RESEARCH ARTICLE

Botany
American Journal of
BOTANICAL SOCIETY OF AMERICA

Rare and widespread: integrating Bayesian MCMC approaches, Sanger sequencing and Hyb-Seq phylogenomics to reconstruct the origin of the enigmatic Rand Flora genus *Camptoloma*

Victoria Culshaw[1] | Tamara Villaverde[1,2] | Mario Mairal[3] | Sanna Olsson[4] | Isabel Sanmartín[1]

[1]Real Jardín Botánico (RJB), CSIC, Plaza de Murillo, 2, Madrid 28014, Spain

[2]Department of Botany, Universidad de Almeria, Carretera Sacramento, La Cañada de San Urbano Almeria 04120, Spain

[3]Department of Biodiversity, Ecology and Evolution, Universidad Complutense de Madrid, Madrid, Spain

[4]Department of Forest Ecology and Genetics, Forest Research Centre, INIA-CIFOR, Carretera de la Coruña km 7.5, Madrid 28040, Spain

Correspondence
Victoria Culshaw and Isabel Sanmartín, Real Jardín Botánico (RJB), CSIC, Plaza de Murillo, 2, Madrid 28014, Spain.
Email: vickycul@hotmail.com and isanmartin@rjb.csic.es

Abstract

Premise: Genera that are widespread, with geographically discontinuous distributions and represented by few species, are intriguing. Is their achieved disjunct distribution recent or ancient in origin? Why are they species-poor? The Rand Flora is a continental-scale pattern in which closely related species appear codistributed in isolated regions over the continental margins of Africa. Genus *Camptoloma* (Scrophulariaceae) is the most notable example, comprising three species isolated from each other on the northwest, eastern, and southwest Africa.

Methods: We employed Sanger sequencing of nuclear and plastid markers, together with genomic target sequencing of 2190 low-copy nuclear genes, to infer interspecies relationships and the position of *Camptoloma* within Scrophulariaceae by using supermatrix and multispecies-coalescent approaches. Lineage divergence times and ancestral ranges were inferred with Bayesian Markov chain Monte Carlo (MCMC) approaches. The population history was estimated with phylogeographic coalescent methods.

Results: *Camptoloma rotundifolium*, restricted to Southern Africa, was shown to be a sister species to the disjunct clade formed by *C. canariense*, endemic to the Canary Islands, and *C. lyperiiflorum*, distributed in the Horn of Africa–Southern Arabia. *Camptoloma* was inferred to be sister to the mostly South African tribes Teedieae and Buddlejeae. Stem divergence was dated in the Late Miocene, while the origin of the extant disjunction was inferred as Early Pliocene.

Conclusions: The current disjunct distribution of *Camptoloma* across Africa was likely the result of fragmentation and extinction and/or population bottlenecking events associated with historical aridification cycles during the Neogene; the pattern of species divergence, from south to north, is consistent with the "climatic refugia" Rand Flora hypothesis.

KEYWORDS
aridification, biogeographic disjunctions, climate change, extinction, relict species, Scrophulariaceae

Intracontinental biodiversity disjunctions, in which climatic barriers isolate disjunct sister taxa within a continent, provide ideal scenarios for testing past climatic change hypotheses (Crisp and Cook, 2007; Mairal et al., 2017). One of the best studied examples is the Rand Flora (RF) pattern, a floristic biogeographic pattern where unrelated plant lineages with subhumid or xeric affinities share a similar disjunct distribution on the edges of the African continent and adjacent islands (Christ, 1892; Bramwell, 1985; Sanmartín et al., 2010; Pokorny et al., 2015; Mairal et al., 2015a, 2017; Villaverde et al., 2018; Riina et al., 2020). Recent studies using phylogenetic, divergence time, paleoclimate, and niche-modeling data indicate that the RF pattern was formed over the last 20 million years through cycles of range contraction (i.e., by ecological vicariance and

population extinction) and range expansion via short and medium-range dispersal. The origin lies in aridification-tropicalization processes that affected different parts of the African continent at different times (Mairal et al., 2015a, 2017; Pokorny et al., 2015).

The current aridification that affects the Mediterranean region and northern Africa, which is a change from wetter to drier climates with increasing annual temperatures and decreasing precipitation, is a major scientific and societal concern (Park et al., 2018; Berdugo et al., 2020). This trend is not recent, it has been ongoing for the last 25 million years, driven by global and regional geotectonism (Senut et al., 2009; Trauth et al., 2008; Pokorny et al., 2015). Understanding how lineages, such as the RF, reacted to these past climate change events, through persistence in situ, adaptation, geographic displacement, or extinction, is crucial to increase the accuracy of our predictions for future climate change–driven outcomes (Willis and McDonald, 2011; Svenning et al., 2015; Meseguer et al., 2018).

Two biogeographic distributions are typically found within RF lineages: (1) an east/west northern disjunction between Macaronesia and northwestern Africa (Western Sahara) on one side, and the region including Horn of Africa, Southern Arabia, and the Socotra Archipelago on the other. Taxa like the genus *Canarina* (Campanulaceae; Mairal et al., 2015a, 2017) or the *Euphorbia balsamifera* complex (Euphorbiaceae; Villaverde et al., 2018; Riina et al., 2020) show this type of disjunction. (2) A south/east disjunction between southern Africa and the

Ethiopian Highlands and Eastern African mountain ranges can be found in the *Euphorbia aphyllis* group (Pokorny et al., 2015). However, some Rand Flora lineages exhibit an even wider disjunction, spanning the entire African continent, a west/east/south biogeographic disjunct pattern. Examples are found in genus *Kleinia* (Asteraceae; Pelser et al., 2007) and *Plocama* (Rubiaceae; Backlund et al., 2007; Rincón-Barrado et al., 2021).

Genus *Camptoloma* Benth (1846), belonging to Scrophulariaceae (Kornhall et al., 2001), is the most paradigmatic example of the east-west-south RF pattern because it only includes three species: *Camptoloma canariense* (Webb & Berthel) Hilliard is endemic to Gran Canaria in the Canary Archipelago; *Camptoloma lyperiiflorum* (Vatke) Hilliard is native to the Horn of Africa–Southern Arabian region and Socotra Island; while *Camptoloma rotundifolium* Benth. (*Camptoloma* type species) is restricted to Namibia (Figure 1). Therefore these three species are separated by thousands of kilometers and by geographic and climatic barriers in the form of the open ocean, desert land, and tropical lowlands (Figure 1). Populations often exhibit restricted distributions, occupying humid microhabitats within a xeric geographic template. For example, *C. canariense* ("saladillo de risco") is a small, perennial woody herb that grows in shaded, vertical crevices on coastal slopes and cliffs in southwest and eastern Gran Canaria. *Camptoloma rotundifolium* ("blondhaim") is a dwarf shrub with succulent, pubescent leaves and geranium-like white to whitish-rose colored flowers in terminal racemes; it occurs in moist areas in the Namibia western escarpments (e.g., Brandberg Mountains) within the
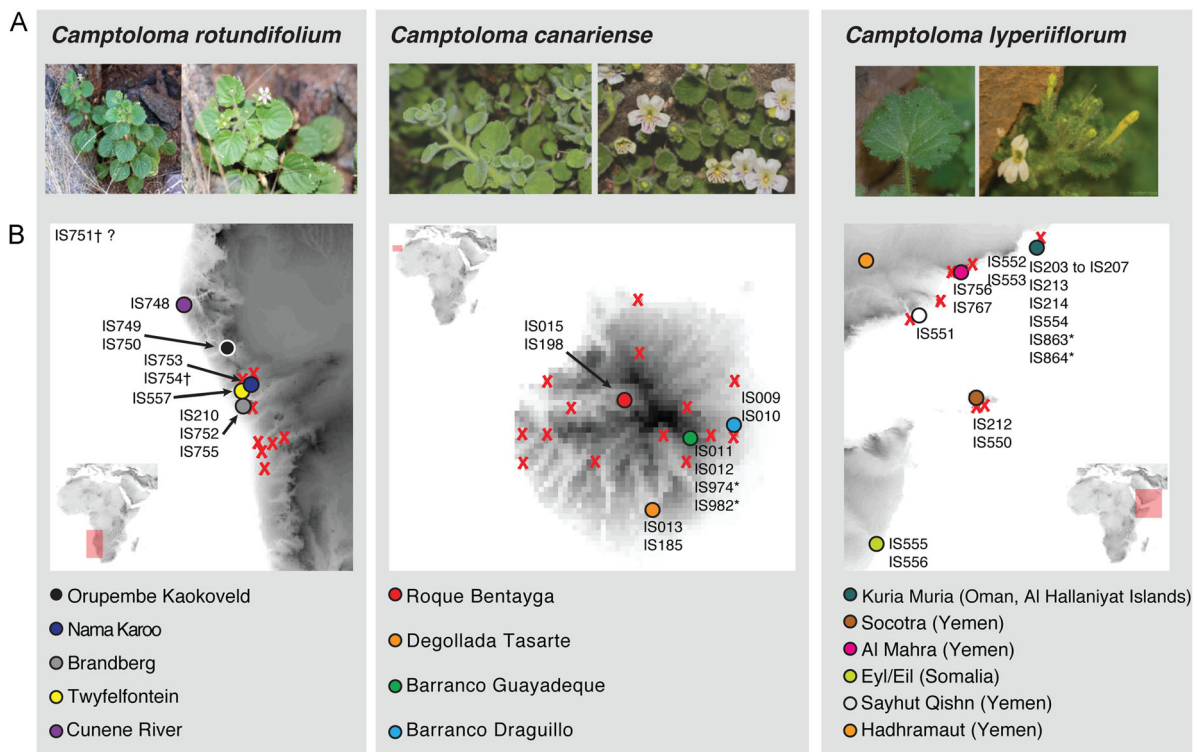
FIGURE 1 (A) Photographs of each of the three species of genus *Camptoloma*: *C. rotundifolium* (photo courtesy of Martin Weigand), *C. canariense* (photo courtesy of Mario Mairal), and *C. lyperiiflorum* (photo courtesy of Morton Ross). (B) Geographic distribution of genus *Camptoloma*, showing the east-west-south disjunction among the three species, spanning thousands of kilometers across the African continent. Red Xs represent reliable location geographical coordinates from GBIF, while colored dots represent locations recorded from herbarium vouchers with the corresponding DNA extraction codes

Karoo-Namib floristic region. *Camptoloma lyperiiflorum* also inhabits a special microclimate in the Socotra Archipelago, characterized by higher humidity and lower seasonality than nearby regions in the Arabian Peninsula, Oman, and Yemen (Domina et al., 2012). Little is known about the dispersal mode or reproductive strategies of *Camptoloma* species, except that pollination is probably by generalist insects in *C. canariense* (Naranjo Suárez and Adenda Santana López, 2006).

Previous attempts to reconstruct species-level relationships within *Camptoloma* have used molecular plastid markers (*rps16* intron, *ndhF* gene, and *trnL-F* intron-spacer) and included only one sequence per species (Kornhall et al., 2001; Oxelman et al., 2005). These studies recovered southern African *C. rotundifolium* as the sister group of a northern clade formed by *C. canariense* and *C. lyperiiflorum*. Genus *Camptoloma* was placed as sister to tribes Buddlejeae and Teedieae. The first includes the cosmopolitan genus *Buddleja* L. (~108 species) and four other genera with a mostly South African distribution (*Chilianthus, Nicodemia, Gomphostigma,* and *Emorya*) have been recently synonymized within a larger *Buddleja* (Chau et al., 2017). The second is composed of six genera (*Teedia, Freylinia, Oftia, Dermatobrothys, Phygelius,* and *Ranopisoa*) and only 12 species distributed in southern Africa. Using the same molecular dataset as Oxelman et al. (2005), Pokorny et al. (2015) estimated the crown age of *Camptoloma* as 5.45 Ma, in the Miocene-Pliocene boundary, while stem-clade *Camptoloma* was dated much older in the Late Miocene (10.19 Ma). However, unlike Oxelman et al. (2005)'s timetree, Pokorny's recovered *C. canariense* as sister to the clade formed by southern African *C. rotundifolium* and eastern African *C. lyperiiflorum*. The South African monotypic genus *Phygelius* E. Mey. ex Benth. was placed as *Camptoloma*'s sister group, whereas Oxelman et al. (2005) recovered *Phygelius capensis* as sister to tribe Teedieae. None of these relationships, however, received strong clade support. The Miocene-Pliocene age for the *Camptoloma* RF disjunction estimated by Pokorny et al. (2015) agree well with the aridification-driven vicariance hypothesis; the large time interval (5 million years) between stem and crown-node *Camptoloma* supports high extinction rates prior to the extant radiation (Pokorny et al., 2015).

In this study, we reconstruct the spatiotemporal evolution of the enigmatic genus *Camptoloma* using both plastid and nuclear markers and a broader sampling, including a comprehensive representation of outgroup taxa at the tribal level for Scrophulariaceae s.s. and population-level sampling for each *Camptoloma* species. To clarify the position of *Camptoloma* within the Scrophulariaceae family, we used Sanger sequencing for seven noncoding plastid markers, including five additional markers compared to previous studies (Oxelman et al., 2005), and the nuclear ribosomal internal transcribed spacer (ITS) region. To increase phylogenetic resolution within *Camptoloma*, we complemented this analysis with a high-throughput sequencing (HTS) approach. The target sequencing technique Hyb-Seq (hybrid enrichment with genome skimming; Weitemier et al., 2014) has gained popularity in recent years because it can sequence hundreds to thousands of low-copy nuclear orthologous genes (LCNGs) across the genome for multiple specimens at an affordable cost, and has proven efficient for both fresh and dried herbarium specimens (Villaverde et al., 2018; Viruel et al., 2020). Additionally, high-copy plastid DNA can be obtained as a by-product (Villaverde et al., 2018). We used a new bait kit targeting 2190 low-copy nuclear loci, spanning ~3.7 million base pairs (bp) that we developed for family Scrophulariaceae (unpublished data). Phylogenetic relationships within *Camptoloma* and Scrophulariaceae were inferred using concatenated (supermatrix) approaches and multispecies-coalescent methods under maximum likelihood and Bayesian inference frameworks.

The resultant phylogenetic hypotheses were used to infer population and species-level divergence times and ancestral ranges, using Markov chain Monte Carlo (MCMC) approaches. At the genus and family level, we employed the Dispersal-Extinction-Cladogenesis biogeographic model (Ree and Smith, 2008) implemented in a Bayesian framework in the RevBayes software (Höhna et al., 2016; Landis et al., 2018). At the population level, we used phylogeographic methods based on a continuous-time, Markov chain (CTMC) process with discrete-states, which assumes an unstructured coalescent process (Lemey et al., 2009) or a Bayesian structured coalescent approximation (BASTA; De Maio et al., 2015); both are implemented in the BEAST 2 software (Bouckaert et al., 2014). Our study aims were to (1) clarify species relationships within *Camptoloma* and its phylogenetic position in Scrophulariaceae; and (2) combine information from population-level and species-level analyses to disentangle the temporal and spatial origins of the Rand Flora disjunction in *Camptoloma*. In particular, we wanted to examine whether the disjunction could be explained by fragmentation and extinction and/or population bottlenecking events associated with historical aridification cycles, in line with the "climatic refugia" hypothesis (Mairal et al., 2015a; Pokorny et al., 2015).

MATERIALS AND METHODS

Taxon sampling

DNA was extracted from silica gel–dried leaves obtained through several field expeditions (2009–2018) to Gran Canaria, Spain for *C. canariense*, and from loans of dried material from different herbaria: Real Jardín Botánico, Madrid, Spain (MA), Royal Botanic Garden of Edinburgh, Scotland (E), Uppsala Museum of Evolution Herbarium, Sweden (U) and Pretoria National Herbarium, South Africa (PRE), in the case of *C. lyperiiflorum* and *C. rotundifolium*. For Sanger sequencing, we used a total of 35 specimens of *Camptoloma* representing 15 populations that span the genus's geographic range: eight specimens of *C. canariense*, 17 specimens of *C. lyperiiflorum*, and 10 specimens of *C. rotundifolium*. Additionally, 38 samples representing 24 species from 22 genera and seven tribes within Scrophulariaceae (Myoporeae, Leucophylleae, Aptosimeae, Scrophularieae, Limoselleae, Buddlejeae, and Teedieae) were included as outgroup taxa (Oxelman et al., 2005). A sample of the *Plantago* genus from Plantaginaceae, sister to Scrophulariaceae, was used as the most external outgroup. In total, our Sanger dataset included 74 specimens.

For Hyb-Seq genomics, *Camptoloma* was represented by two individuals per species (Appendix S1), obtained from either new field work, as in *C. canariense*, or dried herbarium material, for *C. rotundifolium* and *C. lyperiiflorum*; some of the latter corresponded to individuals used also in Sanger sequencing. We added one sample of *Freylinia lanceolata* and one sample of *Buddleja polystsachya* as outgroup taxa. In all, our Hyb-Seq dataset comprised eight samples. Complete voucher information on all Sanger and Hyb-Seq samples, including taxon names, country, and locality of collection, geographical coordinates, and National Center for Biotechnology Information (NCBI) accession numbers can be found in Appendix S1.

Sanger sequencing

DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN Inc., California, USA) according to the manufacturer's instructions. Sequences for seven noncoding plastid regions —the intergenic spacers *trnL-trnF*, *trnS-trnG*, *rpl32-ndhF*, *psbJ-petA*, *petB-petD*, *trnT-trnL* and the *rps16* intron— and the multicopy nuclear marker ITS were obtained using universal plant primers and newly-designed primers for difficult herbarium specimens. See "Extended Material and Methods" in Appendix S2 and Appendix S3 for more details on primers and specific polymerase chain reaction (PCR) protocols. We were unable to sequence several specimens for a few markers, especially for the outgroup taxa. For some of these, we used equivalent sequences from the same marker and species downloaded from GenBank, specifically *trnL-trnF* for *Hebenstretia dentata* and *trnL-trnF* for *Leucophyllum texanum*. For *Plantago*, we constructed a chimera sequence by concatenating different species sequences obtained from GenBank of ITS, *rpl32-ndhF*, *petB-petD*, *rps16*, *trnL-trnF*, and *trnS-trnG*, see Appendix S1 for accession numbers. In all, we generated 521 new sequences for 73 specimens: ITS (71 sequences), *rpl32-ndhF* (73 sequences), *petB-petD* (59 sequences), *psbJ-petA* (62 sequences), *rps16* intron (67 sequences), *trnL-trnF* (64 sequences), *trnS-trnG* (62 sequences), and *trnT-trnL* (66 sequences).

Hyb-Seq genomic sequencing

For the gene capture with Hyb-Seq, we used a 60,000 bait kit designed by us in collaboration with Arbor Biosciences (Ann Arbor, Michigan, USA). This is a large-scale project aimed to improve phylogenetic resolution within family Scrophulariaceae s.s. (unpublished data). Analysis of the Scrophulariaceae dataset is still ongoing, and we only report here results for *Camptoloma*. The custom probe kit targeted 2190 LCNGs and approximately 3.7 million base pairs, and was designed using MarkerMiner (Chamala et al., 2015) based on transcriptomic resources available through the 1KP initiative (http://www.onekp.com/public_data.html; Matasci et al., 2014) of Scrophulariaceae species *Anticharis glandulosa* (code EJBY), *Buddleja* sp. (GRFT), *Buddleja davidii* (XRLM), *Verbascum arcturus* (=Celsia

arcturus, SIBR, and *Verbascum* sp. (XXYA). As the reference genome to define intron and exon boundaries, we used *Erythranthe guttata* (=*Mimulus guttatus*) in Phrymaceae), another family from the order Lamiales available through http://www.phytozome.net (Chan et al., 2010). Wet lab work was carried out by AllGenetics & Biology SL (A Coruña, Spain). A modified Consortium of European Taxonomic Facilities (CETAF) protocol was used for genomic DNA extraction. DNA was purified to remove potential inhibitors, and was quantified using the Qubit dsDNA HS assay (Thermo Fisher Scientific, Waltham, Massachusetts, USA). As DNA quantities and concentrations were above the minimum required, all DNA were used as input for library preparation and bait capture. Libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, Massachusetts, USA) following the manufacturer's instructions, but using half of the recommended volume for all reactions. Libraries were indexed for postsequencing demultiplexing. Following library preparation, specific regions of the genome were hybridized to biotinylated oligonucleotide probes (MyBaits, Arbor Biosciences, Ann Arbor, Michigan, USA), and then captured using magnetic beads coated with streptavidin (Dynabeads MyOne Streptavidin C1, Thermo Fisher Scientific). The libraries were quantified using the Qubit dsDNA HS assay (also from Thermo Fisher Scientific) and pooled in equimolar amounts according to the origin of the sample (herbarium or silica) and its phylogenetic relationship (ingroup or outgroup). These pools were sequenced in two different runs of an Illumina MiSeq PE300.

Phylogenetic analyses

Sanger sequencing

Sequences were aligned using MUSCLE v 3.8.31 (www.drive5.com/muscle; Edgar, 2004) with a maximum of eight iterations and adjusted manually with respect to the event-based criteria of Morrison (2006) in Geneious Pro 5.6.7 (https://www.geneious.com). We constructed three sequence datasets for phylogenetic analyses: (1) the "all specimens" dataset included all 73 individuals sequenced in our study; (2) the "outgroup" dataset included all 38 sequenced outgroup taxa, plus the *Plantago* chimera, and two individuals (representing different populations) within each species of *Camptoloma*—a total of 45 taxa; (3) the three "population-level" datasets included all individuals sequenced within each species of *Camptoloma*: "*C. canariense*" (eight sequences), "*C. lyperiiflorum*" (17 sequences), and "*C. rotundifolium*" (10 sequences).

Phylogenetic inference was performed under an MCMC Bayesian framework using MrBayes v3.2.6 (Ronquist et al., 2012). We used the "all specimens" dataset (73 sequences), with the GenBank sequences of *Plantago* as outgroup to root the tree. Each marker was first analyzed separately to assess congruence among the resultant tree topologies. The software jModelTest v. 2.1.10 (Darriba et al., 2012) was used to select the best-fit nucleotide substitution model—these were GTR+I+G for ITS; TVM+G (*rpl32-ndh*F); TIM3+G (*petB-pet*D), GTR+G

(*psb*J-*pet*A), TVM+G (*rps*16), *trn*L-*trn*F (), TrN+G (*trn*S-*trn*G), and TVM+G (*trn*T-*trn*L). Some of these models are not implemented in MrBayes (e.g., TVM+G) and were replaced by models with similar parameter complexity: GTR+I+G was used for ITS and GTR+G for all plastid markers. Indel (gap) information was included in each aligned matrix using the Simmons and Ochoterena (2000) simple coding algorithm in SeqState (Müller, 2005); this was modeled as a single partition under the restriction site (F81) model according to the MrBayes manual.

Comparison among gene-tree topologies (results not shown) revealed little resolution; there were no incongruent clades among gene trees that also received significant Bayesian clade support (>95%). Given the lack of well supported incongruence, we concatenated individual markers. Additionally, we carried out a sensitivity analysis in which we compared different molecular datasets in terms of their fit to the data: (1) nuclear marker (ITS) dataset; (2) concatenated plastid dataset partitioned by marker; (3) concatenated nuclear-plastid dataset partitioned by genome (ITS vs. plastid); and (4) concatenated nuclear-plastid dataset partitioned by marker. We compared the model marginal likelihood of each dataset and partition strategy using Bayes Factor comparisons with path sampling and stepping-sampling power posteriors in MrBayes. The selected model was the genome-partitioned nuclear-plastid dataset (3), which was used as the basis of the subsequent dating and biogeographic analyses. We also report the results for the nuclear and plastid-only datasets (1, 2), following the jModelTest results, the concatenated nuclear-plastid dataset was analyzed with substitution models GTR+I+G (nuclear), GTR+G (plastid), and F81 for indels; the plastid-only dataset was analyzed under a GTR+G for each marker partition; and the ITS dataset under a more complex GTR+G+I model. See Appendix S2 for more details. The detailed results of the jModelTest are given in Archive A1 (all archived code is deposited in GitHub, https://github.com/vickycul/Archive-Culshaw-et-al-2021).

Each MrBayes analysis was run for 10,000,000 generations, with sampling every 1000th generation and four chains in two parallel searches. Convergence and effective mixing were assessed by ensuring that the effective sample size (ESS) for each parameter reached 200 in Tracer v 1.7 (Rambaut et al., 2018) and the Potential Scale Reduction Factor (PSRF) approached in MrBayes; differences in topology among runs were assessed through the split frequencies in MrBayes (<0.1). After removing 25% of samples (burn-in), the posterior probability tree distribution was summarized in a 50% majority-rule consensus tree with 95% credibility intervals, and visualized in FigTree v 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree). The script to run the concatenated final analysis is given in Archive A1.

Phylogenomic analysis

Demultiplexed reads were quality-filtered using Trimmomatic (Bolger et al., 2014) to remove adapter sequences. We used the HybPiper pipeline (v 1.0; Johnson et al., 2016) to assemble loci into contigs and identify potential paralog copies. DNA contigs, aligned with MAFFT v 7.222 (Katoh and Standley, 2013). Phylogenetic analyses based on the genomic dataset of eight samples used two different approaches. We used a "supermatrix" approach in which all loci (contigs) were concatenated into a single matrix; this was analyzed using the maximum likelihood software IQ-TREE v 1.4.2 (Nguyen et al., 2015) with the TIM3 + G4 model (selected after an automatic model selection using ModelFinder; Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstrap replicates (UFBoot; Minh et al., 2013; Hoang et al., 2018). Additionally, we used a multispecies-coalescent (MSC) approach using the software ASTRAL-II v 2.4.7.7 (Mirarab and Warnow, 2015) with default parameters. The MSC approach accounts for incongruence among gene trees (genealogies), and between the gene trees and the species tree (phylogeny), due to different coalescent times and incomplete lineage sorting. We first generated trees for each individual marker using the ML method RAXML v 8.2.9 (Stamatakis, 2014) using the GTRCAT model with 200 fast bootstraps, followed by slow ML optimization. We then used the best tree found in the ML search, with bootstrap support values, as input gene trees to estimate a species tree by using quartet amalgamation analysis in ASTRAL-II, which has been shown to be algebraically consistent under the MSC model for moderate levels of ILS (Mirarab and Warnow, 2015). Clade support in the species tree was assessed by estimating local posterior probabilities (LPP) in ASTRAL-II (Sayyari and Mirarab, 2016).

Divergence time estimation

Lineage divergence times were estimated using Bayesian relaxed molecular clock models implemented in BEAST v 1.8.2 (Drummond et al., 2012). Our molecular sampling in the "all specimens" dataset includes both population-level (*Camptoloma*), and species-level sampling (Scrophulariaceae). For such a heterogeneous molecular dataset, a two-step analysis is often used: first inferring the crown age of each *Camptoloma* species using the more inclusive Scrophulariaceae dataset; and second, then using these estimates as calibration points to infer population-level divergence. This might, however, result in large levels of uncertainty (Ho et al., 2005). Instead, we used here the "nested-dating" approach developed by Pokorny et al. (2011) and used by Mairal et al. (2015a) to estimate lineage divergence times across different evolutionary levels (populations, species, and genera).

For the nested-dating approach, we input in BEAST as four unlinked partitions: (1) the species-level "outgroup" dataset representing relationships among genera and tribes in Scrophulariaceae, and species within *Camptoloma*, and (2) the three "population-level" datasets composed of all individuals sequenced within *Camptoloma*. Each of these four unlinked partitions was allowed to evolve under their own molecular model (GTR+G). The "outgroup" partition was assigned a birth-death prior with incomplete taxon

sampling; this is appropriate when dealing with scales of millions of years where extinction is likely to have removed tips from the phylogeny (Stadler, 2013). The three "population-level" partitions were modeled using unlinked constant-size coalescent tree priors, which are appropriate for population-level processes (Ho et al., 2005). In other words, the four molecular partitions were unlinked for both the tree and molecular priors—the only prior that was linked across partitions was the molecular clock. The "outgroup" dataset was calibrated using external evidence (see below), whereas the population-level datasets were not. Because the latter partitions share some samples with the outgroup dataset and are linked through the molecular clock, the molecular clock rate inferred for the species-level dataset can be used to constrain the linked population-level datasets without assuming a single branching process (Mairal et al., 2015a). The results are four different time trees.

Because there are no known fossils of Scrophulariaceae, the "outgroup" dataset was calibrated using secondary age estimates from a fossil-calibrated angiosperm timetree (Magallón et al., 2015). The following nodes were assigned normal distribution priors spanning the 95% high posterior density (HPD) credibility intervals obtained in that study: (1) the stem-node of Scrophulariaceae, i.e., the divergence between Scrophulariaceae and its sister-family Plantaginaceae, here represented by genus *Plantago* (see Appendix S1; mean = 48.6 Ma, 95% HPD = 37.9–62.5); (2) the crown-node of Scrophulariaceae (mean = 42.12 Ma, 95% HPD = 29.5–58.5 Ma); and (3) the split between *Scrophularia* and *Verbascum* (mean = 19 Ma, HPD = 8.1–36.4 Ma). We used an uncorrelated lognormal (UCLN) prior, with *ucld.sd* (Lognormal[mean = R0.8, standard deviation = 0.1], where "R" means "in real space") and *ucld.mean* (Lognormal[R0.002, 0.2]). Mixing and convergence were assessed in Tracer (each analysis was run until ESS reached at least 200). A maximum clade credibility (MCC) tree was constructed in TreeAnnotator v 1.8.2 (Drummond et al., 2012) after removing 25% posterior trees as burn; Appendix S2 provides more details, and the script to run this analysis is given in Archive A2 (in GitHub).

Biogeographic analysis

To reconstruct the biogeographic history of *Camptoloma* in relation to its allies within Scrophulariaceae, we used the Dispersal-Extinction-Cladogenesis model (DEC; Ree and Smith, 2008) implemented in a Bayesian framework in RevBayes (Höhna et al., 2016) as described in Landis et al. (2018). Using Bayesian inference allows for incorporating the uncertainty in biogeographic parameter estimation through the computing of marginal posterior probabilities for ancestral ranges and rates. We ran two analyses of 10,000 generations, with sampling every 10th generation on the MCC tree obtained from the "nested-dating" analysis of the "outgroup" dataset, pruned to include only one individual per species. The dispersal rate was modeled using a lognormal prior between 10^{-4} and 10^{-1} events per million

years, whereas all other priors were set as default in the RevBayes Tutorial (https://revbayes.github.io/tutorials/biogeo/biogeo_simple.html). Cladogenetic range evolution was modeled as resulting from two events with equal probability, using a simplex (1, 1) prior: allopatry (vicariance or peripheral isolate speciation; Ree et al., 2005) for widespread taxa, and subset sympatry for single areas. We did not use the DEC+J model allowing for jump dispersal (j parameter) because of recent criticisms of this model ignoring the contribution of branch length information (Ree and Sanmartín, 2018). The script to run this analysis is provided in Archive A3.

Six operational areas were defined based on the distribution patterns of *Camptoloma* and other Scrophulariaceae with similar distribution: Gran Canaria (GC); Namibia (N); western half of South Africa (W); eastern half of South Africa (E); eastern Africa (EA); and southern Arabia (Oman/Yemen/Socotra/Somalia) (OYSS). A seventh area was included to cover the distribution of the outgroup taxa in the rest of the world, named as "Not in Africa" (NA). This was done because the distribution of outgroup taxa was not well represented in our analysis (see below) and we were mostly interested in migration events within Africa and the nearest "Rand Flora" regions. See Archive A6 in GitHub for area coding in the "outgroup" dataset.

Appendix S4 lists the taxa included in the biogeographic analysis and their geographic distribution, grouped by major geographical regions. Distribution data for species and genera were obtained from online open source databases (e.g., Global Biodiversity Information Facility (GBIF), https://www.gbif.org), online plant guidebooks (e.g., http://southernafricanplants.net and https://species.nbnatlas.org), and herbarium data for our DNA vouchers. For the outgroup taxa, we did not use the distribution of the representative species, but instead coded species for the entire geographic range of the genus (e.g., *Scrophularia arguta* (Soland.); Archive A2, A3). We considered this to be a more conservative approach because sampling was uneven among outgroup genera. Even those represented by several species (e.g., *Myoporum* Banks & Sol. ex G.Forst, *Manulea* L. or *Buddleja* L.) were not sampled across their entire geographic range (see Appendix S4: "Represented distribution in our DNA samples" and "Genus Distribution").

Phylogeographic analysis

To infer the phylogeographic history of the three *Camptoloma* species, we used two different Bayesian MCMC ancestral-state inference methods. In all analyses, we used as input files the "population-level" molecular datasets; these were calibrated with the crown age inferred in the nested-dating analysis; the topology and branch lengths were also fixed to be identical to the MCC trees obtained in the nested analysis. We used a finer-scale geographic definition than in the biogeographic analysis: in this case, the phylogeographic

states represented all sampled populations within each species.

Discrete Trait Analysis (DTA, Lemey et al., 2009), implemented in BEAST v 1.8.2, allows inference of ancestral ranges and rates of migration between geographic locations from DNA sequences and their associated geographic information. The model is analogous to the Bayesian Island Biogeographic model developed by Sanmartín et al. (2008), in which phylogeographic states are treated as discrete traits and migration rates between areas are modeled as rates of substitution between nucleotide states using a continuous-time Markov chain (CTMC) process and Bayesian MCMC inference. For the DTA analysis, we used a symmetric (reversible) instantaneous migration rate matrix (Q), which includes $\frac{(K(K-1))}{2}$ dispersal rate parameters (K = number of phylogeographic states/populations). Identically distributed, independent gamma priors (alpha = 1) were used to model the transition rates between populations, while the overall migration rate was modeled using the default CTMC prior (Ferreira and Suchard, 2008). Other settings included independent substitution models implemented for the plastid (GTR+G) and nuclear (GTR+I) markers; a coalescent constant-size population model for the tree growth prior, and a lognormal uncorrelated relaxed clock for the molecular clock prior, with the same parameter values as in the nested-dating analysis above (mean in real space (M), ucld.mean $M = 0.8$, stdev = 0.1; ucld.sd $M = 0.002$; stdev = 0.3); a relative strict clock model was used for the phylogeographic trait. We used Bayesian stochastic variable selection to identify the migration transition rates in the CTMC matrix receiving nonnegligible support (Lemey et al., 2009). The scripts to run this analysis are provided in Archive A4.

DTA has been criticized because of the unrealistic treatment of the migration-mutation process ("mugration"; De Maio et al., 2015). The effect of migration on the effective population size is not modeled in the likelihood of the coalescent (i.e., the dataset is assumed to belong to a single panmictic (geographically unstructured) population), so marginal posterior probability values on ancestral ranges tend to be overestimated, and the method is highly sensitive to unequal sampling effort among areas (De Maio et al., 2015; Müller et al., 2017). Here, we used an alternative method: the Bayesian structured coalescent approximation (BASTA, De Maio et al., 2015), which implements an approximation to the structured coalescent process to model migration among independent subpopulations, each located in a different geographic state and with their own effective population size. The model is a modification of the Multi-TypeTree (MTT) structured coalescent model of Vaughan et al. (2012), but computationally more efficient in handling a large number of populations and areas because it integrates over individual migration histories as in DTA (De Maio et al., 2015). To compare BASTA inferences with DTA, we ran the model on each Camptoloma population dataset using BEAST v 2.4.7 (Bouckaert et al., 2014). We

implemented a migration-mutation Volz model with symmetric (reversible) transition rates and equal population sizes for all geographic states. For the location rates, we implemented multivariate gamma priors, normalized with the inverse of geographic distance between centroid population coordinates. As in DTA, the topology and branch lengths were fixed to the MCC trees from the nested-dating analysis; the other priors followed the DTA analysis. The scripts to run this analysis are provided in Archive A5.

RESULTS

Phylogenetic inference

Sanger sequencing

Table 1 summarizes some statistics of the markers sequenced for the different datasets. The "all specimens" dataset included 521 sequences and the matrix length was 7945 nucleotide sites; the species-level "outgroup" dataset included 317 sequences and a total matrix length of 7901 sites; the number of sequences in the three "population-level" datasets was 64 (C. canariense), 121 (C. lyperiiflorum), and 66 (C. rotundifolium), respectively, with the total number of sites in each matrix ranging between 6415 and 6555 (Table 1).

Figure 2 shows the consensus tree obtained from the MrBayes analysis based on the "all specimen" concatenated (nuclear-plastid) dataset obtained by Sanger sequencing. The figures in Appendix S5 and S6 show the consensus trees generated from the plastid-only dataset and the nuclear (ITS)-only dataset, respectively. All three trees recovered genus Camptoloma as a monophyletic group (posterior probabilities (PP) =1), sister to tribes Teedieae and Buddlejeae. The monotypic South African genus Phygelius was placed as sister-group to tribe Teedieae (PP = 1). Support was high for many nodes, with most tribes and genera recovered as monophyletic (PP = 1). Tribe Limoselleae (including Manulea, Sutera Roth, Zaluzianskia Benth & Hook., Pseudoselago Hillard, Glumicalyx Hiern, Hebenstreitia L. and Lismosella L.) was recovered as sister to tribe Scrophularieae (Scrophularia L., Verbascum L.). Myoporeae (Myoporum) forms a clade with Leucophylleae (Capraria L., Leucophyllum Bonpl., although the latter is not supported as monophyletic, and Aptosimeae (Peliostomum E. Mey. ex Benth., Aptosimum Burch. ex Benth.) (Figure 2). Within Camptoloma, C. canariense is recovered as sister to the African species C. rotundifolium and C. lyperiiflorum in the concatenate nuclear-plastid tree (Figure 2) and in the ITS tree (Appendix S6), although this relationship receives low clade support: $PP = 0.64$ and $PP = 0.51$, respectively. The plastid-only tree (Appendix S5) supports a different relationship, with C. rotundifolium as sister to C. canariense–C. lyperiiflorum, with $PP = 0.77$. In each of the three trees, the conspecific specimens are grouped together, which supports monophyly for each of the three species (PP = 1, Figure 2).

TABLE 1 Summary statistics for Sanger sequencing of the nuclear (ITS) and seven plastid markers generated from the species-level ("outgroup") dataset, and three "population-level" datasets representing each individual species of *Camptoloma*; see text for more details

"Outgroup" dataset (outgroups + 2 specimens of each *Camptoloma* species)	ITS	*rpl32-ndhF*	*petB-petD*	*psbJ-petA*	*rps16 intron*	*trnL-trnF*	*trnS-trnG*	*trnT-trnL*
No. of sequences	42	44	35	37	41	38	39	41
No. of taxa	24	26	21	21	24	22	23	22
Length matrix alignment (No. of base pairs)	721	785	1058	1256	961	1054	1055	1011
No. of constant sites	417	526	675	772	654	773	636	678
No. of parsimony-informative sites	231	169	159	239	184	170	216	191
Camptoloma canariense								
No. of sequences	8	8	8	8	8	8	8	8
No. of populations	4	4	4	4	4	4	4	4
Length matrix alignment								
(No. of base pairs)	642	623	932	1063	822	926	731	676
No. of constant sites	642	618	932	1061	819	926	729	675
No. of parsimony-informative sites	0	5	0	2	3	0	1	1
Camptoloma lyperiiflorum								
No. of sequences	17	17	15	13	17	13	14	15
No. of populations	6	6	6	6	6	6	6	6
Length matrix alignment								
(No. of base pairs)	650	679	932	1093	824	932	754	691
No. of constant sites	626	649	917	1065	817	927	743	685
Parsimony-informative characters	15	5	6	6	3	1	4	1
Camptoloma rotundifolium								
No. of sequences	10	10	7	10	7	7	7	8
No. of populations	5	5	5	5	5	5	5	5
Length matrix alignment								
(No. of base pairs)	649	615	932	1106	824	927	732	673
No. of constant sites	645	613	932	1082	821	926	727	673
No. of parsimony-informative sites	1	1	0	5	0	0	1	0

Hyb-Seq-phylogenomics

Appendix S7 shows capture success and other summary statistics obtained with AMAS (Borowiec, 2016) for the eight samples analyzed. The raw reads were deposited in GenBank under BioProject PRJNA701536. Capture success was highest in *C. canariense* and lowest for some samples of *C. rotundifolium*. The average number of loci was 1418 (3,272,444 bp). For some samples, HybPiper indicated the presence of paralog copies; these were not included in the final analyses. Figure 2 (inset) shows phylogenetic relationships among species of *Camptoloma* obtained with IQ-TREE, showing Southern African *C. rotundifolium* as sister to the west/east Northern African clade formed by *C. canariense* and *C. lyperiiflorum*, which is similar to the plastid-only Sanger tree (Appendix S5). These relationships received the maximum clade support (BS = 100). ASTRAL-II (1721 individual trees) recovered the identical topology, again with very high support (LPP = 1.0). Conspecific samples were grouped together, supporting monophyly of all three species. All three species in *Camptoloma* were recovered as reciprocal monophyletic.
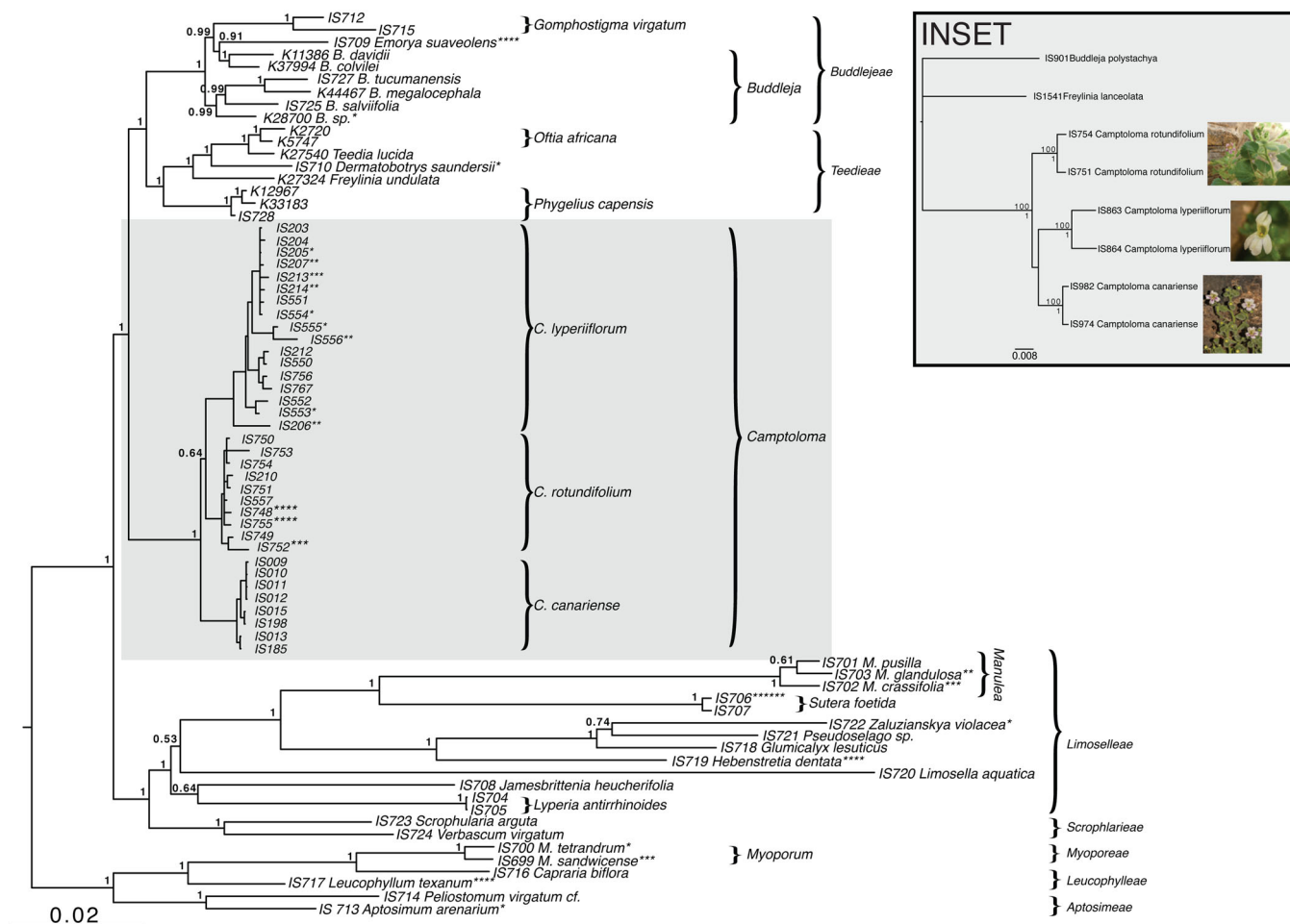
FIGURE 2 Bayesian phylogeny of *Camptoloma* and representative outgroups in Scrophulariaceae. Majority-rule consensus tree obtained by MCMC Bayesian inference in MrBayes using the "all specimens" dataset composed of DNA sequences of nuclear (ITS) and plastid (*trn*L-*trn*F, *trn*S-*trn*G, *rpl*32-*ndh*F, *psb*J-*pet*A, *pet*B-*pet*D, *trn*T-*trn*L and *rps*16 intron) markers. Numbers above branches indicate clade posterior probability (*PP*) values. Symbol * next to a taxon name indicates the number of markers missing for that taxon. Inset: Species-level phylogeny of *Camptoloma* based on 2129 low-copy nuclear genes analyzed with the supermatrix maximum likelihood method IQ-TREE. Nodal values above branches indicate clade Ultrafast bootstrap support from IQ-TREE. Values below branches indicate local posterior probabilities obtained for the same nodes, using the multispecies-coalescent method ASTRAL-II. Because the branch of *Plantago* was so long and uninformative, the tree was pruned to remove its branch. Note that genera *Gomphostigma* and *Emorya* are now synonymized within genus *Buddleja* (Chau et al., 2017). *: Samples used in the Hyb-Seq phylogenomic study, †: Samples used for both Sanger and Hyb-Seq phylogenomic studies

Divergence time estimation

The MCC tree obtained from the "nested-dating" analysis of the "outgroup" dataset shows a topology similar to the MrBayes tree for most tribal and generic relationships, although generally with higher clade support (Figure 3A). The main conflict was the relationship among the three species of *Camptoloma*. Contrary to the MrBayes topology, the BEAST MCC tree (Figure 3A) shows South African *C. rotundifolium* as sister to the northern clade formed by *C. canariense* and *C. lyperiiflorum*; this relationship received moderate support (*PP* = 0.83). Other differences affected the relationships between Myoporeae, Leucophylleae, and Aptosimae.

Estimates of lineage divergence times (Figure 3A) placed the age of stem-node *Camptoloma*, i.e., the divergence of *Camptoloma* from sister tribes Teedieae and Buddlejeae, in the Miocene (21.1 Ma, 95% HPD: 11.52–34.05). The age of crown-node *Camptoloma*, i.e., the divergence of *C. rotundifolium* from the other species, was estimated in the Pliocene, at 4.26 Ma (95% HPD: 1.29–9.09 Ma), while the split between *C. canariense* and *C. lyperiiflorum* was dated as 3.42 Ma (1.05–7.70, Figure 3A). The nested-dated trees from the population-level datasets (Figure 3B) indicate an Early Pleistocene age for the start of population divergence within *C. rotundifolium* (1.85 Ma; 0.61–4.24, Figure 3). The species "crown age" was dated as older, close to the Plio–Pleistocene boundary (3.09 Ma; 1.28–6.11), in the case of *C. lyperiiflorum*. This age was considerably younger (Late Pleistocene) in *C. canariense* (0.76 Ma; 0.23–1.69).
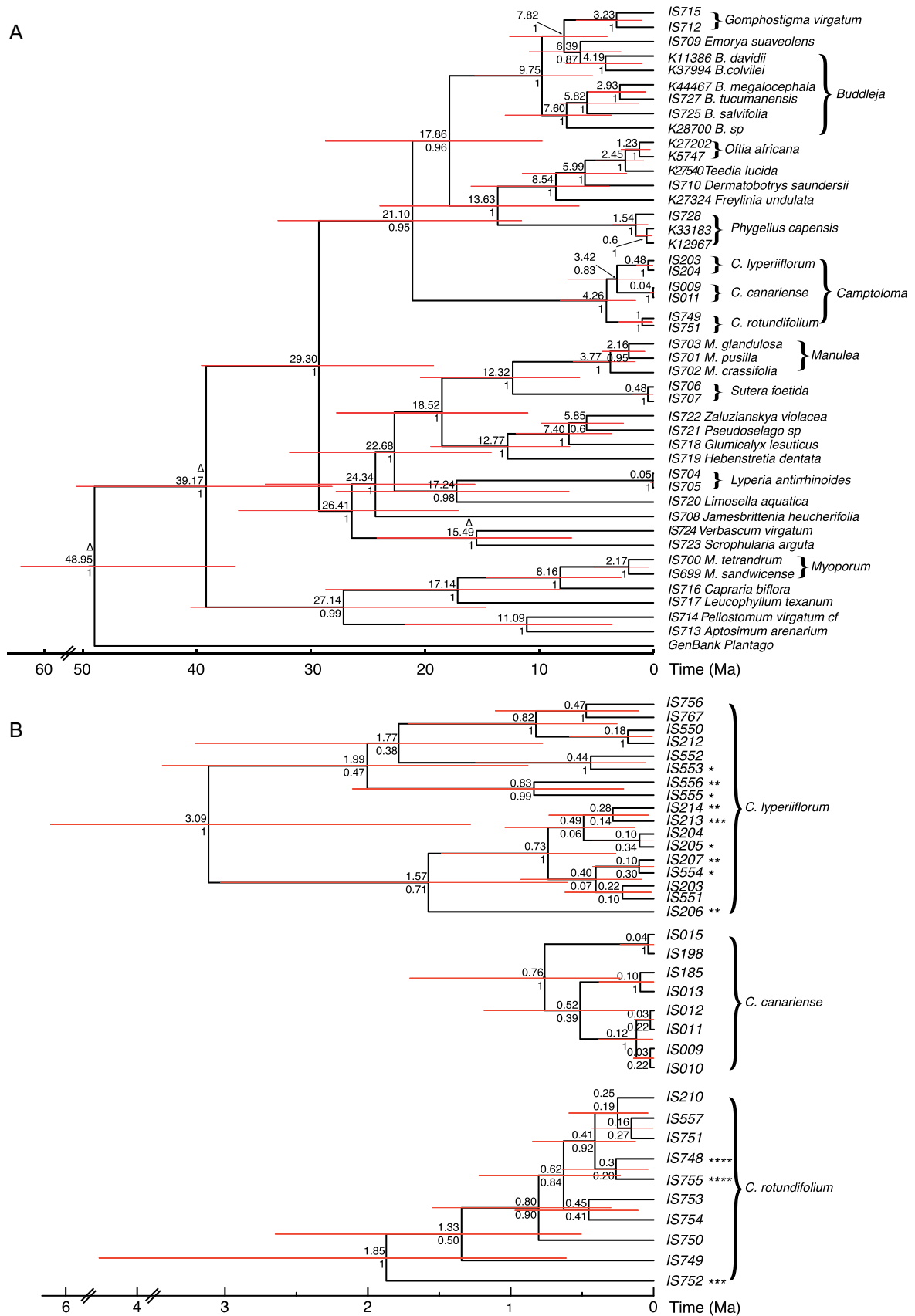
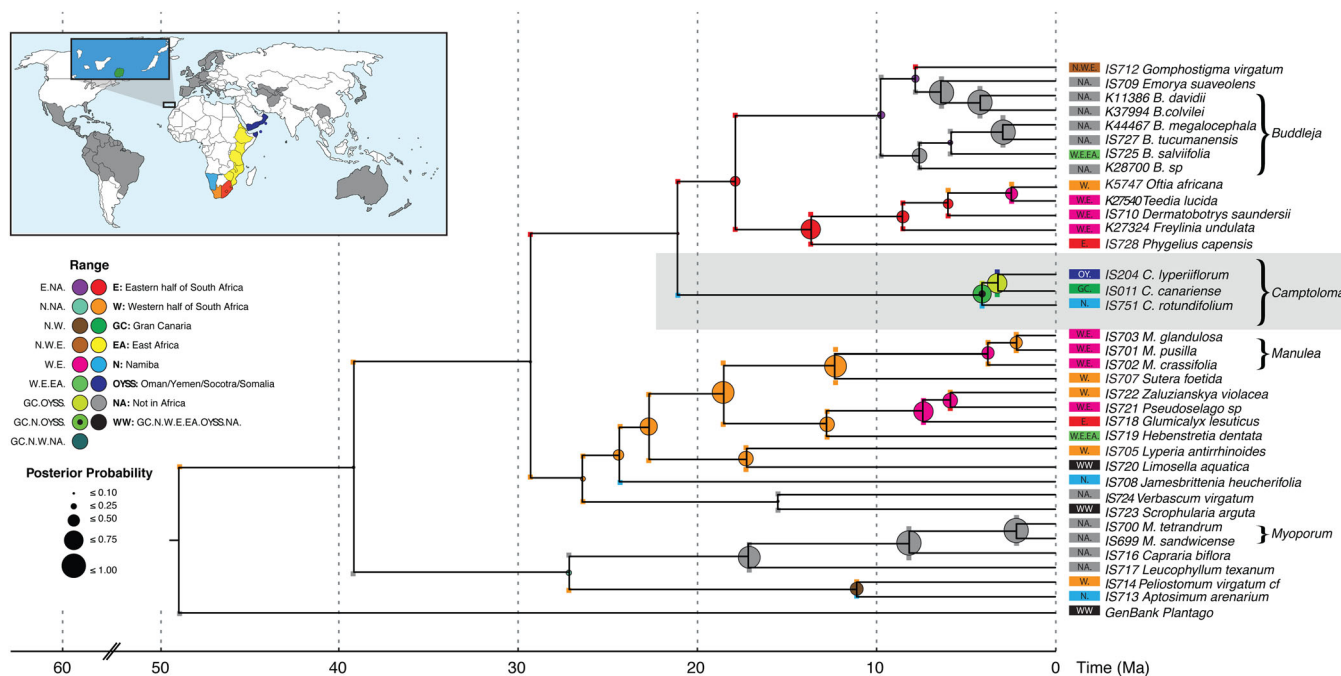**FIGURE 3** (See caption on next page)

FIGURE 4 Bayesian biogeographic reconstruction of *Camptoloma* and representative outgroups in Scrophulariaceae using the Dispersal-Extinction-Cladogenesis (DEC) model implemented in RevBayes. The phylogeny is the MCC tree from the nested-dating analysis of the outgroup dataset (Figure 3A), where the tree was pruned to leave one individual per species. Colored rectangles close to taxon names indicate the current distribution of each taxon (see inset map and legend for color codes); widespread distributions are represented by combining single-area colors, e.g., the pink W.E. is represented as red and orange in the map. Range inheritance scenarios are presented at each node as squares: the size of the circles is proportional to the marginal posterior probability of the inferred ancestral range (see inset legend). Note that genera *Gomphostigma* and *Emorya* are now synonymized within genus *Buddleja* (Chau et al., 2017)

Biogeographic and phylogeographic analyses

The RevBayes-DEC analysis (Figure 4) reconstructed the most recent common ancestor (MRCA) of *Camptoloma* and its sister tribes Teedieae and Buddlejeae as occurring in eastern South Africa. This was followed by migration to the west in Teedieae, and to the west and north in Buddlejeae; the current cosmopolitan distribution of *Buddleja* is explained by dispersal from southern African ancestors. The stem-ancestor of *Camptoloma* was inferred as occurring in Namibia (western South Africa, Figure 4), implying a dispersal event from the east. Sister tribes Limoselleae and Scrophularieae are also of western South African origin. However, inferences for many of these backbone nodes received very low marginal posterior probabilities (<0.01), a result of the limited taxon sampling, long internal branches, and the cosmopolitan distribution of Scrophulariaceae. Crown-node *Camptoloma* was reconstructed as already occupying the three regions where the three extant species occur (>0.75), implying a dispersal event northwest (Gran

Canaria) and eastward (Somalia, Oman, Yemen and Socotra). The current disjunct distribution of the genus is explained by a first vicariance event between southern and northern ancestors, followed by vicariance between west and east across the Sahara (Figure 4).

The DTA analysis returned ancestral location inferences with high marginal posterior probability values (Figure 5A), while BASTA generally recovered much lower marginal probabilities (Figure 5B). For *C. canariense*, results were very similar between the two methods, depicting geographically structured patterns, with adjacent populations clustered together. For *C. lyperiiflorum*, and especially for *C. rotundifolium*, BASTA favored ancestral locations and source areas of migration events in areas underrepresented or occupying a derived position in the tree. For example, Twilfelfonteyn was inferred as the ancestral area for every node in *Camptoloma rotundifolium*, while specimens from Sayhut-Quishn (Yemen) and Socotra Island (Yemen), both occupying more derived ("embedded") positions within the tree, were inferred as the origin of many ancestral nodes in *C. lyperiiflorum* (Figure 5B). Many of these inferences

FIGURE 3 Bayesian timetree of *Camptoloma* and representative outgroups within Scrophulariaceae. (A) Maximum clade credibility (MCC) tree showing lineage divergence times estimated by "nested-dating" in BEAST of the species-level "outgroup" dataset, including Camptoloma and representative genera of tribes within Scrophulariaceae. (B) MCC trees obtained by nested-dating, using the "population-level" datasets for the three species of *Camptoloma*. Numbers above branches indicate mean ages, with 95% HPD credibility intervals represented by the blue bars; those below branches indicate posterior probability values, with Δ showing nodes constrained by secondary time calibrations. Symbol * next to a taxon name indicates the number of markers that failed sequencing for that taxon. Note that genera *Gomphostigma* and *Emorya* are now synonymized within genus *Buddleja* (Chau et al., 2017)
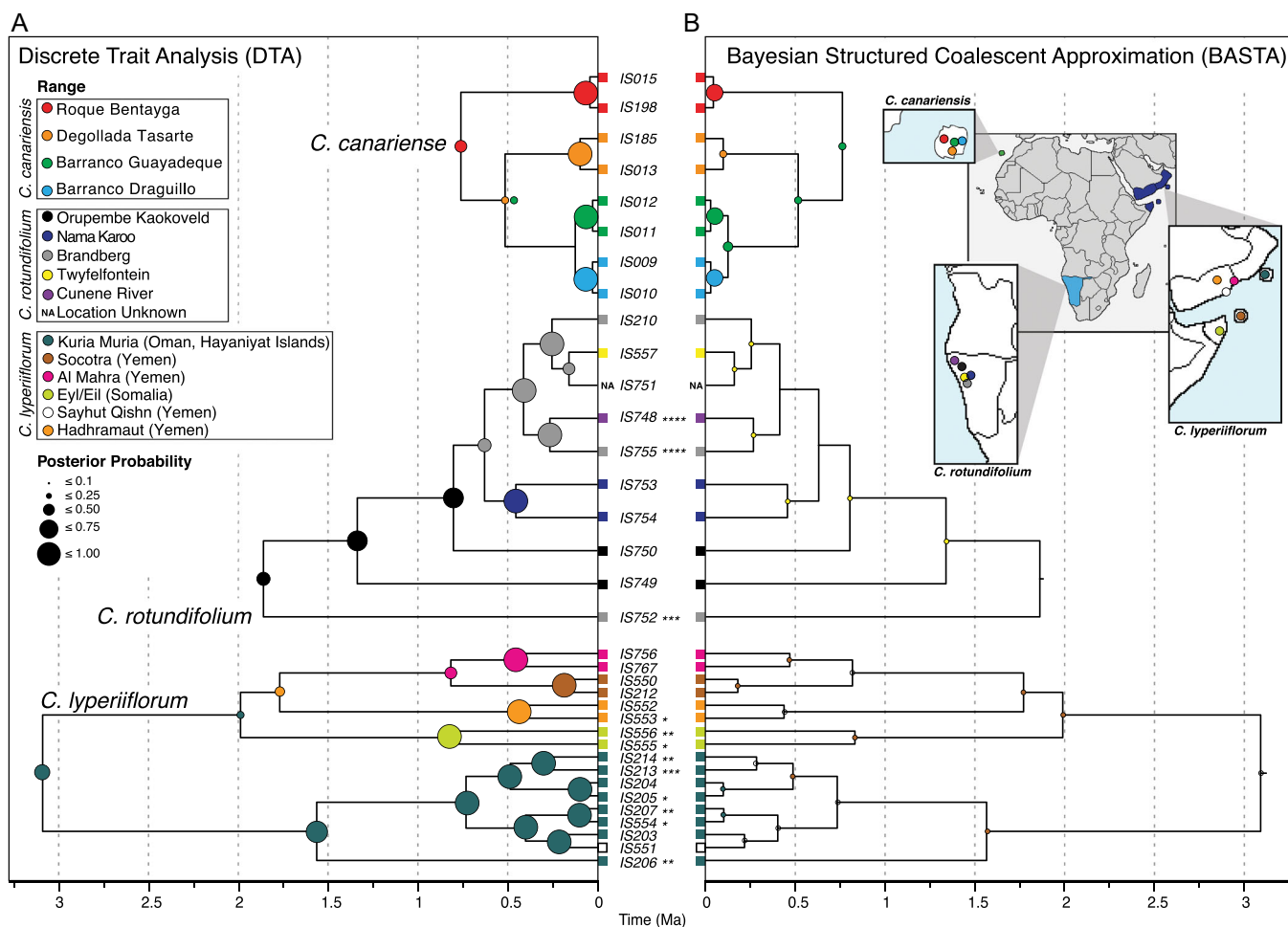
FIGURE 5 Bayesian phylogeographic reconstruction of population history within each *Camptoloma* species, obtained with (A) Discrete trait analysis (DTA), and (B) Bayesian structured coalescent approximation (BASTA). The tree topology and branch lengths were constrained to follow the nested-dating MCC trees shown in Figure 3B. Maps on the right show locations of current populations, with color codes and legends indicated in the left inset legend. Colored squares at tips indicate the distribution of each sample. Circles at nodes depict the inferred ancestral range; the size is proportional to the marginal posterior probability (see left inset legend). Symbol * next to a taxon name indicates the number of markers that are missing for that taxon

were supported by low marginal posterior probabilities (*PP* < 0.1). In contrast, DTA depicted much more conserved geographical patterns, implying a lower frequency of migration events (Figure 5A).

DISCUSSION

Systematics of genus *Camptoloma*

Our molecular dataset, which includes the high-copy nuclear marker ITS, sequenced for the first time in *Camptoloma* (*trnS-trnG, psbJ-petA, petB-petD, trnT-trnL*, and *rpl32-trnL*) provides phylogenetic resolution for the systematic position of the genus within Scrophulariaceae. Although some small differences can be seen at the level of generic relationships, our combined nuclear-chloroplast phylogeny (Figure 2) is in agreement with tribal relationships obtained by Oxelman et al. (2005) using plastid markers. The

individual gene trees show that most phylogenetic resolution for tribal and basal relationships comes from the concatenated plastid dataset (Appendix S6), while ITS provided little resolution. *Camptoloma* is recovered as the sister-group of the mostly South African tribes Teedieae and Buddlejeae, with *Phygelius capensis* as sister to Teedieae, in accordance with Oxelman et al. (2005) but contrary to Pokorny et al. (2015). A recent multilocus phylogenetic study on Buddlejeae (Chau et al., 2017, based on a different set of plastid and nuclear markers, also depicts a sister-group relationship between *Phygelius* E. Mey ex Benth. and Teedieae (*Oftia*). Our results also support the Chau et al. (2017) synonymization of the small South African genera *Gomphostigma* (2 species) and *Emorya* (1 species) with a larger cosmopolitan *Buddleja*. These synonyms were recently confirmed by a target sequencing approach using transcriptomic and genomic resources of *Buddleja* (Chau et al., 2018). The nested-dating timetree (Figure 3) shows similar relationships to the MrBayes tree for tribal

relationships (Figure 2) and is in agreement with Oxelman et al. (2005) plastid phylogeny.

In contrast, the relationship among the three *Camptoloma* species shows conflict among trees generated by different methods and molecular datasets. The MrBayes tree based on the concatenated ITS-nuclear-plastid dataset (Figure 2) failed to retrieve a strongly supported topology for *Camptoloma*, with *C. canariense* as sister to a weakly supported clade ($PP = 0.61$) formed by the African species *C. lyperiiflorum–C. rotundifolium*. This relationship is not found in the Sanger plastid tree (Appendix S6), and it is only weakly supported ($PP = 0.61$) in the ITS phylogeny (Appendix S5). On the other hand, the BEAST timetree (Figure 3A) shows the South African *Camptoloma rotundifolium* as sister to the northern African clade *C. canariense–C. lyperiiflorum*, with moderate support ($PP = 0.83$). We believe this relationship, used as the basis for the biogeographic analysis and the discussion below, is the true one. First, this topology is in agreement with the relationships found by Kornhall et al. (2001) and Oxelman et al. (2005), based on a partly different set of plastid markers. Second, and more important, this relationship is the one recovered by the phylogenomic analysis of 2190 orthologous low-copy nuclear genes spanning ~3.7 million bp (Figure 2 inset). Both the MSC method ASTRAL-II, and the supermatrix method IQ-TREE, support the topology (*C. rotundifolium* (*C. lyperiiflorum–C. canariense*)) with the maximum support (LPP = 1, UFBoot = 100; Figure 2 inset). Given that the only marker supporting the MrBayes relationships is the multicopy marker ITS, we argue that the sister-group relationship of *C. lyperiiflorum* and *C. rotundifolium* is likely an artifact caused by incomplete lineage sorting in ITS. Multispecies-coalescent methods using multiple gene trees such as those implemented in ASTRAL-II have been shown to better account for this source of incongruence compared to those based on single genes or standard concatenation approaches (but see Roch et al., 2019 and Jiang et al., 2020 for a discussion on the accuracy of summary coalescent methods).

Spatiotemporal evolution and the origin of the Rand Flora disjunction

Two main hypotheses have been postulated to explain the Rand Flora disjunction (Sanmartín et al., 2010): (1) vicariance: extant species are the remnants of a more widespread African flora that went partly extinct as a result of aridification from the Miocene onwards (Axelrod and Raven, 1978); (2) dispersal: the species' present disjunct distributions are the result of more recent long-distance dispersal events between already geographically isolated regions, followed by in situ diversification (Thulin, 1994; Francisco-Ortega et al., 1999; Coleman et al., 2003; Galley et al., 2007; Pelser et al., 2012). The current view (based on phylogenetic and ecological niche-modeling evidence) is that Rand Flora disjunctions are not of the same age. Instead, the pattern

was formed over the Neogene, the last 30 million years, through successive aridification waves, which favored climatic extinction and range fragmentation, followed in some cases by dispersal (reconnection) during periods of wet and humid climate (Mairal et al., 2015a, 2017, 2018; Pokorny et al., 2015; Villaverde et al., 2018). In many groups, the eastern/western disjunction across North Africa is estimated to be older than the eastern–southern African disjunction (Pokorny et al., 2015). This is explained by initial vicariance in the north (i.e., due to the Late Miocene formation of the Sahara Desert; Senut et al., 2009), followed by stepping-stone dispersal from eastern to southern Africa along the temperate-climate corridor formed by the uplift of the Eastern Arc Mountains in the Pliocene (Sepulchre et al., 2006; Mairal et al., 2015a; Rincón-Barrado et al., 2021).

Our spatiotemporal reconstruction of the origin of the Rand Flora pattern in *Camptoloma* agrees partly with these hypotheses. It supports the climatic vicariance hypothesis for the currently wide disjunct distribution, but unlike other genera, the southern African endemic is reconstructed as the sister-group of the northern African clade (inset in Figures 2 and 3), pointing to an even older history (Figure 4). The stem-node of *Camptoloma* is dated in the Early Miocene (~21 Ma, Figure 3A), and reconstructed, albeit with high uncertainty, as occurring in southwestern Africa (Figure 4). With the exception of the genus *Buddleja* (the closest relatives of *Camptoloma*), genera in tribes Teedieae and Buddlejeae are all restricted to southern Africa (Appendix S1). Part of the biogeographic uncertainty is caused by the long branch of more than 15 million years separating this ancestor from the most recent common ancestor (MRCA) of the extant species. A period without diversification events in the reconstructed phylogeny may be explained by stasis (low speciation rates) or by high extinction rates removing early-diverging lineages prior to the extant radiation (Sanmartín and Meseguer, 2016; López-Estrada et al., 2019). With only three species, we could not reliably estimate diversification rates in *Camptoloma*. However, palaeobotanical evidence and comparison with diversification rates in the family Scrophulariaceae (Pokorny et al., 2015) suggest that the second explanation is more likely.

During the Neogene, changes in global atmospheric circulation profoundly transformed the climate of the African continent, which could have limited biotic connections between northern and southern Africa (Sanmartín et al., 2010). Following breakup from Gondwana ~100 Ma ago, Africa drifted northeastwards, approaching the Equator. The collision of the Arabian plate with Eurasia in the Early Miocene closed the Tethys Seaway, interrupting the connection between the Atlantic and Indian Oceans. These events, together with the gradual uplift of Eastern Africa, which climaxed in the Miocene–Pliocene, introduced a drier, more arid climate to the continent (Sepulchre et al., 2006; Trauth et al., 2009; Liu et al., 2018). Aridification probably started in the southwest (17–16 Ma), where the current Namib Desert stands, and advanced northeastwards

during the Mid–Late Miocene, culminating with the formation of the northern Sahara Desert (~8–7 Ma, Senut et al., 2009). These aridification events have been linked to high extinction rates in the fossil record in African plants (Plana, 2004; Pan et al., 2006; Denk et al., 2014).

Uplift of the Eastern Arc Mountains started in the Late Miocene, ~7–8 Ma, and reached a maximum during the Pliocene (5–3 Ma), with the formation of semiarid plateaus (the Ethiopian Highlands) and the desertification of low-lying areas in the Horn of Africa (Sepulchre et al., 2006). These mountain ranges could have acted as biogeographic corridors, allowing migration of Rand Flora lineages and other Afrotemperate taxa, between southern and northern Africa (Galley et al., 2007; Popp et al., 2008; Roquet et al., 2009; Barres et al., 2013; Mairal et al., 2015a; Meseguer et al., 2015; Rincón-Barrado et al., 2021). The crown-ancestor of *Camptoloma* (~4.3 Ma, Figure 3A) was inferred to occur in all three regions where the genus is currently present by the Early Pliocene (Figure 4). It is thereby possible that the stem-ancestor of *Camptoloma* was endemic to southern Africa, as its closest relatives, tribes Buddlejeae and Teedieae, are, and that the genus expanded its range eastwards and northwards during the wetter periods of the Late Miocene–Early Pliocene (Beerling et al., 2009; Fedorov et al., 2013), that is, during the time interval spanned by the branch length between the stem and crown nodes (Figure 3A). Afterwards, these ancestors could have gone extinct in part of this range, leaving current taxa in isolated enclaves at the edges of the original distribution: Gran Canaria, Somalia-Yemen-Oman, and southwestern Africa (Figure 4). A similar scenario has been proposed for *Monsonia* (Geraniaceae; García-Aloy et al., 2017), with an origin of the genus in southwestern Africa in the Early Miocene (~21 Ma), followed by northeastward dispersal (4–6 Ma); this was favored by the general cooling trend of the Late Miocene and the uplift of the Eastern Arc Mountains in the Early Pliocene (Sepulchre et al., 2006; García-Aloy et al., 2017). Other authors have linked the distribution of *C. lyperiiflorum* in Somalia-Socotra and *C. rotundifolium* in Namibia with the "Dry Pleistocene Corridor" hypothesis, a pattern linking southwest African taxa with Eastern African–Southern Arabian xeric floristic elements, which is attributed to dispersal during the Pleistocene glacial cycles (Bellstedt et al., 2012). However, the fact that the distribution of the genus does not extend into other deserts, such as the southern Kalahari or the northern Sahara, as it happens in other "Arid Corridor" taxa (*Ceropegia, Zygophyllum, Fagonia*; Bellstedt et al., 2012), and the older Pliocene divergence estimated here for *C. rotundifolium–C. lyperiiflorum* (4.3 Ma, Figure 3A), does not lend support to the Dry Pleistocene Corridor hypothesis.

The Early Pliocene, the period between 5 and 4 Ma, was characterized by a warm, temperate climate, after which temperatures dropped and drier conditions became dominant (Fedorov et al., 2013) During the global climate-warming event of the Mid Pliocene (~3.5 Ma; Zachos et al., 2001), aridification intensified in the southwest Palearctic and led to the onset of the modern Mediterranean climate (Thompson et al., 2005). The Mid Pliocene Warming Event (MPWE) was a consequence of both global events, i.e., the closing of the Panama Isthmus and the interruption of the Central American seaway, which brought colder sea surfaces into the Central Atlantic (Billups et al., 1998), and regional tectonic events, i.e., the uplift of the East Africa Rift System, which created a rain shadow between Central and West Africa and the drier east African plateau (Sepulchre et al., 2006). Mairal et al. (2017), using ecological niche models and paleoclimate projections, predicted that the MPWE would have interrupted connections between northwestern and northeastern populations of Rand Flora taxa. In our reconstruction, the estimated date for the divergence of *C. lyperiiflorum* and *C. canariense* at the edges of the northern African range, is indeed congruent with the MPWE (3.42 Ma), although there is large uncertainty in the age estimates (Figure 3A).

The scenario described above supports the "climatic refugia" hypothesis, according to which the current disjunct distribution of *Camptoloma* across Africa is the result of fragmentation and extinction events associated to a historical aridification process (Sanmartín et al., 2010; Mairal et al., 2015a, 2017; Pokorny et al., 2015; Villaverde et al., 2018; Rincón-Barrado et al., 2021). This scenario does not preclude the possibility of dispersal events. Indeed, some short-to-medium range dispersal scenario is needed to explain the current disjunction, i.e., along the Eastern Arc corridor or across North Africa, as it has been suggested in other Rand Flora taxa (*Canarina*, Mairal et al., 2015a; *Plocama*, Rincón-Barrado et al., 2021). Yet, the deep temporal divergences (Early–Mid Pliocene) among the three *Camptoloma* species observed do not support a recent, long-distance dispersal explanation (Coleman et al., 2003; Désamoré et al., 2010). In other examples of wide-ranged disjunct distributions in Africa (García-Aloy et al., 2017; Pirie et al., 2018), dispersal over long distances was followed by niche release and rapid diversification resulting in large numbers of species. Rand Flora taxa, however, often comprise a low number of species, exhibit small and geographically isolated populations, and long temporal gaps between stem and crown ages (Mairal et al., 2015a, 2018; Pokorny et al., 2015; Rincón-Barrado et al. 2021). All this evidence supports a paleoendemic scenario with historically high extinction rates and population-contraction events. The abovementioned studies were based on Sanger sequencing of a few chloroplast and nuclear loci, which sometimes translated in low levels of resolution; our study, and other recent studies by our team (Villaverde et al., 2018; Riina et al., 2020), show that the use of genomic data can be a complementary approach that can help us clarify patterns that were previously less specific.

Population-level phylogeographic history

Our results indicate that the start of population divergence in *C. canariense* occurred much later (0.76 Ma; Figure 3B) than in the African species *C. lyperiiflorum* and *C. rotundifolium* (3.09–1.85 Ma). A similar pattern has been found in other Rand

Flora taxa, such as *Canarina* (Mairal et al., 2015a), the *Euphorbia balsamifera* complex (Villaverde et al., 2018), or *Plocama* (Rincón-Barrado et al., 2021). The divergence of *C. canariense* from *C. lyperiiflorum* is dated around the MPWE (3.42 Ma), with the establishment of Mediterranean seasonality (yearly summer droughts) on the Canaries. It is possible that *C. canariense* had originally a wider distribution across the Archipelago and that its current restricted range in Gran Canaria is the result of contraction due to aridification: the eastern Canary Islands, which are dated as old as the Early Miocene (~20 Ma) and were moister in the past, may have harbored early-diverging populations of *C. canariense* that went extinct in more recent times. Conversely, the Late Pleistocene estimate of the crown age of *C. canariense* may indicate that the Macaronesian component of the Rand Flora is of recent origin. Mairal et al. (2015a) hypothesized that Rand Flora Canarian endemics are the result of dispersal events from a northwestern African population during Pleistocene glacial cycles, which later went extinct leaving a spatial "gap" between the northwestern and eastern Rand Flora species.

Regarding patterns of phylogeographic variation, DTA and BASTA gave conflicting results. DTA supported a geographically conserved population structure in all three species of *Camptoloma* (Figure 5A), while BASTA inferred a pattern with frequent dispersal events for *C. rotundifolium* and *C. lyperiiflorum*, often having as source the least represented (most undersampled) population (Figure 5B), and with low clade support ($PP < 0.01$). The high clade support for inferred ancestral ranges in DTA agrees with the overconfidence bias reported by De Maio et al. (2015). Müller et al. (2017) provided an explanation on the bias observed in BASTA ancestral ranges and general lack of clade support. In the BASTA model, probabilities of a lineage being in a certain state (geographic location) along the branch between migration and branching events are independent of the coalescent process (time, mutation rate). If there is asymmetry in coalescent rates between states, for example, because one location is undersampled relative to the other, and subtended by a shorter branch (i.e., embedded within a younger clade), the "derived" undersampled state will be assigned a lower migration rate "out" and a higher migration rate "into"; the opposite occurs for the more frequently sampled "basally-diverging" state. The result is that the derived undersampled state becomes an absorbing state and is reconstructed as the ancestral range for the root and majority of nodes in the phylogeny.

This is exactly the case for Twyfelfontein and Sayhut-Quishn in *C. rotundifolium* and *C. lyperiiflorum,* respectively, which are reconstructed as the ancestral range for most nodes in the phylogeny and are the least represented locations, occupying derived ("embedded") positions within the tree (Figure 5B). In *C. canariense*, each location is roughly represented by the same number of individuals, so this bias is not observed. The bias is stronger under an overall migration rate that is much lower than the coalescent rate, as is often the case in nonviral phylogenies, and under strongly asymmetric coalescent rates. The shape of the tree has also an influence. In balanced trees such as

C. canariense, all terminal branches are similar in length, whereas in unbalanced trees (*C. lyperiiflorum*, *C. rotundifolium*), differences among branch lengths become larger (Figure 5B). In fact, simulations in our lab indicate that under low overall migration rate and with a highly unbalanced tree, BASTA exhibits anomalous bimodal Bayesian posterior distributions (Cornuault and Sanmartín, 2019). Notice that the aberrant pattern is more apparent for *C. rotundifolium* than *C. lyperiiflorum*; the latter shows a slightly more balanced tree and therefore some geographical structuring of genetic patterns, for example, in the case of Socotra and Kuria-Muria (Figure 5B).

In short, although DTA is known to overestimate clade support for ancestral ranges (and is sensitive to uneven sampling; De Maio et al., 2015), we think results from this analysis, depicting geographically structured populations, are more reliable than those from BASTA. Moreover, despite the large differences, both methods agree on some interesting results. Within *C. canariense*, the inference of dispersal events from populations located in central (BASTA, Figure 5B) or western (DTA, Figure 5A) Gran Canaria agree with the idea of long-persistence of plant lineages in these locations, which have a recent history of geological stability (Emerson 2003; Mairal et al., 2015b). In *C. lyperiiflorum*, long-term persistence has been argued for southern Arabian island systems such as Al-Hallaniyah (Oman) or Socotra (Yemen), which are inferred as the source areas of dispersal events (Figure 5A, B); these areas probably maintained a milder climate than surrounding arid regions during the Pleistocene glacial cycles (Domina et al., 2012).

Rabinowitz (1981) defined "narrow and rare" endemic species as those displaying low genetic diversity and geographically restricted distributions, stemming from historically high extinction rates (Kruckeberg and Rabinowitz, 1985). Currently, the three species of *Camptoloma* occupy milder climate "microrefugia" within a harsher, more xeric geographical template, i.e., shaded vertical crevices along the western and southern coastal cliffs of Gran Canaria in *C. canariense* (Naranjo Suárez and Adenda Santana López, 2006); rocky crevices in the Brandberg mountain range along the Namibian coast in *C. rotundifolium* (Craven and Craven, 2000), or areas of high humidity and lower seasonality in the Socotra Archipelago in *C. lyperiiflorum* (Domina et al., 2012). *Camptoloma canariense* is a protected endemism included in the Spanish and EU catalog/Red List of species under special protection ("VU loc"): there are 13 populations registered, but each population includes just a few individuals (Naranjo Suárez and Adenda Santana López, 2006). The species is also included as a "trigger species" in 10 key priority biodiversity areas in Gran Canaria, which should receive special protection under the European Union's Regional Ecosystem profile for the Macaronesian Region (Bañares et al., 2010). No such protection figure exists for *C. rotundifolium* or *C. lyperiiflorum*. Yet, the small population sizes, narrow geographic distributions, and high habitat specificity of these endemics suggest that they are vulnerable to stochastic demographic fluctuations such

as population bottlenecks and effective population size declines, resulting from anthropogenic activities; this has been inferred in other Rand Flora taxa (Mairal et al., 2018). *Camptoloma*, however, is unique in that species are narrow endemics but the genus as a whole fits the definition of "rare and widespread" ("large and narrow rarity;" Rabinowitz, 1981); it is a genus with a continental-wide distribution but which is composed of a sparse number of species, geographically isolated by thousands of kilometers. This makes the genus of even higher conservation value.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

I.S. designed the study; M.M. and V.C. collected samples; V.C., with help from S.O., performed the molecular work; T.V. carried out the phylogenomic analyses; V.C. and I.S. performed all other analytical steps; I.S. and V.C. wrote the manuscript, with major contributions from M.M.; all authors revised and approved the final draft.

DATA AVAILABILITY STATEMENT

All archived scripted code from this study is deposited in the public repository GitHub at https://github.com/vickycul/Archieve-Culshaw-et-al-2021.

ORCID

Victoria Culshaw http://orcid.org/0000-0002-3228-4322
Tamara Villaverde https://orcid.org/0000-0002-9236-8616
Mario Mairal https://orcid.org/0000-0002-6588-5634
Sanna Olsson https://orcid.org/0000-0002-1199-4499
Isabel Sanmartín https://orcid.org/0000-0001-6104-9658

REFERENCES

Axelrod, D. I., and P. H. Raven. 1978. Late Cretaceous and Tertiary vegetation history of Africa. In: M. J. A. Werger (ed), Biogeography and Ecology of Southern Africa, pp. 77–130. Springer, the Netherlands.

Backlund, M., B. Bremer, and M. Thulin. 2007. Paraphyly of Paederieae, recognition of Putorieae and expansion of Plocama (Rubiaceae-Rubioideae). *Taxon* 56: 315–328. https://www.jstor.org/stable/25065790

Bañares, A., G. Blanca, J. Güemes, J. C. Moreno, and S. Ortiz. 2010. Atlas y libro rojo de la flora vascular amenazada de España. Adenda 2010. Dirección General de Medio Natural y Política Forestal (Ministerio de Medio Ambiente, y Medio Rural y Marino)-Sociedad Española de Biología de la Conservación de Plantas, 170. http://www.jolube.es/pdf/afa_index.htm

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

Beerling, D. J., R. A. Berner, F. T. Mackenzie, M. B. Harfoot, and J. A. Pyle. 2009. Methane and the CH_4 related greenhouse effect over the past 400 million years. *American Journal of Science* 309: 97–113. https://www.ajsonline.org/content/309/2/97.abstract

Bellstedt, D. U., C. Galley, M. D. Pirie, and H. P. Linder. 2012. The migration of the palaeotropical arid flora: zygophylloideae as an example. *Systematic Botany* 37: 951–959.

Berdugo, M., M. Delgado-Baquerizo, S. Soliveres, R. Hernández-Clemente, Y. Zhao, J. J. Gaitán, N. Gross, et al. 2020. Increasing aridity promotes sequential, systemic, and abrupt thresholds in drylands worldwide. *Science* 367: 787–790.

Billups, K., A. C. Ravelo, and J. C. Zachos. 1998. Early Pliocene climate: A perspective from the western equatorial Atlantic warm pool. *Paleoceanography* 13: 459–470.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Borowiec, M. L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, et al. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Bramwell, D. 1985. Contribucion a la biogeografia de las islas canarias. *Botinca Macaronesica* 14: 3–34.

Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

Chan, A. P., J. Crabtree, Q. Zhao, H. Lorenzi, J. Orvis, D. Puiu, A. Melake-Berhan, et al. 2010. Draft genome sequence of the oilseed species Ricinus communis. *Nature Biotechnology* 28: 951–956.

Chau, J. H., N. O'Leary, W.-B. Sun, and R. Olmstead. 2017. Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers. *Botanical Journal of the Linnean Society* 184: 137–166. https://academic.oup.com/botlinnean/article/184/2/137/3865471

Chau, T. H., W. A. Rahfeldt, and R. G. Olmstead. 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences* 6: e1032.

Christ, H. 1892. Exposé sur le rôle que joue dans le domaine de nos flores la flore dite ancienne africaine. *Journal of Archaeological Science* 3: 369–374.

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

Cornuault, J., and I. Sanmartín. Comparative phylogeography: the origin of variation in dispersal patterns. International Biogeographic Society Meeting, Malaga. January 8-12, 2019.

Craven, P., and D. Craven. 2000. The flora of the Brandberg, Namibia. *Cimbebasia Memoir* 9: 49–67.

Crisp, M. D., and L. G. Cook. 2007. A congruent molecular signature of vicariance across multiple plant lineages. *Molecular Phylogenetics and Evolution* 43: 1106–1117.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

De Maio, N., C.-H. Wu, K. M. O'Reilly, and D. Wilson. 2015. New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genetics* 11: e1005421.

Denk, T., H. T. Güner, and G. W. Grimm. 2014. From mesic to arid: leaf epidermal features suggest preadaptation in Miocene dragon trees (Dracaena). *Review of Palaeobotany and Palynology* 200: 211–228.

Désamoré, A., B. Laenen, N. Devos, M. Popp, J. M. González-Mancebo, M. A. Carine, and A. Vanderoorten. 2010. Out of Africa: north-westwards Pleistocene expansions of the heather Erica arborea. *Journal of Biogeography* 38: 164–176. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2699.2010.02387.x

Domina, G., P. Marino, V. Spadaro, and F. M. Raimondo. 2012. Vascular flora evolution in the major Mediterranean islands. *Biodiversity Journal* 3: 337–342.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Emerson, B. C. 2003. Genes, geology and biodiversity: faunal and floral diversity on the island of Gran Canaria. *Animal Biodiversity and Conservation* 26: 9–20. http://abc.museucienciesjournals.cat/volume-26-1-2003-abc/genes-geology-and-biodiversity-faunal-and-floral-diversity-on-the-island-of-gran-canaria-2/?lang=en

Fedorov, A. V., C. M. Brierley, K. T. Lawrence, Z. Liu, P. S. Dekens, and A. C. Ravelo. 2013. Patterns and mechanisms of early Pliocene warmth. *Nature* 496: 43–49.

Ferreira, M. A. R., and M. A. Suchard. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics* 36: 355–368.

Francisco-Ortega, J., L. Goertzen, A. Santos-Guerra, A. Benabid, and R. Jansen. 1999. Molecular systematics of the Asteriscus Alliance (Asteraceae: Inuleae) I: Evidence from the Internal Transcribed Spacers of Nuclear Ribosomal DNA. *Systematic Botany* 24: 249–266. https://www.jstor.org/stable/2419551

Galley, C., B. Bytebier, D. U. Bellstedt, and H. P. Linder. 2007. The Cape element in the Afrotemperate flora: from Cape to Cairo? *Proceedings of the Royal Society B: Biological Sciences* 274: 535–543.

García-Aloy, S., I. Sanmartín, G. Kadereit, D. Vitales, A. M. Millanes, C. Roquet, P. Vargas, et al. 2017. Opposite trends in the genus Monsonia (Geraniaceae): specialization in the African deserts and range expansions throughout Eastern Africa. *Scientific Reports* 7: 9872.

Ho, S. Y., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic over-estimation of recent divergence times. *Molecular Biology and Evolution* 22: 1561–1568. https://pubmed.ncbi.nlm.nih.gov/15814826/

Hoang, D. T., A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65: 726–736.

Jiang, X., S. V. Edwards, and L. Liu. 2020. The multispecies coalescent model outperforms concatenation across diverse phylogenomic datasets. *Systematic Biology* 69: 795–812.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

Katoh, K., and D. M. Standley. 2013. MAFFT: multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780. https://academic.oup.com/mbe/article/30/4/772/1073398

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

Kruckeberg, A. R., and D. Rabinowitz. 1985. Biological aspects of endemism in higher plants. *Annual Review of Ecology and Systematics* 16: 447–479.

Landis, J. B., D. E. Soltis, Z. Li, H. E. Marx, M. S. Barker, D. C. Tank, and P. S. Soltis. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* 105: 348–363.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. *PLoS Computational Biology* 5: e1000520.

Liu, H., S. Li, A. Ugolini, F. Momtazi, and Z. Hou. 2018. Tethyan closure drove tropical marine biodiversity: Vicariant diversification of intertidal crustaceans. *Journal of Biogeography* 45: 941–951.

López-Estrada, E. K., I. Sanmartín, M. García-París, and A. Zaldívar-Riverón. 2019. High extinction rates and non-adaptive radiation explains patterns of low diversity and extreme morphological disparity in North American blister beetles (Coleoptera, Meloidae). *Molecular Phylogenetics and Evolution* 130: 156–168.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207: 437–453.

Mairal, M., J. Caujapé-Castells, L. Pellissier, R. Jaén-Molina, N. Álvarez, M. Heuertz, and I. Sanmartín. 2018. A tale of two forests: ongoing aridification drives population decline and genetic diversity loss at continental scale in Afro-Macaronesian evergreen-forest archipelago endemics. *Annals of Botany* 122: 1005–1017.

Mairal, M., L. Pokorny, J. J. Aldasoro, M. Alarcón, and I. Sanmartín. 2015a. Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus Canarina (Campanulaceae). *Molecular Evolution* 24: 1335–1354. https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13114

Mairal, M., I. Sanmartín, J. J. Aldasoro, V. Culshaw, I. Manolopoulou, and M. Alarcón. 2015b. Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic Canarina canariensis (Campanulaceae). *Molecular Ecology* 24: 3944–3963. https://onlinelibrary.wiley.com/doi/10.1111/mec.13282

Mairal, M., I. Sanmartín, and L. Pellissier. 2017. Lineage-specific climatic niche drives the tempo of vicariance in the Rand Flora. *Journal of Biogeography* 44: 911–923.

Matasci, N., L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3. http://doi.org/10.1186/2047-217x-3-17

Meseguer, A. S., J. M. Lobo, J. Cornuault, D. Beerling, B. R. Ruhfel, C. C. Davis, E. Jousselin, and I. Sanmartín. 2018. Reconstructing deep-time palaeoclimate legacies in the clusioid Malpighiales unveils their role in the evolution and extinction of the boreotropical flora. *Global Ecology and Biogeography* 27: 616–628.

Meseguer, A. S., J. M. Lobo, R. Ree, D. J. Beerling, and I. Sanmartín. 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: The case of Hypericum (Hypericaceae). *Systematic Biology* 64: 215–232.

Minh, B. Q., M. A. T. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30: 1188–1195.

Mirarab, S., and T. Warnow. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44–i52.

Morrison, D. A. 2006. Multiple sequence alignment for phylogenetic purposes. *Australian Systematic Botany* 19: 479–539. https://www.publish.csiro.au/SB/SB06020

Müller, K. 2005. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4: 65–69.

Müller, N. F., D. A. Rasmussen, and T. Stadler. 2017. The structured coalescent and its approximations. *Molecular Biology and Evolution* 34: 2970–2981.

Naranjo Suárez, J., and I. Adenda Santana López. 2006. VU Scrophulariaceae Camptoloma canariensis (Webb & Berthel.) Hilliard. In: Bañares, Á., G. Blanca, J. Güemes, J. C. Moreno, and

S. Ortiz. (eds.). Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Adenda 2006. Dirección General para la Biodiversidad-Sociedad Española de Biología de la Conservación de Plantas. Madrid, Spain: 62–63.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Oxelman, B., P. Kornhall, R. G. Olmstead, and B. Bremer. 2005. Further disintegration of Scrophulariaceae. *Taxon* 54: 411–425.

Pan, A. D., B. F. Jacobs, J. Dransfield, and W. J. Baker. 2006. The fossil history of palms (Arecaceae) in Africa and new records from the Late Oligocene (28–27 mya) of north-western Ethiopia. *Botanical Journal of the Linnean Society* 151: 69–81. https://onlinelibrary.wiley.com/doi/10.1111/j.1095-8339.2006.00523.x

Park, C.-E., S.-J. Jeong, M. Joshi, T. J. Osborn, C.-H. Ho, S. Piao, D. Chen, et al. 2018. Keeping global warming within 1.5°C constrains emergence of aridification. *Nature Climate Change* 8: 70–74. https://www.nature.com/articles/s41558-017-0034-4

Pelser, P. B., R. J. Abbott, H. P. Comes, J. J. Milton, M. Möller, M. E. Looseley, G. V. Cron, et al. 2012. The genetic ghost of an invasion past: colonization and extinction revealed by historical hybridization in Senecio. *Molecular Ecology* 21: 369–387.

Pelser, P. B., B. Nordenstam, J. W. Kadereit, and L. E. Watson. 2007. An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of Senecio L. *Taxon* 56: 1077–1104. https://www.jstor.org/stable/25065905

Pirie, M., M. D. Kandziora, N. M. Nürk, N. C. Le Maitre, A. L. Mugrabi de Kuppler, B. Gehrke, E. G. H. Oliver, and D. U. Bellstedt. 2018. Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by Erica. *BMC Ecology and Evolution* 19: 222. https://bmcecolevol.biomedcentral.com/articles/10.1186/s12862-019-1545-6

Plana, V. 2004. Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. *The Philosophical Transactions of the Royal Society of London. Series B* 359: 1585–1594. https://royalsocietypublishing.org/doi/10.1098/rstb.2004.1535

Pokorny, L., G. Oliván, and A. J. Shaw. 2011. Phylogeographic patterns in two Southern Hemisphere species of Calyptrochaeta (Daltoniaceae, Bryophyta). *Systematic Botany* 36: 542–553.

Pokorny, L., R. Riina, M. Mairal, A. S. Meseguer, V. Culshaw, J. Cendoya, M. Serrano, et al. 2015. Living of the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. *Frontiers in Genetics* 6: 154.

Popp, M., A. Gizaw, S. Nemomissa, J. Suda, and C. Brochmann. 2008. Colonization and diversification in the African 'sky islands' by Eurasian Lychnis L. (Caryophyllaceae). *Journal of Biogeography* 35: 1016–1029. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2699.2008.01902.x

Rabinowitz, D. 1981. Seven forms of rarity. In: Synge, H. (ed). The biological aspects of rare plant conservation, John Wiley & Sons, Chichester, United Kingdom, 205–217.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59: 2299–2311.

Ree, R. H., and I. Sanmartín. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* 45: 741–749. https://onlinelibrary.wiley.com/doi/10.1111/jbi.13173

Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

Riina, R., T. Villaverde, M. Rincón-Barrado, J. Molero, and I. Sanmartín. 2020. More than one sweet tabaiba: disentangling the systematics of the succulent dendroid shrub Euphorbia balsamifera. *Journal of Systematics and Evolution* 59: 490–503. https://onlinelibrary.wiley.com/doi/abs/10.1111/jse.12656

Rincón-Barrado, M., S. Olsson, T. Villaverde, B. Moncalvillo, L. Pokorny, Forrest L., et al. 2021. Ecological and geological processes impacting speciation modes drive the formation of wide-range disjunctions in tribe Putorieae (Rubiaceae). *Journal of Systematics and Evolution*. https://doi.org/10.1111/jse.12747

Roch, S., M. Nute, and T. Warnow. 2019. Long-branch attraction in species tree estimation: inconsistency of partitioned likelihood and topology-based summary methods. *Systematic Biology* 68: 281–297.

Ronquist, F., M. Teslenko, P. van Der Mark, D. L. Ayres, L. Darling, S. Höhna, B. Larget, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.

Roquet, C., I. Sanmartín, N. Garcia-Jacas, L. Sáez, A. Susanna, N. Wikström, and J. J. Aldasoro. 2009. Reconstructing the history of Campanulaceae with a Bayesian approach to molecular dating and dispersal–vicariance analyses. *Molecular Phylogenetics and Evolution* 52: 575–587.

Sanmartín, I., C. L. Anderson, M. Alarcón, F. Ronquist, and J. J. Aldasoro. 2010. Bayesian island biogeography in a continental setting: The Rand Flora case. *Biology Letters* 6: 703–707.

Sanmartín, I., and A. S. Meseguer. 2016. Extinction in phylogenetics and biogeography: from timetrees to patterns of biotic assemblage. *Frontiers in Genetics* 7: 35.

Sanmartín, I., P. van Der Mark, and F. Ronquist. 2008. Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography* 35: 428–449. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2699.2008.01885.x

Sayyari, E., and S. Mirarab. 2016. Anchoring quartet-based phylogenetic distances and applications to species tree reconstruction. *BMC Genomics* 17: 101–113.

Senut, B., M. Pickford, and L. Ségalen. 2009. Neogene desertification of Africa. *Comptes Rendus Geoscience* 341: 591–602.

Sepulchre, P., G. Ramstein, F. Fluteau, M. Schuster, J. J. Tiercelin, and M. Brunet. 2006. Tectonic uplift and Eastern Africa aridification. *Science* 313: 1419–1423.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analysis. *Systematic Biology* 49: 369–381.

Stadler, T. 2013. How can we improve accuracy of macroevolutionary rate estimates? *Systematic Biology* 62: 321–329.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post- analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. https://academic.oup.com/bioinformatics/article/30/9/1312/238053

Svenning, J.-C., W. L. Eiserhardt, S. Normand, A. Ordonez, and B. Sandel. 2015. The influence of palaeoclimate on present-day patterns in biodiversity and ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 46: 551–572. https://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-112414-054314

Thompson, K., A. P. Askew, J. P. Grime, N. P. Dunnett, and A. J. Willis. 2005. Biodiversity, ecosystem function and plant traits in mature and immature plant communities. *Functional Ecology* 19: 355–358. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.0269-8463.2005.00936.x

Thulin, M. 1994. Aspects of disjunct distributions and endemism in the arid parts of the Horn of Africa, particularly Somalia. In: J. H. Seyani and A. C. Chikuni (eds), Proceedings of the XIIIth plenary meeting of AETFAT, Zomba, Malawi. 2, National Herbarium and Botanic Gardens of Malawi, Zomba. pp. 1109–1115.

Trauth, M. H., J. C. Larrasoaña, and M. Mudelsee. 2009. Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews* 28: 399–411.

Vaughan, T., P. Drummond, and A. Drummond. 2012. Within-host demographic fluctuations and correlations in early retroviral infection. *Journal of Theoretical Biology* 295: 86–99.

Villaverde, T., L. Pokorny, S. Olsson, M. Rincón-Barrado, M. G. Johnson, E. M. Gardner, N. J. Wickett, et al. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.

Viruel, J., M. B. Kantar, R. Gargiulo, P. Hesketh-Prichard, N. Leong, C. Cockel, F. Forest, et al. 2020. Crop wild phylorelatives (CWPs): phylogenetic distance, cytogenetic compatibility and breeding system data enable estimation of crop wild relative gene pool classification. *Botanical Journal of the Linnean Society* 195: 1–33. https://academic.oup.com/botlinnean/article/195/1/1/5903667

Weitemier, K., S. C. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: apps.1400042.

Willis, K. J., and G. M. MacDonald. 2011. Long-term ecological records and their relevance to climate change predictions for a warmer world. *Annual Review of Ecology, Evolution, and Systematics* 42: 267–287. https://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-102209-144704

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Bullups. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to Present. *Science* 292: 686–693.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Appendix S1. Taxon names, location, voucher information, and GenBank accession numbers for all taxa included in this study. *: Samples used in the Hyb-Seq phylogenomic study, †: Samples used for both Sanger and Hyb-Seq phylogenomic studies.

Appendix S2. Extended Materials and Methods. This section provides details on Materials and Methods not included in the main article.

Appendix S3. Primers and specific protocols used for PCR amplification of markers.

Appendix S4. Geographic distribution of all Scrophulariaceae genera included in this study. Outgroup genera were represented by one or several species, whose distribution may not cover the entire distribution of the genus (e.g. *Buddleja*, *Myoporum*).

Appendix S5. Bayesian phylogeny of *Camptoloma* and representative outgroups in Scrophulariaceae. Majority-rule consensus tree obtained by MCMC Bayesian inference in MrBayes using the "all specimens" dataset composed of DNA sequences of plastid-only (trnL-trnF, trnS-trnG, rpl32-ndhF, psbJ-petA, petB-petD, trnT-trnL, and rps16 intron) markers. Numbers above branches indicate clade posterior probability (PP) values. Note that genera *Gomphostigma* and *Emorya* are now synonymized within genus *Buddleja* (Chau et al., 2017).

Appendix S6. Bayesian phylogeny of *Camptoloma* and representative outgroups in Scrophulariaceae. Majority-rule consensus tree obtained by MCMC Bayesian inference in MrBayes using the "all specimens" dataset comprising DNA sequences of the nuclear (ITS) marker. Numbers above branches indicate clade posterior probability (PP) values. Note that genera *Gomphostigma* and *Emorya* are now synonymized within genus *Buddleja* (Chau et al., 2017).

Appendix S7. Summary statistics for the phylogenomic dataset generated with Hyb-Seq for *Camptoloma*. Information on vouchers is given in Appendix S1. The initial target was 2190 low-copy nuclear genes, covering ~3.7 million base pairs. The table shows several statistics depicting capture success, as well as a number of paralogs detected by HybPiper. Number of genes 25% target: number of genes reaching 25% target length.