Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae)

M. Mairal,* L. Pokorney,* J. J. Aldasoro,† M. Alarcón† and I. Sanmartín*

*Real Jardín Botánico (RJB-CSIC), 28014, Madrid, Spain, †Institut Botànic de Barcelona (IBB-CSIC), 08038 Barcelona, Spain

Abstract

Transoceanic distributions have attracted the interest of scientists for centuries. Less attention has been paid to the evolutionary origins of ‘continent-wide’ disjunctions, in which related taxa are distributed across isolated regions within the same continent. A prime example is the ‘Rand Flora’ pattern, which shows sister taxa disjunctly distributed in the continental margins of Africa. Here, we explore the evolutionary origins of this pattern using the genus *Canarina*, with three species: *C. canariensis*, associated with the Canarian Laurisilva, and *C. eminii* and *C. abyssinica*, endemic to the Afromontane region in East Africa, as case study. We infer phylogenetic relationships, divergence times and the history of migration events within *Canarina* using Bayesian inference on a large sample of chloroplast and nuclear sequences. Ecological niche modelling was employed to infer the climatic niche of *Canarina* through time. Dating was performed with a novel nested approach to solve the problem of using deep time calibration points within a molecular dataset comprising both above-species and population-level sampling. Results show *C. abyssinica* as sister to a clade formed by disjunct *C. eminii* and *C. canariensis*. Miocene divergences were inferred among species, whereas infraspecific divergences fell within the Pleistocene–Holocene periods. Although *C. eminii* and *C. canariensis* showed a strong genetic geographic structure, among-population divergences were older in the former than in the latter. Our results suggest that *Canarina* originated in East Africa and later migrated across North Africa, with vicariance and aridification-driven extinction explaining the 7000 km/7 million year divergence between the Canarian and East African endemics.

Keywords: Bayesian biogeography, climate-driven extinction, continental islands, long-distance dispersal, nested phylogenetic dating, vicariance

Received 22 October 2014; revision received 10 February 2015; accepted 11 February 2015

Introduction

Transoceanic disjunct distributions have long attracted the attention of biogeographers (von Humboldt & Bonpland 1805; Hooker 1867; Raven & Axelrod 1972; Donoghue & Smith 2004). A prime example is the Gondwanan distribution exhibited by groups like ratites or marsupials, in which sister lineages are scattered across continents now isolated by thousands of kilometres of oceanic waters (Treviranus 1803; Hooker 1853). Fragmentation of an ancient widespread distribution by plate tectonics (vicariance) and long-distance dispersal events have alternatively been postulated to explain this pattern (Givnish & Renner 2004; Sanmartín & Ronquist 2004).

In contrast, less attention has been paid to the evolutionary origins of ‘continent-wide’ disjunctions, in which related taxa are distributed across geographically isolated regions within the same continent. Transoceanic disjunctions are explained either by tectonic-induced vicariance (i.e. continental drift) followed by biotic division (Raven & Axelrod 1972; Sanmartín et al. 2001) or by LDD (Renner 2004). Within-continent disjunctions, on the other hand, can be explained by LDD...
et al. (2003; Pelser et al. 2012) but are often attributed to large-scale climatic events, such as global climate cooling or aridification that would have extirpated a once continuous biota from part of its area of distribution, leaving relict taxa in refugia or ‘continental islands’ (Axelrod & Raven 1978; Wiens & Donoghue 2004; Crisp & Cook 2007). The barrier that caused the range division in this case is not the opening of an ocean basin, but an environmental change that creates stretches of inhospitable land that are outside the climatic tolerances of the organism (Wiens & Donoghue 2004). Within-continent disjunctions are thus interesting to explore the role of climate-driven extinction in the assembly of biodiversity patterns (Linder 2014).

A prime example of this type of disjunction is the African ‘Rand Flora’ pattern (from the German word meaning rin, aka ‘flora from the edge’), in which distantly related plant families show a similar disjunct distribution, with sister taxa inhabiting geographically isolated regions in the continental margins of Africa— that is north-west Africa, Horn of Africa–southern Arabia, eastern Africa and Southern Africa, and adjacent islands, Macaronesia, Socotra, Madagascar (Christ 1892; Engler 1910; Lebrun 1961; Quèzel 1978; Andrus et al. 2004; Sanmartín et al. 2010 for a historical review). Although they differ in aspects such as morphology, habit or phenoology, Rand Flora lineages share some degree of adaptation to subtropical or temperate environments, so that the tropical lowlands of central Africa or the arid terrains of the Sahara and Sino-Arabic Deserts in the north and the Kalahari desert in the south constitute for them effective climatic barriers to dispersal. Traditionally, this pattern has been explained by vicariance, the fragmentation of an ancient widespread African flora by aridification events during the Late Neogene, leaving relict taxa that survived and diversified in ‘climatic refuges’ at the margins of the continent (Axelrod & Raven 1978; Bramwell 1985). However, the advent of molecular phylogenetics and the possibility of obtaining estimates of divergence times have shown that, for some lineages, these disjunctions can be better explained in terms of recent independent dispersal events among the Rand Flora regions, followed by in situ diversification (Fiz et al. 2008; Meseguer et al. 2013). Because continental disjunct patterns such as the Rand Flora are explained by the appearance of a climatic barrier that causes range division (e.g. the formation of the Sahara desert in the Late Miocene), ecological niche modelling techniques (ENMs) might also be useful to examine the evolutionary origins of Rand Flora lineages. By reconstructing the potential climatic niche of a species and projecting it backwards in time, we can identify areas that were in the past within the organism’s range of climatic tolerances but are inhospitable today due to large-scale climate change (Yesson & Culham 2006; Smith & Donoghue 2010; Meseguer et al. 2014).

One of the strongest connections within the Rand Flora pattern links the Macaronesian Islands to East Africa. Genera such as Campyloma (Kornhall et al. 2001), Aeonium (Mort et al. 2002), Campylandthus (Thivi et al. 2010) or Euphorbia (Riina et al. 2013) harbour Macaronesian endemics, whose sister groups are found along eastern Africa and southern Arabia. In a recent meta-analysis of Rand Flora lineages, Sanmartí et al. (2010) found a comparatively high rate of historical dispersal between these two regions (i.e. NW Africa vs. E Africa/S Arabia), suggesting a long history of biotic connections across the Sahara. Here, we focus on one of the most striking examples of this disjunction, which has never been studied before. The bellflower genus Canarina (family Campanulaceae) is a small angiosperm genus of three species, one endemic to the Canary Islands (Canarina canariensis (L.) Vatke. (1874)) and two other distributed exclusively in the montane regions of eastern Africa: Canarina emini Aschers. ex Schweinf. (1892) and Canarina abyssinica Engl. (1902). Canarina canariensis is associated with the Canarian laurisilva, the highly endemic laurel forest present in the western and central Canary Islands. Canarina emini is an epiphyte endemic to the forests belts of the Afrotasmane region, while C. abyssinica occurs in the upland open forests of eastern Africa (Fig. 1; see Supplementary Text ‘Study Group’ for a more detailed description of the morphology, biology and geographic distribution of each species). Both the Canarian laurisilva and the Afrotasmane region – a series of isolated areas forming an archipelago-like centre of endemism in the mountains of East and West Africa (White 1983) – are traditionally considered as examples of the refugium–fragmentation theory: the remnants of a subtropical flora that once was widespread through Africa but became later extinct due to climatic aridification events (Axelrod & Raven 1978; Bramwell 1985). Therefore, Canarina represents not only a wide continental disjunction of nearly 7000 km across the Sahara, but also a potential relict of an ‘ancient pan-African flora’ (Axelrod & Raven 1978) and a prime candidate to test the climatic vicariance theory in the origins of the Rand Flora pattern. Moreover, the particular distribution of Canarina in the Canary Islands and in the fragmented Afrotasmane forests offers us a unique opportunity to study patterns of colonization in true ‘oceanic islands’ vs. ecological ‘mountain islands’ (aka ‘sky islands’, McCormack et al. 2009). The high-altitude mountain regions in the Afrotasmane region of East Africa have often been equated to ecological islands (Hedberg 1961; Popp et al. 2008; McCormack © 2015 John Wiley & Sons Ltd
et al. 2009), isolated from one another by stretches of dry land or tropical lowlands.

_Canarina_ belongs to tribe Platycodoneae, a basal lineage within family Campanulaceae (Roquet et al. 2009) that includes six other genera endemic to the mountains of central and eastern Asia. Recent molecular studies have reconstructed the phylogeny of the tribe (Wang et al. 2013; Zhou et al. 2013), but a phylogeny of _Canarina_ is still missing due to the difficulty obtaining material for the East African endemic species poses. Here, we present
the first species-level phylogeny of the genus using evidence from the nuclear ribosomal (nrDNA) ITS region and seven noncoding chloroplast (cpDNA) markers and a large sample of infraspecific sequences – covering the entire geographic range of *C. eminii* and *C. canariensis* – as well as a representative sample of genera within Platycodoneae. Bayesian inference methods were used to estimate lineage divergence times and to infer ancestral ranges and the main history of migration events within *Canarina*. Lack of fossils in plant phylogenetic studies often results in deep calibration points being applied to a broader data set, which sometimes includes both above-species and below-species level sampling (Blanco-Pastor *et al.* 2013; Nolasco-Soto *et al.* 2014). This can result in uncertain or even biased time estimates as we move from species level to coalescent dynamics (Ho *et al.* 2008, 2011). Here, we used a novel ‘nested-dating’ Bayesian approach to accommodate the expected change in molecular rates and tree growth model resulting from heterogeneous species-population sampling schemes. Finally, we used ENMs and paleoenvironmental data to estimate the climatic niche of *Canarina*, which when projected backward in time allowed us to detect climatically suitable areas that might have formed part of its geographic distribution in the past or acted as climatic dispersal corridors.

**Materials and methods**

**Taxon sampling and DNA sequencing**

Throughout several field campaigns in eastern Africa and the Canary Islands (2009–2012), fresh material for 29 individuals, representing different populations within *Canarina*, was collected and included in the analysis (Table S1, Supporting information): one sample of *Canarina abysinica* (from the Ethiopian Highlands), seven samples of *Canarina eminii* and 21 of *Canarina canariensis*, covering the entire range of distribution of the last two species. The low number of samples in *C. abys- sinica* reflects the difficulty to collect this species, which has apparently disappeared from many of the original localities where it was first described by Hedberg (1961; see Data S1 for a description on the sampling effort and current conservation status of *C. abysinica*). Nine species representing additional genera within tribe Platycodoneae (*Campanumaea*, *Codonopsis*, *Cyananthus*, *Cyclocodon*, *Ostroswicka* and *Platycodon*), as well as related tribe Campanulaceae (*Campanula* L.) and Campanulaceae subfamilies Lobelioideae (*Lobelia* L.) and Cyphioideae (*Cyphia* P.J. Bergius) were used as alternative outgroups in the phylogenetic and biogeographic analyses. In all, 256 sequences were generated for this study and 11 downloaded from GenBank. Species names, voucher information and GenBank accession numbers for all sequences are provided in Table S1 (Supporting information).

We selected seven noncoding plastid regions exhibiting high levels of genetic variation, the intergenic spacers *psb*-petA, *rpl32-trnL*, *trnL-trnF*, *trnS-trnG* and 3′*trnV-ndhC* (Shaw *et al.* 2005, 2007) and the *trnG* and the petD group II introns (petB-petD, Borsch *et al.* 2009). Details on PCR amplification and sequence editing and alignment are given in Data S1 and Table S2 (Supporting information). Two data sets were constructed to address different objectives. The ‘Platycodoneae data set’ (*n* = 12) included samples of all aforementioned out-group genera plus one accession of each *Canarina* species and was used to reconstruct the phylogeny of the tribe and provide additional calibration points in the dating analyses. The ‘*Canarina* data set’ (*n* = 29) included one accession of each population sampled within *Canarina*, plus one sequence of genera *Ostroswicka* and *Cyclocodon*, which were identified in a prior Campanulaceae study as closely related to *Canarina* (Mansion *et al.* 2012). This latter data set was used to infer the population and phylogeographic history of *C. canariensis* and *C. eminii*.

**Phylogenetic inference**

Phylogenetic relationships were estimated for each marker separately using Bayesian inference implemented in MrBayes (Ronquist *et al.* 2012). Additional analyses were run using maximum likelihood implemented in the software RAXML (Stamatakis *et al.* 2008). The Platycodoneae data set was rooted using *Lobelia* as the outgroup taxon; the *Canarina* data set was rooted using *Ostroswicka* as the outgroup, except for the interspacer 3′*trnV-ndhC* and the *trnG* intron for which *Ostroswicka* and *Cyclocodon* sequences were missing, in which case we used *Platycodon*. Details on these analyses are provided in the Data S1.

Before concatenating the genes into a combined data set, we checked for topological congruence in the inferred relationships by examining the Bayesian consensus trees and searching for well-supported clades (PP > 0.95) in the consensus tree of one marker that were not present in the consensus trees of the other markers (Antonelli & Sannimartín 2011). All analysed genes recovered similar phylogenetic relationships at the generic level, but significant incongruence was found in species relationships within *Canarina* for both the Platycodoneae (*Fig.* S1, Supporting information) and the *Canarina* data sets (*Fig.* S2, Supporting information) for the plastid markers. Three cpDNA genes (*rpl32*, *trnSG* and *trnV-ndhC*) grouped *C. eminii* with *C. abys-
with C. canariensis as their sister group, while the rest of markers either failed to resolve relationships (trnG2G) or placed C. abyssinica as sister to a clade formed by C. canariensis and East African C. eminii (petBD, trnLF, pslB-petA). The latter relationship was also recovered by the single nuclear marker ITS (Figs S1–S2, Supporting information). The same relationships were also obtained using ML although with lower support values (Fig. S2, Supporting information).

Incongruent relationships between gene trees can be attributed to different phenomena, including paralogy, concerted evolution, incomplete lineage sorting (ILS), homoplasy or noise resulting from substitutional saturation or PCR artefacts. Paralogy and concerted evolution are not expected in plastid markers since, unlike multiple-copy nuclear markers like ITS, cpDNA genes are thought to be single copy and behave as a single, linked genome. Multispecies coalescent models (Heled & Drummond 2010) can address ILS but require infraspecific sampling for each species, whereas we only had one sequence for C. abyssinica and all outgroup genera. Instead, we used BUCKy (Larget et al. 2010) to estimate the Bayesian support for alternative topologies among different genes when analysed in a concatenated data set. BUCKy makes no assumption about the reason for discordance among gene trees but instead estimates the dominant history of sampled individuals and how much of the genome supports each relationship, using Bayesian concordance analysis. Groups of genes sharing the same tree are detected (while accounting for uncertainty in gene tree estimates) and then combined to gain more resolution on their common tree (Ané et al. 2007; Larget et al. 2010). Using BUCKy with default settings (α = infinity, allowing genes to evolve independently) showed that inclusion of rpl32 in a concatenated cpDNA Platycodonaceae data set was responsible for significant topological changes in the phylogeny of Canarina, but that this was not the case with other incongruent markers such as trnSG, which consistently grouped C. eminii with C. canariensis, and C. abyssinica as their sister species (Table S3, Supporting information).

Noncoding intergenic spacer regions, such as rpl32-trnL, have become very popular for solving relationships at low taxonomic levels because of their high sequence variability (Shaw et al. 2007), but recent studies have pointed out that this variability is not necessarily correlated with phylogenetic usefulness and can lead to higher levels of homoplasy (Korotkova et al. 2011). To test whether higher levels of homoplasy and substitutional saturation might explain topology differences among cpDNA genes, we plotted uncorrected pairwise distances against maximum-likelihood distances among sequences estimated in PAUP* v4.0b10 (Swofford 2002) and looked for deviation from linearity in saturation plots (Fig. S3, Supporting information). All plots showed a strong fit to a linear regression but rpl32 showed slight levels of saturation at the deepest divergences (Fig. S3, Supporting information). Furthermore, a MrBayes analysis of a cpDNA concatenate data set of Canarina estimated gene-specific rate multiplier that was four times higher in rpl32 than in any other region (Table S3, Supporting information), while the total tree length was two times higher in rpl32 (TL = 1.374) compared with other plastid markers (trnSG: TL = 0.668; petBD: TL = 0.448, Table S3, Supporting information), suggesting faster higher mutation rates. These phenomena were not observed in trnV or trnSG, which showed rate multipliers and tree-length estimates similar to petBD (Table S3, Supporting information). Moreover, plastid rpl32 also exhibited the largest proportion of indels in relation to substitutions than any other marker (35.06%; Table 1).

Given this possible level of homoplasy, we decided to exclude rpl32 from further analyses. Additionally, we excluded trnG2G because of lack of variability (Figs S1 and S2, Supporting information), and the 3’trnV-ndhC interspacer because it showed slight levels of saturation (Fig. S3, Supporting information) and we lacked sequences for all outgroup taxa except Platycodon (Figs S1 and S2, Supporting information); it has been shown that outgroup composition can have a strong influence on the ingroup topology and support values (Rothfels et al. 2012). On the other hand, we kept the trnSG gene in our analyses because – although it supported the same species topology as rpl32 – it did not show evidence of saturation or accelerated substitution rates like the latter marker (Table 1, Fig. S3, Supporting information). Therefore, for the final analyses of the Canarina data sets, we concatenated the four regions, pslB-petA, petB-petD, trnL-trnF and trnS-trnG into a combined cpDNA matrix, which was analysed in conjunction with the nuclear ITS, as the latter marker supported the same topology as the combined cpDNA data set and no evidence of multiple copies were found. The concatenated data matrix was analysed under the GTR + G model, partitioned by gene and allowing the overall mutation rate to differ among partitions using the MrBayes command prset rate = variable.

Divergence time estimation

Lineage divergence times were estimated using the Bayesian relaxed-clock models implemented in BEAST v.1.7 (Drummond & Rambaut 2007). Choice of model priors was based on Bayes factor comparisons using the path sampling (PS) and stepping stone (SS) sampling methods in BEAST, which have been shown to...
Table 1 Summary statistics of the chloroplast and nuclear regions analysed here for the Canarina data set (no outgroups). Fragment length is given in base pairs (bp); alignment length includes the indels

<table>
<thead>
<tr>
<th>Fragment length</th>
<th>rpl32-trnL</th>
<th>3′trnV-ndhC</th>
<th>psbF-petA</th>
<th>petB-petID</th>
<th>TrnL-trnF</th>
<th>TrnS-trnG</th>
<th>trnG intron</th>
<th>ITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragments (bp)</td>
<td>581-647</td>
<td>756-855</td>
<td>822-840</td>
<td>753-798</td>
<td>683-918</td>
<td>666-688</td>
<td>661-676</td>
<td>604-734</td>
</tr>
<tr>
<td>Alignment length</td>
<td>647</td>
<td>855</td>
<td>841</td>
<td>808</td>
<td>933</td>
<td>690</td>
<td>677</td>
<td>734</td>
</tr>
<tr>
<td>Constant sites</td>
<td>590</td>
<td>822</td>
<td>814</td>
<td>766</td>
<td>902</td>
<td>666</td>
<td>666</td>
<td>648</td>
</tr>
<tr>
<td>Variable sites</td>
<td>57</td>
<td>33</td>
<td>27</td>
<td>42</td>
<td>31</td>
<td>24</td>
<td>11</td>
<td>86</td>
</tr>
<tr>
<td>Indel (%)</td>
<td>35.06</td>
<td>–</td>
<td>20.13</td>
<td>11.2</td>
<td>8.13</td>
<td>6.98</td>
<td>–</td>
<td>10.28</td>
</tr>
</tbody>
</table>

outperform the harmonic mean estimator in terms of consistency and reduced variance (Baele et al. 2012). The Yule model and the uncorrelated lognormal distribution (UCLD) were selected, respectively, as the tree and clock model priors for all the analyses (Table S4, Supporting information). Two MCMCs were run for 20 million generations, sampling every 1000th generation. We used Tracer v.1.6 (Rambaut et al. 2013) to monitor convergence and EES values (>200) for all parameters, and TreeAnnotator v. 1.7 (Rambaut & Drummond 2013) to construct a maximum clade credibility tree from the posterior distribution after discarding 10% samples as burn-in.

There are no known fossils of Canarina, so we relied on two approaches to estimate lineage divergence times. First, we used a standard ‘secondary calibration approach’ in which a more inclusive, higher-level data set is used to estimate divergence times within the ingroup. We estimated divergence times among Platycodonae genera using the cpDNA data set with a GTR + G model (we did not include ITS to avoid potential artefacts derived from simultaneously dating plastid and nuclear genomes, which might have very different divergence rates at this level, see Wolfe et al. 1987). We used a uniform prior for the ucld.mean within values commonly observed in plant plastid markers (10^-4-10^-1 substitutions/site/Ma, Wolfe et al. 1987) and a default exponential prior for the standard deviation (SD). As calibration points, we used secondary age constraints drawn from the fossil-rich, angiosperm-wide phylogenetic analysis of Bell et al. (2010). The split between Lobelia and Campanulaceae was calibrated using a normal distribution spanning the confidence interval in the aforementioned study (mean = 56 Ma, SD = 7.5, 95% high posterior density (HPD) = 41–67 Ma), whereas the split between Campanulaceae (Campanula) and Platycodonae was set to mean = 43 Ma (SD = 8, 95% HPD = 28–56 Ma). The ages estimated in this analysis were used to calibrate two nodes in the Canarina data set: the divergence between Cyclocoodon and Ostrowskia (mean = 20.83 Ma, SD = 6.0) and the divergence between Canarina and Ostrowskia (mean = 13.7 Ma, SD = 3.5). The cpDNA + ITS data set was used for this analysis, because at this level differences in mutation rates are minor. Although BEAST selected the UCLD prior (Table S4, Supporting information), Tracer revealed poor mixing and low EES values for the ucld.mean and ucld.stdev parameters, which did not improve after increasing the run length. We thus used the model with the next lowest marginal likelihood, a Yule strict clock model, for the analysis. The mean clock rate was assigned a broad uniform distribution prior (10^-6–10^-1), with default prior settings for the rest of parameters.

Heterogeneous molecular data sets spanning both species- and population-level sampling such as the Canarina data set pose a set of problems in the estimation of lineage divergence times. First, there is the need to apply just one tree prior to the entire phylogeny, from the older deep-time branches to the younger infraspecific events towards the tips. A stochastic branching prior like Yule is likely to overestimate the date of the most recent divergence events, as for short time scales genetic divergence may precede species divergence (Ho et al. 2011), and the opposite effect is expected for coalescent demographic priors. Multispecies coalescent models such as those implemented in *BEAST (Heled & Drummond 2010) can address this problem but require infraspecific sampling for each species, whereas we only had one sequence for C. abyssinica and each outgroup genus. Second, Ho et al. (2005) demonstrated that when deep-time calibration points are used in heterogeneous species-population sampling schemes, extrapolation of molecular rates across the species-population boundary might yield biased estimates of the rate of molecular variation. In our case, the root and stem nodes in the Canarina data set are both constrained with deep-time calibration events (>10 Ma). One consequence of this is the need to use ‘all-encompassing’ priors for the mean clock rate (e.g. Blanco-Pastor et al. 2013; Nolasco-Soto et al. 2014) that accommodate the expected change as we move from the slow, long-term substitution rates at the base of the tree (above-species level) to the rapid mutation rates towards the tips (infraspecific sampling), which might result in uncertain time estimates with broad confidence intervals.
To solve this problem, we used here a nested-dating partitioned approach – first proposed by Pokorny et al. (2011) – in which a higher-level data set calibrated with external evidence (the Platycedoneae data set) is used to constrain the molecular clock rate of additional linked data sets containing population-level data. For this, we constructed two data sets containing all accessions of ITS and plastid markers (petBD, psbJ, trnLF and trnSG) for every sampled population within Canarina eminii \( (n = 7) \) and C. canariensis \( (n = 21) \). These two data sets were not constrained by any calibration point, but their molecular clock was drawn from the mutation rate of the higher-level Platycedoneae partition, that is the ‘clock model’ was linked across partitions and assigned a UCLD prior. The ‘tree model’ was unlinked to accommodate the fact that not all markers and taxa were represented equally across partitions, that is the Platycedoneae data set included only data for the plastid markers and one accession each within C. canariensis and C. eminii. This allowed us to assign a branching Yule tree prior to the above-species level (Platycedoneae) partition and a coalescent constant-size prior to the infraspecific Canarina partitions, the latter selected by Bayes factor PS and SS comparisons.

**Ancestral area reconstruction**

The Bayesian discrete phylogeographic approach of Lemey et al. (2009), implemented in BEAST v.1.7, was used to infer ancestral ranges and trace the history of migration events across space and time in Canarina. This is a continuous-time Markov chain (CTMC) model with the discrete states being the areas or geographic locations of the sequences and the transition rates between states and the migration rates between areas (Sanmartín et al. 2008). Bayesian MCMC inference is used to estimate simultaneously the posterior distribution of phylogenetic relationships, branch lengths and geographic ancestral states, while accounting for uncertainty in all of these parameters, including the estimation of ancestral frequencies for the root (Lemey et al. 2009). Migration rates between areas and the geodispersal rate scalar \( \mu \) were modelled using default gamma prior distributions (Lemey et al. 2009). Two replicate searches of 20 million generations each, sampling every 1000th generation, were combined in TreeAnnotator, after removing the 10% burn-in, to produce a maximum clade credibility (MCC) tree. Bayesian stochastic variable selection (BSVS, Lemey et al. 2009) was used to infer the migration events that are best supported by the data. We run two different analyses. To reconstruct the biogeographic history of the genus, we used the Canarina data set with identical settings to the ‘secondary calibration’ dating analysis and four discrete areas: East Asia, central Asia, East Africa and Canary Islands. To reconstruct phylogeographic patterns within C. eminii and C. canariensis, we used the population-level data sets and a constant-size coalescent model, with the root node in each analysis calibrated with the divergence time estimates obtained from the nested analysis, and a finer-scale definition of areas (Fig. 1b). For C. canariensis, six discrete areas were defined corresponding to the islands in the Canarian Archipelago where the species is present: Gran Canaria (GC), La Gomera (GO), La Palma, and El Hierro (EH) and Tenerife, with the latter divided into two areas: eastern Tenerife (TFE) and western Tenerife (TFW), following previous phylogeographic studies pointing out to an east–west division within the island (Juan et al. 2000). For C. eminii, we divided the montane regions of eastern Africa following the criterion of Gehrke & Linder (2014), except that we subdivided the Ethiopian plateaus into north-west and south-east Ethiopia as several studies have shown phylogeographic disjunctions across the Ethiopian Rift (e.g. Assefa et al. 2007; Wondimu et al. 2014). In all, we have defined four areas, whose limits are shown in Fig. 1a,b: the Abyssinian plateau (the highlands located west of the Ethiopian Rift), Harar plateau (highlands east of the Ethiopian Rift), Imatong-Usumbara (scattered ‘sky islands’ from south Sudan to Tanzania) and Kivu–Rwenzori (northern part of the Albertine Rift). We also ran an additional analysis in which each plateau and sky island has been considered as an independent region (areas = 5).

**Ecological niche modelling**

To understand whether the wide geographic disjunct distribution in Canarina might have been caused by environmental change, we constructed a species distribution model for the genus, using extant occurrence data from two species at the western and eastern side of the disjunction for which we had enough data. In all, we used 122 records: 67 for C. canariensis and 54 for C. eminii (Table S5, Supporting information), covering the entire distributional range of these two species. Data points were obtained from published monographs and inventories (Hedberg 1961; Fernández-López 2014), online databases (www.jardincanario.org/flora-de-gran-canaria; www.gbif.org, www.anthos.es), and data compiled through fieldtrips. Climatic data for current conditions were obtained from WorldClim (www.worldclim.org; Hijmans et al. 2005). For past climate scenarios, we used two global Hadley Centre general circulation models that incorporate the effect of changes in atmospheric CO\(_2\) and that have been previously used to represent major changes in global climate (Meseguer et al. 2014): a 280-ppm CO\(_2\) Late Miocene
simulation (Bradshaw et al. 2012) and a 560-ppm CO₂ Mid-Pliocene simulation (Beerling et al. 2009). Simulations were cropped to include only Africa and surrounding areas. To model the distribution of Canarina, we combined the available 122 occurrences with a set of six bioclimatic variables that could be estimated for past scenarios: total annual precipitation, maximum and minimum monthly precipitation, annual mean temperature, and maximum and minimum monthly temperature. We ran the analyses considering two 4-month periods that cover the two seasons with more accentuated differences in precipitation: November to February and June to September; using as geographic boundaries, the grid included within 28°N to -10°S, for both paleoclimate and present-day simulations. Pseudoabsences were generated by selecting 5000 random points across an area that covers slightly further than the current latitudinal range of Canarina (latitude 40°N–20°S; longitude 30°W–50°E). We used ensemble modelling (a procedure integrating the results from multiple modelling techniques, Araújo & New 2007) to generate our predictions. Four modelling techniques – generalized linear models (GLM), generalized additive models (GAM), general boosting method (GBM) and random forests (RF) – were run and summarized using R packages bioMod2, foreign, raster, SDMTools, rms, gbm, gam, rJava, dismo and randomForest (references for R packages are given in the Data S1).

Results

Phylogenetic relationships and molecular dating

Table 1 summarizes some statistics of the genomic regions studied. Figure 2 shows the results of the Bayesian analysis of the Platycodoneae data set. Most nodes received a clad support (PP) > 0.95, and the phylogeny was congruent between plastid and nuclear markers (Fig. 2a,b). Ostrowskia is recovered as the sister group of Canarina, with Cyclocodon and Platycodon diverging next. Genera Cyananthus, Codonopsis and Campanumoca form the sister clade (Fig. 2a). Analysis of the Canarina concatenated nuclear-plastid data set (Fig. 2c) recovered a monophyletic Canarina (PP = 1.0), with Canarina abyssinica as sister to a clade formed by Canarina eminii and Canarina canariensis with high support (PP = 1.0, ML bootstrap = 80). Geographically structured subclades were recovered within each species with varying levels of support. In general, sequence variation among populations was higher in C. eminii than in C. canariensis (Fig. 2c).

BEAST analysis of the Platycodoneae data set resulted in a phylogeny (Fig. 3a) that was congruent with the MrBayes MCC tree (Fig. 2). Divergence of Campanulaceae and Platycodoneae is dated in the Late Eocene (41.9 Ma, 95% HPD = 28.6–54.7, Table S6, Supporting information), with the first divergence within the tribe dated as Oligocene 29.1 Ma (95% HPD = 18.2–42 Ma). Canarina and Ostrowskia diverged in the Mid-Miocene (13.8 Ma, 6.6–21.7), while the crown age of Canarina is dated as Late Miocene (8.2 Ma, 3.3–14.1). Within Canarina, the ‘standard’ and ‘nested’ BEAST approaches gave divergence time estimates with overlapping confidence intervals (Figs 3 and 4; Fig. S4, Table S6, Supporting information). Species divergences (stem ages) were dated in the Late Miocene (8.4–6.5 Ma), whereas crown ages in C. eminii and C. canariensis (the first population divergences) were dated much younger, in the Early–Mid-Pleistocene (1.76–0.76 Ma, Figs 3 and 4). Population ages were generally older in C. eminii (1.76–1.28 Ma) than in C. canariensis (0.81–0.76) (Fig. S4, Table S6, Supporting information). The nested approach (Fig. 4) resulted in generally younger age estimates for infraspecific events and older ages for the basal, backbone nodes compared to the standard approach (Fig. 3b); for example, the eastern subclade of C. canariensis is dated as 0.38 Ma (0.094–0.891) in the nested tree and 0.59 Ma (0.23–1.05) in the non-nested tree, whereas the opposite pattern is seen for the Canarina–Ostrowskia divergence (13.9 vs. 11.6 Ma) and the divergence between Ostrowskia and Cyclocodon (20.9 vs. 14.1 Ma, Figs 3 and 4). There was also a difference in the geographic structuring of the populations: the two populations in the Abyssinian plateau were grouped in a clade with Elgon and Rwenzori in the standard approach (Fig. 3b), but placed in a separate clade in the nested approach, although the latter with weak support (Fig. 3b).

Phylogeography and colonization history

Bayesian ancestral area reconstruction (Fig. 5) supports an origin of Canarina in East Africa, although there is considerable uncertainty due to the existence of long basal branches and the different geographic origin of the two outgroups (PP = 0.58). A prior migration event from East Asia to East Africa (PP = 0.41) is inferred along the branches separating Canarina from the most closely related genera Cyclocodon and Ostrowskia, although central Asia is another possibility (PP = 0.22, Fig. 5a). The ancestral area of C. eminii is reconstructed as East Africa (PP = 0.59), implying a migration event from East Africa along the long branch (7.9–1.0 Ma) leading to C. canariensis (Fig. 5a). Within each species, several migration events are inferred (Fig. 5b–c). In C. eminii, the Imatong–Usambara is inferred as the source area, although with low probability (PP = 0.3). Considering plateaus and each sky island as separate

© 2015 John Wiley & Sons Ltd
areas (Fig. S5, Supporting information) resulted in the Abyssinian plateau being inferred as the source area ($P_{\text{P}} = 0.23$), but marginal probabilities for ancestral areas were generally much lower (i.e. there was higher uncertainty because of a lower ratio area/data). In *C. canariensis*, colonization of East Tenerife is followed by an early separation between eastern and western Teneriffean clades (0.8 Ma), and several events of inter-island colonization to the east and west involving Tenerife. Migration from western Tenerife (Teno, Adeje) to La Gomera and to La Palma was inferred within the western subclade, with later migration from La Palma to El Hierro (Fig. 5c). Migration to the east from Tenerife (Tope del Carnero) to Gran Canaria is inferred within the eastern subclade, although colonization in the opposite direction is also possible. At least two other independent events of back colonization from Gran Canaria to Tenerife are inferred, involving the populations of Badajoz, Ruiz and Anaga (Fig. 5c). Constraining the dispersal rates according to geographic
distance resulted in a very similar reconstruction, except that Gran Canaria rather than Tenerife was inferred as the ancestral area of the eastern clade of *C. canariensis*, albeit with very low support (TFE = 0.288, GC = 0.290). The geodispersal rate scalar $\mu$ (number of dispersal events per site per million year) was considerably higher in *C. canariensis* (3.6) than in *C. eminii* (1.8).
Ecological niche modelling

Our climate niche projections predict that the geographical area with favourable climatic conditions for *Canarina* experienced a reduction from the Late Miocene to the present (Fig. 6). A climatic ‘corridor’ with suitable conditions can be observed in the Late Miocene projection, connecting east and western North Africa. This connection is interrupted in the Mid-Pliocene simulation, which shows fragmentation into isolated pockets of climatically favourable conditions. The inferred potential distribution for the present largely coincides with the extant distribution, showing an extreme reduction in range at both sides of the Sahara desert.

Discussion

Secondary calibration vs. nested-dating approach

A standard problem in plant phylogenetic dating studies is the lack of fossil calibration points. This is especially important in Rand Flora groups because of the limited number of macrofossils known from North Africa and the Canary Islands (Whittaker et al. 2008; but see Anderson et al. 2009). The most common solution to this problem has been to use a secondary calibration approach, in which age constraints derived from the analysis of a higher-level phylogeny including the group of interest (e.g. the Platycodoneae data set), itself calibrated with the fossil record or with other external evidence (e.g. Bell et al. 2010’s analysis), is used to provide calibration points for the dating of a less inclusive data set, for example the *Canarina* data set. This often translates into a loss of precision in the age estimates due to the need to use an uninformative, broad mean rate prior. Second, if the data set used to estimate divergence times spans both inter- and infraspecific divergences, this might result in biased age estimates, for example, when the phylogeny combines a dense population sampling for one species on one hand, embedded within a tree in which the rest of taxa, at species or above-species level, are represented by one sequence each, on the other (Nolasco-Soto et al. 2014).

The change in the model of molecular evolution as we move from phylogenetic substitution rates at interspecific relationships to the coalescent dynamics...
characteristic of infraspecific evolution might overestimate the age of the most recent events, due to the time dependency in molecular rates and to the fact that gene coalescent events often precede species divergences at the population level (Ho et al. 2005, 2011). This is especially problematic if deep-time calibration points are used to date the basal nodes that require the inclusion of distantly related outgroup taxa.

Fig. 5 Results from the BEAST Bayesian ancestral range reconstruction (Lemey et al. 2009). Coloured branch lengths (see legend) represent for each lineage the ancestral range with the highest posterior probability. Pie charts at nodes represent uncertainty in the estimation, with black colour representing ancestral areas receiving <0.1 posterior probabilities. (a) MCC tree from the analysis of the Canarina data set (standard secondary calibration approach; stem age highlighted inside a square and crown ages highlighted inside circles). (b) MCC tree of the Canarina eminii population-level data set (nested-dating approach). (c) MCC tree of the Canarina canariensis population-level data set (nested-dating approach). Numbers above branches indicate mean ages and numbers below branches indicate Bayesian PP. Lines in maps represent migration events that receive significant support from the data, as recovered by the BSVS procedure. Colour intensity and thickness of these lines proportional to relative strength (the thicker the line, the higher the dispersal rate) and support (the more intense the colour, the stronger the support: purple > yellow > white). Maps have been modified from satellite pictures in Google Earth.

Fig. 6 Geographic projections of the climatic niche model of Canarina over three different time periods: present, Mid-Pliocene and Late Miocene. Blue circles indicate extant occurrences and represent the entire current distribution. Soft yellow-coloured regions indicate low climatic suitability values; conversely, dark red indicate high suitability areas.
To reconcile deep calibration and species demographic history, Ho et al. (2008) proposed an approach in which independent demographic (coalescent) priors were applied to each species, while the basal nodes connecting the clades in the tree are modelled according to a stochastic branching tree prior. The approach followed here, based on Pokorny et al. (2011), is slightly different as we do not have infraspecific sampling for all taxa in the phylogeny (e.g. the outgroup taxa are represented by one sequence each). Instead, we used different partitions, sharing some of the taxa and markers, in which the ‘calibrated’ higher-level partition informs the molecular clock from which the molecular rates of the lower-level partitions are drawn from. Our approach is also different to the ‘multi-species coalescent’ model in *BEAST* (Heled & Drummond 2010) because the latter focuses on coestimating a species tree from multiple gene trees across closely related species, while accounting for coalescent-based phenomena that might cause discrepancy between species and gene trees, such as ILS. Heled & Drummond (2010)’s approach requires infraspecies sampling for each species (3–9 gene copies per lineage) to accurately estimate population parameters like effective population sizes (McCormack et al. 2011). In our analysis, only two species include population-level data (*Canarina canariensis, Canarina eminii*); whereas *C. abyssinica* and the outgroup taxa are represented by one sequence each. Also, ongoing gene flow is unlikely to be a problem for the deepest divergences in our phylogeny, such as the splits between *Canarina* and its closest relatives and between the outgroup taxa. The discussion below focuses on the results from this nested-dating analysis.

**Early evolutionary history of canarina**

Our phylogeny for Platycodoneae is congruent with previous studies, supporting a close relationship of *Platycodon*, and *Cyclocodon* with *Canarina* (the ‘Platycodon clade’, Wang et al. 2013) and confirming the monotypic genus *Ostrowskia* as the sister group of *Canarina* (Mansion et al. 2012). The origin of Platycodoneae is dated around the Late Eocene–Early Oligocene (29 Ma) in agreement with Roquet et al. (2009). *Canarina* is unique within Platycodoneae because of its African distribution. Our time estimates for the divergence with the central Asian *Ostrowskia* (14–11 Ma) suggest that *Canarina*’s ancestors could have taken advantage of the collision of the Arabian Plate with Eurasia (c. 16 Ma, Sanmartin 2003; see Allen & Armstrong 2008 for an earlier date) to migrate into eastern Africa from central–west Asia. This migration could also have been favoured by the uplift of the Red Sea margins (c. 14–13 Ma, Goudie 2005) and a dramatic change in climatic conditions around this period. Starting in the Mid-Miocene, a progressive aridification of the African continent – resulting from both global tectonic changes (e.g. the closing of the Tethys Seaway) and the uplift of eastern Africa (Trauth et al. 2009) – led to the gradual replacement of lowland rainforests by woodland savannah in the central and northern Sahara and in parts of South Africa, and later expansion of grasslands and open steppe habitats in south-west Asia and eastern Africa (Bonnefille et al. 1990; Coetzee 1993; Maley 1996; Plana 2004; Senut et al. 2009). It has been suggested that this created a dispersal route that was used by other non-tropical plant lineages – usually with adaptations to more continental conditions – to migrate from west Asia into Africa (Fiz et al. 2008; Popp et al. 2008; Roquet et al. 2009; Barres et al. 2013; Meseguer et al. 2013). A similar hypothesis has been argued for several East African ‘sky island’ species, which could have used the Arabian mountains as ‘stepping stones’ to reach East Africa (Assefa et al. 2007; Popp et al. 2008). Dispersal from central–west Asia to eastern Africa is also supported by the fact that the fruits of the sister genus of *Canarina* and *Ostrowskia*, are spherical capsules, which when dry are able to release multiple small slight seeds that can be easily dispersed by wind (Zhaparova 1996; Kamelina & Zhinkina 1998). The subsequent isolation of *Canarina* from its Asian ancestors could have been reinforced by the absence of post-Miocene Red Sea land bridges (Fernandes et al. 2006) and a global increase in aridification around 8–6 Ma, coincident with a new period of tectonic activity in eastern Africa and the expansion of grasslands in the Horn of Africa (Cerling et al. 1997; Sepulchre et al. 2006). This event could also account for the divergence of *C. abyssinica* from the ancestor of *C. eminii* and *C. canariensis*, which is estimated around this time in our analysis (8–7 Ma). *Canarina eminii* is commonly associated with well-preserved closed forests, while *C. abyssinica* occurs in open upland forests, so it is possible that habitat specialization driven by Late Miocene climate aridification explains the divergence between these two species.

An alternative topology, showing *C. eminii* and *C. abyssinica* as sister species to *C. canariensis*, was supported by chloroplast markers such as rpl32 and 3′trnV–ndhF. Although incongruence among genes might be attributed to several biological phenomena, in the case of rpl32 it is likely that homoplasy related to higher levels of molecular variation (i.e. saturation at deep phylogenetic levels) and difficulties in alignment due to a high indel/substitution ratio (Table S3, Supporting information) had misled the phylogenetic analysis. For 3′trnV–ndhF, the lack of a closely related outgroup could be the explanation, as when this marker is included in a concatenated cpDNA data set rooted with *Ostrowskia*, we recovered the ‘right’ topology grouping.
C. eminii and C. canariensis with relatively high support (PP = 0.98, ML = 77; Fig. S6, Supporting information). In contrast, chloroplast intron regions like the petD II intron possess characteristics, such as high phylogenetic signal per informative character and a well-known secondary structure and molecular evolution, that make them an ideal choice for solving phylogenetic relationships at species level in Campanulaceae (Borsch et al. 2009; Mansanet et al. 2012). This was also the marker for which we have sequences for all outgroup taxa. Moreover, petBD was, after ITS, the marker showing in our analyses the lowest levels of substitutional saturation and the largest number of potentially informative characters i.e., number of mutations per sequenced nucleotide (Korotkova et al. 2011). Therefore, although we recognize that inclusion of additional plastid and nuclear markers is desirable, we believe that the topology grouping C. eminii with C. canariensis as sister to C. abyssinica accurately reflects the evolutionary relationships among the species.

**Long-distance dispersal vs. vicariance and climate-driven extinction**

The vicariance–refugium hypothesis posits that the Rand Flora pattern was formed by the fragmentation of a once continuous flora by aridification events, leaving relicts at the eastern and western sides of the geographic disjunction. In Canarina, this hypothesis would predict a pattern of ‘reciprocal monophyly’ between the disjunct taxa, with eastern Africa and Canarian taxa recovered as sister groups (Couvreur et al. 2008; Thiv et al. 2010), and an age for the disjunction that must predate the barrier that caused the range division, that is the origin of the present Sahara desert. Conversely, the long-distance dispersal (LDD) hypothesis implies the expectation that the taxa at one extreme of the disjunction (i.e. the Canarian endemic) would be embedded within a clade formed by taxa from the other side (i.e. an eastern African clade) and that the disjunction should clearly postdate the formation of the barrier.

At first, the pattern found here, with C. canariensis nested within a clade of two East African endemics, agrees better with the LDD hypothesis. Canarina species are characterized by the presence of fleshy fruits, with passerine bird- and lizard-mediated zoohchory reported for C. canariensis (Rodríguez et al. 2008). A dispersal event across the 7000 km of the Sahara probably requires other dispersal vectors, such as long-distance migratory birds. For example, Popp et al. (2011) argued that a recent (Holocene) single long-distance dispersal by a bird could explain the extreme bipolar distribution of crowberries (Emetrum), and similar LDD explanations have been proposed to explain wide range disjunctions between South Africa and North Africa/Canary Islands in Senecio (Coleman et al. 2003; Pelser et al. 2012). Nevertheless, the long temporal gap separating C. canariensis and C. eminii, with a stem age predating the formation of the Sahara, c. 6 Ma agrees better with a climate-driven vicariance explanation. Interestingly, the alternative topology recovered by rpl32, grouping C. eminii and C. abyssinica as sister to C. canariensis, would actually reinforce the vicariance explanation, as the divergence between C. canariensis and the East African endemics would probably be dated even earlier (>8–7 Ma), substantially predating the age of origin of the Sahara.

What could be the cause behind this vicariance (allopatric) event? Paleontological reconstructions show a wetter North Africa at least until the Late Miocene (Griffin 2002), which became increasingly more arid as a result of successive aridification events related to a variety of factors, including the opening of the Drake Passage, the closing of the Tethys Seaway and the uplift of eastern Africa (Sepulchre et al. 2006; Trauth et al. 2009). The first recorded signs of aridification in the Sahara date back to the end of the Miocene, ca. 7–6 Ma ago (Senut et al. 2009), which is roughly in agreement with the split between C. eminii and C. canariensis (6.5 Ma). Nevertheless, the rapid alternation of arid and humid periods starting in the Miocene–Pliocene boundary (Trauth et al. 2009; Micheels et al. 2009) might have allowed repeated events of isolation and reconnection across both sides of the Sahara (Désamoré et al. 2011). We do not have evidence of any of these recent events of reconnection in the phylogeny of Canarina. Instead, the 6.4 Ma divergence estimated here between the Canarian and East African endemics is roughly in agreement with the age estimated for the disjunction of other Rand Flora lineages, for example, Campylanthus (Thiv et al. 2010) or Plocama.

In addition, our ecological niche models and paleoclimate projections support the hypothesis of a more widespread distribution of Canarina across north–central Africa in the past, which became fragmented by climate change. They show a more or less continuous ‘climatic corridor’ across North Africa during the Late Miocene period, which became interrupted during the more arid Mid-Pliocene period. The latter shows the presence of isolated patches of climatic suitability (Fig. 6), which could have acted as potential ‘stepping stones’ for dispersal across the Sahara, or as climatic refugia once aridification started. Worsening climate conditions, with increasing aridity at the Plio–Pleistocene boundary (Senut et al. 2009), might have caused the extinction of intermediate populations across central North Africa, leaving the current species as the only remnants (relicts) of a past widespread distribution. Similar scenarios...
have been hypothesized in other Rand Flora lineages for which supporting fossil evidence exists, such as Dracaena (Denk et al. 2014). Whether Canarina was ever continuously distributed across North Africa, with uninterrupted gene flow between both extremes of the disjunction, or whether, alternatively, the pattern is the result of gradual range expansion, westwards across the Sahara, is difficult to discern with the current evidence. The vicariance hypothesis, for example, predicts also range expansion across the Sahara prior to the allopatric (vicariant) event. Interestingly, the lower levels of genetic diversity found in C. canariensis compared to the East African C. eminii agree with a more recent dispersal event, perhaps from a now extinct and geographically closer, North African (Moroccan) population. What our evidence does suggest is that Canarina could have a wider distribution across north-central Africa in the past and that there has been a long history of isolation between the two extremes of the disjunction. The long stem between the stem divergence of C. canariensis and the start of infraspecies (population) divergence can be interpreted as evidence of extinction of the intermediate populations (Antonelli & Sanmartín 2011). Alternatively, it could be understood as the result of strong purifying selection with little population differentiation – driven perhaps by climatic change – and, followed by a recent demographic expansion. We favour extinction over purifying selection because the latter is expected to affect one gene but not to produce congruent patterns across genes (Williamson & Orive 2002). Although population-level studies are needed to test this hypothesis, an interesting corollary of our results suggests that Canarina could have a wider distribution across north-central Africa in the past and that there has been a long history of isolation between the two extremes of the disjunction. The long stem between the stem divergence of C. canariensis and the start of infraspecies (population) divergence can be interpreted as evidence of extinction of the intermediate populations (Antonelli & Sanmartín 2011). Alternatively, it could be understood as the result of strong purifying selection with little population differentiation – driven perhaps by climatic change – and, followed by a recent demographic expansion. We favour extinction over purifying selection because the latter is expected to affect one gene but not to produce congruent patterns across genes (Williamson & Orive 2002). Although population-level studies are needed to test this hypothesis, an interesting corollary of our study is that the age of divergence of an island endemic from its continental sister species is not necessarily equivalent to the age of colonization of the island as it is often assumed in island studies (Kim et al. 2008), especially if extinction has been high in the continent.

**Geographic oceanic islands vs. ‘Ecological’ mountain islands**

Canarina, with its distribution in true oceanic islands and mountain ‘sky islands’, offers an interesting comparison on the role of geographic vs. ecological barriers in structuring plant genetic variation. It is well known that oceanic islands are able to cope with large climatic changes better than continental landmasses because of the tempering effect created by the ocean to which they are exposed. The sky islands of the Afromontane regions in East Africa (i.e. high plateaus and mountains in Ethiopia and subtropical East Africa) probably acted in a similar way, allowing species and communities to migrate altitudinally and thus avoid the thermal and hydric stress produced by aridification episodes (Fjeldså & Lovett 1997). Paleobotanical and phylogeographical evidence suggest that the slopes of these montane regions were covered by forests until recently (Bonnefille et al. 1990; Kuper & Kröpelin 2006). During the glacial arid periods of the Late Pliocene and Pleistocene, these forests probably became separated, and later reconnected during the humid, warmer interglacial periods (Coetzee 1964; Maley 1996; Kebede et al. 2007; Popp et al. 2008). In more recent times, land use and deforestation might have contributed to further isolation of these forest patches (EFAF 1994; FAO 2001). The relatively old infraspecific divergences estimated here for C. eminii, ranging from 700 000 years between Elgon and Rwenzori to a few thousand years between Giffa and Dembecha (Fig. 4), suggest that population divergence in this montane species was more likely driven by Pleistocene climatic events than by forest fragmentation after the expansion of agriculture. Moreover, our results support other phylogeographic studies in Afromontane taxa (Knox & Palmer 1998; Kebede et al. 2007) that pointed to the Ethiopian Rift Valley as an important geographic barrier, segregating populations to the east and west of this barrier. In contrast, the fact that the eastern subclade in C. eminii (0.4 Ma, PP = 1) groups together populations as far away as Harenna Forest and Yirga, in southern Ethiopia, and the Aberdare Range, in Kenya, suggests that the eastern range of the Rift has been less isolated than the west, probably due to the existence of better connections between forest patches on this side of Rift (Coetzee 1964; Hedberg 1969; Kebede et al. 2007).

The oldest extant Canary Islands emerged ca. 20 Ma (Fernández-Palacios et al. 2011), but our time estimates place population divergence in C. canariensis within the last 800 000 years, considerably younger than in C. eminii. The first recovered divergence event is one of within-island segregations between east and west Tenerife. This pattern has been reported in other endemic organisms (Juan et al. 2000) and attributed to the geological origin of Tenerife, which resulted from the merging of three paleoislands c.a. 1 Ma ago (Ancochea et al. 1990). Subsequent events, such as a central eruptive episode ca. 0.8 Ma and giant landslides on the northern flank of Tenerife (Kralet et al. 2001), might have later prevented reconnections between east and west C. canariensis populations. Interisland dispersal events from Tenerife to the east and west are also reconstructed, in agreement with the role of the central islands as a source of migration within the archipelago (Sanmartín et al. 2008), but these are all dated after the divergence within Tenerife, indicating that probably within-island catastrophic/geological events have been a more important barrier to dispersal for C. canariensis populations than the ocean waters separating the islands.
Conclusions

Continental-scale disjunct distribution patterns, such as the Rand Flora, are especially interesting in the context of the present biodiversity crisis because they are often attributed to climate-driven extinction that would have extirpated a once continuous biota from part of its distributional range (Axelrod & Raven 1978; Crisp & Cook 2007). Here, we show that in the case of genus Canarina, this disjunction predates the origin of the Sahara and might be explained by climate-driven vicariance and extinction. The potential ancient age of within-continent disjunctions (Crisp & Cook 2007) implies that we often do not have fossil taxa close to the group of interest. We benefit here from a nested-dating approach that implements two different tree models (birthdeath vs. coalescent) for simultaneous phylogenetic analysis of data at different levels of organization. Our study emphasizes the importance of climate-driven extinction in the assembly of regional biodiversity patterns, in particular in the context of the ongoing aridification of the Mediterranean Basin.

Acknowledgements

We are grateful to Richard Abbott and three anonymous reviewers, whose comments helped to significantly improve the manuscript. We thank Fatima Durán and Guillermo Sanjuebanito for laboratory assistance. Field work could not have been conducted without the cooperation of Juan Ojeda, Oscar Saturno and the staff at the Jardín Botánico Canario Viera y Clavijo (Gran Canaria); Cabildo de Tenerife; Jacinto Leralta and Angel Fernández (La Gomera); Félix Manuel Medina from the Cabildo of La Palma and the Cabildo of El Hierro are thanked for help with accommodation and sampling logistics during field expeditions. We thank to Kenya National Commission for Science, Technology and Innovation, the Bale Mountains and the Harenna Forest National Park (Ethiopia) and Rwenzori Mountains National Park (Uganda) for its collaboration during fieldwork. We also thank Alejandro González, Juli Caújapé and Moisés Soto for providing fresh samples, David Beerling for providing climatic data and Loïc Pellissier, Javier Fuertes, Mike Thiv and José Luis Blanco Pastor for help during different stages of this work. This work was funded by the Spanish Ministerio de Economía y Competitividad (Projects CGL2006-09696, CGL2009-1332-C03-01, CGL2012-40129-C02-01) the JAE-Doc programme (CSIC/PSE) to MA, and a PhD research grant (BES-2010-037261) to MM. LP was funded by a research contract under CGL2012-40129-C02-01.

References


© 2015 John Wiley & Sons Ltd


Table S4 Model likelihood estimators obtained using the Path sampling (PS) and Stepping-Stone (SS) sampling methods implemented in BEAST.

Table S5 Geographical coordinates used in the Ecological Niche Modelling of Canarina.

Table S6 Mean ages and 95% HPD confidence intervals for the different BEAST analyses represented in Fig. 3a–b–c: (a) Platycodoneae dataset; (b) Canarina dataset, standard dating; (c) nested dating of the C. eminii and C. canariensis datasets.

Fig. S1 Bayesian Majority-Rule consensus trees obtained by MrBayes from the single-gene analyses of the Platycodoneae dataset.

Fig. S2 Bayesian Majority-Rule consensus trees inferred from the single-gene analyses of the Canarina dataset.

Fig. S3 Saturation plots for the single-gene nuclear (ITS) and chloroplast markers, showing the uncorrected pairwise distances (p distance) against corrected maximum likelihood distances (ML distance) derived in PAUP using the appropriate model of substitution.

Fig. S4 Maximum clade credibility (MCC) trees obtained from different BEAST analyses, showing nodes with mean ages and 95% HPD confidence intervals (values specified in Table S4): (a) Platycodoneae dataset; (b) standard dating of the Canarina dataset; (c) nested dating analysis of all three linked datasets: Platycodoneae (left) and population-level C. eminii and C. canariensis (right).

Fig. S5 Bayesian phylogeographic analysis of the C. eminii population-level dataset (nested dating approach) using an alternative coding of the geographic areas: Elgon and Ruwenzori considered as different OTUs.

Fig. S6 Bayesian Majority-Rule consensus tree obtained by MrBayes from the Canarina concatenated chloroplast and nuclear dataset rooted with Ostrowskia (ITS, psbF-petA, trnL-trnF, petB-petD, trnS-trnG, trnV-ndhC).