The most common chromosome number is  $x = 16$ , but some species have  $x = 12$ , 13 and 14 (Table 1). Besides this dysploid series, some species display different ploidy levels (diploid and tetraploid); and one of the counts ( $2n = 58$ ) was explained in terms of hybridization between species having different basic chromosome numbers (Hellwig, 1994).

An investigation of phylogenetic relationships would allow the re-examination, and hopefully, resolution of a number of controversial taxonomic hypotheses regarding the identity of certain taxa [e. g., *V. lippii* Cass. vs *V. tubuliflora* (Murb.) Sennen, *V.*

1 *abyssinica* vs *V. somalensis* (Oliv. & Hiern) C. Jeffrey; cf. Murbeck, 1897; Wagenitz,  
 2 1989; Wood, 1997], together with a re-evaluation of morphological similarities between  
 3 eastern and western species [e. g., *V. belouini* (Humbert) Maire and *V. abyssinica* (A.  
 4 Rich.) C. Jeffrey ex Cufod. (Humbert, 1927)]. Furthermore, while some subspecies can  
 5 be distinguished on the basis of well-defined morphological characters (e. g. *V.*  
 6 *abyssinica* group in Wagenitz, 1991), the validity of others is debatable [e. g. *V.*  
 7 *maroccana* (Barratte & Murb.) Maire vs *V. crupinoides* (Desf.) Maire, *V. sinaica* (DC.)  
 8 Wagenitz vs *V. saharae* (L. Chevall.) Wagenitz, and *V. socotrensis* Wagenitz vs *V.*  
 9 *dhofarica* Wagenitz (Murbeck, 1897; Chevallier, 1905; Kilian & Hein, 2006)].

10 The biogeographical history of *Volutaria* is also poorly understood. Hellwig (2004)  
 11 suggested an Asian origin and posterior westward expansion and radiation across North  
 12 Africa and Macaronesia, and, based on their perennial mode of life, an ancient origin for  
 13 the eastern African and Arabian species. Susanna & Garcia-Jacas (2009) also proposed  
 14 an eastern origin for these plants based on the biogeography of the entire *Volutaria*  
 15 group. All the remaining genera of the group are Asian, and only *Volutaria* has reached  
 16 the Mediterranean and Macaronesian regions. This hypothesis was partly supported by  
 17 the work by Barres & al. (2013), but to date, there still has been no comprehensive  
 18 biogeographical analysis of the distribution pattern of *Volutaria*.

19 Here, we present a molecular study using nuclear ITS and ETS sequences, both  
 20 directly sequenced and cloned, in combination with plastid markers *ndhF*, *rpl32F*-  
 21 *trnL*<sup>UAG</sup> and *trnL-trnL-F*. All these regions have been instrumental in unraveling the  
 22 phylogeny of subtribe Centaureinae (Susanna & al., 2011) and tribe Cardueae (Barres &  
 23 al., 2013), and the combination of plastid and nuclear markers can also facilitate the  
 24 detection of introgression (e.g., in *Centaurea* sect. *Phrygia*, cf. López-Alvarado & al.,  
 25 2014). Nuclear ribosomal DNA sequences (both ITS and ETS) also provide additional

information in that they occur as multi-copy arrays (Wendel & al., 1995; Baldwin & Markos, 1998). Such copies are usually homogenized via concerted evolution, but sometimes other copies persist. When these have been acquired through hybridization, ITS and ETS are useful in detecting reticulation (Nieto Feliner & Rosselló, 2007). For this reason, both regions have been used to study the related genus *Centaurea* L. (Garcia-Jacas & al., 2009). However, causes other than reticulation must also be considered when interpreting polymorphic DNA patterns: in particular incomplete lineage sorting, which can also be the cause of DNA polymorphisms (Twyford & Ennos, 2012).

We sequenced nuclear and plastid regions of a comprehensive sample of *Volutaria* with the aim to: a) establish a phylogenetic framework for the genus, reconstruct its interspecific evolutionary relationships and test its monophyly; b) examine the relevance of hybridization and reticulation in the evolution of the genus; c) establish a temporal framework and correlate the main dispersal events in the history of *Volutaria* with the major tectonic and climatic events that occurred in the Mediterranean, Macaronesia and eastern African regions from the early Miocene; and d) identify whether the biogeographical processes (such as dispersal and fragmentation) responsible for the Rand Flora pattern, and observed in *Volutaria*, occurred as a single event, or repeatedly.

## MATERIAL AND METHODS

**Plant material .** — All currently accepted *Volutaria* species were sampled for much of their geographical range (Appendix S1 and Fig. 1). Field sampling was conducted in Southwest Asia (Yemen, Socotra), eastern and North Africa (Ethiopia, Kenya, Morocco), Macaronesia (Canary Islands) and the Iberian Peninsula. Sampling



was completed with the use of herbarium specimens and DNA aliquots provided by several institutions, as well as individual cultivated specimens [*V. muricata* (L.) Maire from southwestern Morocco, *V. tubuliflora* (Murb.) Sennen from southeastern Spain and *V. canariensis* Wagenitz from Gran Canaria]. All fresh leaf samples were dried and stored in the presence of silica-gel until laboratory analyses were conducted.

Up to eleven Cardueae species were used as outgroups: *Amberboa turanica* Iljin, *Carduus carlinoides* Gouan, *Cirsium palustre* (L.) Scop., *Cousinia microcarpa* Boiss., *Cynara cornigera* Lind., *Cynara humilis* L., *Goniocaulon indicum* C.B. Clarke, *Plagiobasis centauroides* Schrenk, *Russowia sogdiana* B. Fedtsch., *Saussurea maximowiczii* Herder and *Tricholepis tibetica* Hook. f. & Thomson ex C.B. Clarke. All these species were selected on the basis of previous work (Susanna & al., 2006; Susanna & al., 2011; Barres & al., 2013) which had also provided their respective DNA sequences, with the exception of *Amberboa turanica* and *Plagiobasis centauroides* whose ETS region was first sequenced here. Voucher data and GenBank sequence accession numbers are provided in Appendix S1.

**DNA Extraction, Amplification, and Sequencing.** — Total genomic DNA was extracted using the CTAB method (Doyle & Dickson, 1987) as modified by Cullings (1992) and Tel-Zur & al. (1999). Double-stranded DNA of the ITS region was amplified using ITS1 as the forward primer and ITS4 as the reverse primer (White & al., 1990). The profile used for PCR amplification follows the protocol described by Susanna & al. (2006). The ETS region was amplified with ETS1F as the forward primer (Linder & al., 2000) and 18SETS as the reverse primer (Baldwin & Markos, 1998). In some cases, AST-1 and AST-2 were also used as internal primers (Markos & Baldwin, 2001). The profile used for PCR amplification was described by Galbany-Casals & al.

(2009). In both regions, reactions were performed on 25 µl volumes with 10% 10× AmpliTaq buffer, 10% 25 mM MgCl<sub>2</sub>, 10% 2 mM dNTPs mix, 4% of each primer at 5 µM, 0.5 µl of DMSO (dimethyl sulfoxide; Sigma-Aldrich, St. Luis, MO, USA), 1 U AmpliTaq DNA polymerase (Applied Biosystems, Foster City, CA, USA) and 2 µl of template DNA of unknown concentration. Reaction vessels were topped up to 25 µl with sterile, distilled, water.

Double-stranded cpDNA of the *trnL5'-trnF* region was amplified using the *trnL*-c, forward, and *trnL*-f, reverse, primers (Taberlet & al., 1991). In some cases, *trnL*-d, reverse, and *trnL*-e, forward, were also used. The conditions used for amplification of this region were described by Susanna & al. (2006). The *rpl32-trnL<sup>UAG</sup>* intergenic spacer was amplified using the primers *rpl32F* as forward, and *trnL-UAG* as reverse (Shaw & al., 2007). PCR did not work for DNA extractions of some herbarium specimens. Therefore, two internal primers were used (*rpl32-580F* 5'TCGGAATTACTATATGGATT3' and *rpl32-580R* 5'AATCCATATAGTAATTCCGA3') following identical thermocycler settings (Susanna & al., 2011), but lowering the annealing temperature to 52°C. The major part of the *ndhF* encoding region was amplified using a set of four primers. The 5'- end portion of the gene was not used for analysis because of its low substitution level (Kim & Jansen, 1995). Overlapping sequence fragments were obtained by amplifying the 3' end-portion of the gene in two pieces. For the 5' quarter, we used 3'F as forward primer (Eldenäs & al., 1999) and 1783R (Barres & al., 2013) as reverse primer. For the 3' quarter, we used 1626F (Barres & al., 2013) as forward primer and +607 (Kim & Jansen, 1995) as reverse primer. The conditions for amplifications were as outlined by Kim & al. (2002). The PCR reactions were performed in accordance with the protocol used for the nuclear regions, but with the addition of 2.5 µl 400 ng/µl BSA (bovine

serum albumin; New England Biolabs, NE, USA). Plastid and nuclear PCR products were purified using ExoSAP-IT (USB Corp., Cleveland, OH, USA). Direct sequencing of the amplified DNA segments was performed at the University of Florida ICBR Core Facility, using an ABI 3730xl (Applied Biosystems) and using BigDye Terminator Cycle Sequencing v3.1 (Applied Biosystems), and following the manufacturer's protocol. Nucleotide sequences were edited using BioEdit v7.0.5.3 (Hall, 1999) and aligned manually.

**Cloning.** — In some cases, where the presence of more than one copy was suspected, PCR products of the ITS region of certain individuals of *V. albicaulis*, *V. dhofarica*, *V. saharae*, and *V. somalensis*, and those of the ETS region of some individuals of *V. muricata*, were cloned using a TOPO TA Cloning kit following the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). Where possible, eight to sixteen positive colonies from each reaction were screened with direct PCR using T7 and M13R universal primers following the amplification conditions described by Vilatersana & al. (2007). Eight to sixteen PCR products of the same size were selected for sequencing in both directions using the same primers. Unique substitutions in clones from a single accession were excluded. Consensus sequences were generated for each accession and region, condensing the single base-pair differences among clones. This approach reduces the size of the matrices as well as the impact of PCR artefacts, such as chimeric sequences and Taq errors (Cline & al., 1996; Popp & Oxelman, 2001).

**Phylogenetic analyses.** — Two nuclear (ITS and ETS) and three plastid markers (*trnL-F*, *rpl32F-trnL*<sup>UAG</sup>, *ndhF*) were used to assemble five separate sequence datasets (available at [ftp://descargas:Dscargas\\_15@www.ibb.bcn-csic.es](ftp://descargas:Dscargas_15@www.ibb.bcn-csic.es), folder Volutaria). For the ETS dataset, we omitted the first 168 bp of the 3' end because this highly-variable

region was difficult to align unambiguously with the species of the outgroup and with *Volutaria muricata*. Dataset 1 was used to build a robust phylogeny of *Volutaria* and consisted of a combined ITS and ETS matrix of 63 sequences for 16 *Volutaria* species and two subspecies, including ITS cloned copies of *V. saharae* (L. Chevall.) Wagenitz and a cloned copy of the ETS region of *V. muricata*. The outgroup comprised three species: *Plagiobasis centauroides*, *Russowia sogdiana*, and *Tricholepis tibetica* (Fig. 2). Dataset 2 was used for comparison with nuclear Dataset 1, with the aim of detecting potential incongruencies and help in the identification of hybridization events. It contained the three chloroplast regions combined in a new matrix of 87 samples in total, comprising 17 *Volutaria* species and two subspecies, and the same outgroup of data set 1 (Fig. 3). Dataset 3 was used to date divergence events in *Volutaria*. It comprised 56 ITS sequences in total, 45 of 17 *Volutaria* species and two subspecies and 2–7 cloned copies of *V. albicaulis* (Deflers) Wagenitz, *V. dhofarica* Wagenitz, *V. saharae* and *V. somalensis* (Oliv. & Hiern) C. Jeffrey. The outgroup consisted of eleven species: *Amberboa turanica*, *Carduus carlinoides*, *Cirsium palustre*, *Cousinia microcarpa*, *Cynara cornigera*, *Cynara humilis*, *Goniocaulon indicum*, *Plagiobasis centauroides*, *Russowia sogdiana*, *Saussurea maximowiczii*, and *Tricholepis tibetica* (Fig. 4). Dataset 4 consisted of 93 ITS sequences for 17 *Volutaria* species and two subspecies plus an outgroup integrated by *Plagiobasis centauroides*, *Russowia sogdiana*, and *Tricholepis tibetica*. Dataset 5 consisted of an ETS matrix of 64 sequences comprising 16 *Volutaria* species and two subspecies using the same outgroup as datasets one, two and four. The phylograms from datasets 4 and 5 are shown in supporting information (Electr. Suppl.: Figs. S1--S2).

Phylogenetic analyses were conducted using Maximum Parsimony (MP) and Bayesian inference (BI), as implemented in PAUP v.4.0b10 (Swofford, 2002) and

MrBayes v.3.2 (Ronquist & al., 2012; <http://mrbayes.sourceforge.net/>), respectively. Prior to BI analysis, the Akaike information criterion was used for model selection using jModeltest v.2.1.1. (Darriba & al., 2012). The selected models were SYM+G (for ITS), GTR+G (for ETS, *rpl32F-trnL*<sup>UAG</sup>, *ndhF*) and F81+G (for *trnL-F*). The coded indels of dataset 2 were assumed to follow a model having equal forward and backward transition rates. MrBayes was run for  $20 \times 10^6$  generations for all datasets, starting from different trees and saving one from every 1000 trees. The consensus tree was computed from the posterior distribution after discarding the first 25% of trees as burn-in. Only moderate ( $\geq 0.90$ ) or significant ( $\geq 0.95$ ) Bayesian posterior probabilities are shown on the phylograms (Buerki & al., 2009).

Maximum Parsimony analyses were conducted for datasets 1, 2, 4 and 5. They always began with heuristic searches using tree bisection reconnection (TBR) branch swapping with character states specified as unordered and unweighted. The indels were coded as missing data. All of the most parsimonious trees (MPTs) were saved. In order to sample different islands of MPTs (Maddison, 1991), 1000 replications with random taxon addition and TBR branch swapping were performed. Preliminary analyses of dataset 1 (ITS and ETS), dataset 4 (ITS) and dataset 5 (ETS) revealed that the tree limit of PAUP was reached very quickly, and therefore we conducted a heuristic search using 1000 replicates and random taxon addition, saving only 500 trees per replicate. Bootstrap (BS) analyses were performed using 100 replicates and heuristic search with the default options for dataset 2 (combined plastid data). For bootstrapping datasets 1, 4 and 5, the tree limit of PAUP was reached very quickly, so thereafter we followed the approach of Lidén & al. (1997), using 1000 replicates, random taxon addition with 20 replicates, and no branch swapping. The resulting consistency index (CI), retention index (RI) and homoplasy index (HI) are given in Table 2. Bootstrap values (BS)

exceeding 70% are shown on the phylograms, and those nodes with  $BS \geq 75\%$  are considered to have relevant support.

**Network analysis.** – Character incongruence in the ITS dataset caused by reticulation, the presence of pseudogenes, recombinant sequences or PCR artefacts were examined with a distance network analysis (split graphs). The neighborNet (NN) algorithm (Bryant & Moulton, 2004) was used as implemented in SplitsTree4 v4.11.3 (Huson & Bryant, 2006) with the criterion set to uncorrected pairwise (p) distances, and excluding constant and non-informative characters.

**Molecular dating.** – Divergence times for *Volutaria* were estimated using a relaxed molecular clock as implemented in BEAST v1.8.2 (Drummond & Rambaut, 2007). The ITS matrix (dataset 3) was used, and six monophyletic groups were defined in BEAUti, in accordance with Barres & al. (2013): i) “all species of data set 3”; ii) “*Cynara cornigera* and *Cynara humilis*”; iii) “*Cousinia* and *Saussurea*”; iv) “*Carduus carlinoides* and *Cirsium palustre*”; v) “*Volutaria* group”, which includes *Volutaria* together with *Amberboa turanica*, *Goniocaulon indicum*, *Plagiobasis centauroides*, *Russowia sogdiana* and *Tricholepis tibetica*; vi) “*Volutaria* genus” (on the basis of our ITS-ETS phylogram).

A calibration of our ITS tree was then conducted based on the fossil-based age estimates of Barres & al. (2013). This work provides a complete phylogeny of tribe Cardueae dated rigorously from five fossils, which reinforces the reliability of the dating results of the *Volutaria* phylogeny. Nodes 47, 51, 63, 72, and 84 of Barres & al. (2013) were used as calibration points relating, respectively, to the first, second, third, fourth and fifth monophyletic groups of our time-calibrated ITS tree (Table 3). The age

of the first monophyletic group (i, “all species of dataset 3”) with prior normal distribution, was restricted to a mean of 30.53 million years ago (Ma), and a standard deviation (stdev) of 1.7; the age of the second group (ii, “*Cynara* species”) with prior normal distribution, was restricted to a mean of 11.90 Ma and a stdev of 1.9; the age of the third group (iii, “*Cousinia* and *Saussurea*”) with prior normal distribution, was restricted to a mean of 19.96 Ma and a stdev of 1.8; the age of the fourth group (iv, “*Carduus* and *Cirsium*”) was restricted to a minimum of 14 Ma with prior lognormal distribution based on fossil achenes identified as *Cirsium* (Mai, 1995 ); and the age of the fifth group (v, “*Volutaria* group”) with the prior normal distribution, was 22.1 Ma, and the stdev was 1.9. The lognormal relaxed clock and random starting tree under “Yule speciation” process were selected after testing alternative options by “path sampling” and “stepping-stone sampling” (Baele & al., 2012; Baele & al., 2013). All priors were left at their default values, but parameters ucl.d.mean and ucl.d.stdev were changed to normal distribution, which improved the resulting model with all effective sample size (EES) values over 1000 according to Tracer v.1.4. BEAST was run for 40 x 10<sup>6</sup> generations, and sampling occurred every 10<sup>3</sup> generations. Trees were summarized in a maximum clade credibility tree implemented in TreeAnnotator v.1.8.2 and visualized in FigTree v.1.3.1.

**Biogeographical inference.** — We defined nine geographic regions based on the concentration of species of *Volutaria*, the geographic isolation of taxa and the whole chorology of the outgroup species: Arabian Peninsula, Asia, East Africa, Horn of Africa, Iberian Peninsula, Macaronesia (Canary islands), North Africa and Socotra (Figs. 1, 4). Using the time calibrated tree (the maximum-credibility tree) obtained from Beast (cf. *Molecular dating*), we applied six different biogeographical models (DEC, DEC+J, DIVA, DIVA+J, BayArea, BayArea+J) as implemented in the R package

BioGeoBEARS (Matzke, 2013; Matzke, 2014) under a maximum likelihood framework, and compared their statistical fit using AIC. The three main models allow for different biogeographical possibilities; the free parameter "+J" enables one daughter lineage to disperse to an area outside the ancestral range (i.e. without adjoining it to the ancestral range), thus incorporating the process of founder-event speciation (allowing long-distance dispersal events). All analyses were run with a time and dispersal restriction applied to Macaronesia (which here corresponds only to the Canary Islands), in which we fixed at 0 the dispersal probability to this region before 23 Ma, in order to incorporate the fact that the oldest Canary island emerged around that time (van den Bogaard, 2013).

## RESULTS

**Monophyly and diversity of *Volutaria*.** — Nuclear markers strongly supported the monophyly of *Volutaria* (BS = 100%, BPP = 1; Fig. 2) and established the presence of four major clades within the genus. The first three clades included five North African species [*V. crupinoides* (Desf.) Maire, *V. maroccana* (Barratte & Murb.) Maire, *V. muricata*, *V. sinaica* (DC.) Wagenitz and *V. saharae*], most of which have low chromosome numbers ( $2n = 24, 26, 28$ ), except for *V. saharae* (Figs. 1--2, Table 1). All these species are strongly supported as monophyletic (BS = 100%, BPP = 1), except for *V. saharae*, which also groups together with *V. lippii* (Fig. 2).

Clade 4 (BS = 100%, BPP = 1) contains the remaining species, which have higher chromosome numbers,  $2n = 32$  (Fig. 2, Table 1). It comprises a well-supported eastern African and Arabian subclade (clade 5) among the western Africa species *V. belouini* (Humbert) Maire, *V. lippii* Cass., and *V. saharae*, the Macaronesian endemics *V.*



*canariensis* Wagenitz and *V. bollei* (Sch. Bip. ex Bolle) G. Kunkel, and the widespread *V. tubuliflora*.

Clade 5 (BPP = 1) comprises three further subclades containing six species in total. *Volutaria albicaulis* samples from the Djebel Urays (Yemen) are strongly supported in one clade (BS = 98%, BPP = 1), but combined nuclear markers do not show differences between *V. abyssinica* and *V. boranensis* (Cufod.) Wagenitz, nor between *V. dhofarica* and *V. djiboutensis* Wagenitz. This fifth clade also includes two species: namely, populations of *V. albicaulis* (Deflers) Wagenitz from the Arabian Peninsula (excluding Djebel Urays) and *V. somalensis* from northern Somalia, only represented in Figs. 3--4.

Chloroplast markers strongly confirm and support the monophyly of *Volutaria* (BS = 100%, BPP = 1), but they provide a poorly resolved phylogram (Fig. 3), even with the inclusion of coded indels. Most of the species appear unresolved in a single large clade. Chloroplast markers also show some incongruencies compared with the nDNA data (Figs. 2--4): i), with some sequences of eastern *V. boranensis* (Cufod.) Wagenitz, *V. dhofarica* and *V. djiboutensis* appearing as sister to sequences of western *V. maroccana* and *V. crupinoides* from Morocco, Tunisia and Jordania, and sequences of *V. sinaica* from Morocco (Electr. Suppl.: Table S1); ii). Populations of *V. crupinoides*, *V. lippii* and *V. maroccana* are segregated into different clades (Electr. Suppl.: Table S2); iii), whereas there is strong support for grouping the two Macaronesian species *V. bollei* and *V. canariensis* together (BS = 97%, BPP = 1; Fig. 3).

Chloroplast markers plus cloning of ITS (Figs. 3, 4; Electr. Suppl.: Figs. S1, S3) also suggest a hybrid origin for *Volutaria saharae* ( $2n = 64$ ). This species is grouped with *V. sinaica* in the cpDNA tree (Fig. 3), but possesses two different ITS copies, one identical to a Tunisian sample of *V. lippii* subsp. *lippii* with  $2n = 32$ , whereas the other is identical to an Arabian sample of *V. sinaica* with  $2n = 28$  (Fig. 4 clades 1, 5, Table 1).

*Volutaria djiboutensis*, based on its chromosome number ( $2n = 58$ ), may represent another hybrid taxon related to *V. crupinoides* ( $2n = 26$ ) and *V. dhofarica* ( $2n = 32$ ). Moreover, *V. djiboutensis* and *V. dhofarica* group together in the ITS tree (Figs. 4; Electr. Suppl.: Figs. S1, S3), and also with *V. crupinoides*, together with *V. maroccana* and *V. sinaica*, in the cpDNA tree (Fig. 3).

ITS sequences suggest the probable existence of a new species. Herbarium specimens originally identified as *V. albicaulis* appear in our nDNA trees in three different supported clades (Fig. 4 clades 8, 10--11; Electr. Suppl.: Figs. S1, S3). Moreover, cloning of Arabian *V. albicaulis* yielded ITS copies similar (at least in part) to those of *V. abyssinica*, *V. boranensis* and to cloned copies of *V. somalensis* (Fig. 4 clades 11--10), but not to *V. albicaulis* from the type locality, namely Djebel Urays, in Yemen (Fig. 4 clade 8). Thus, henceforth, the Arabian Peninsula populations of *V. albicaulis* (extra Djebel Urays) will be referred to as *V. albicaulis*\*

The ITS trees (Fig. 4; Electr. Suppl.: Fig. S1) also reveals identical nDNA copies shared between non-sympatric species. *Volutaria somalensis* from the mountains of northern Somalia shares an ITS copy with *V. belouini*, currently separated from that taxon by the Sahara Desert (Fig. 4 clades 6). Moreover, *Volutaria somalensis* shares ITS copies with species separated by the Red Sea: *V. abyssinica*, *V. albicaulis*\*, and *V. boranensis* (Fig. 4 clades 10--11). The four species also show strong reticulation (Electr. Suppl.: Fig. S3). Likewise, *Volutaria dhofarica* contains some ITS copies related to *V. djiboutensis* and *V. socotrensis* (Fig. 4 clade 14), together with copies related without significant support with *V. albicaulis* and *V. albicaulis*\* (Fig. 4 clades 7, 9; Electr. Suppl.: Figs. S1, S3).

**Timing and diversification in *Volutaria*.** — Our ITS chronogram places the origin of *Volutaria* at the beginning of the Miocene, 21.66 Ma (node 63, Fig. 4, Table 3), when it diverged from a sister clade leading to the formation of its Asian relatives *Amberboa*, *Goniocaulon*, *Plagiobasis*, *Russowia* and *Tricholepis* (Fig. 4). The chronogram strongly supports (BPP = 1) four main diversification events (Fig. 4).

The first diversification occurred in the middle-lower Miocene 15.05 Ma (node 68) and resulted in the origin of four clades, hereafter considered to be “early diverging clades” that contain: *V. sinaica* (clade 1), *V. crupinoides* and *V. maroccana* (clade 2), *V. muricata* (clade 3), and a fourth clade containing the remaining species of the genus.

The second and major diversification event occurred in the middle-upper Miocene (9.13 Ma, node 76) with the separation of ten clades (clades 5 to 14) comprising species currently found in Macaronesia, western Africa, eastern Africa and southern Arabia.

The third diversification event took place throughout the late Miocene and the transition from Pliocene to Pleistocene. First, at 6.53 Ma, *V. maroccana* split from *V. crupinoides* (clade 2). Then, between 4.11 and 2.19 Ma, more species arose in clade 4: i) the narrow endemic Macaronesian *V. bollei* segregated from *V. lippii* and a clone of *V. saharae* (clade 5). *Volutaria bollei* represents the colonization of the Canary Islands, which is independent from the arrival of the other Macaronesian species *V. canariensis*, whose timing remains unresolved (clade 13); ii) *Volutaria albicaulis*\* and its clones diverged from *V. abyssinica*, *V. boranensis* and *V. somalensis* (clades 10--11, nodes 96, 101); iii) Some samples of *V. socotrensis* segregated from a taxon related to *V. djiboutensis* and *V. dhofarica* (clade 14). Finally, throughout the Pleistocene, clones of *V. somalensis* separated from *V. belouini* (clade 6, node 82), but also occur grouped with *V. abyssinica* and *V. boranensis* (clade 11, node 103). Similarly, clones of *V. saharae* suggest that this species was once closely linked to two different North African

species, *V. sinaica* and *V. lippii* (clades 1, 5). Hereafter, all clades derived within clade 4 (Fig. 4) and related species that originated throughout the Pliocene and Pleistocene will be referred to as “late diverging clades or species” since they originate from younger ancestors (*i.e.* nodes  $\geq 76$ , with intermediate or terminal positions in the tree), as compared with clades 1--3 and related species derived from the basal node 68 (Fig. 4).

**Biogeographical reconstruction.** – The BayArea+J model fitted the data best ( $\Delta AIC = 0$ ), DEC+J being the next best ( $\Delta AIC = 29.49$ ). Figure 4 and Table 3 show the relative probability ( $p$ ) of each reconstruction for each supported node based on the Beast analysis and the reconstruction obtained for BayArea+J (Data are available upon request from the corresponding author). According to this model, the ancestor of *Volutaria* and its relatives (*Amberboa*, etc.) originally occurred in Asia (node 63;  $p = 0.92$ ). Africa emerges as the ancestral area for the first diversification pulse of *Volutaria* during the early-middle Miocene (Crown node of Fig. 4 clades 1--4, node 68;  $p = 1$ ). Subsequently, the ancestral area of those taxa involved in the second diversification pulse also includes North Africa (Crown node of Fig. 4 clades 5--14, node 76;  $p = 1$ ). Throughout the Pliocene and Pleistocene, three further changes occurred: i) The ancestor of the northwestern African species *V. belouini* diverged from the eastern African species *V. somalensis*, probably by vicariance (Fig. 4 clade 6, node 82;  $p = 0.99$ ); ii) North Africa appears to be the most probable ancestral area for the Macaronesian species *V. bollei* (Fig. 4 clade 5, node 107;  $p = 1$ ); iii) and Socotra would have been the geographical source not only of *V. socotrensis*, but also of Arabian *V. dhofarica* and African *V. djiboutensis* (Fig. 4 clade 14, node 88;  $p = 0.59$ ).

## DISCUSSION

**Phylogenetic diversity.** — Nuclear and plastid markers (Figs. 2--3) support the monophyly of all the investigated species of *Volutaria*. Consequently, the proposal that morphological convergence of cryptic taxa (at generic level) occurs on both sides of the Sahara should be discounted (Andrus & al., 2004). A close relationship to the Asian genus *Amberboa*, within which many *Volutaria* species have been included for so long, can also be discounted, as had been suggested by Wagenitz (1989, 1991).

*Volutaria* species are unevenly distributed in four major clades. The first three contain early-diverging taxa that mainly occur throughout North Africa (Figs. 2, 4). The fourth clade comprises species that represent the remaining diversity of the genus, and this clade splits into a polytomy consisting of several clades. The most evident is a robust clade comprising all the eastern species (east Africa and south Arabia; Clade 5, Fig. 2). The remaining clades of this polytomy largely represent species that currently have a western distribution (northwest Africa and Macaronesia).

Monophyly of each of the western species in the fourth clade is supported by nuclear markers, which also help to clarify certain interspecific relationships and highlight the need for further phylogenetic research. Both Macaronesian endemic species *V. bollei* and *V. canariensis* stand alone as well-defined taxa (Fig. 2; Electr. Suppl.: Figs. S1--S2) that share plastid sequences (Fig. 3). However, the phylogenetic relationships between these, and indeed, with continental species, need to be studied in greater detail. *Volutaria bollei* shares an ancestor with African *V. lippii*. However, *V. lippii* and *V. canariensis* are not closely related according to molecular data (Figs. 2--4; Electr. Suppl.: Figs. S1--S3), despite clear morphological similarities (Wagenitz, 1991).

*Volutaria lippii* is another well-defined taxon, clearly separated from the very similar *V. tubuliflora* (Fig. 2), over which some confusion has long existed (Murbeck, 1897, 1923; Wagenitz, 1989). However, the two accepted subspecies of *Volutaria lippii*

(Wagenitz, 1991) have different interspecific relationships. *Volutaria lippii* subsp. *medians* (Maire) Wagenitz shares plastid polymorphisms with sympatric *V. crupinoides*, *V. maroccana*, and *V. tubuliflora* (Fig. 3; Electr. Suppl.: Table S2), whereas *V. lippii* subsp. *lippii* is involved in the hybrid origin of *V. saharae*, together with *V. sinaica* (Figs. 2, 4); see section *Hybrids and interspecific introgression*.

Compared with their western counterparts, eastern species of the fourth clade are not as well supported. *Volutaria abyssinica*, *V. boranensis*, *V. somalensis*, and the newly discovered Arabian taxon, *V. albicaulis*\*, forms an unresolved complex (Figs. 2, 4) of mountain-dwelling taxa centered around the Aden Gulf (Horn of Africa plus the Arabian Peninsula), at altitudes exceeding 1450 m a.s.l. (Table 1). *Volutaria abyssinica* cannot easily be segregated from *V. boranensis* in our phylogeny (Fig. 2). The main differences between these taxa are the longer tips of the phyllaries and the smaller hermaphrodite flowers and capitula of *V. boranensis* (Wagenitz, 1991; Mesfin Tadesse, 2004; Calleja & al., in prep.). Within this complex, *V. albicaulis*\* (Fig. 4) has long been misidentified as *V. albicaulis* or *V. abyssinica* (Calleja & al., in prep.). The latter, however, does not exist in the Arabian Peninsula, and the only specimens that we consider to be *V. albicaulis sensu stricto* are those from the type locality (Deflers, 1896), namely, Djebel Urays and the surrounding mountains, at altitudes of 900--1400 m a. s. l. (Fig. 1, Table 1). Conversely, specimens that we provisionally named *V. albicaulis*\* (including the cloned specimen), show some obvious morphological differences (Calleja & al., in prep), and were collected at higher altitudes across a much wider geographical range, encompassing Southern Arabia and Yemen (Fig. 1, Table 1). The origin of *Volutaria albicaulis*\* (Arabian Peninsula) may be related to that of *V. somalensis* (Somalia), especially since both species share an ancient relationship with *V. abyssinica* and *V. boranensis* (Fig. 4 clades 10--11), which is supported by interspecific

reticulation (Electr. Suppl.: Fig. S3). However, when one considers that *V. somalensis* might possibly share an ancestor with the western narrow endemic *V. belouini* (Fig. 4 clade 6), its origin becomes increasingly complex.

A second group of species from the same subclade of eastern taxa, and occurring in the same area (Horn of Africa and Arabian Peninsula), but at altitudes of less than 1500 m a. s. l., is represented by *V. dhofarica* (Yemen and Oman), *V. djiboutensis* (Djibouti) and *V. socotrensis* (Socotra). *Volutaria socotrensis* has recently been separated from *V. dhofarica*, based on its larger habit and its occurrence on Socotra Island (Kilian & Hein, 2006). However, all the available herbarium specimens from Socotra morphologically resemble *V. dhofarica* (Calleja & al., in prep) and appear to be grouped in a well-supported clade, together with *V. dhofarica* and *V. djiboutensis* (Figs. 2, 4). *Volutaria djiboutensis* shows distinctive morphological traits (Wagenitz & Hellwig, 1996b, Calleja & al., in prep.), but molecular markers also failed to distinguish it from *V. socotrensis* and *V. dhofarica*. Furthermore, the position of ITS clones in the phylogenetic trees (Fig. 4; Electr. Suppl.: Figs. S1) revealed that there is a case here for considering *V. dhofarica* to be a complex taxon: it is related to *V. djiboutensis* and *V. socotrensis*, but it also conserves ITS copies as unrelated polymorphisms.

Further detailed integrative studies should now focus on disentangling the diversity of this genus in the Aden Gulf. We recognized the occurrence of two different clades (*V. abyssinica* and relatives, and *V. dhofarica* and relatives) growing in the same region, but at different altitudinal ranges. These clades do not overlap geographically with a third clade, represented by the micro-endemic *V. albicaulis* in southwestern Yemen.

**Hybrids and interspecific introgression.** — Cloned sequences of ITS help both to unravel the origin of some *Volutaria* species and to address their phylogenetic and biogeographical relationships. *Volutaria saharae* is derived from the North African

lineage represented by *V. lippii* subsp. *lippii*, and the North African and Arabian Peninsula lineages represented by *V. sinaica* (Figs. 2--4), which would have been the maternal parent, and is morphologically identical to *V. saharae*, differing mainly in the position of the hilum on the achenes (Chevallier, 1905; Wagenitz, 1991). Its unique ability to colonize the Sahara Desert might be a consequence of its allopolyploid origin (Rieseberg & al., 2003), and the karyological incongruence with its putative parents (see Table 1) may be due to descending dysploidy in *V. sinaica* (Wagenitz & Hellwig, 1996a).

*Volutaria djiboutensis* may well be another hybrid species. Hellwig (1994; Table 1) suggested an allopolyploid origin for this taxon because its chromosome number is equal to the sum of the chromosome numbers of *V. dhofarica* and *V. crupinoides*. Our results support this hypothesis in that *V. djiboutensis* shares nDNA and cpDNA polymorphisms with both species (Figs. 3--4; Electr. Suppl.: Figs. S1, S3). If this is so, then this hybrid origin should be to the result of an ancient hybridization event, because the purported parental species are currently non-sympatric.

Regarding introgression, *V. crupinoides*, *V. maroccana* and *V. sinaica*, placed by our analysis in early diverging clades (Fig. 4), share plastid polymorphisms with the currently non-sympatric African and Arabian *V. boranensis* and *V. djiboutensis* (Fig. 3; Electr. Suppl.: Table S1). As *V. crupinoides*, *V. maroccana* and *V. sinaica* appear at terminal nodes (or tips *sensu* Schaal & al., 1998) of the tree (Fig. 3), this unexpected relationship could provide evidence for the retention of ancestral polymorphisms (Schaal & al., 1998). However, ancient hybridization events cannot be entirely ruled out, since the shared plastid polymorphisms are found in all three plastid markers (Electr. Suppl.: Table S1; see McKinnon, 2005; Twyford & Ennos, 2012).



Similarly, certain individuals representing four species and two subspecies (*V. crupinoides*, *V. lippii* subsp. *lippi*, *V. l.* subsp. *medians*, *V. maroccana*, *V. muricata*, and *V. tubuliflora*) share plastid polymorphisms (Figs. 2--3; Electr. Suppl.: Table S2). These molecular similarities may be the result of incomplete lineage sorting of ancestral polymorphisms or plastid introgression by interspecific hybridization (McKinnon, 2005; Twyford & Ennos, 2012). There is more support for introgression between *V. lippii* subsp. *medians* and *V. maroccana*, because these taxa are currently sympatric and share polymorphisms in all three plastid markers (Electr. Suppl.: Table S2), which is difficult to explain with incomplete lineage sorting (McKinnon, 2005; Twyford & Ennos, 2012). Conversely, the shared polymorphisms of *V. crupinoides* and *V. lippii* subsp. *medians*, or those of *V. lippii*, *V. maroccana*, *V. muricata*, *V. sinaica*, and *V. tubuliflora*, do not follow any geographical pattern, and they are not present in all the makers. In these cases, incomplete lineage sorting of ancestral polymorphisms cannot be discounted. (McKinnon, 2005; Twyford & Ennos, 2012).

**Biogeography.** — The phylogeny of *Volutaria* confirms that the genus exhibits a distribution that agrees with that of the Rand Flora, having two diversity poles at either side of the Sahara Desert, one in northwestern Africa and Macaronesia, and the other on the Horn of Africa and in southern Arabia, including Socotra. Only two species, *V. sinaica* and *V. tubuliflora*, currently display wide distributions that encompass both areas. Since the closest known relatives of *Volutaria* occur in Asia (Susanna & Garcia-Jacas, 2009; Barres & al., 2013), the hypothesis that *Volutaria* originally occurred across northern Africa, and that its present distribution is due to fragmentation followed by western-eastern vicariance, is rejected. Instead, some authors suggested that *Volutaria* could have immigrated from western Asia and successfully radiated into

1 North and western Africa, before reaching Macaronesia and southwestern Europe  
2 (Hellwig, 2004; Susanna & Garcia-Jacas, 2009). This westward expansion reflects the  
3 migration hypotheses inferred and confirmed for many other taxa (Davis & Hedge,  
4 1971; Roquet & al., 2009; Barres & al., 2013; Mairal & al., 2015). Hellwig (2004) also  
5 suggested that this westward migration is supported by the predominantly perennial  
6 habit of the Arabian and eastern African species. However, our data show a much more  
7 complex biogeographical history.

8 The origin of the genus and the two current poles of biodiversity are a consequence  
9 of different migration and diversification episodes. Migration, radiation and extinction  
10 events are linked to the principal tectonic and climatic changes in western Asia and  
11 North Africa during the Neogene and Pleistocene. *Volutaria* initially originated, and  
12 subsequently diverged, from its Asian relatives (*Amberboa*, *Goniocaulon*, *Plagiobasis*,  
13 *Russowia* and *Tricholepis*) in western Asia during the lower Miocene (20 Ma). At that  
14 time, Asia and the Tethys were subjected to an aridification event that resulted in the  
15 opening of new habitats. These changes triggered the radiation of the Asian flora and  
16 allowed it to migrate throughout North Africa and southern Europe, until it eventually  
17 reached Macaronesia (Hellwig, 2004; Roquet & al., 2009; Barres & al., 2013). The first  
18 diversification event for *Volutaria* subsequently occurred during the middle Miocene  
19 (15 Ma), and this was facilitated by the connecting of western Asia to southern Arabia  
20 and northeastern Africa (Allen & Armstrong, 2008), and the availability of large, open  
21 habitats (Feakins & Demenocal, 2010). The ancestors of the early diverging clades (Fig.  
22 4 clades 1--3) will have migrated and colonized new areas reaching North Africa, later  
23 giving rise to *V. crupinoides*, *V. muricata*, and *V. sinaica*. Simultaneously, it is possible  
24 that clade 4 might have reached the same area, but also spread to the Horn of Africa and  
25 the Arabian Peninsula. Fruit dispersal mechanisms have been largely ignored, but the

occurrence of ectozoochory and wind dispersal in *Volutaria* should be considered, based on the presence of a semi-rigid, denticulate pappus and an elaiosome, in conjunction with the lightness of the achenes (Wagenitz & Hellwig, 1996a).

The time-calibrated tree reveals that the ancestors of present-day Arabian and eastern African species (*V. abyssinica*, *V. albicaulis*, *V. boranensis*, *V. dhofarica*, *V. djiboutensis*, *V. socotrensis*, and *V. somalensis*) originated later than those of western (*V. muricata*) or more widespread species (*V. crupinoides* and *V. sinaica*; Fig. 4). The Asian ancestors of this first westward migration wave would have declined in their original eastern habitats. Extinction events should be considered in palaeobiogeographical reconstructions (Jakob & Blattner, 2006; Susanna & al., 2011; Mairal & al., 2015), especially for arid regions where species turnover rates are much higher compared to those of more mesophilous habitats (Stebbins, 1952). Although this proposal is very speculative, the cloned copies of *V. dhofarica* (Southern Arabia) of unsupported relationships might also be considered as proof of extinction of ancestral lineages. On the basis of these various pieces of evidence, the hypothesis that extant, perennial, eastern species are ancestral (Hellwig, 1994) should be rejected, firstly, because eastern species do not represent ancient lineages (Fig. 4), and secondly, because both eastern and western species can be both annual or perennial (Table 1).

The time-calibrated ITS tree shows a second and major radiation event that occurred during the Late Miocene (9.13 Ma), and that resulted in the formation of ten new lineages, with six taxa radiating in East Africa (Fig. 4). This second diversification event is concomitant with the onset of orogenic pulses and subsequent changes in climate and vegetation across Africa during the Late Miocene (Sepulchre & al., 2006; Feakins & Demenocal, 2010). The increasingly arid and semi-arid conditions may have offered newly available habitats for diversification events (Hellwig, 2004; Barres & al.,

2013) throughout the potential ancestral area indicated by BioGeoBEARS, which encompasses North Africa (Fig. 4 clade 4). The time-calibrated tree indicates that this wide distribution could have been reached in parallel with the westward expansion of clades 1--4 (Fig. 4). The potentially wide range of *Volutaria* during the Miocene (clades 1--4) might explain the links revealed by molecular data that still exist today between geographically separated species. The incongruencies revealed by the chloroplast phylogeny (Fig. 3) related to shared polymorphisms (Electr. Suppl.: Table S1) suggest the occurrence of ancient introgression between the ancestor of the eastern African and Arabian species *V. abyssinica*, *V. boranensis*, *V. dhofarica*, and *V. djiboutensis* and the western or widespread species *V. crupinoides*, *V. maroccana*, and *V. sinaica* (Fig. 4).

A more detailed view is offered by the nuclear phylogeny (Fig. 2), which demonstrates the large diversification pulse of clade 4 in eastern (east African and south Arabia) and western clades (northwest Africa, the Iberian Peninsula and Macaronesia). The extreme aridity of North Africa during the Messinian 7--5 Ma (Schuster & al., 2006) may have fragmented and restricted the range of many species on both sides of Africa (Axelrod and Raven, 1978; Sanmartín & al., 2010; Barres & al., 2011; Pokorný & al., 2015), thus promoting an eastern-western disjunction of clade 4, and subsequent speciation by vicariance (Thiv & al., 2010; Mairal & al., 2015).

The third diversification pulse took place on both sides of the distribution range of the genus throughout the Pliocene-Pleistocene transition; a period characterized by extreme climate variability, with arid episodes across North Africa and Macaronesia (Senut & al., 2009). At the eastern edge, the ongoing opening of the Red Sea and the Aden Gulf (Bosworth & al., 2005) may also have promoted speciation (Bruyns & al., 2011), splitting an ancient lineage of mountain-dwelling taxa into a number of different species, for example, *V. abyssinica* in the high plateaus of Ethiopia, *V. albicaulis*\* in the

mountains of the southeastern Arabian Peninsula, and *V. somalensis* in the mountain region of Somaliland (Fig.1 and Table 1). Later, the expansion of semi-arid environments in eastern Africa (Trauth & al., 2005; Senut & al., 2009) may have triggered the separation of *V. boranensis* from *V. abyssinica*. By this time, *V. boranensis* had reached as far as Kenya and Tanzania, the southeastern limit of the genus in eastern Africa (Fig. 1). Also in the same region (Red Sea and Aden Gulf), but at lower elevations (< 1500 m), another lineage began to diversify during the Pliocene-Pleistocene transition. It involved at least three very closely related taxa, namely *V. dhofarica* (southern Arabian Peninsula), *V. djiboutensis* (Djibouti) and *V. socotrensis* (Socotra island). As a by-product, the date assigned to this diversification confirmed that Socotra was colonized by *Volutaria* long after its separation from nearby continents (D'Acremont & al., 2010). The lack of support within clade 14 of Fig. 4 renders irrelevant the result by BioGeoBEARS that Socotra was the ancestral region for this clade.

On the western edge, the climatic changes of the Pliocene-Pleistocene transition (Trauth & al., 2009; Fernández-Palacios & al., 2011) would also have favored colonization and speciation events in northwest Africa and Macaronesia. The arrival of the ancestor of *V. bollei* in the Canaries indicates a successful colonization event, together with that of *V. canariensis*, but unfortunately, the timing of this cannot be resolved satisfactorily (Fig. 4). It is possible that the ancestral areas for both species might be North Africa, the closest geographical source supported by BioGeoBEARS, as also suggested for other Macaronesian taxa (Hellwig, 2004; Caujapé-Castells, 2011). This colonization agrees with those of other taxa commonly found in open habitats in semi-desert climatic regimes (Kim & al., 2008; Barres & al., 2011). Migration was probably triggered by the shift from a wet tropical regime to a seasonal climatic pattern,

with recurrent aridity crises and strong westward winds in the Canary Islands (Trauth & al., 2009; Fernández-Palacios & al., 2011 and references therein). Unlike other genera (Kim & al., 2008; Barres & al., 2011), *Volutaria* did not successfully radiate in Macaronesia. BioGeoBEARS indicates that *V. bollei* and *V. lippii* share an African ancestor (Fig. 4 clade 5). Speciation of *V. lippii* could have been initiated by the onset of the Mediterranean climate (Suc, 1984), since this taxon currently prefers open habitats in dry Mediterranean areas.

Environmental fluctuations, coupled with Pleistocene glaciations, might also have driven the subsequent expansion, fragmentation, and diversification processes of *Volutaria* in North Africa, and similar results have been obtained for other taxa, such as *Aeonium* Webb & Berthel., *Campanula* L. and *Hypericum* L. (Pokorny et al. 2015). ITS cloning (Fig. 4) reveals a common ancestor for *V. somalensis* (Somalia) and *V. belouini* (Morocco), that may have resulted from ulterior vicariant speciation on both sides of the Sahara Desert. Morphologically, *V. belouini* more closely resembles species that occur on the Horn of Africa than those that currently inhabit Northwest Africa (Calleja & al., in prep).

Glacial cycles might also have promoted diversification in *Volutaria* by precipitating hybridization events between temporally sympatric species (Nieto Feliner, 2014). ITS cloning reveals the hybrid origin of *V. saharae*, formed from two species that currently do not co-exist, namely, *V. lippii* and *V. sinaica*. The potential connections between them could have been mediated by recurrent expansions and contractions of semi-desert and Mediterranean plant communities (Hooghiemstra & al., 1992).

As well as climatic shifts, human activities might also have indirectly driven the expansion of some species (Marini & al., 2012). *Volutaria tubuliflora* shows by far the

widest range (Fig. 1), and has recently been reported from desert regions of Chile (Teillier & al., 2014) and California (C.J. McDonald, pers. comm). It invariably occurs alongside roads or close to irrigated crop fields in arid regions (e. g. Arabian deserts or southern Morocco, J.A. Calleja, University Autònoma of Barcelona, pers. obs.), where alien species are able to withstand a wide range of otherwise adverse climatic regimes, including semi-desert and desert environments (Abd El-Ghani & El-Sawaf, 2004).

## ACKNOWLEDGEMENTS

Financial support provided by the Spanish Ministry of Science and Technology (project CGL2009-13322-C03-03), the “Generalitat de Catalunya” (“Ajuts a Grups de Recerca Consolidats” 2009/SGR/00439 and 2014/SGR/00514), Alliance 4 Universities, the University Autònoma of Barcelona, the Linnean Society of London and the Systematics Association is gratefully acknowledged. We thank L. Barres, M. Galbany, J. López-Alvarado, S. López-Vinyallonga, and R. Vilatersana for help with laboratory procedures and L. Sáez for his assistance at Montpellier Herbarium, as well as J. Molero, J.J. Aldasoro, A. Hilpold, L. Barres and Á. Marrero for collecting particular species. We greatly appreciate the help of CIRAD Montpellier, Herbarium of Montpellier University, Herbarium of the National Museum of Natural History of Paris, Herbarium Berolinense, Herbarium of the Natural History Museum in Wien, Royal Botanic Garden of Edinburgh, Royal Botanic Gardens, Kew, Uppsala University, Herbarium of the University of Salamanca, Botanic Institute of Barcelona, Royal Botanic Garden of Madrid and the project *Flora Iberica* (CGL 2008-02982-C03) in providing herbarium specimens and/or DNA samples.

**Table 1.** Sample sizes, chromosome numbers, life form, habitat preferences and altitudinal range of the studied taxa. Samples = Number of sampled populations.  $2n$  = Diploid chromosome numbers after Hellwig (1994). A = annual, P = perennial; uppercase letter refers to the more common life form whereas lowercase refers to the less common life form (based on information available from herbarium specimens, Floras and personal observations). <sup>1</sup> Hellwig (1994) published the chromosome number based on a herbarium specimen determined as *V. albicaulis*, but our molecular results for the same specimen indicate a new species (*V. albicaulis*\*). <sup>2</sup> Hellwig (1994) published the chromosome number of an unidentified herbarium specimen found at Forêt du Day (Djibouti) where only *V. djiboutensis* occurs.

Taxon	Samples	$2n$	Life form	Habitat and altitude (m)
<i>V. abyssinica</i> (A.Rich.) C. Jeffrey ex Cufod.	4	-	A, P	High open plateau with <i>Acacia</i> sp. pl., 1500 -- 2400 m
<i>V. albicaulis</i> (Deflers) Wagenitz	2	-	A, P	Open shrublands on volcanic areas, 1000 -- 1300 m
<i>V. albicaulis</i> *	2	32 <sup>1</sup>	P	Rocky open areas, 1400 -- 2450 m
<i>V. belouini</i> (Humbert) Maire	1	-	P, a	Rocky open areas, 900 -- 1400 m
<i>V. bollei</i> (Sch.Bip. ex Bolle) G. Kunkel	1	-	A, P	Volcanic rocks with scattered herbaceous vegetation, 10 -- 400 m
<i>V. boranensis</i> (Cufod.) Wagenitz	4	32	A, p	Disturbed savannas and roadsides, 1350 -- 2050 m
<i>V. canarensis</i> Wagenitz	7	32 + 1b	A	Open shrublands on volcanic rocky areas; 20 -- 550 m
<i>V. crupinoides</i> (Desf.) Maire	7	26	A	Disturbed rocky areas, roadsides, field crops and dry river beds, 50 -- 1800 m
<i>V. dhofarica</i> Wagenitz	4	32	A, p	Sandy soils, 500 -- 1300 m
<i>V. djiboutensis</i> Wagenitz	1	58 <sup>2</sup>	A	Clay soils, < 1450 m
<i>V. lippii</i> Cass.	1	32	A	Open shrublands, disturbed habitats, roadsides, field crops and dry river banks, 0 -- 150 m
<i>V. lippii</i> subsp. <i>medians</i> (Maire) Wagenitz	4	32	A	Open shrublands, disturbed habitats, roadsides and field crops; 0 -- 150 m
<i>V. maroccana</i> (Barratte & Murb.) Maire	3	-	A	Open shrublands, disturbed habitats, roadsides and field crops, 0 -- 180 m
<i>V. muricata</i> (L.) Maire	5	24	A, p	Open shrublands and woodlands, disturbed habitats, roadsides and field crops, 100 -- 700 m
<i>V. saharae</i> (L.Chevall.) Wagenitz	1	64	A	Rocky soils, 150 -- 600 m
<i>V. sinaica</i> (DC.) Wagenitz	5	28	A, p	Rocky soils, open shrublands, 650 -- 1700 m
<i>V. socotrensis</i> Wagenitz	2	-	A, P	Rocky soils, open shrublands, 400 -- 600 m
<i>V. somalensis</i> (Oliv. & Hiern) C. Jeffrey	2	-	P	Rocky grounds, open shrublands, 1800 m
<i>V. tubuliflora</i> (Murb.) Sennen	14	32 & 64	A	Disturbed habitats, roadsides, field crops, 0 -- 760 m



**Table 2.** Data set information and tree parameters: number of taxa and sequences, regions, number of bases, gaps and informative characters, Consistence and Retention indices, and models of DNA evolution.

	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
Tree	Figure 2	Figure 3	Figure 4 (time calibrated)	Supplementary material Figures S1, S3	Supplementary material Figure S2
N° of Taxa	16 <i>Volutaria</i> species + 2 subspecies + clones + 3 Outgroup	17 <i>Volutaria</i> species + 2 subspecies + 3 Outgroup	17 <i>Volutaria</i> species + 2 subspecies + clones + 11 Outgroup	17 <i>Volutaria</i> species + 2 subspecies + clones + 3 Outgroup	16 <i>Volutaria</i> species + 2 subspecies + 3 Outgroup
N° sequences	66	87	56	96	67
Regions	ITS + ETS	<i>trnL-F</i> + <i>rpl32-trnL_UAG</i> + <i>ndhF</i> + coded indels	ITS + clones	ITS + clones	ETS
N° cloned species + N° clonal sequences	2 + 4	-	4 + 14	4 + 15	1 + 2
Aligned length	624 + 1000	735 + 927 + 1250 + 18	635	624	1000
N° of gaps	50	18	20	14	27
Parsimony analysis					
Parsimony informative characters	225	53	---	76	158
N° of most parsimony trees	498500	218	---	400000	395100
N° of steps	406	68	---	135	287
Consistency index	0,6405	0,6628	---	0,5592	0,5915
Retention index	0,9148	0,9152	---	0,9134	0,8900
Homoplasy index	0,3595	0,3372	---	0,4408	0,4085
Bayesian inference					
Model of molecular evolution	SYM+G; GTR+G	F81+G; GTR+G; GTR+G; F81	SYM+G	SYM+G	GTR+G

**Table 3.** Age estimates and reconstructed ancestral ranges for each of the nodes in the chronogram represented in Figure 4. Bayesian posterior probabilities (>0.9), Mean / median crown ages and 95% highest posterior density (HPD) intervals (millions of years = Ma) based on a relaxed molecular-clock analysis of ITS sequences in BEAST. Letters correspond to the following ancestral areas or combination of areas with a relative probability (BayArea+j Model) equal or greater 0.05: A = Asia, E = East Africa, H = Horn of Africa, I = Iberian Peninsula, M = Macaronesia (Canary Islands), N = North Africa, R = Arabian Peninsula, S = Socotra, U = Europe (except Iberian Peninsula). Asterisks (\*) at 47, 51, 63, 72, and 84 refer to the nodes of Barres & al. (2013) on the basis of which the ITS tree was calibrated.

Clade	Node	Bayesian posterior probabilities	Age (Ma)	95% HPD interval	BayArea+j Model probabilities
47*	57		30.53 / 30.53	26.29 -- 34.93	
	59	0.99	20.42 / 20.19	13.60 -- 27.61	A:0.57; AEHIMNSRU:0.29; N:0.14
72*	60	1	14 / 14	14 -- 18.26	AU:0.67; A:0.33
51*	61	1	11.90 / 11.90	7.26 -- 17.73	N:1
63*	62	1	19.96 / 19.96	15.31 -- 24.97	A:1
84*	63	1	22.10 / 22.10	17.38 -- 27.51	A:0.92; AEHIMNSRU:0.08
	65	0.99	11.78 / 11.73	5.66 -- 17.87	A:1
	66	0.94	13.56 / 13.59	7.94 -- 19.96	A:1
	67	0.97	10.07 / 11.73	4.53 -- 15.70	A:1
	68	1	15.05 / 14.96	9.69 -- 20.32	N:1
Clade 1	69	1	1.87 / 1.47	0.08 -- 4.81	N:1
Clade 1	70	0.99	0.45 / 0.27	0.00 -- 1.46	N:1
Clade 2	72	1	6.53 / 6.1	2.16 -- 11.83	N:1
Clade 2	73	1	1.97 / 1.67	0.18 -- 4.48	N:1
Clade 2	74	1	1.93 / 1.58	0.01 -- 4.68	N:1
Clade 3	111	1	1.50 / 1.05	0.01 -- 4.27	N:0.97
Clade 4	76	1	9.13 / 8.88	5.32 -- 13.39	N:0.99
Clade 5	107	1	3.29 / 2.97	0.81 -- 6.53	N:1
Clade 5	109	0.96	1.77 / 1.54	0.27 -- 3.81	N:1
Clade 5	110	0.99	0.72 / 0.56	0.01 -- 1.86	N:1
Clade 6	82	1	1.84 / 1.56	0.15 -- 4.15	N:0.99
Clade 7	80	1	1.93 / 1.70	0.27 -- 4.06	R:1
Clade 8	81	1	0.76 / 0.47	0.00 -- 2.48	R:1
Clade 9	100	1	0.96 / 0.76	0.01 -- 2.49	R:1
Clade 10	96	0.97	2.93 / 2.75	1.02 -- 5.17	H:0.5; R:0.5

Clade 10	98	1	1.20 / 1.02	0.07 -- 2.65	R:1
Clade 10	99	1	0.92 / 0.74	0.01 -- 2.28	H:1
Clade 11	101	0.99	4.11 / 3.92	1.73 -- 6.78	R:0.5; H:0.49
Clade 11	102	1	1.79 / 1.59	0.24 -- 3.79	R:1
Clade 11	103	1	1.90 / 1.72	0.35 -- 3.72	H:0.97
Clade 12	86	1	0.64 / 0.38	0.00 -- 2.11	IRNM:1
Clade 13	93	1	1.26 / 0.98	0.01 -- 3.29	M:1
Clade 14	88	1	2.19 / 1.99	0.53 -- 4.26	S:0.59; H:0.21; R:0.19

1

2

## Figure legends:

**Fig. 1.** Distribution of the 17 *Volutaria* species and two subspecies included in this study.

**Fig. 2.** Phylogenetic relationships of 63 samples (including three clonal copies) representing 16 species and two subspecies of *Volutaria* based on the combined analysis of nrDNA regions ITS and ETS. The majority-rule consensus phylogram obtained in the Bayesian analysis is shown. Bayesian posterior probabilities ( $\geq 0.90$ ) and maximum parsimony bootstrap values ( $\geq 70\%$ ) are shown above and below branches, respectively. Sample labels are followed by identification number and geographic origin. Abbreviations: CAN = Canary islands; cl = cloned copy, clcons = consensus of similar cloned copies.

**Fig. 3.** Phylogenetic relationships of 87 samples representing 17 species and two subspecies of *Volutaria* based on the combined analysis of cpDNA regions *trnL-F*, *ndhF* and *rpl32*. The majority-rule consensus tree obtained in the Bayesian analysis is shown. Bayesian posterior probabilities ( $\geq 0.90$ ) and maximum parsimony bootstrap values ( $\geq 70\%$ ) are shown above and below branches, respectively. Specimen labels are followed by identification number and geographical origin. Specimens in **bold** show incongruencies when compared to the nDNA results.

**Fig. 4.** Molecular dating and biogeographic analysis. Maximum clade credibility tree from the relaxed molecular-clock analysis of ITS sequences in BEAST. It comprises 56 samples (including 14 clonal copies) of 17 species and two subspecies of *Volutaria*. Asterisks (\*) at numbers 47, 51, 63, 72, and 84 refer to the nodes of Barres & al. (2013) on the basis of which the phylogeny was calibrated. Sequence labels are followed by identification number and geographical origin. Abbreviations: CAN = Canary Islands; cl = clonal copy; clcons = consensus of similar clonal copies. Pie charts (shown only for supported nodes) show the relative probability of each area or combination of areas

being ancestral, according to the ancestral area reconstructions based on the BayArea+J model implemented in BioGeoBEARS. Letters correspond to the following ancestral areas or combination of areas: A = Asia, E = East Africa, H = Horn of Africa, I = Iberian Peninsula, M = Macaronesia (Canary Islands), N = North Africa, R = Arabian Peninsula, S = Socotra, U = Europe (except Iberian Peninsula).









