This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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

M. MAIRAL[*1], L. POKORNY[1], J. J. ALDASORO[2], M. ALARCÓN[2], AND I. SANMARTÍN[*1]

[1] *Real Jardín Botánico (RJB-CSIC), 28014 Madrid, Spain*

[2] *Institut Botànic de Barcelona (IBB-CSIC), 08038 Barcelona, Spain*

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

Correspondence to be sent to: Real Jardín Botánico (RJB-CSIC), 28014 Madrid, Spain; Tel: +34914203017; Fax: +34914200157; e-mail: mariomairal@gmail.com, isanmartin@rjb.csic.es

Running title: Drivers of within-continent disjunctions

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 to 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 that deep calibration points within a molecular dataset comprising both above-species and population-level sampling poses. 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.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 kilometers 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 either explained 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 (Coleman *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; Crisp & Cook 2007; Wiens & Donoghue 2004). 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; Pokorny *et al*. in review).

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

A prime example of this type of disjunction is the African "Rand Flora" pattern (from the German word meaning *rim*, 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—i.e., northwest Africa, Horn of Africa-Southern Arabia, Eastern Africa, and Southern Africa, and adjacent islands, Macaronesia, Socotra, Madagascar (Christ 1892; Engler 1910; Lebrun 1961; Quézel 1978; Andrus *et al.* 2004; Sanmartín *et al.* 2010; see Pokorny *et al.* (in review) for a historical review). Although they differ in aspects such as morphology, habit, or phenology, 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 (Pokorny *et al.* in review). 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 modeling 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' range of climatic tolerances but are inhospitable today

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 like *Camptoloma* (Kornhall *et al*. 2001), *Aeonium* (Mort *et al*. 2002), *Campylanthus* (Thiv *et al*. 2010), or *Euphorbia* (Riina *et al*. 2014) harbor Macaronesian endemics, whose sister groups are found along Eastern Africa and Southern Arabia. In a recent meta-analysis of Rand Flora lineages, Sanmartín *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: *C. eminii* Aschers. ex Schweinf. (1892) and *C. abyssinica* Engl. (1902). *Canarina canariensis* is associated to the Canarian laurisilva, the highly endemic laurel forest present in the western and central Canary Islands. *Canarina eminii* is an epiphyte endemic to the forests belts of the Afromontane region, while *C. abyssinica* occurs in the upland open forests of Eastern Africa (Figure 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 Afromontane 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 Afromontane forests offers us a unique opportunity to study patterns of colonization in *true* "oceanic islands" versus ecological "mountain islands" (aka "sky-islands", McCormack *et al*. 2009). The high-altitude mountain regions in the Afromontane region of East Africa have often been equated to ecological islands (Hedberg 1961; Popp *et al*. 2008; McCormack *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 that 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 non-coding 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 dataset, 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

the expected change in molecular rates and tree growth model resulting from heterogenous 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.

MATERIAL 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*, were collected and included in the analysis (Table S1): one sample of *C. abyssinica* (from the Ethiopian Highlands), seven samples of *C. eminii,* and twenty-one of *C. canariensis*, covering the entire range of distribution of the last two species. The low number of samples in *Canarina abyssinic*a 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 Supplementary Text for a description on the sampling effort and current conservation status of *C. abyssinica*). Nine species representing additional genera within tribe Platycodoneae (*Campanumoea*, *Codonopsis*, *Cyananthus*, *Cyclocodon*, *Ostrowskia*, and *Platycodon*), as well as related tribe Campanuleae (*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 (see below). 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.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

We selected seven non-coding plastid regions exhibiting high levels of genetic variation, the intergenic spacers *psb*J–*pet*A, *rpl*32–*trn*L, *trn*L–*trn*F, *trn*S–*trn*G, and 3'*trn*V–*ndh*C (Shaw *et al*. 2005, 2007), and the *trn*G and the *pet*D group II introns (*pet*B–*pet*D, Borsch *et al*. 2009). Details on PCR amplification and sequence editing and alignment are given in Supplementary Text and Table S2. Two datasets were constructed to address different objectives. The "*Platycodoneae* dataset" (n = 12) included samples of all aforementioned outgroup 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* dataset" (n = 29) included one accession of each population sampled within *Canarina*, plus one sequence of genera *Ostrowskia* and *Cyclocodon*, which were identified in a prior Campanulaceae study as closely related to *Canarina* (Mansion *et al*. 2012). This latter dataset 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* dataset was rooted using *Lobelia* as the outgroup taxon; the *Canarina* dataset was rooted using *Ostrowskia* as the outgroup, except for the interspacer *3´trn*V-*ndh*C and the *trn*G intron for which *Ostrowskia* and *Cyclocodon* sequences were missing, in which case we used *Platycodon*. Details on these analyses are provided in the Supplementary Text.

Before concatenating the genes into a combined dataset, 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

were not present in the consensus trees of the other markers (Antonelli & Sanmartín 2011). All analyzed genes recovered similar phylogenetic relationships at the generic level, but significant incongruence was found in species relationships within *Canarina* for both the *Platycodoneae* (Figure S1) and the *Canarina* datasets (Figure S2) for the plastid markers. Three cpDNA genes (*rpl*32, *trn*SG, and *trn*V-*ndhC*) grouped *C. eminii* with *C. abyssinica* with *C. canariensis* as their sister-group, while the rest of markers either failed to resolve relationships (*trn*G2G) or placed *C. abyssinica* as sister to a clade formed by *C. canariensis* and East African *C. eminii* (*pet*BD, *trn*LF, *psb*J-*pet*A). The latter relationship was also recovered by the single nuclear marker ITS (Figs. S1-S2). The same relationships were also obtained using ML although with lower support values (Fig. S2).

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 artifacts. 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 analyzed in a concatenated dataset. 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

to evolve independently) showed that inclusion of *rpl*32 in a concatenated cpDNA *Platycodoneae* dataset was responsible for significant topological changes in the phylogeny of *Canarina*, but that this was not the case with other incongruent markers such as *trn*SG, which consistently grouped *C. eminii* with *C. canariensis*, and *C. abyssinica* as their sister species (Table S3).

Non-coding intergenic spacer regions, such as *rpl*32-*trn*L, 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). All plots showed a strong fit to a linear regression but *rpl*32 showed slight levels of saturation at the deepest divergences (Figure S3). Furthermore, a MrBayes analysis of a cpDNA concatenate dataset of *Canarina* estimated gene-specific rate multiplier that were four times higher in *rpl*32 than in any other region (Table S3), while the total tree length was two times higher in *rpl*32 (TL=1.374) compared to other plastid markers (*trn*SG: TL=0.668; *pet*BD: TL=0.448, Table S3), suggesting faster higher mutation rates. These phenomena were not observed in *trn*V or *trn*SG, which showed rate multipliers and tree-length estimates similar to *pet*BD (Table S3). Moreover, plastid *rpl*32 also exhibited the largest proportion of indels in relation to substitutions than any other marker (35.06%; Table 2).

Given this possible level of homoplasy, we decided to exclude *rpl*32 from further analyses. Additionally, we excluded *trn*G2G because of lack of variability (Figs. S1, S2), and the *3'trn*V-*ndh*C interspacer because it showed slight levels of saturation (Fig. S3) and

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

we lacked sequences for all outgroup taxa except *Platycodon* (Figs. S1, S2); 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 *trn*SG 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). Therefore, for the final analyses of the *Canarina* datasets, we concatenated the four regions, *psb*J-*pet*A, *pet*B-*pet*D, *trn*L-*trn*F and *trn*S-*trn*G into a combined cpDNA matrix, which was analyzed in conjunction with the nuclear ITS, since the latter marker supported the same topology as the combined cpDNA dataset (see below) and no evidence of multiple copies were found. The concatenated data matrix was analyzed 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 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). Two MCMC chains 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 burnin.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 dataset is used to estimate divergence times within the ingroup. We estimated divergence times among Platycodoneae genera using the cpDNA dataset with a GTR + G model (we did not include ITS to avoid potential artifacts 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 Campanuleae (*Campanula*) and Platycodoneae 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* dataset: the divergence between *Cyclocodon* 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 dataset were used for this analysis, since at this level differences in mutation rates are minor. Although BEAST selected the UCLD prior (Table S4), 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.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

297 Heterogeneous molecular datasets spanning both species and population-level
298 sampling such as the *Canarina* dataset pose a set of problems in the estimation of lineage
299 divergence times. First, there is the need to apply just one tree prior to the entire phylogeny,
300 from the older deep-time branches to the younger infraspecific events towards the tips. A
301 stochastic branching prior like Yule is likely to overestimate the date of the most recent
302 divergence events, since for short time scales genetic divergence may precede species
303 divergence (Ho *et al*. 2011), and the opposite effect is expected for coalescent demographic
304 priors. Multispecies coalescent models such as those implemented in *BEAST (Heled &
305 Drummond 2010) can address this problem but require infraspecific sampling for each
306 species, whereas we only had one sequence for *C. abyssinica* and each outgroup genus.
307 Secondly, Ho *et al*. (2005) demonstrated that when deep-time calibration points are used in
308 heterogeneous species-population sampling schemes, extrapolation of molecular rates across
309 the species-population boundary might yield biased estimates of the rate of molecular
310 variation. In our case, the root and stem nodes in the *Canarina* dataset are both constrained
311 with deep-time calibration events (> 10 Ma). One consequence of this is the need to use "all-
312 encompassing" priors for the mean clock rate (e.g., Blanco-Pastor *et al*. 2013; Nolasco-Soto
313 *et al*. 2014) that accommodate the expected change as we move from the slow, long-term
314 substitution rates at the base of the tree (above-species level) to the rapid mutation rates
315 towards the tips (infraspecific sampling), which might result in uncertain time estimates with
316 broad confidence intervals.

317 To solve this problem, we used here a *nested-dating* partitioned approach —first
318 proposed by Pokorny *et al*. (2011)— in which a higher-level dataset calibrated with external
319 evidence (the *Platycodoneae* dataset) is used to constrain the molecular clock rate of
320 additional linked datasets containing population-level data. For this, we constructed two
321 datasets containing all accessions of ITS and plastid markers (*pet*BD, *psb*J, *trn*LF, and

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

*trn*SG) for every sampled population within *Canarina eminii* (n = 7) and *C. canariensis* (n = 21). These two datasets were not constrained by any calibration point, but their molecular clock was drawn from the mutation rate of the higher-level *Platycodoneae* partition, i.e., 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, i.e., the *Platycodoneae* dataset 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 (*Platycodoneae*) 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 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 μ were modeled 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)

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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* dataset 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 datasets 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 northwest and southeast Ethiopia since several studies have shown phylogeographic disjunctions across the Ethiopian Rift (e.g., Assefa *et al*. 2007; Wondimu *et al*. 2014; Mairal *et al.* in review). In all, we have defined four areas, whose limits are shown in Figure 1b: a): the Abyssinian plateau (the highlands located west of the Ethiopian Rift), Harar plateau (highlands east of the Ethiopian Rift), Imatong-Usambara (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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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), 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 CO_2 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 four-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 28ºN to -10ºS, 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 40ºN-20ºS; longitude 30ºW-50ºE). We used ensemble modeling (a procedure integrating the results from multiple modeling techniques, Araújo & New, 2007) to generate

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

our predictions. Four modeling 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 Supplementary Text).

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* dataset. Most nodes received a clade 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 *Campanumoea* form the sister-clade (Fig. 2a). Analysis of the *Canarina* concatenated nuclear-plastid dataset (Fig. 2c) recovered a monophyletic *Canarina* (PP = 1.0), with *C. abyssinica* as sister to a clade formed by *C. eminii* and *C. canariensis* with high support (PP = 1, ML BS = 60). 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 dataset resulted in a phylogeny (Fig. 3a) that was congruent with the MrBayes MCC tree (Fig. 2). Divergence of Campanuleae and Platycodoneae is dated in the Late Eocene (41.9 Ma, 95% HPD = 28.6–54.7, Table S6), 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

421 "standard" and "nested" BEAST approaches gave divergence time estimates with
422 overlapping confidence intervals (Fig. 3, Fig 4, Fig. S4, Table S6). Species divergences
423 (stem-ages) were dated in the Late Miocene (8.4–6.5 Ma), whereas crown-ages in *C. eminii*
424 and *C. canariensis* (the first population divergences) were dated much younger, in the Early-
425 Mid Pleistocene (1.76–0.76 Ma, Fig. 3, Fig. 4). Population ages were generally older in *C.*
426 *eminii* (1.76–1.28 Ma) than in *C. canariensis* (0.81–0.76) (Fig. S4, Table S6). The nested
427 approach (Fig. 4) resulted in generally younger age estimates for infraspecific events and
428 older ages for the basal, backbone nodes compared to the standard approach (Fig. 3b); for
429 example, the eastern subclade of *C. canariensis* is dated as 0.38 Ma (0.094–0.891) in the
430 nested tree and 0.59 Ma (0.23–1.05) in the non-nested tree, whereas the opposite pattern is
431 seen for the *Canarina-Ostrowskia* divergence (13.9 vs. 11.6 Ma) and the divergence between
432 *Ostrowskia* and *Cyclocodon* (20.9 vs. 14.1 Ma, Fig. 3, Fig. 4). There was also a difference in
433 the geographic structuring of the populations: the two populations in the Abyssinian Plateau
434 were grouped in a clade with Elgon and Rwenzori in the standard approach (Fig. 3b), but
435 placed in a separate clade in the nested approach, although the latter with weak support (Fig.
436 3b)
437
438 *Phylogeography and Colonization History*
439 Bayesian ancestral area reconstruction (Figure 5) supports an origin of *Canarina* in East
440 Africa, although there is considerable uncertainty due to the existence of long basal branches
441 and the different geographic origin of the two outgroups (PP = 0.58). A prior migration
442 event from East Asia to East Africa (PP = 0.41) is inferred along the branches separating
443 *Canarina* from the most closely related genera *Cyclocodon* and *Ostrowskia*, although
444 Central Asia is another possibility (PP = 0.22, Fig. 5a). The ancestral area of *C. eminii* is
445 reconstructed as East Africa (PP = 0.59), implying a migration event from East Africa along

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 areas (Fig. S5) resulted in the Abyssinian Plateau being inferred as the source area (PP=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, though 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 μ (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 favorable climatic conditions for *Canarina* experienced a reduction from the Late Miocene to the present (Figure 6). A climatic “corridor” with suitable conditions can be observed in the Late

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 favorable 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 Versus 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 dataset), 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 dataset, e.g.,the *Canarina* dataset. 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 dataset used to estimate lineage 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,

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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.

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 modeled according to a stochastic branching tree prior. The approach followed here, based on Pokorny *et al*. (2011), is slightly different since 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 "multispecies coalescent" model in *BEAST (Heled & Drummond 2010) because the latter focuses on co-estimating 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 and Drummond (2010)'s approach requires infraspecies sampling for each species (3-9 gene copies per lineage) in order to accurately estimate population parameters like effective population sizes (McCormack et al. 2010). In our analysis, only two species include population-level data (*C. canariensis*, *C. eminii*); *C. abyssinica* and the outgroup taxa do not. 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.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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, Sanmartín 2003; see Allen & Amstrong 2008 for an earlier date) to migrate into Eastern Africa from Central-West Asia. This migration could also have been favored 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 southwest 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

546 fruits of the sister-genus of *Canarina*, *Ostrowskia*, are spherical capsules, which when dry
547 are able to release multiple small slight seeds that can be easily dispersed by wind (Kamelina
548 & Zhinkina 1998, Zhaparova 1996). The subsequent isolation of *Canarina* from its Asian
549 ancestors could have been reinforced by the absence of post-Miocene Red Sea land bridges
550 (Fernandes *et al*. 2006) and a global increase in aridification around 8–6 Ma, coincident with
551 a new period of tectonic activity in Eastern Africa and the expansion of grasslands in the
552 Horn of Africa (Cerling *et al*. 1997; Sepulchre *et al*. 2006). This event could also account for
553 the divergence of *Canarina abyssinica* from the ancestor of *C. eminii* and *C. canariensis*,
554 which is estimated around this time in our analysis (8–7 Ma). *Canarina eminii* is commonly
555 associated to well-preserved closed forests, while *C. abyssinica* occurs in open upland
556 forests, so it is possible that habitat specialization driven by Late Miocene climate
557 aridification explains the divergence between these two species.

558 An alternative topology, showing *C. eminii* and *C. abyssinica* as sister-species to *C.*
559 *canariensis*, was supported by chloroplast markers such as *rpl*32 and *3´trn*V-*ndh*C.
560 Although incongruence among genes might be attributed to several biological phenomena, in
561 the case of *rpl*32 it is likely that homoplasy related to higher levels of molecular variation
562 (i.e., saturation at deep phylogenetic levels) and difficulties in alignment due to a high
563 indel/substitution ratio (Table S3), had misled the phylogenetic analysis. For *3´trn*V-*ndh*C,
564 the lack of a closely related outgroup could be the explanation, since when this marker is
565 included in a concatenated cpDNA dataset rooted with *Ostrowskia*, we recovered the "right"
566 topology grouping *C. eminii* and *C. canariensis* with relatively high support (PP = 0.98, ML
567 = 77; Fig. S6). In contrast, chloroplast intron regions like the *petD* II intron possess
568 characteristics, such as high phylogenetic signal per informative character and a well-known
569 secondary structure and molecular evolution, that make them an ideal choice for solving
570 phylogenetic relationships at species level in Campanulaceae (Borsch *et al*. 2009; Mansion

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

571 *et al*. 2012). This was also the marker for which we have sequences for all outgroup taxa.
572 Moreover *petBD* was, after ITS, the marker showing in our analyses the lowest levels of
573 substitutional saturation and the largest number of potentially informative characters —
574 number of mutations per sequenced nucleotide (Korotkova *et al*. 2011) —. Therefore,
575 although we recognize that inclusion of additional plastid and nuclear markers is desirable,
576 we believe that the topology grouping *C. eminii* with *C. canariensis* as sister to *C.*
577 *abyssinica*, accurately reflects the evolutionary relationships among the species.
578
579 *Long-Distance Dispersal versus Vicariance and climate-driven extinction*
580 The vicariance-refugium hypothesis posits that the Rand Flora pattern was formed by
581 the fragmentation of a once continuous flora by aridification events, leaving relicts at the
582 eastern and western sides of the geographic disjunction. In *Canarina*, this hypothesis would
583 predict a pattern of "reciprocal monophyly" between the disjunct taxa, with Eastern Africa
584 and Canarian taxa recovered as sister groups (Couvreur *et al*. 2008; Thiv *et al*. 2010), and an
585 age for the disjunction that must predate the barrier that caused the range division, i.e., the
586 origin of the present Sahara Desert. Conversely, the long-distance dispersal (LDD)
587 hypothesis, implies the expectation that the taxa at one extreme of the disjunction (i.e., the
588 Canarian endemic) would be embedded within a clade formed by taxa from the other side
589 (i.e., an Eastern African clade), and that the disjunction should clearly postdate the formation
590 of the barrier.
591 At first, the pattern found here, with *C. canariensis* nested within a clade of two East
592 African endemics, agrees better with the LDD hypothesis. *Canarina* species are
593 characterized by the presence of fleshy fruits, with passerine bird- and lizard-mediated
594 zoochory reported for *C. canariensis* (Rodríguez *et al*. 2008). A dispersal event across the
595 7000 km of the Sahara probably requires other dispersal vectors, such as long-distance

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 (*Empetrum*), 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, would also allow for agrees better with a climate-driven vicariance explanation. Interestingly, the alternative topology recovered by *rpl*32, grouping *C. eminii* and *C. abyssinica* as sister to *C. canariensis*, would actually reinforce the vicariance explanation, since the divergence between *C. canariensis* and the East African endemics would probably be dated even earlier (> 8-7 Ma), considerably 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

estimated for the disjunction of other Rand Flora lineages, e.g, *Campylanthus* (Thiv *et al*. 2010) or *Plocama* (Pokorny *et al*. in review).

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 if, 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 favor 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 Versus "Ecological" Mountain Islands

Canarina, with its distribution in *true* oceanic islands and mountain “sky-islands”, offers an interesting comparison on the role of geographic versus 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 (Fjeldsaå and 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 & Kropelin, 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

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

more recent times, land use and deforestation might have contributed to further isolation of these forest patches (EFAP 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 Gifta 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; Mairal *et al.* in review) 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 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 segregation 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 (Krastel *et al.* 2001), might have later prevented reconnections between east and west *C. canariensis* populations. Inter-island 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),

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 Sanjuanbenito 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 Ángel Fernández (La Gomera); Félix Manuel Medina from the Cabildo of La Palma, and the Cabildo of El

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

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 authorities of 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 Caujapé 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), and a PhD research grant (BES-2010-037261) to MM. LP was funded by a research contract under CGL2012-40129-C02-01.

References

Allen MB, Armstrong HA (2008) Arabia–Eurasia collision and the forcing of Mid-Cenozoic global cooling. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **265**, 52–58.

Ancochea E, Fuster J, Ibarrola E, Cendrero A, Coello J, Hernan F, Cantagrel JM, Jamond C (1990) Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. *Journal of Volcanology and Geothermal Research*, **44**, 231–249.

Anderson CL, Channing A, Zamuner AB (2009) Life, death and fossilization on Gran Canaria–implications for Macaronesian biogeography and molecular dating. *Journal of Biogeography,* **36**, 2189–2201.

Andrus N, Trusty J, Santos-Guerra A, Jansen RK, Francisco-Ortega J (2004) Using molecular phylogenies to test phytogeographical links between East/South Africa–Southern Arabia and the Macaronesian islands—a review, and the case of *Vierea* and *Pulicaria* section *Vieraeopsis* (Asteraceae). *Taxon*, **53**, 333–346.

Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution*, **24**, 412–426.

Antonelli A, Sanmartín I (2011) Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. *Systematic Biology*, **60**, 596–615.

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Assefa A, Ehrich D, Taberlet P, Nemomissa S, Brochmann C (2007) Pleistocene colonization of afro-alpine 'sky islands' by the arctic-alpine *Arabis alpina*. *Heredity*, **99**, 133–142.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

Axelrod DI, Raven PH (1978) Late Cretaceous and Tertiary vegetation history of Africa. *Biogeography and ecology of southern Africa*, 77-130. Springer Netherlands.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution,* **29**, 2157–2167.

Barres L, Sanmartín I, Anderson CL, Susanna A, Buerki S, Galbany-Casals M, Vilatersana R (2013) Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany*, **100**, 867–882.

Beerling D, Berner RA, Mackenzie FT, Harfoot MB, Pyle JA (2009) Methane and the CH_4 related greenhouse effect over the past 400 million years. *American Journal of Science*, **309**, 97–113.

Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-revisited. *American Journal of Botany*, **97**, 1296–1303.

Blanco-Pastor JL, Fernández-Mazuecos M, Vargas P (2013) Past and future demographic dynamics of alpine species: limited genetic consequences despite dramatic range contraction in a plant from the Spanish Sierra Nevada. *Molecular ecology*, **22**, 4177–4195.

Bonnefille R, Roeland JC, Gulot J (1990) Temperature and rainfall estimates for the past 40000 years in equatorial Africa. *Nature*, **346**, 347–349.

Borsch T, Korotkova N, Raus T, Lobin W, Lohne C (2009) The petD group II intron as a species level marker: utility for tree inference and species identification in the diverse genus *Campanula* (Campanulaceae). *Willdenowia*, **39**, 7–33.

Bradshaw CD, Lunt DJ, Flecker R, Salzmann U, Pound MJ, Haywood AM, Eronen JT (2012) The relative roles of CO_2 and palaeogeography in determining Late Miocene climate: results from a terrestrial model-data comparison. *Climate of the Past Discussions*, **8**, 715–786.

Bramwell D (1985) Contribución a la biogeografía de las Islas Canarias. *Botánica Macaron*ésica, **14**, 3–34.

Cerling TE, Harris JM, MacFadden BJ, Leakey MG, Quade J, Eisenmann V, Ehleringer JR (1997) Global vegetation change through the Miocene/Pliocene boundary. *Nature*, **389**, 153–158.

Christ, H (1892) Exposé sur le rôle que joue dans le domaine de nos flores la flore dite *ancienne africaine*. *Archives des Sciences Physiques et Naturelles Genève* **3**, 369–374.

Coetzee JA (1964) Evidence for a considerable depression of the vegetation belts during the Upper Pleistocene on the East African mountains.

Coetzee JA (1993) African flora since the terminal Jurassic. *Biological Relationships between Africa and South America*, **37**, 61.

Coleman M, Liston A, Kadereit JW, Abbott RJ (2003) Repeat intercontinental dispersal and Pleistocene speciation in disjunct Mediterranean and desert *Senecio* (Asteraceae). *American Journal of Botany*, **90**, 1446–1454.

Couvreur TL, Chatrou LW, Sosef MS, Richardson JE (2008) Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC biology*, **6**, 54.

Crisp MD, Cook LG (2007) A congruent molecular signature of vicariance across multiple plant lineages. *Molecular Phylogenetics and Evol*ution **43**, 1106–1117.

Denk T, Güner HT, Grimm GW (2014) From mesic to arid: Leaf epidermal features suggest preadaptation in Miocene dragon trees (*Dracaena*). *Review of Palaeobotany and Palynol*ogy, **200**, 211–228.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

804 Désamoré A, Laenen B, Devos N, Popp M, González-Mancebo JM, Carine MA,
805 Vanderpoorten A (2011) Out of Africa: north-westwards Pleistocene expansions of the
806 heather *Erica arborea*. *Journal of Biogeography*, **38**, 164–176.
807 Donoghue MJ, Smith SA (2004) Patterns in the assembly of temperate forests around the
808 Northern Hemisphere. *Philosophical Transactions of the Royal Society of London. Series*
809 *B: Biological Sciences*, **359**, 1633–1644.
810 Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling
811 trees. *BMC Evolutionary Biol*ogy, **7**, 214.
812 EFAP (1994) Ethiopia Forestry Action Program. Final Report, Ministry of Natural
813 Resources Development and Environmental Protection, Addis Ababa.
814 Engler A (1910) Die Pflanzenwelt Afrikas inbesondere seiner tropischen Gebiete.
815 Gründzuge der Pflanzenverbreitung in Afrika un die Characterpflanzen Afrikas," in *Die*
816 *Vegetation der Erde*, eds. Engler, A. and Drude, O. (Leipzig: Verlag vonc Wilhelm
817 Engelmann), 1–1030.
818 Fernandes CA, Rohling EJ, Siddall M (2006) Absence of post-Miocene Red Sea land
819 bridges: biogeographic implications. *Journal of Biogeography*, **33**, 961–966.
820 Fernández-López AB (2014) Mapa de *Canarina canariensis* para la flora vascular de la isla
821 de la Gomera. Parque Nacional de Garajonay.
822 Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García del Rey E, Arévalo
823 JR, Whittaker RJ (2011) A reconstruction of Palaeo-Macaronesia, with particular
824 reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of*
825 *Biogeogr*aphy, **38**, 226–246.
826 Fiz O, Vargas P, Alarcón M, Aedo C, García JL, Aldasoro JJ (2008) Phylogeny and
827 historical biogeography of *Geraniaceae* in relation to climate changes and pollination
828 ecology. *Systematic Bot*any, **33**, 326–342.
829 Fjeldsaå J, Lovett JC (1997) Geographical patterns of old and young species in African
830 forest biota: the significance of specific montane areas as evolutionary centres.
831 *Biodiversity and Conserv*ation, **6**, 325–346.
832 FAO (2001) Food and Agricultural Organization of the United Nations. Forest Resource
833 Assessment, 2000 (FAO, Rome). *Forestry Paper* 140.
834 Gehrke B, Linder HP (2014) Species richness, endemism and species composition in the
835 tropical Afroalpine flora. *Alpine Botany*, **124**, 165–177.
836 Givnish TJ, Renner SS (2004) Tropical intercontinental disjunctions: Gondwana breakup,
837 immigration from the boreotropics, and transoceanic dispersal. *International Journal of*
838 *Plant Sciences*, **165**, S1–S6.
839 Goudie AS (2005) The drainage of Africa since the Cretaceous. *Geomorphology*, **67**, 437-
840 456.
841 Griffin, D L (2002) Aridity and humidity: two aspects of the late Miocene climate of North
842 Africa and the Mediterranean. *Palaeogeography, Palaeoclimatology,Palaeoecology,* **182**,
843 65–91.
844 Hedberg O (1961) Monograph of the genus *Canarina* L. (Campanulaceae). *Svensk Botanisk*
845 *Tidskrift*, **55**, 17–62.
846 Hedberg O (1969) Evolution and speciation in a tropical high mountain flora. *Biological*
847 *Journal of the Linnaean Society*, **1**, 135–148.
848 Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data.
849 *Molecular Biology and Evol*ution, **27**, 570–580.
850 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution
851 interpolated climate surfaces for global land areas. *International Journal of Climatology*,
852 **25**, 1965–1978.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Ho SY, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PW, Shapiro B (2008) Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters*, **4**, 370–374.

Ho SY, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A (2011) Time-dependent rates of molecular evolution. *Molecular Ecol*ogy, **20**, 3087–3101.

Hooker JD (1853) Botany of the Antartic Voyage of H.M. Discovery ships "Erebus" and "Terror" in the years 1831-43. *Introductory essay to the flora of New Zealand*. Reeve, London, pp. 209–223.

Hooker JD (1867) On insular floras: a lecture. *Journal of Botany*, **5**, 23–31.

Kamelina OP, Zhinkina NA (1998) On the embryology of *Ostrowskia magnifica*. The ovule and seed. *Botanicheskii Zhurnal,* **83**, 9–20.

Kebede M, Ehrich D, Taberlet P, Nemomissa S, Brochmann C (2007) Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology,* **16**, 1233–1243.

Kim SC, McGowen MR, Lubinsky P, Barber JC, Mort ME, Santos-Guerra A (2008) Timing and tempo of early and successive adaptive radiations in Macaronesia. *PLoS One*, **3**, e2139.

Knox EB, Palmer JD (1998) Chloroplast DNA evidence on the origin and radiation of the giant lobelias in eastern Africa. *Systematic Botany,* **23**, 109–149.

Kornhall P, Heidari N, Bremer B (2001) Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. *Plant Systematics and Evol*ution, **228**, 199–218.

Korotkova N, Borsch T, Quandt D, Taylor NP, Müller KF, Barthlott W (2011) What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic Rhipsalideae (Cactaceae). *American Journal of Botany*, **98**, 1549–1572.

Krastel S, Schmincke HU, Jacobs CL, Rihm R, Le Bas TP, Alibés B (2001) Submarine landslides around the Canary Islands. *Journal of geophysical Research: Solid Earth* (1978–2012), **106**, 3977–3997.

Kuper R, Kröpelin S (2006) Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution. *Science*, **313**, 803–807.

Larget BR, Kotha SK, Dewey CN, Ané C (2010) BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, **26**, 2910–2911.

Lebrun J, (1961) Les deux flores d'Afrique tropicale. Royale de Belgique. *Classe des Sciences. Memoires.* Collection in-8. 2éme série **32**, 1–82.

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS Computational Biol*ogy, **5**, e1000520.

Linder H P (2014) The evolution of African plant diversity. *Frontiers in Ecology and Evolution*, **2**, 38.

Mairal M, Aldasoro JJ, Sanmartín I, Vargas P, Alarcón M (in review). Phylogeography of *Canarina eminii*: a strong genetic structure across the Rift Valley (Africa). *Molecular Phylogenetics and Evolution.*

Maley J (1996) The African rain forest–main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary.*Proceedings of the Royal Society of Edinburgh. Section B. Biological Sciences*, **104**, 31–73.

Mansion G, Parolly G, Crowl AA, Mavrodiev

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

902 E, Cellinese N, Oganesian M, Fraunhofer K, Kamari G, Phitos D, Haberle R (2012) How to
903 handle speciose clades? Mass taxon-sampling as a strategy towards illuminating the
904 natural history of *Campanula* (Campanuloideae). *PLoS ONE*, **7**, e50076.
905 McCormack JE, Huang H, Knowles LL (2009) Sky Islands. Pp. 841–843 in: Gillespie R.G.,
906 Clague D.A. (Eds.) *Encyclopedia of Islands*. Berkeley and Los Angeles: University of
907 California Press, Ltd.
908 McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL (2011) Calibrating
909 divergence times on species trees versus gene trees: implications for speciation history of
910 *Aphelocoma* jays. *Evolution*, *65*(1), 184-202.
911 Meseguer AS, Aldasoro JJ, Sanmartín I (2013) Bayesian inference of phylogeny,
912 morphology and range evolution reveals a complex evolutionary history in St. John's
913 wort (*Hypericum*). *Molecular phylogenetics and evolution*, **67**, 379–403.
914 Meseguer AS, Lobo JM, Ree R, Beerling DJ & Sanmartín I (2014) Integrating Fossils,
915 Phylogenies, and Niche Models into Biogeography to reveal ancient evolutionary history:
916 the Case of *Hypericum* (Hypericaceae). *Systematic biology*, syu**088**.
917 Micheels A, Eronen J, Mosbrugger V (2009) The Late Miocene climate response to a
918 modern Sahara desert. *Global and Planetary Change*, **67**, 193–204.
919 Mort ME, Soltis DE, Soltis PS, Francisco-Ortega J, Santos-Guerra A (2002) Phylogenetics
920 and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and
921 chloroplast sequence data. *Systematic Bot*any, **27**, 271–288.
922 Nolasco-Soto J, González-Astorga J, Nicolalde-Morejón F, Vergara-Silva F, de los
923 Monteros AE, Medina-Villarreal A (2014) Phylogeography and demographic history of
924 *Zamia paucijuga* Wieland (Zamiaceae), a cycad species from the Mexican Pacific
925 slope. *Plant Systematics and Evolution*, 1–15.
926 Osborne CP (2008) Atmosphere, ecology and evolution: what drove the Miocene expansion
927 of C4 grasslands? *Journal of Ecol*ogy, **96**, 35–45.
928 Pelser PB, Abbott RJ, Comes HP, Milton JJ, Moeller M, Looseley ME, ... Kadereit JW
929 (2012) The genetic ghost of an invasion past: colonization and extinction revealed by
930 historical hybridization in *Senecio.Molecular ecology*, **21**, 369–387.
931 Plana V (2004) Mechanisms and tempo of evolution in the African Guineo–Congolian
932 rainforest. *Philosophical Transactions of the Royal Society of London*, **359**, 1585–1594.
933 Pokorny L, Oliván G, Shaw A (2011) Phylogeographic patterns in two southern hemisphere
934 species of *Calyptrochaeta* (Daltoniaceae, Bryophyta). *Systematic Botany*, **36**, 542–553.
935 Pokorny L, Riina R, Mairal M, Culshaw V, Meseguer A, Serrano M, Carbajal R, Ortiz S,
936 Heuertz M, Sanmartín I (in review) Living on the edge: timing of Rand Flora pattern
937 disjunctions compatible with ongoing aridification in Africa. *Frontiers in Genetics.*
938 Popp M, Gizaw A, Nemomissa S, Suda J, Brochmann C (2008) Colonization and
939 diversification in the African 'sky islands' by Eurasian *Lychnis* L. (Caryophyllaceae).
940 *Journal of Biogeogr*aphy, **35**, 1016–1029.
941 Popp M, Mirré V, Brochmann C (2011) A single Mid-Pleistocene long-distance dispersal by
942 a bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*).
943 *Proceedings of the National Academy of Sciences*,**108**, 6520–6525.
944 Quézel P, 1978. Analysis of the Flora of Mediterranean and Saharan Africa. *Annals of the*
945 *Missouri Botanical Garden,* **65**, 479–534.
946 Rambaut A, Drummond AJ (2013) TreeAnnotator v1. 7.0.
947 Rambaut A, Drummond AJ, Suchard M (2013) Tracer v1. 6.
948 Raven PH, Axelrod DI (1972) Plate Tectonics and Australasian Paleobiogeography. *Science*,
949 **176**, 1379–1386.
950 Renner S (2004) Plant dispersal across the tropical Atlantic by wind and sea currents.
951 *International Journal of Plant Sciences*, **165**, S23–S33.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

952 Riina R, Carneiro-Torres DS, Peirson JA, Berry PE, Cordeiro I (2014) Further support for
953 the Crotoneae phylogeny: A new species of Brasiliocroton (Euphorbiaceae) based on
954 morphological, geographical, and molecular evidence. *Systematic Botany*, **39**, 227–234.
955 Rodríguez A, Nogales M, Rumeu B, Rodríguez B (2008) Temporal and spatial variation in
956 the diet of the endemic lizard *Gallotia galloti* in an insular Mediterranean
957 scrubland. *Journal of Herpetology*, **42**, 213–222.
958 Roquet C, Sanmartín I, Garcia-Jacas N, Sáez L, Susanna A, Wikström N, Aldasoro JJ (2009)
959 Reconstructing the history of Campanulaceae with a Bayesian approach to molecular
960 dating and dispersal–vicariance analyses. *Molecular Phylogenetics and Evolution*, **52**,
961 575–587.
962 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,
963 Suchard, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference
964 and model choice across a large model space. *Systematic biology*, **61**, 539–542.
965 Rothfels CJ, Larsson A, Kuo L-Y, Korall P, Chiou W-L, Pryer KM (2012) Overcoming deep
966 roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II
967 ferns. *Systematic Biology*, **61**, 490–509.
968 Ryan WB, Carbotte SM, Coplan JO, O'Hara S, Melkonian A, Arko R, ... Zemsky R (2009)
969 Global Multi-Resolution Topography synthesis.*Geochemistry, Geophysics,*
970 *Geosystems*, **10**, 3.
971 Sanmartín I, Enghoff H, Ronquist F (2001) Patterns of animal dispersal, vicariance and
972 diversification in the Holarctic. *Biological Journal of the Linnean Society*, **73**, 345–390.
973 Sanmartín I (2003) Dispersal vs. vicariance in the Mediterranean: historical biogeography of
974 the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography*, **30**,
975 1883–1897.
976 Sanmartín I, Ronquist F (2004) Southern Hemisphere biogeography inferred by event-based
977 models: plant versus animal patterns. *Systematic Biol*ogy, **53**, 216–243.
978 Sanmartín I, Van Der Mark P, Ronquist F (2008) Inferring dispersal: a Bayesian approach to
979 phylogeny-based island biogeography, with special reference to the Canary Islands.
980 *Journal of Biogeogr*aphy, **35**, 428–449.
981 Sanmartín I, Anderson CL, Alarcon M, Ronquist F, Aldasoro JJ (2010) Bayesian island
982 biogeography in a continental setting: the Rand Flora case. *Biology Letters*, **6**, 703–707.
983 Senut B, Pickford M, Ségalen L (2009) Neogene desertification of Africa. *Comptes Rendus*
984 *Geoscience,* **341**, 591–602.
985 Sepulchre P, Ramstein G, Fluteau F, Schuster M, Tiercelin JJ, Brunet M (2006) Tectonic
986 uplift and Eastern Africa aridification. *Science*, **313**, 1419–1423.
987 Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling
988 EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding
989 chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**,
990 142–166.
991 Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast
992 genome sequences to choose noncoding regions for phylogenetic studies in angiosperms:
993 the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.
994 Smith SA, Donoghue MJ (2010) Combining historical biogeography with niche modeling in
995 the Caprifolium clade of *Lonicera* (Caprifoliaceae, Dipsacales). *Systematic Biology*, **59**,
996 322–341.
997 Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML
998 web servers. *Systematic biology*, **57**, 758–771.
999 Swofford D (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods)
1000 v4. Sunderland, MA: Sinauer Associates.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

Thiv M, Thulin M, Hjertson M, Kropf M, Linder HP (2010) Evidence for a vicariant origin of Macaronesian–Eritreo/Arabian disjunctions in *Campylanthus* Roth (Plantaginaceae). *Molecular Phylogenetics and Evolution*, **54**, 607–616.

Trauth MH, Larrasoaña JC, Mudelsee M (2009) Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews*, **28**, 399–411.

Treviranus GR (1803) Biologie, oder Philosophie der lebenden. *Natur*, 2 volumes, Röwer, Göttingen.

von Humboldt A, Bonpland A (1805) Essai sur la géographie des plantes; accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe, pendant les années 1799, 1800, 1801, 1802 et 1803. Paris: chez Levrault, Schoell et compagnie, libraries.

Wang Q, Zhou SL, Hong DY (2013) Molecular phylogeny of the platycodonoid group (Campanulaceae s. str.) with special reference to the circumscription of *Codonopsis*. *Taxon*, **62**, 498–504.

Whittaker RJ, Triantis KA, Ladle RJ (2008) A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, **35**, 977-994.

White F (1983) The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa by F White. *Natural Resources Research Report XX, UNESCO, Paris, France*.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, **19**, 639–644.

Williamson S, Orive ME (2002) The Genealogy of a Sequence Subject to Purifying Selection at Multiple Sites. *Molecular Biology and Evol*ution, **19**, 1376–1384.

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, **84**, 9054–9058.

Wondimu T, Gizaw A, Tusiime FM, Masao CA, Abdi AA, Gussarova G, Popp M, Nemomissa S, Brochmann C (2014) Crossing barriers in an extremely fragmented system: two case studies in the afro-alpine sky island flora. *Plant Systematics and Evolution*, **300**, 415–430.

Yesson C, Culham A (2006). Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Systematic Biology*, **55**, 785–802.

Zhaparova NK (1996) Root development of *Ostrowskia magnifica* Regel, a rare species in Kazakhstan. *Acta Phytogeographica Suecica*, **81**, 88–91.

Zhou Z, Hong D, Niu Y, Li G, Nie Z, Wen J, Sun H (2013). Phylogenetic and biogeographic analyses of the Sino-Himalayan endemic genus *Cyananthus* (Campanulaceae) and implications for the evolution of its sexual system. *Molecular phylogenetics and evolution*, **68**, 482–497.

DATA ACCESSIBILITY.

DNA sequences: Genbank accessions KP761423 to KP761687

GenBank accessions, sampling locations and/or online-only appendices uploaded as online supplemental material

Original script input file used to perform the nested BEAST approach: Dryad doi: 10.5061/dryad.5jc73.

NEXUS files for the single and concatenated dataset: Dryad doi:10.5061/dryad.5jc73.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

AUTHOR CONTRIBUTIONS

I.S. and M.M. designed the study. M.M., M.A., and J.J. gathered the data. M.M., L.P. and I.S. analyzed the data. M.M. and I.S. wrote the paper with help from L.P.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

1053 **Table 1.** Summary statistics of the chloroplast and nuclear regions analyzed here for the
1054 *Canarina* dataset (no outgroups). Fragment length is given in base pairs (bp); alignment
1055 length includes the indels.

	*rpl*32-*trn*L	3´*trn*V-*ndh*C	*psb*J-*pet*A	*pet*B-*pet*D	*Trn*L-*trn*F	*Trn*S-*trn*G	*trn*G intron	ITS
Fragment length	581–647	756–855	822–840	753–798	683–918	666–688	661–676	604–734
Alignment length	647	855	841	808	933	690	677	734
Constant sites	590	822	814	766	902	666	666	648
Variable sites	57	33	27	42	31	24	11	86
Indel (%)	35.06%	–	20.13%	11.2%	8.13%	6.98%	–	10.28%

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

Figure Captions

Figure 1. a) Worldwide distribution of tribe Platycodoneae (Campanulaceae), showing the geographic disjunction between the single African genus (*Canarina*) and the remaining members of the tribe, which are endemic to the mountains of Asia. b) Geographic distribution of the three species of *Canarina*; the distribution of the East African species, *C. eminii* and *C. abyssinica* has been modified from Hedberg (1961). Numbers correspond to the sampled populations, with codes given in Table S1. Maps have been modified from GeoMapApp (Ryan *et al*. 2009; www.geoMapApp.org).

Figure 2. Bayesian Majority-Rule consensus trees obtained by MrBayes from the: a) the *Platycodoneae* concatenated chloroplast dataset (*psb*J-*pet*A, *trn*L-*trn*F, *pet*B-*pet*D); b) the *Platycodoneae* nuclear ribosomal (ITS) dataset; c) the *Canarina* concatenated chloroplast and nuclear dataset (ITS, *psb*J-*pet*A, *trn*L-*trn*F, *pet*B-*pet*D, *trn*S-*trn*G). Numbers above branches indicate Bayesian credibility values (PP); numbers below branches indicate maximum likelihood bootstrap support values. Codes for *Canarina* populations correspond to those shown in Table S1.

Figure 3. MCC tree with 95% HPD confidence intervals for main phylogenetic relationships and lineage divergence times obtained in BEAST (stars indicate constrained nodes) for the: a) Platycodoneae dataset (*psb*J-*pet*A, *trn*L-*trn*F, *pet*B-*pet*D). b) *Canarina* dataset (*psb*J-*pet*A, *trn*L-*trn*F, *pet*B-*pet*D, *trn*S-*trn*G, ITS).

Figure 4. Nested analyses of all three linked datasets: Platycodoneae (left) and *C. eminii* and *C. canariensis* (right) (see text for more details). Numbers above branches indicate mean ages and numbers below branches indicate Bayesian PP. Codes for *Canarina* populations correspond to those shown in Table S1. Mean ages and confidence intervals of all analyses are indicated in Fig. S4 and Table S6.

Figure 5. Results from the BEAST Bayesian ancestral range reconstruction (Lemey *et al.* 2009). Colored 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 less than 0.1 posterior probabilities. a) MCC tree from the analysis of the *Canarina* dataset (standard secondary calibration approach; stem age highlighted inside a square and crown ages highlighted inside circles). b) MCC tree of the *C. eminii* population-level dataset (nested dating approach). c) MCC tree of the *C. canariensis* population-level dataset (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. Color 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 color the stronger the support: purple> yellow> white). Maps have been modified from satellite pictures in Google Earth.

Figure 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-colored regions indicate low climatic suitability values; conversely, dark red indicate high suitability areas.

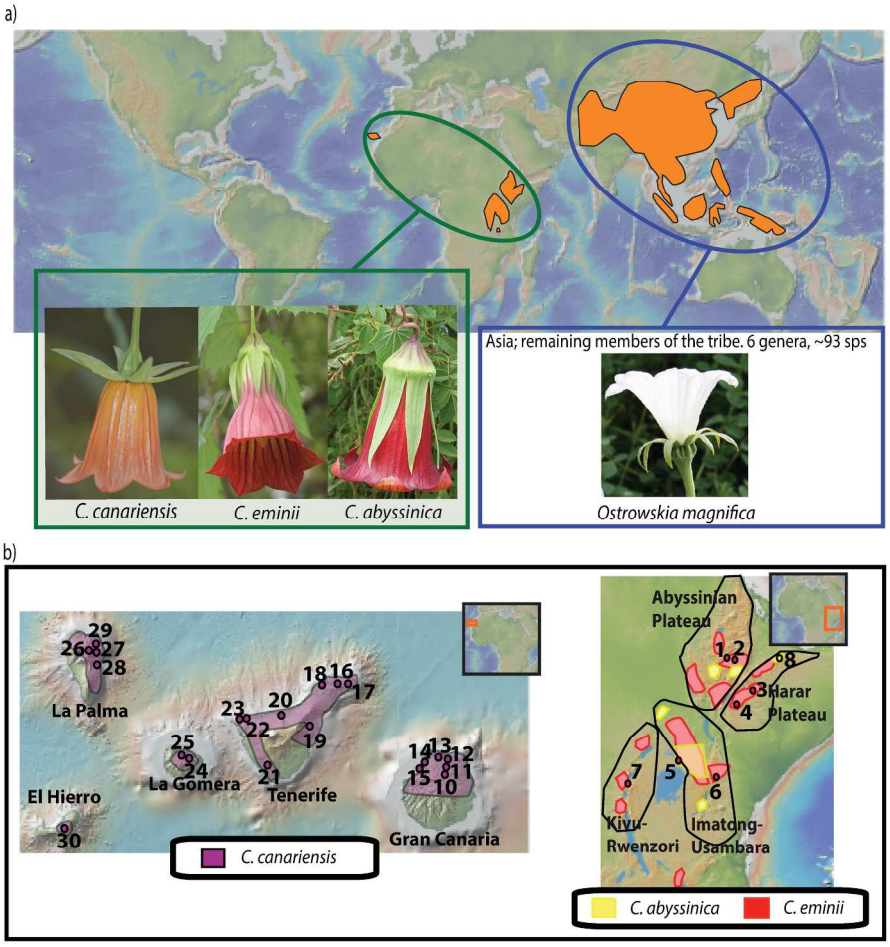
This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

SUPPORTING INFORMATION

Additional supporting information (Supplementary Text) may be found in the online version of this article.

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

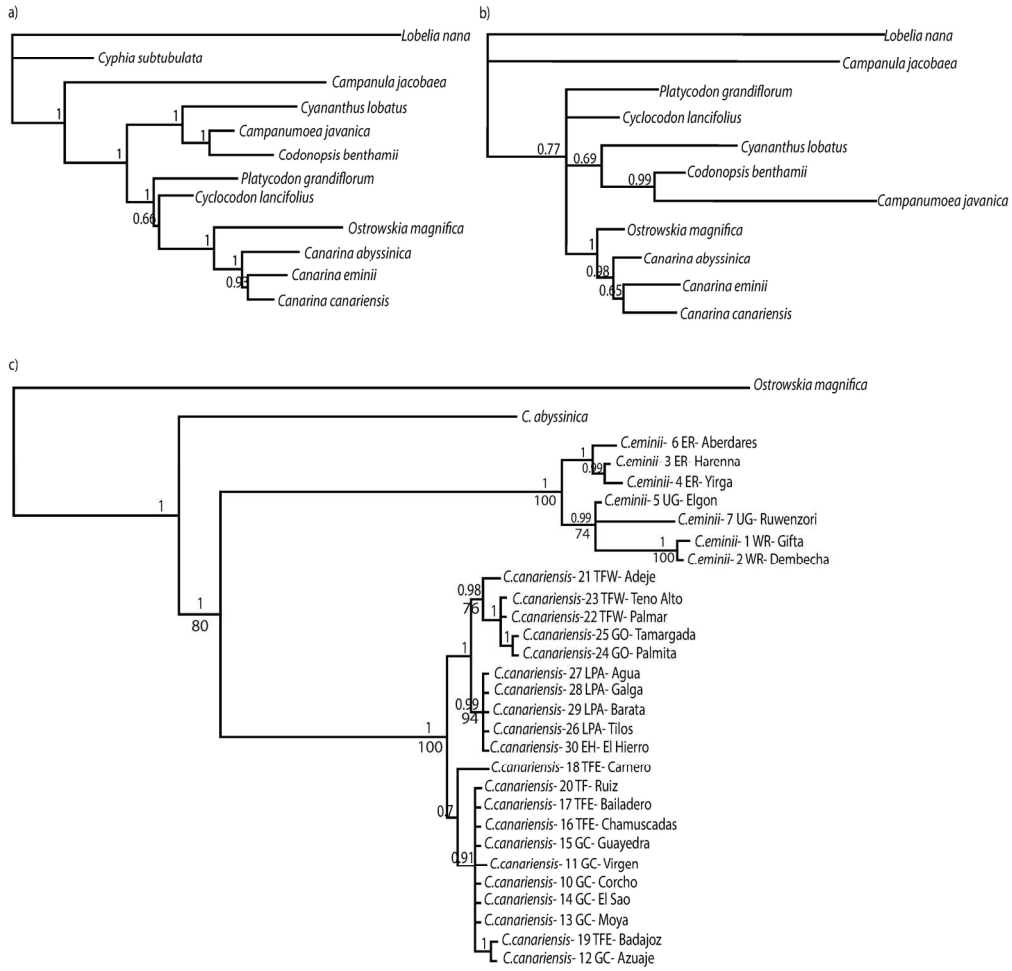
Figure 1



a) Worldwide distribution of tribe Platycodoneae (Campanulaceae), showing the geographic disjunction between the single African genus (Canarina) and the remaining members of the tribe, which are endemic to the mountains of Asia. b) Geographic distribution of the three species of Canarina; the distribution of the East African species, C. eminii and C. abyssinica has been modified from Hedberg (1961). Numbers correspond to the sampled populations, with codes given in Table S1. Maps have been modified from GeoMapApp (Ryan et al. 2009; www.geoMapApp.org).
221x296mm (300 x 300 DPI)

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

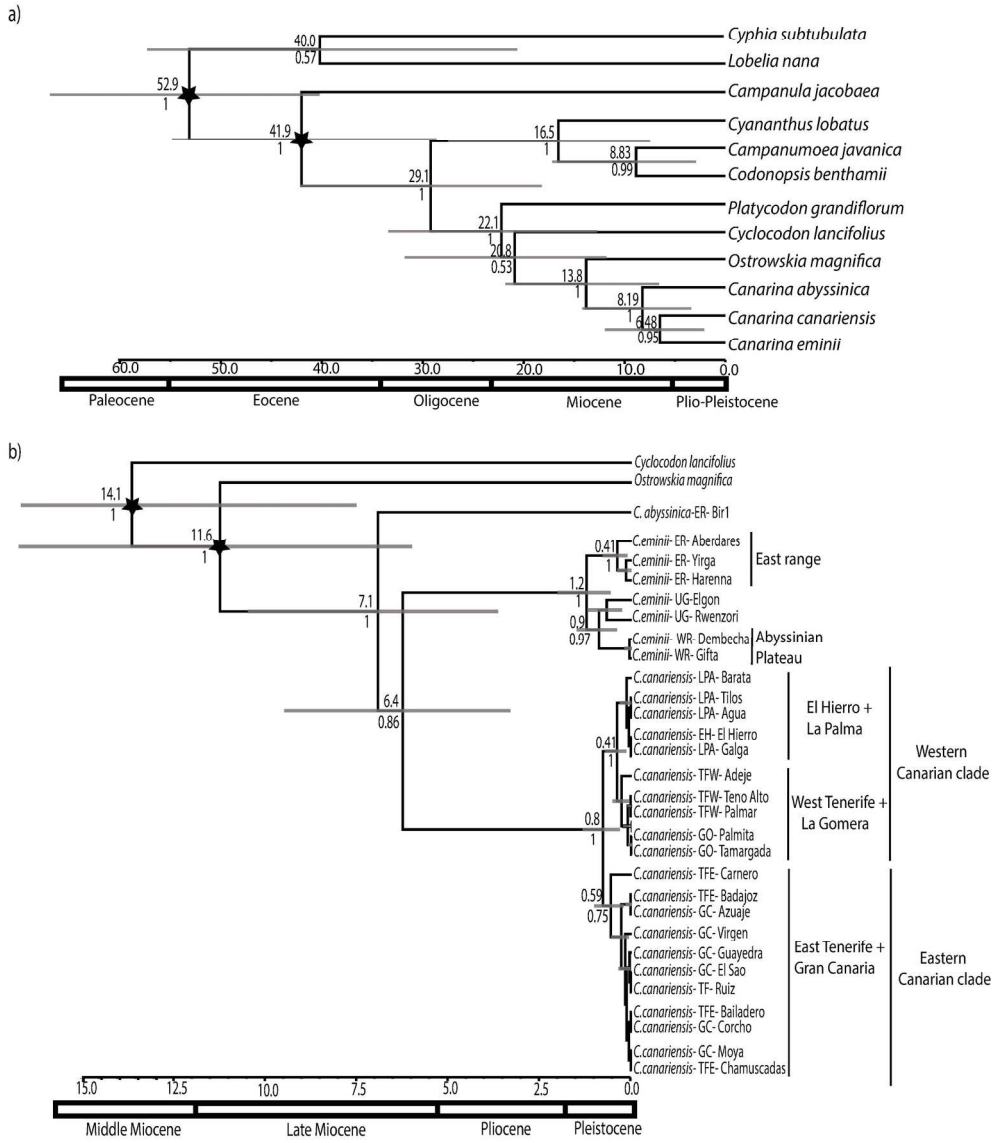
Figure 2



Bayesian Majority-Rule consensus trees obtained by MrBayes from the: a) the Platycodoneae concatenated chloroplast dataset (psbJ-petA, trnL-trnF, petB-petD); b) the Platycodoneae nuclear ribosomal (ITS) dataset; c) the Canarina concatenated chloroplast and nuclear dataset (ITS, psbJ-petA, trnL-trnF, petB-petD, trnS-trnG). Numbers above branches indicate Bayesian credibility values (PP); numbers below branches indicate maximum likelihood bootstrap support values. Codes for Canarina populations correspond to those shown in Table S1.
169x171mm (300 x 300 DPI)

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

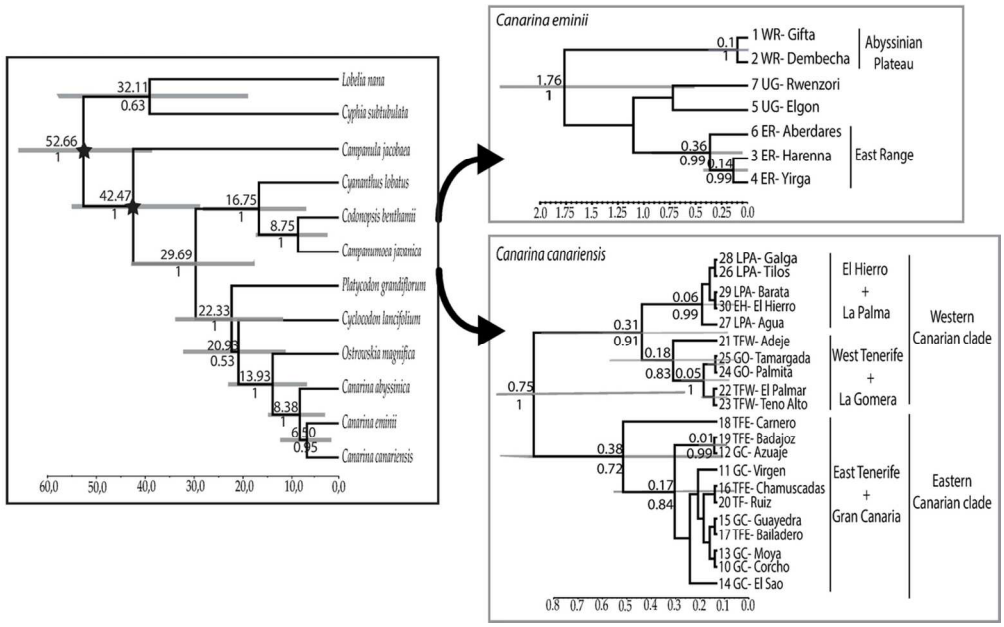
Figure 3



MCC tree with 95% HPD confidence intervals for main phylogenetic relationships and lineage divergence times obtained in BEAST (stars indicate constrained nodes) for the: A) Platycodoneae dataset (psbJ-petA, trnL-trnF, petB-petD). B) Canarina dataset (psbJ-petA, trnL-trnF, petB-petD, trnS-trnG, ITS).
180x218mm (300 x 300 DPI)

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

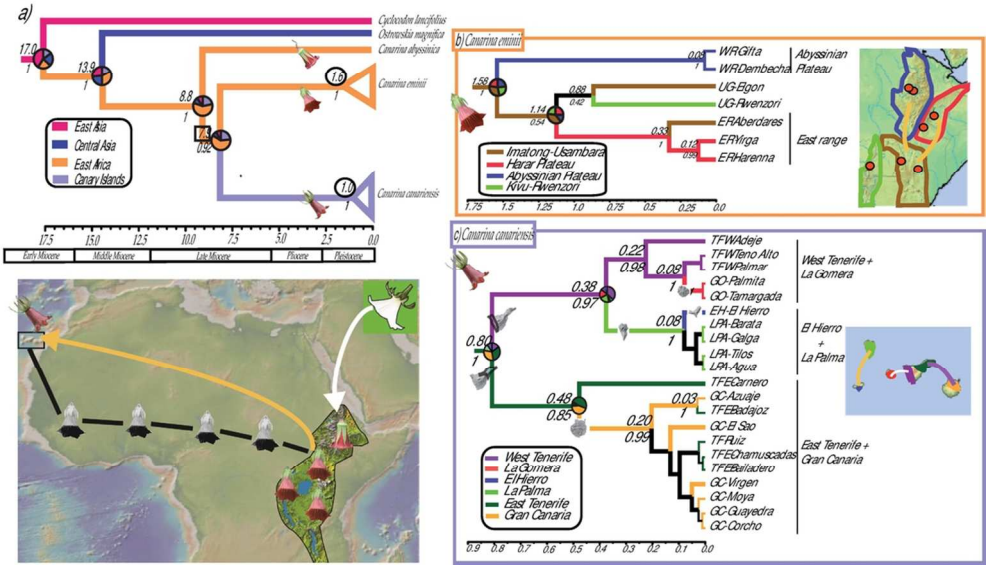
Figure 4



Nested analyses of all three linked datasets: Platycodoneae (left) and C. eminii and C. canariensis (right) (see text for more details). Numbers above branches indicate mean ages and numbers below branches indicate Bayesian PP. Codes for Canarina populations correspond to those shown in Table S1. Mean ages and confidence intervals of all analyses are indicated in Fig. S4 and Table S6.
108x70mm (300 x 300 DPI)

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

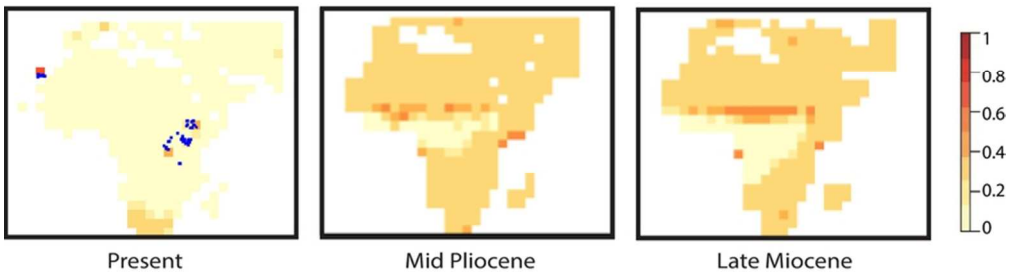
Figure 5



Results from the BEAST Bayesian ancestral range reconstruction (Lemey et al. 2009). Colored 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 less than 0.1 posterior probabilities. a) MCC tree from the analysis of the Canarina dataset (standard secondary calibration approach; stem age highlighted inside a square and crown ages highlighted inside circles). b) MCC tree of the C. eminii population-level dataset (nested dating approach). c) MCC tree of the C. canariensis population-level dataset (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. Color 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 color the stronger the support: purple> yellow> white). Maps have been modified from satellite pictures in Google Earth.
100x59mm (300 x 300 DPI)

This is an Accepted Manuscript of an article published in Molecular Ecology on 16 March 2015, available online: http://dx.doi.org/10.1111/mec.13114

Figure 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-colored regions indicate low climatic suitability values; conversely, dark red indicate high suitability areas.
74x22mm (300 x 300 DPI)