

# Phylogenomics sheds new light on the drivers behind a long-lasting systematic riddle: the figwort family Scrophulariaceae

Tamara Villaverde<sup>1,8</sup> , Isabel Larridon<sup>2,3</sup>, Toral Shah<sup>2,4</sup>, Rachael M. Fowler<sup>5</sup>, John H. Chau<sup>6</sup>, Richard G. Olmstead<sup>7</sup> and Isabel Sanmartín<sup>1</sup> 

<sup>1</sup>Real Jardín Botánico (CSIC), Plaza de Murillo, 2, Madrid 28014, Spain; <sup>2</sup>Royal Botanic Gardens, Kew, Richmond, TW9 3AE, UK; <sup>3</sup>Systematic and Evolutionary Botany Lab, Department of Biology, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium; <sup>4</sup>Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, SL5 7PY, UK; <sup>5</sup>School of BioSciences, The University of Melbourne, Parkville, Vic. 3010, Australia; <sup>6</sup>Department of Zoology, Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Auckland Park 2006, South Africa; <sup>7</sup>Department of Biology and Burke Museum, University of Washington, Seattle, WA 98155, USA; <sup>8</sup>Present address: Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, C/ Tulipán s.n., Móstoles, Madrid 28933, Spain

## Summary

Author for correspondence:

Tamara Villaverde

Email: [tamara.villaverde@urjc.es](mailto:tamara.villaverde@urjc.es)

Received: 13 October 2022

Accepted: 23 February 2023

New Phytologist (2023)

doi: 10.1111/nph.18845

**Key words:** *Androya*, biogeography, *Camptoloma*, *Phygelius*, phylogenomic, Scrophulariaceae, Southern Africa, target capture.

- The figwort family, Scrophulariaceae, comprises *c.* 2000 species whose evolutionary relationships at the tribal level have proven difficult to resolve, hindering our ability to understand their origin and diversification.
- We designed a specific probe kit for Scrophulariaceae, targeting 849 nuclear loci and obtaining plastid regions as by-products. We sampled *c.* 87% of the genera described in the family and use the nuclear dataset to estimate evolutionary relationships, timing of diversification, and biogeographic patterns.
- Ten tribes, including two new tribes, Androyeae and Camptolomeae, are supported, and the phylogenetic positions of *Androya*, *Camptoloma*, and *Phygelius* are unveiled. Our study reveals a major diversification at *c.* 60 million yr ago in some Gondwanan landmasses, where two different lineages diversified, one of which gave rise to nearly 81% of extant species. A Southern African origin is estimated for most modern-day tribes, with two exceptions, the American Leucophylleae, and the mainly Australian Myoporeae. The rapid mid-Eocene diversification is aligned with geographic expansion within southern Africa in most tribes, followed by range expansion to tropical Africa and multiple dispersals out of Africa.
- Our robust phylogeny provides a framework for future studies aimed at understanding the role of macroevolutionary patterns and processes that generated Scrophulariaceae diversity.

## Introduction

Scrophulariaceae are a popular and economically important plant family for their horticultural value, for example, *Buddleja* Houtt. ex L., *Myoporum* Banks & Sol. ex G.Forst., and *Nemesia* Vent. (e.g. Gray, 1895; Dirr, 1990). Additionally, species of some genera have been used in traditional herbal medicines, for example, *Eremophila* R.Br., *Scrophularia* Tourn. ex L., and *Verbascum* L. (e.g. Richmond, 1993; Richmond & Ghisalberti, 1995; John Sadgrove *et al.*, 2014), as well as shown to be promising new drug leads (Gericke *et al.*, 2020).

However, within angiosperms, Scrophulariaceae s.l. represent one of the most prominent cases of a taxonomic ‘hodgepodge’, in which the traditional circumscription of the family, based only on morphological characters, turned out to unite a widely polyphyletic group of plants (Olmstead & Reeves, 1995; Olmstead *et al.*, 2001). In their previous definition, Scrophulariaceae s.l. used to be the largest family (> 5000 spp.; Fischer, 2004) of the order Lamiales. The polyphyly of the family was first reported by

Olmstead & Reeves (1995), who recovered two unrelated clades based on [chloroplast sequence](#) data. The first clade, indicated as ‘scroph I’, was later referred to as Scrophulariaceae s.s. (Olmstead *et al.*, 2001), while the second clade, indicated as ‘scroph II’, corresponds to Plantaginaceae (APGIII, 2009; APGIV, 2016). Subsequent studies (Young *et al.*, 1999; Olmstead *et al.*, 2001; Oxelman *et al.*, 2005; Rahmzadeh *et al.*, 2005) identified additional lineages corresponding all or in part to 10 of the currently recognized families in Lamiales (Tank *et al.*, 2006; APGIII, 2009).

Intrafamilial classifications of Scrophulariaceae s.l. were initially based on a range of morphological characters, such as floral aestivation (Bentham 1846; reviewed by Von Wettstein, 1891, which was one of the most widely accepted treatments), stamen morphology (Van Tieghem, 1903), and nectary traits (Bellini, 1907). These early classification systems were not entirely congruent, indicating difficulties for recognizing natural groups in Scrophulariaceae s.l. (Olmstead & Reeves, 1995; Olmstead *et al.*, 2001). Although characters derived from their life forms or

phenotypic traits have helped define some clades, the lack of morphological synapomorphies is still hampering morphology-based classification in the family. Scrophulariaceae consists mostly of perennial or annual herbs, shrubs, and some tree species (e.g. *Myoporum semotum* Heenan & de Lange). Flower forms are diverse, especially of the corolla, encompassing for instance bell-shaped, tubular, or corollas with one or two spurs. Such display exemplifies a dependency on pollinators (interested in nectar, pollen, or oil). Bees, butterflies, birds, wasps, and flies are the most common pollinators (Kampny, 1995), with a lizard species (*Gallotia stehlini*) an exceptional flower visitor (Ortega-Olivencia *et al.*, 2012). Pollination systems in Scrophulariaceae are also aligned with corolla color and the production of scents to attract different visitors (e.g. Gong *et al.*, 2015).

The current taxonomic circumscription of Scrophulariaceae, Scrophulariaceae s.s. (henceforth Scrophulariaceae; Olmstead *et al.*, 2001; APGII, 2003; Oxelman *et al.*, 2005), makes it sister to the clade indicated as 'higher core Lamiales' (Schäferhoff *et al.*, 2010; Refulio-Rodríguez & Olmstead, 2014). The family now comprises 56 genera and *c.* 1800–2000 species, spp. (Supporting Information Table S1). Scrophulariaceae include eight major clades recognized as tribes (Oxelman *et al.*, 2005): Aptosimeae Benth. & Hook.f., Buddlejaceae Bartl., Hemimerideae Benth., Leucophylleae Miers, Limoselleae Dumort., Myoporeae Rchb., Scrophulariaceae Dumort., and Teedieae Benth. However, tribal relationships lack a firm phylogenetic footing which has prevented exploring macroevolutionary questions, for example, about spatiotemporal and morphological evolution, at the family level.

Scrophulariaceae are predominantly distributed in the Southern Hemisphere. Four of their eight tribes (Aptosimeae, Hemimerideae, Limoselleae, and Teedieae) are mainly distributed in Southern Africa, while tribe Myoporeae is predominantly Australian. The three remaining tribes, Buddlejaceae Bartl., Leucophylleae, and Scrophulariaceae, are mainly distributed in the Northern Hemisphere, though they also include African species. In addition, three species-poor genera remain unassigned to any tribe in the present classification: *Androya* H. Perrier, a monotypic genus endemic to Madagascar; *Camptoloma* Benth., with only three species characterized by an Africa-Macaronesian ('Rand Flora') disjunct pattern of distribution (Pokorný *et al.*, 2015; Culshaw *et al.*, 2021) (*C. rotundifolium* Benth. (Namibia and Angola), *C. lyperiiflorum* (Vatke) Hilliard (southern Arabia and Horn of Africa), and *C. canariense* (Webb & Berthel.) Hilliard (Gran Canaria in the Canary Islands)); and *Phygellus* E.Mey. ex Benth, with two species distributed in Southern Africa.

Though several recent phylogenetic studies have explored the spatiotemporal evolution of less inclusive groups, tribes, or genera, within Scrophulariaceae (Gándara & Sosa, 2013; Navarro-Pérez *et al.*, 2013; Chau *et al.*, 2017; Culshaw *et al.*, 2021), there has been no attempt yet to reconstruct the biogeographic origins and lineage divergence times across the entire family. Some authors have ventured hypotheses regarding the origin of the family or specific taxa. For example, Oxelman *et al.* (2005) defined Scrophulariaceae as a mainly Southern Hemisphere family with Southern African ancestors, for which northern

hemisphere Scrophulariaceae is the main exception. Chau *et al.* (2017) found a strong geographical signal at the continental level in infrageneric relationships in Buddlejaceae, in which Southern African species formed a basal grade within the tribe, with two major clades distributed in the Old and New Worlds. The only family-level biogeographic analysis is Culshaw *et al.* (2021). However, their focus was on the African genus *Camptoloma* and sampling was rather limited for 'non-African' clades; this, together with the use of molecular (chloroplast) markers with low molecular variability, led to considerable uncertainty around the ancestral distribution of Scrophulariaceae (though a Southern African origin for *Camptoloma* and tribes Buddlejaceae, Teedieae, and Limoselleae, received moderate or high support). Therefore, Southern Africa appears as a plausible geographic origin for the family, with centers of diversity in other continents being the result of subsequent range expansion.

The combination of species-rich lineages (e.g. tribe Scrophulariaceae, six genera, 530–750 spp.) and species-poor lineages (e.g. tribe Leucophylleae, three genera, 21 spp.; tribe Teedieae, five genera, 15 spp.) is an often-repeated pattern in evolution (e.g. Magallón & Sanderson, 2001; Domínguez Lozano & Schwartz, 2005; Smith *et al.*, 2011), and raises interesting questions about the underlying processes. Some authors have linked heterogeneity in diversification rates to the appearance of morphological, anatomical, or physiological 'key innovations' in certain clades, leading to accelerated speciation rates (e.g. Donoghue & Sanderson, 2015). In other cases, rapid diversification has been linked to the colonization of new geographic regions, perhaps by the acquisition of novel climatic preferences (Spalink *et al.*, 2016; Meseguer *et al.*, 2018). Without a solid phylogenetic framework for the family, none of these hypotheses can be addressed.

In this study, we aimed to generate the first phylogenomic hypothesis, using a target capture sequencing approach (Hyb-Seq), in family Scrophulariaceae in order to reconstruct its spatiotemporal evolution, and use this information to explain differences in distribution and diversity patterns among the extant clades. Using available transcriptomic and genomic resources, we designed a probe kit to capture 849 loci, using 2189 orthologous low-copy nuclear sequences (*c.* 3.7 Mbp), and targeted a nearly complete sampling of Scrophulariaceae at the genus level (*c.* 84%), with all eight tribes and the three previously unassigned genera included. We also estimated the ages and ancestral geographic ranges for all tribes to infer patterns of lineage diversification; we used molecular clock relaxed dating methods and model-based biogeographic and diversification rate analyses, implemented within a Bayesian framework to account for uncertainty in model parameters. Our objectives were: to test the utility of the developed probe kit to resolve systematic boundaries at different phylogenetic levels, infra species, species, and above-species; to provide new phylogenetic resolution, based on a large genomic dataset, for evolutionary relationships among tribes and genera in Scrophulariaceae, especially the position of the three enigmatic unassigned genera; and to reconstruct temporal and geographical diversification patterns in the family Scrophulariaceae in order to evaluate the support for an African origin and an 'out-of-Africa' biogeographic pattern.

## Materials and Methods

### Taxon sampling

Our strategy for taxonomic sampling was dual: to maximize diversity (tribal, generic, and species in a few genera) and to test the efficiency of probe kit at different evolutionary levels. To obtain a comprehensive sampling of Scrophulariaceae (Table S1), we included representatives of all currently recognized tribes based on the taxonomic treatments of Von Wettstein (1891), Barringer (1993), Olmstead *et al.* (2001), Kornhall *et al.* (2001), Oxelman *et al.* (2005), Chinnock (2007), Chau *et al.* (2017), Scheunert & Heubl (2017), Fowler *et al.* (2020), Bayly *et al.* (2020), Zhao *et al.* (2022), and others. Tribal representation included Aptosimeae (3 samples), Buddlejaceae (9), Hemimerideae (7), Leucophylleae (2), Limoselleae (24), Myoporeae (8), Scrophulariaceae (6), and Teediaceae (4). We also included 10 samples belonging to genera that have not been formally assigned to tribes: *Androya* (1), *Camptoloma* (7), and *Phygelius* (2). We included 49 out of 56 currently accepted genera in the family (Table S1). In addition to tribal, generic, and species representation, our dataset also included several samples for some species (for which we had available material from different localities); this was done in order to test our probe kit at the intraspecific level, following Villaverde *et al.* (2018): *Buddleja polystachya* Fersen. (3 samples); *Buddleja virgata* L.f. (2); *Camptoloma canariense* (2); *Camptoloma lyperiiflorum* (3), and *Camptoloma rotundifolium* (2). Additionally, we incorporated three samples as outgroup: *Plocama pendula* Aiton (1; Rubiaceae) and *Campylanthus salsoloides* (L.f.) Roth (2; Plantaginaceae). Our 76 sample-dataset includes 73 ingroup samples representing 66 species (48 genera) of Scrophulariaceae, and three outgroup samples representing two genera and families. Only seven Scrophulariaceae genera were missing from our sampling: *Leucophyllum* Bonpl. (Leucophylleae), *Ranopisoa* J.-F.Leroy (Teediaceae), *Nathaliella* B.Fedtsch. and *Rhabdotosperma* Hartl (Scrophulariaceae), *Barthlottia* E.Fisch., *Limosella* L., and *Melanospermum* Hilliard (Limoselleae). For tribe Myoporeae, fresh leaf tissue was collected from the field or from cultivated collections. For the remaining species, samples were obtained from herbarium collections at K, MO, and PRE. Voucher information is summarized in Table S2.

### Probe design

To design the probe kit, we used five Scrophulariaceae transcriptomes available through the 1KP initiative ([www.onekp.com/public\\_data.html](http://www.onekp.com/public_data.html); Matasci *et al.*, 2014): *Anticharis glandulosa* Asch. (code EJBY), *Buddleja davidii* Franch. (GRFT), *Buddleja davidii* (XRLM), *Verbascum arcturus* L. (= *Celsia arcturus*, SIBR), and *Verbascum* sp. (XXYA). As the reference genome to estimate intron/exon boundaries, we used *Erythranthe guttata* (DC.) G.L.Nesom (= *Mimulus guttatus* DC., Phrymaceae) available in MARKERMINER v.1.0 (Chamala *et al.*, 2015); downloaded from PLAZA 2.5 database (Van Bel *et al.*, 2012). MARKERMINER v.1.0 was employed to identify orthologous low-copy nuclear genes, which were used to develop the gene target probes. We selected

2189 sequences that target 849 orthologous low-copy nuclear loci (Table S3), ranging from 900 to 6287 bp and representing a total length of 3724 226 bp. Daicel Arbor Biosciences (Ann Arbor, MI, USA; <http://www.arborbiosci.com>) manufactured a target enrichment kit with in-solution biotinylated probes. The 120-mer probes (60 060 total probes) were tiled at 2× in those loci present in two or more transcriptomes (1652 sequences), or 3× tiling-densities for those loci found in only one transcriptome (537 sequences). The Scrophulariaceae probe kit is available through the public online repository DIGITAL.CSIC (<http://hdl.handle.net/10261/291126>).

### DNA extraction, library preparation, target enrichment, and sequencing

Dried leaf tissue was weighed to obtain 20–40 mg per sample. Genomic DNA was extracted using a modified CTAB protocol (Shee *et al.*, 2020). CTAB extracted samples were purified using Agencourt AMPure XP Bead Clean-up (Beckman Coulter, Indianapolis, IN, USA). The quality of the extraction was checked with Qubit™ 3.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). For tribe Myoporeae, samples were extracted following Fowler *et al.* (2020).

DNA samples were sent to AllGenetics (A Coruña, Spain) for library preparation, target enrichment, and sequencing. Only one DNA sample (*Myoporum bontiodides* (Siebold & Zucc.) A. Gray, from fresh material) was fragmented to a fragment size of 550 bp using the dsDNA NEBNext Fragmentase enzyme mix (New England Biolabs, Ipswich, MA, USA). Libraries were constructed using 400 ng of genomic DNA and the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) using half of the recommended volume for all reactions. Library quantities were checked using the Qubit Fluorometer and then pooled in groups of samples (Table S2), aiming for a quantity of 500 ng per group. Samples from the same tribe, or closely related, were pooled together. Following library preparation, specific regions of the genome were hybridized to biotinylated probes specifically designed for Scrophulariaceae (MyBaits; Daicel Arbor Biosciences) following the manufacturer's protocol. The libraries were pooled and sequenced in four different runs, two of them in a fraction (¼ and ¾, respectively) of an Illumina MiSeq PE300 (Illumina, San Diego, CA, USA).

### Raw data processing

Raw reads were cleaned using TRIMMOMATIC v.0.39 (Bolger *et al.*, 2014) to remove Illumina adapters and to trim low-quality nucleotides using the setting LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50. We processed the reads using HYBPIPER v.1.3.1. (Johnson *et al.*, 2016) with default settings, BWA v.0.7.17 (Li & Durbin, 2009) and the Scrophulariaceae-specific target loci DNA sequence file as the reference (Table S3). Mapped reads were assembled into contigs with SPADes v.3.13.1 (Bankevich *et al.*, 2012). Subsequently, EXONERATE v.2.2 (Slater & Birney, 2005), available in HYBPIPER, was used to align the assembled contigs to their associated target

sequence and remove intronic regions ('exons dataset'). HYBPIPER flags potential paralogs when multiple contigs are discovered that map well to a single reference sequence. All loci flagged as potential paralogs were removed from downstream analyses. A list of the potential paralogs is provided (Table S4). For each gene, the consensus sequences were used to generate the exons dataset including all accessions for those genes with >10 sequences. Contigs were aligned using MAFFT v.7 (Katoh & Standley, 2013) with the '-auto' option. Summary statistics were calculated using AMAS (Borowiec, 2016).

### Phylogenomic analyses

Two approaches were used to analyze our nuclear exons datasets: a multispecies summary coalescent (MSC) approach and a concatenated maximum likelihood (ML) approach. First, we used a summary method statistically consistent under the MSC in ASTRAL-III v.5.5.11 (Zhang *et al.*, 2018). The aligned exons matrices were first used to infer individual ML gene trees with IQ-TREE v.2.0 (Nguyen *et al.*, 2015; Minh *et al.*, 2020b) with 1000 ultrafast bootstraps (BS) using the '-bb' option (Hoang *et al.*, 2018). Branches with support equal to 10 BS or lower were collapsed using NEWICK UTILITIES (Junier & Zdobnov, 2010), as recommended by Mirarab (2019). Automated outlier detection across gene tree sets was done in TREESHINK (Mai & Mirarab, 2018) to remove excessively long branches beyond a designated threshold. The resulting 'shrunk' trees were used as input to infer species trees in ASTRAL. To output quartet support values, the option '-t 2' was used. Pie charts showing the proportion of quartet values were plotted in R (R Core Team, 2022), using the packages: APE (Paradis & Schliep, 2019), GGIMAGE (Yu *et al.*, 2017), GGTREE (Yu *et al.*, 2017), TREEIO (Wang *et al.*, 2020), and their corresponding dependencies.

Second, we concatenated all individual matrices into an exon supermatrix using AMAS. Phylogenetic trees were then built by ML in IQ-TREE v.2.0, after its automatic model selection using MODELFINDER (Kalyaanamoorthy *et al.*, 2017), and using the '-bb' and '-m TEST' options. Trees were plotted in FIGTREE v.1.4.4 (<https://github.com/rambaut/figtree/releases>). Finally, to investigate gene tree vs species tree concordance, we calculated two measures of genealogical concordance in our dataset, the gene concordance factor (gCF) and the site concordance factor (sCF), using the options '-gcf' and '-scf' in IQ-TREE (Minh *et al.*, 2020a). This approach provides a description of possible disagreement among loci and across sites within the sequence.

For the reconstruction of the plastid phylogeny, we used the off-target reads also recovered with the target enrichment approach. We obtained the coding regions from the annotated plastomes of *Scrophularia dentata* Royle ex Benth., *Scrophularia henryi* Hemsl., and *Scrophularia takesimensis* Nakai (NC036942, NC036943, and NC026202, respectively). A total of 86 genes were used as target sequences in HYBPIPER (Table S5). HYBPIPER was run with default parameters with the '-bwa' option. The exons matrices obtained from the plastid data were analyzed with the same methods and parameters as those used for the nuclear data, omitting the multispecies coalescent analyses as suggested by Doyle (2022).

### Divergence time estimation analyses

We used Bayesian relaxed clocks implemented in the software BEAST v.1.10 (Suchard *et al.*, 2018) to estimate ages using the nuclear exon supermatrix. Full Bayesian dating of targeted sequence datasets is often a computationally inefficient task due to the large dataset size, high percentage of missing data in the captured loci, heterotachy (i.e. heterogeneity over time in molecular rates), and potential incongruence among gene tree topologies stemming from the coalescent process itself (Smith *et al.*, 2018; Walker *et al.*, 2019). To address this, we used a 'gene-shopping' approach in which we selected only those loci (only exons) that had the highest number of quartets that were equally resolved as in the concatenated tree from the IQ-TREE analysis. After this procedure, 36 loci were selected for 76 tips (Table S6). These were concatenated into a single matrix (63 780 bp) and analyzed in BEAST v.1.10.4 using GTR+G substitution, the uncorrelated relaxed lognormal clock model (Drummond *et al.*, 2006), and the birth-death model with incomplete sampling (Stadler, 2009). As the initial starting tree to inform the process, we used the IQ-TREE topology dated in TREEPL (Smith & O'Meara, 2012) with a uniform distribution of 109.821–120.5255 million yr ago (Ma) for the crown of Scrophulariaceae obtained from Ramírez-Barahona *et al.* (2020). We applied two secondary calibration points in BEAST with a normal prior (Ho & Phillips, 2009) which spanned the 95% highest posterior density (HPD) credibility interval from Ramírez-Barahona *et al.* (2020): stem of Scrophulariaceae (divergence *Campylanthus*-ingroup) with a mean  $\pm$  SD of 82.37  $\pm$  5.0 Ma; crown of Scrophulariaceae with a mean  $\pm$  SD of 64.23  $\pm$  11.0 Ma. The MCMC chain was run for 20  $\times$  10<sup>7</sup> generations, sampled every 10 000<sup>th</sup> generation, with a burnin of 20 million states. TRACER v.1.7.1 (Rambaut *et al.*, 2018) was used to ensure adequate mixing and that all parameters had reached an EESS >200. The software TREEANNOTATOR v.1.10 (Drummond *et al.*, 2006) was used to generate the maximum clade credibility (MCC) tree, summarizing clade support as posterior probability values.

### Diversification analyses

We estimated changes in speciation and extinction rates over time and across lineages using Bayesian methods implemented in REVBayes (Höhna *et al.*, 2016). We first used an episodic birth-death model (EBD), in which rates are modeled as a piecewise-constant Brownian process changing at discrete points in time, with the mean-variance autocorrelated between time slices (Magee *et al.*, 2020). Incomplete taxonomic sampling is known to mislead estimates of diversification rates (Höhna, 2014). Different strategies have been proposed to cope with missing species in diversification rate estimates (Höhna *et al.*, 2011). Here, we chose to implement an 'empirical' sampling strategy to account for the uneven taxonomic representation within tribes at the species and genus level (e.g. Teedieae vs Scrophularieae). We used as the reference tree the MCC tree from the BEAST analysis, pruned to exclude the outgroup and leaving a single species per genus.

We then defined ‘clades’ to represent individual tribes as groups of two or three genera belonging to that tribe: for example, ‘Hemimeridae’ grouped genera *Diclis* Benth. and *Colpias* E.Mey ex Benth. The rationale is that we want to estimate the net diversification rate of a clade, for which we need to define a crown node. Next, for each of these ‘tribal clades’, we specified the number of missing species in our phylogeny according to current estimates (Table S1). These missing speciation events will be modeled as if occurring between the crown age of the defined ‘clade’ and the present (Höhna, 2014).

To explore whether changes in diversification rates are linked to specific lineages, we implemented a lineage-specific birth–death shift model, which assumes a multistate State-dependent Speciation and Extinction (SSE) model (Maddison *et al.*, 2007), with each state having its own rates, and where shifts in states (diversification rates) are modeled dynamically (Höhna *et al.*, 2019). These dynamic SSE models have been shown to be mathematically more conservative and exact than other approaches that map shifts into phylogenies such as BAMM (Laudanno *et al.*, 2020). We followed the tutorial [https://revbayes.github.io/tutorials/divrate/branch\\_specific.html](https://revbayes.github.io/tutorials/divrate/branch_specific.html), using six rate categories to discretize the lognormal distribution of the speciation rate variation, and a chain length of 2500 generations. We used a global sampling fraction of  $\rho = 0.02$  (46 tips for a total diversity of 2034 species in the family) and assumed that the extinction rate was the same for all rate categories. We used the R package REV GADGETS v.1.0 (Tribble *et al.*, 2022) to visualize the magnitude and timing of the rate changes. See Methods S1 for more details on the diversification analyses. The scripts to run these analyses are provided in <https://github.com/isabelsanmartin/RevBayes-analyses-Scrophulariaceae>.

## Historical biogeography

We inferred the geographical range evolution of Scrophulariaceae under the Dispersal–Extinction–Cladogenesis model (DEC; Ree & Smith, 2008). We used the Bayesian implementation of the DEC model (Landis *et al.*, 2018) in REV BAYES (Höhna *et al.*, 2016), which gives us the possibility to account for uncertainty in the values of anagenetic parameters and ancestral nodal ranges by estimating marginal posterior probabilities (Landis *et al.*, 2018; Lator *et al.*, 2019). We used the Plants of the World Online (POWO) database (accessed on 28 February 2022; <https://powo.science.kew.org/>) to obtain data on the global geographic distribution of the genera here included for Scrophulariaceae. We also checked the RAINBIO database to incorporate additional data (Dauby *et al.*, 2016) as well as Manning *et al.* (2021) and Cámara-Leret *et al.* (2020). We defined seven biogeographical regions based on distribution patterns in extant Scrophulariaceae species: Southern Africa (S), Tropical Africa (T), Madagascar (M); Paelearctic (E), Indo-Malaysian, including Vietnam, Korea and Japan (Y); Americas (A); and Australia–New Zealand and Pacific Islands (Z). See Methods S1 for more details on area definition. We ran two independent DEC analyses using the nuclear MCC tree, after removing the outgroup and leaving a single tip per genus. Each tip was assigned the full range of the genus

following Plants of the World Online and the revised classification (Table S1). The internal phylogenetic relationships within Myoporeae are unclear; thus, although there are endemic genera (e.g. *Bontia*), we collapsed the entire tribe into one tip with the assigned distribution representing the presumed origin of the tribe (Z; Fowler *et al.*, 2021). Two independent analyses using default priors in REV BAYES were run for 10 000 generations, sampling every 10<sup>th</sup> generation. Stochastic Character Mapping (Freyman & Höhna, 2019) was used to infer the timing and direction of geographic range changes along branches, by dividing the phylogeny into 500 time slices. Results were summarized into a maximum *a posteriori* (MAP) tree, with marginal probabilities for ancestral nodal ranges and branch-mapped dispersal events. The script to run this analysis is provided in <https://github.com/isabelsanmartin/RevBayes-analyses-Scrophulariaceae>.

## Results

### Performance of the specific probe kit of Scrophulariaceae

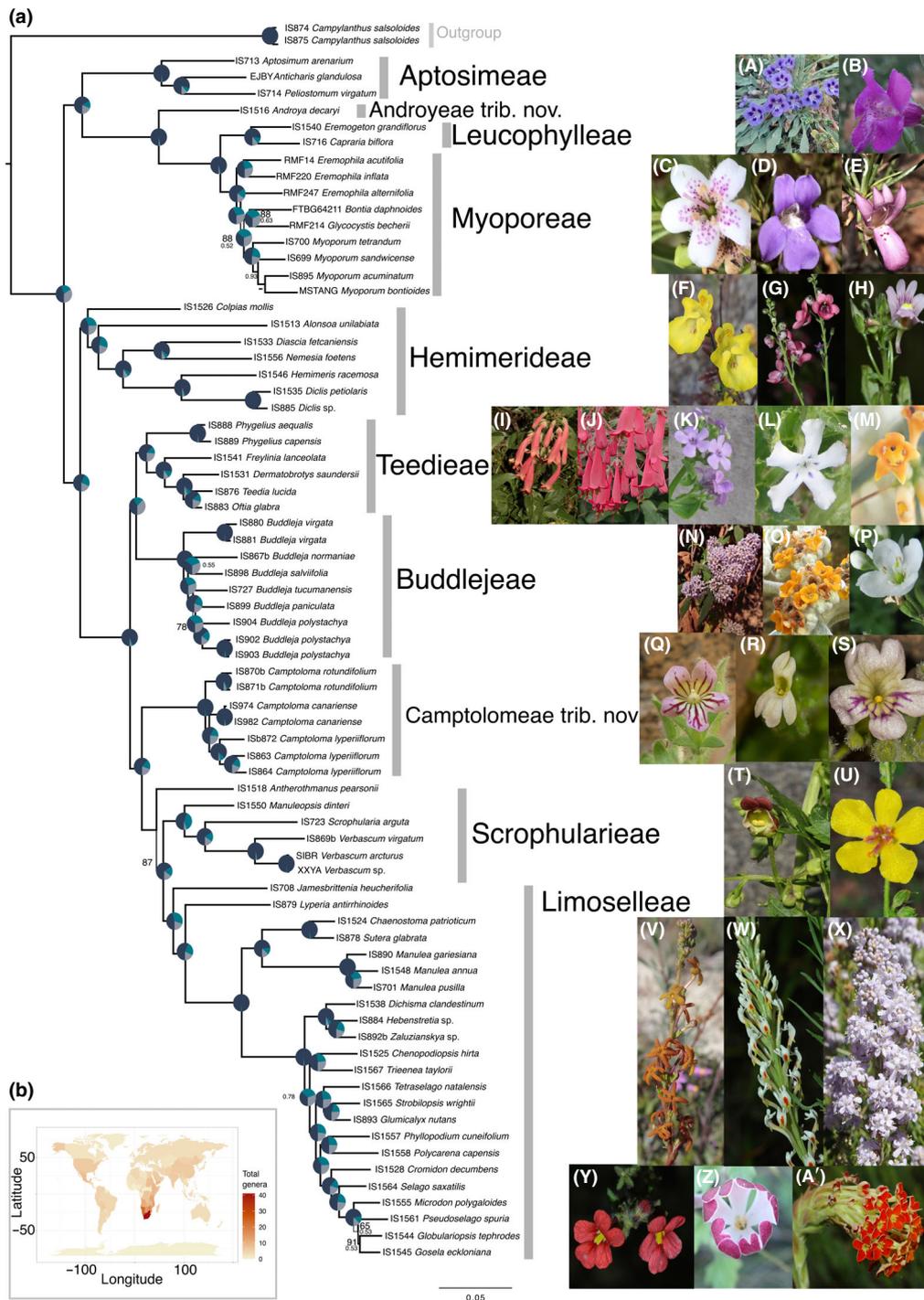
Our designed probe kit was successful in capturing the targeted genes across a large sampling of genera and all tribes within Scrophulariaceae. The increased number of loci used (849 nuclear and 86 plastid loci) compared with those used in previous studies (up to three plastid markers; Olmstead *et al.*, 2001; Oxelman *et al.*, 2005) has resulted in obtaining the most robust phylogeny reported to date in Scrophulariaceae (Fig. 1).

Recovery success per loci, per sample, and total sequence length of the targeted loci are shown in Table S7. The sample *Glekia krebsiana* (IS1543) was removed from subsequent analyses due to the low number of loci recovered. Summary statistics for each exon matrix are available in Table S8. Similar summary statistics have been calculated for matrices including multiple samples for the same genera (i.e. *Buddleja*, *Camptoloma*, *Campylanthus*, *Diclis*, *Eremophila*, *Manulea*, and *Phygellus*).

We were also able to obtain plastid coding regions from off-target reads. Recovery success and sequence length of the targeted loci are in Table S9. Summary statistics for the exon matrices are available in Table S10. Raw FASTQ files for each individual used are archived on the NCBI SRA database (project PRJNA701536).

### Resolving the phylogenetic relationships in Scrophulariaceae

**Nuclear dataset** The monophyly of the eight previously recognized tribes in the family was recovered in the ASTRAL topology with high support, but not in the IQ-TREE topology, where ‘IS1518\_Antherothamnus\_pearsonii’ is recovered as the sister lineage of Scrophulariaceae + Limoselleae, albeit with moderate support (87% BS; Fig. 1). The resulting tree identified two main clades. In the first clade, the three tribes are each found to be monophyletic with very high support. Aptosimeae are retrieved as sister to a clade formed by Leucophylleae, Myoporeae, and *Androya decaryi*. *Androya* is unassigned to any tribe and was recovered as sister to the clade formed by Leucophylleae and



**Fig. 1** Tribal relationships of Scrophulariaceae using nuclear data and map of the current distribution of its genera. (a) Phylogenetic reconstruction of Scrophulariaceae using 76 samples and inferred with 849 orthologous low-copy nuclear loci (exons only) under the concatenation approach (topology from the ML analysis performed with IQ-TREE) and with a multispecies summary coalescent approach (performed with ASTRAL). All branches with very high support (i.e. > 95 bootstrap support from IQ-TREE, BS; > 0.95 local posterior probability from ASTRAL, LPP) unless otherwise marked. Slash indicates missing node in ASTRAL and branch width correspond to BS support. Pie charts visualizing quartet support values at the nodes (dark blue, agreeing loci with the main topology; blue, loci supporting a first alternative topology; gray, loci supporting a second alternative topology). Vertical gray bars depict the outgroup, recognized tribes within the family, and unassigned genera (*Phygellus*, *Androya*, and *Camptoloma*); the last two assigned now to new tribes, Androyeae and Camptolomeae. Photographs illustrate floral diversity in the family: (A) *Aptosimum*; (B) *Leucophyllum*; (C) *Myoporum*; (D) *Eremophila drummondii*; (E) *Eremophila alternifolia*; (F) *Hemimeris*; (G) *Diascia*; (H) *Nemesia*; (I) *Phygellus aequalis*; (J) *Phygellus capensis*; (K) *Teedia*; (L) *Oftia*; (M) *Freylinia*; (N) *Buddleja salvifolia*; (O) *Buddleja tucumanensis*; (P) *Buddleja virgata*; (Q) *Camptoloma rotundifolium*; (R) *Camptoloma hypericiflorum*; (S) *Camptoloma canariense*; (T) *Scrophularia*; (U) *Verbascum*; (V) *Manulea*; (W) *Hebenstretia*; (X) *Selago*; (Y) *Jamesbrittenia*; (Z) *Zaluzianskya*; (A') *Glumicalyx*. Photographs by: (A, J, M, L, O, P, V, X, Z) J. Chau; (C–E) R. Fowler; (B) R. Olmstead; (R) M. Ross; (Q) E. Entenmann; (S) M. Mairal; (G–I, T, U, W, Y, A') M. Luceño. (b) Current distribution of the Scrophulariaceae genera.

Myoporeae in all retrieved topologies with very high support. In the second clade, Hemimerideae are sister to the remaining taxa, which are retrieved in two highly supported clades. One clade groups Teedieae with *Phygelius*, one of the unassigned genera, and together, they are sisters to Buddlejeae. In the other clade, Scrophulariaceae are sister to Limoselleae, with the unassigned genus *Camptoloma* sister to them (Fig. 1).

High gCF and sCF values support each of the crown nodes that contains the unassigned genera: *Camptoloma* (89.6% and 88.2%, respectively), *Phygelius* (84.7% and 95.2%, respectively), and *Androya* (76.5% and 68.2%, respectively). Only three nodes on the exon nuclear dataset have a gCF below 10% and have a relatively low sCF (<41%; Fig. S2); of those, just one is the node retrieving a tribe (i.e. Scrophulariaceae), whereas the other two are within tribes (i.e. within tribe Myoporeae and Limoselleae; Fig. S1). These low values of gCF combined with the low sCF for those nodes might be due to both true conflicting signal or uninformative loci (Minh *et al.*, 2020a,b).

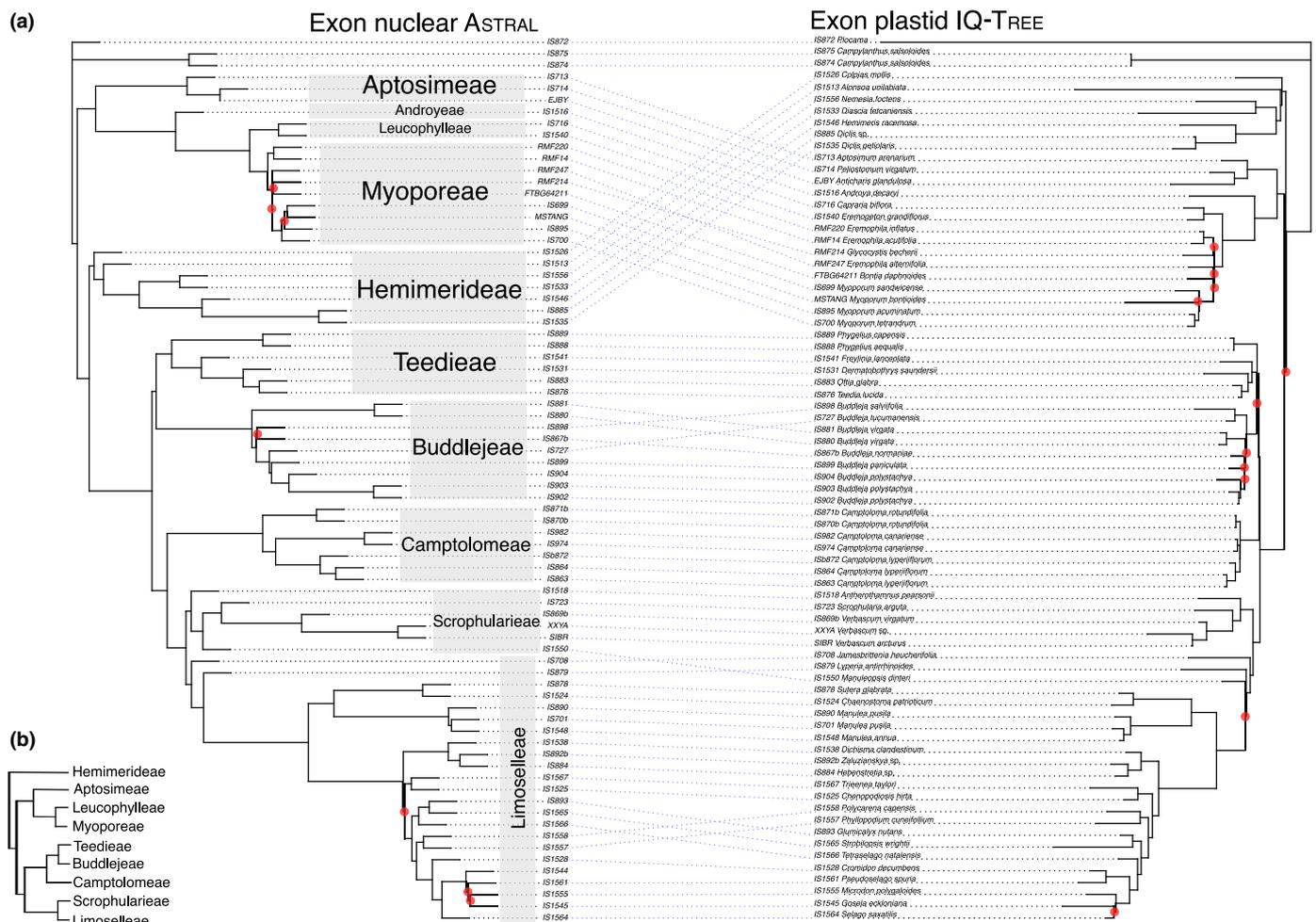
**Plastid dataset** Using the plastid dataset, the backbone of the family lacks support in the ML analysis (i.e. < 95 BS; Figs 2, S1).

However, the monophyly of all tribes receive high clade support in the plastid ML analysis (i.e. > 95% BS), except for Limoselleae (Fig. 2). Hemimerideae appears as sister to all remaining tribes, but this relationship is weakly supported (73 BS). *Antherothamnus* is included in Scrophulariaceae, and *Manuleopsis* occurs within Limoselleae; both placements are consistent with prior plastid analyses (Oxelman *et al.*, 2005). *Androya* is sister to Myoporeae + Leucophylleae and together they are sister to tribe Aptosimeae. *Camptoloma* is sister to a clade comprising *Phygelius* + Teedieae + Buddlejeae, but with low clade support (83 BS).

Two nodes have low gCF (<10%) and both have low sCF (<43.5%; Fig. S3). One of these nodes retrieve all tribes but Hemimerideae and is poorly supported (73% BS). The other node is the crown of tribe Limoselleae, which is highly supported in the IQ-TREE topology (100% BS).

### Diversification and biogeographic analyses

The topology obtained with BEAST was very similar to the one provided by IQ-TREE, including the position of *Antherothamnus*

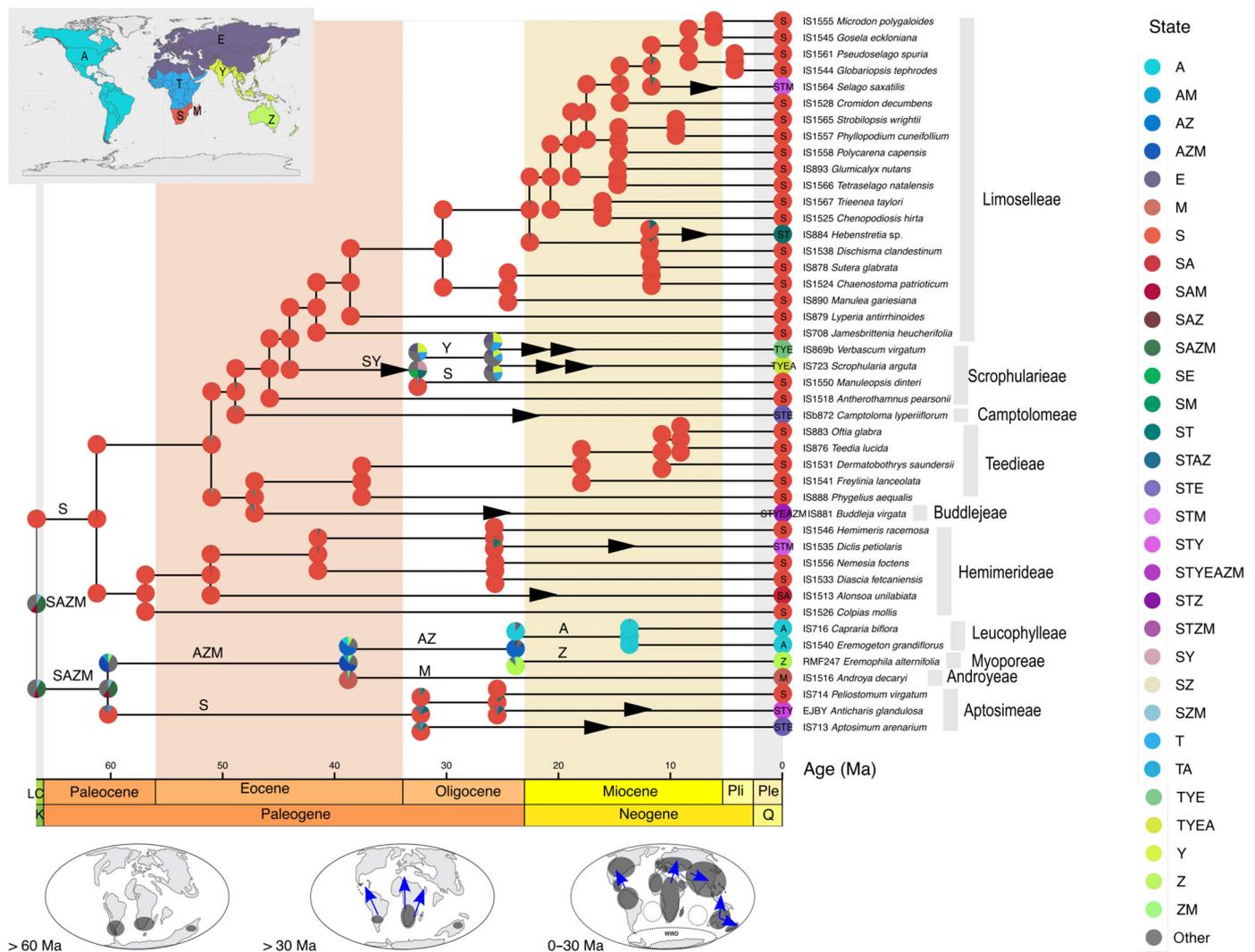


**Fig. 2** Tanglegram comparison of tree topologies obtained for Scrophulariaceae. (a) Comparison of the nuclear species tree and the plastid tree topologies of the Scrophulariaceae relationships with 76 samples. Nuclear topology inferred from individual nuclear gene trees (849 exon trees) using a multispecies summary coalescent approach performed with ASTRAL-III; and plastid topology inferred from a matrix of 86 concatenated exon trees performed with IQ-TREE. All branches with very high support (i.e. > 0.95 LPP, > 95 BS) unless otherwise marked with a red circle. (b) Tree topology reflecting tribal phylogenetic relationships in Oxelman *et al.* (2005).

*pearsonii* as sister to Leucophylleae + Scrophulariaceae. The stem age of family Scrophulariaceae was estimated as 80.90 (95% HPD = 90.25–71.59 Ma; Figs S4, S5). The crown node of Scrophulariaceae was placed in the late Cretaceous (mean 66.66 Ma, 95% HPD = 80.46–53.33 Ma). Posterior and prior probabilities for these two calibrated nodes were largely congruent. Hemimerideae are estimated to be the oldest crown clade recognized as a tribe in Scrophulariaceae, with a mean age of 56.93 Ma, during the Paleocene (95% HPD = 70.01–44.69 Ma).

The Bayesian MAP (Fig. 3) reconstruct family Scrophulariaceae as originating in an ancestral landmass comprising Southern Africa, Australia, South America, and Madagascar (SAZM, Fig. 3), with moderate support (marginal posterior probability, PP = 0.41). The MRCA of one of the two major clades within the family, the ancestor of tribes Aptosimeae, Androyeae, and

Leucophylleae + Myoporeae, is inferred to have the same widespread distribution and received a similar support (PP = 0.39). Biogeographic evolution within this clade involved a series of allopatric speciation events, first isolating South Africa (Aptosimeae), followed by Madagascar (Androyeae), and the latter separation of Australia-South America (Leucophylleae-Myoporeae; Fig. 3); support for these events ranged between PP = 0.69 and 0.98. Dispersal events from Southern Africa to Tropical Africa and Southeast Asia are reconstructed within tribe Aptosimeae in the last 20 Ma. The MRCA of the second major clade, the ancestor of tribes Hemimeridae, Buddlejaceae, Camptolomeae, Teedieae, Scrophulariaceae, and Limoselleae, is reconstructed as originating in Southern Africa with high support (PP = 0.97). The ancestral ranges of the remaining nodes in this clade are also reconstructed with high support (PP > 0.79), except



**Fig. 3** Geographical range evolution of Scrophulariaceae. Inset map on the upper left indicates the seven areas of distribution used in the biogeographical analysis of Scrophulariaceae based on a Bayesian implementation of the Dispersal–Extinction–Cladogenesis model. Color codes for widespread ranges: (A) Americas; (E) Palearctic; (M) Madagascar; (S) South Africa; (T) Tropical Africa and South Arabia; (Y) India, Malaysia, Vietnam, Korea, and Japan; (Z) Australasian and Pacific Islands. Abbreviations correspond to Late Cretaceous (LC), Pliocene (Pli), and Pleistocene (Ple) Periods and Cretaceous (K) and Quaternary (Q) Epochs. Lateral grey bars depict Scrophulariaceae tribes. Black arrows represent the estimated time of dispersal from the stochastic mapping analysis performed in RevBAYES. Lower map figures, redrawn from Buerki *et al.* (2011), summarize the main vicariance and dispersal events in the history of the family with grey ovals. Dashed circles symbolize the subequatorial currents and the West Wind Drift (WWD).

for the MRCA of Scrophulariaceae and the crown nodes of genera *Scrophularia* and *Verbascum*, with marginal probabilities below 0.30 (Fig. 3). Diversification within this second major clade occurred mainly within the Southern African region, but independent dispersal events to other southern and northern hemisphere landmasses are inferred in at least eight genera in five of the six tribes (the exception is Teedieae, which evolved within Southern Africa). Though posterior support for the timing of these events is not high (Figs S6, S7), most of them are inferred after 23 Ma in the MAP tree (Fig. 3). Many dispersal events involved adjoining landmasses (e.g. dispersals from Southern Africa to Tropical Africa and Madagascar in Limoselleae or Hemimerideae), but some were to more distant regions, for example, dispersals to America in Hemimerideae and to Southeast Asia in Buddlejeae. Madagascar was colonized at least in three separate dispersal events in Hemimerideae, Buddlejeae, and Limoselleae (Fig. 3).

The EBD analysis (Fig. 4) shows a pattern of increase in diversification rate, associated with the speciation rate, between 25 and 10 Ma, and a corresponding decrease in the relative extinction rate. The LBDS analysis inferred increases in diversification rates associated with tribes Teedieae and Limoselleae, and to a lesser extent to Androyeae-Leucophylleae-Myoporeae. Given that the tree used for this analysis was resolved to the genus level, these shifts can be interpreted as an increase in the rate of genus origination. For example, Teedieae with six genera has a crown age of

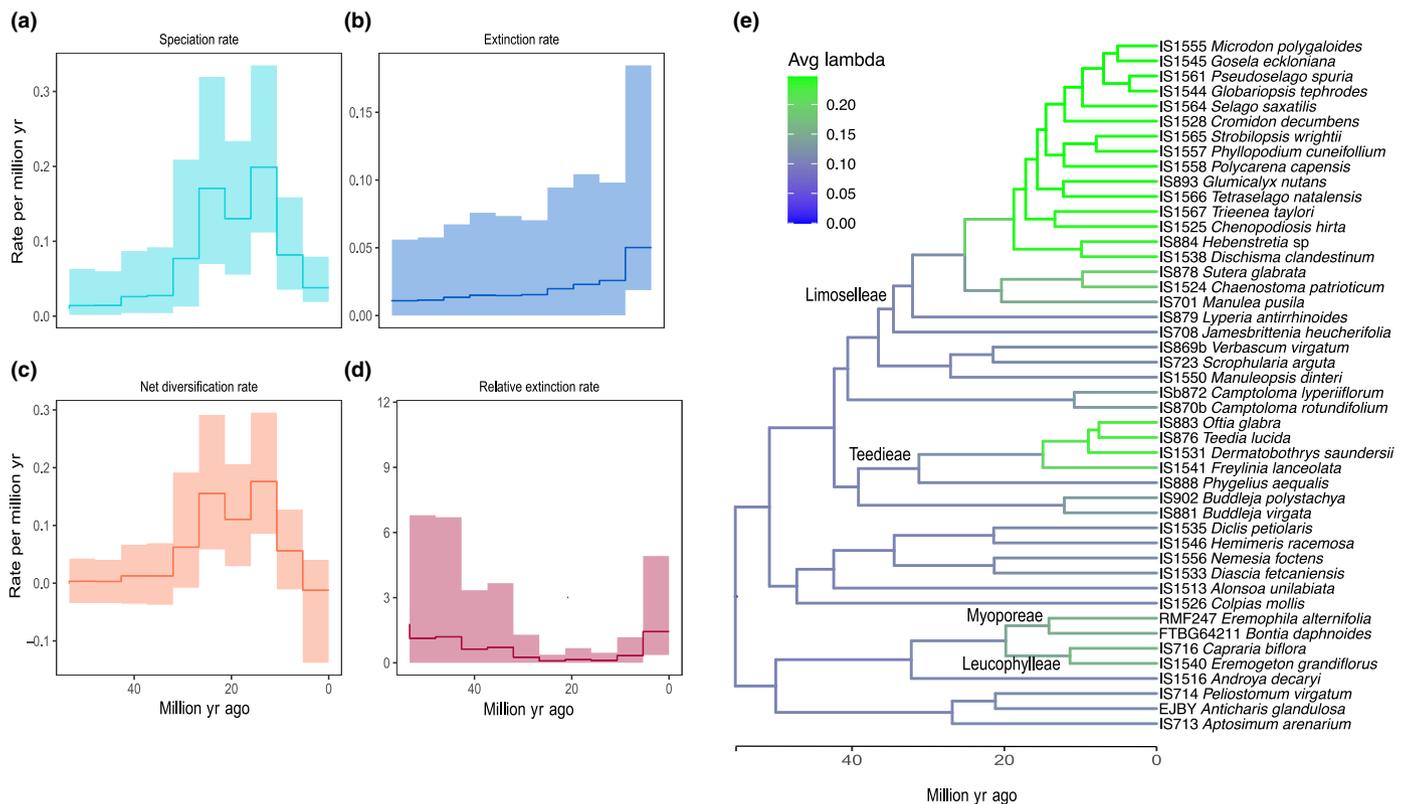
36 Ma (21.6 Ma if *Phygellius* is excluded); by contrast, Hemimerideae has also six genera but they diverged over the last 58 Ma and therefore exhibits a lower rate of generic diversification (Fig. 4).

## Discussion

### Resolving the deep-branching relationships in Scrophulariaceae

Nearly 20 yr after the benchmark study of Oxelman *et al.* (2005), our results recovered fully resolved and strongly supported tribal relationships, providing a long-sought backbone topology and timeline for future studies on Scrophulariaceae evolution. We show the utility of the Scrophulariaceae targeted sequencing kit in resolving relationships at the supra- and infrageneric species level, allowing the resolution of long-standing questions in Scrophulariaceae systematics: the placement of *Androya*, *Camptoloma*, and *Phygellius*.

Our nuclear phylogeny improves our perception of the evolution of the family and will necessitate several changes in its classification, including new tribes to accommodate *Androya* and *Camptoloma* (Androyeae and Camptolomeae) and expanding Teedieae to include *Phygellius*. This new tribal classification makes sense from a morphological and biogeographical perspective. The American tribe Leucophylleae and the mainly Australian tribe Myoporeae are now shown to diverge after the Madagascan



**Fig. 4** Diversification analysis in Scrophulariaceae. (a) Results from the episodic birth–death (EBD) model with 10 time intervals showing changes in speciation and extinction rates, and net diversification (speciation minus extinction) and relative extinction (ratio extinction : speciation) rates over time. Shaded regions show the 90% credible intervals for the rate. (b) Changes in speciation rates among major lineages within Scrophulariaceae, estimated using a multistate SSE birth–death model (LBDS).

Androyeae, Camptolomeae, Scrophulariaceae, and Limoselleae diversified initially in Southern Africa, with several dispersals out of Africa afterward. Finally, the morphological affinities of *Phygellius* with some genera within Teedieae, including woody habit (also shared with Buddlejaceae), leafy inflorescences, and anthers with separate, equal, and often parallel thecae, supports its inclusion in this tribe (Oxelman *et al.*, 2005). All 10 tribes were recovered as monophyletic, except for Scrophulariaceae in some analyses: for example, *Antherothamnus* was placed as sister to Scrophulariaceae and Limoselleae in the nuclear concatenation approach (IQ-TREE, Fig. 1), albeit with low support.

### Systematics of Scrophulariaceae

Our nuclear analyses strongly support two major lineages within the family. One contains 11 genera and *c.* 340 species (*c.* 17% of the total diversity of the family) and consists of Aptosimeae sister to *Androya* + Leucophylleae + Myoporeae, which agrees with previous plastid phylogenies (Olmstead *et al.*, 2001; Oxelman *et al.*, 2005). This lineage is widely distributed in Africa, Madagascar, Americas (mainly South and Central America), Australia-Oceania, and extending into Eastern Asia. The second contains 45 genera and *c.* 1700 species (*c.* 83% of the family) and is composed of two sublineages with a primary distribution in Southern Africa, and a secondary center of diversity in the northern hemisphere. One of these sublineages corresponds to tribe Hemimerideae and the second to *Phygellius* + Teedieae + Buddlejaceae and *Camptoloma* + Scrophulariaceae + Limoselleae. A few genera (*Buddleja*, *Scrophularia*, *Verbascum*, and *Limosella*) have widespread distributions.

We show that there is congruence in the phylogenetic signal coming from the nuclear and plastid genomes. The division of Scrophulariaceae into two main clades and the monophyly of the eight tribes is supported also in the plastid phylogeny (Figs 2, S1). However, the position of the enigmatic genus *Camptoloma* as sister to Teedieae-*Phygellius*-Buddlejaceae is incongruent with its position in the nuclear tree, though it agrees with results based on Sanger sequencing of a few plastid markers (Oxelman *et al.*, 2005; Culshaw *et al.*, 2021). By contrast, the incongruent position of *Manuleopsis dinterii* (IS1550) between the nuclear and plastid results – embedded within Scrophulariaceae in the nuclear trees, but included in Limoselleae in the plastid tree (100% BS) – is supported by both nuclear concatenated and multispecies coalescent approaches (100% BS, 1 LPP). Another case of conflict is *Rhabdotosperma*, not sampled here, which was shown by Dong *et al.* (2022) to be nested within *Verbascum*. Whether these cases of incongruence are due to a lack of information from the plastid genome (which is here obtained as a by-product of a target enrichment approach), or to a biological phenomenon such as ancient hybridization (Stull *et al.*, 2023), remains to be tested. Well-supported conflict between nuclear and plastid trees in the placement of *Manuleopsis* could be explained by plastid transfer from Limoselleae. Studies with a wider sampling, including the genus *Barthlottia*, which was sister to *Manuleopsis* and together placed in Limoselleae in Oxelman *et al.* (2005), are needed to corroborate this hypothesis.

### Spatiotemporal evolution of Scrophulariaceae

The biogeographic scenario recovered here draws a picture of an early evolution of Scrophulariaceae in the southern hemisphere landmasses, and later expansion to the Northern Hemisphere. Despite some uncertainty in the crown nodal reconstruction, our analyses point to Scrophulariaceae originating in the ‘Gondwanan’ landmasses of Africa, South America, Madagascar, and Australia at the Cretaceous–Paleocene boundary (*c.* 66 Ma). Of the two major clades in the family, the least diverse, comprising tribes Aptosimeae, Androyeae, Myoporeae, and Leucophylleae, started to diversify within this ancestral landmass *c.* 60 Ma (Figs 3, S4), during the Paleocene. Divergence within this clade involved sequential vicariance events: *c.* 60 Ma, the Southern African ancestor of Aptosimeae diverged from the rest of the lineage, followed by the Madagascar Androyeae *c.* 40 Ma, and later divergence between Leucophylleae and Myoporeae in the New World and Australian regions at *c.* 24 Ma (Fig. 3). This sequence agrees with the pattern of geological rifting, albeit not with its timing (Sanmartín, 2011 and references therein). Plants are known to be good overseas dispersers, but in addition to long-distance dispersal, there are other migration routes, that might explain the Gondwana-like pattern in the Myoporeae-Androyeae clade. For example, though Africa separated from Gondwana *c.* 120 Ma, connections between Antarctica (and attached landmasses South America and Australia) with Africa-Madagascar-India could have persisted until the Late Cretaceous (80–65 Ma) via the Kerguelen Plateau and the Gunnerus Ridge (Yoder & Nowak, 2006). Similarly, Madagascar reached its current position relative to Africa at *c.* 130–118 Ma, but dispersals via the Davie Ridge, which extends across the Mozambique Channel, could have been possible during the Paleogene (60–40 Ma), supported by molecular evidence (e.g. McCall, 1997; Raxworthy *et al.*, 2002; Yoder & Nowak, 2006).

A different pattern is observed for the MRCA of the most-diverse lineage in the family, comprising tribes Hemimerideae, Buddlejaceae, Teedieae, Camptolomeae, Scrophulariaceae, and Limoselleae. Diversification within this clade seems to have involved evolution in isolation in Southern Africa during most of the Paleogene, with subsequent migration events ‘out of Africa’ during the Neogene (Fig. 3). The two tribes with the greatest generic diversity, Teedieae and Limoselleae (Fig. 4), are entirely or predominantly distributed in Southern Africa (Fig. 3). Four of the six tribes in this clade are inferred to have originated and have their main center of diversity in this continent (Fig. 3). The only exceptions are Scrophulariaceae, which has a more uncertain origin, and Buddlejaceae, which is inferred to have originated in Southern Africa (Fig. 3) but has its greatest diversity outside this continent. In all, 7 of the 10 tribes of Scrophulariaceae defined here, have their ancestral area in Southern Africa.

Dispersals out of Africa and range expansions into other continents did not start until *c.* 23 Ma, in the early Neogene, though some dispersal events are inferred slightly earlier, for example, in the lineage leading to Buddlejaceae (Fig. 3). The start of these ‘out-of-Africa’ dispersals coincides with the increase in net diversification rates observed at *c.* 25 Ma (Fig. 4). Tribal diversification in Scrophulariaceae might thus have been fueled by these events of

geographic expansion. After the cooling event at the end of the Paleogene (34 Ma), an increase in global temperatures at the Chattian-Ruppelian boundary, *c.* 27 Ma, the Late Oligocene Warming Event (LOWE; Bohaty & Zachos, 2003), was followed by the Mid Miocene Climatic Optimum (17–14 Ma, MMCO; Zachos *et al.*, 2008), which brought about the first onset of arid conditions world-wide. Aridification was not synchronous across continents. In Australia and Africa, aridification began in the Mid Miocene, whereas in the Asian interior, it started at the Eocene–Oligocene boundary, *c.* 33.9 Ma. Most Scrophulariaceae tribes started diversifying around these times, between 30 and 12 Ma (Fig. 3), with the exception of Hemimerideae (57 Ma, Figs S4, S5). These events of rapid climate change were driven by geological tectonics (e.g. the opening of the Drake Passage between South America and Antarctica and the collision of Africa–Arabia against Eurasia) and led to the establishment of major ocean currents, such as the south and north subequatorial currents between Africa, Southeast Asia, and Australasia, or the eastward West Wind Drift around Antarctica. These ocean and wind currents might have favored geographic expansion in the family and ‘promoted’ diversification, especially within species-rich tribes Scrophulariaceae and Limoselleae.

In addition, the global introduction of drier climates at the start of the Neogene might have benefited the geographic expansion of Scrophulariaceae. With our present taxon sampling, we cannot reconstruct the ancestral climatic preferences of the family, but if Scrophulariaceae originated indeed in Africa or one of the former ‘Gondwanan’ landmasses, around the Paleocene (60 Ma), its ancestral habitat was probably subtropical. Paleogene Africa (65–34 Ma) was characterized by wet and warm climates, and an evergreen vegetation. This changed with the introduction of arid conditions at the Paleogene–Neogene boundary, coincident with the expansion of drylands world-wide (Pokorny *et al.*, 2015). Some tribes and genera that now exhibit more arid climatic tolerances could have reached their current, out-of-Africa distribution around this period. For example, the majority of species within tribe Myoporeae are arid/desert-adapted and occur predominantly in Australia’s arid zone (e.g. *Eremophila*, 3000 species). Its sister tribe, American Leucophylleae, is also largely desert-distributed (*Leucophyllum* 16 sp. Chihuahan Desert, Mexico). Conversely, the other major clade, formed by Buddlejeae + Teedieae + Camptoolomeae, have more temperate and mesic habitat preferences, which probably resemble those of their ancestors. An exception is Scrophulariaceae (759 species), which presumably colonized the Northern Hemisphere, adapting to both temperate and continental (cold and arid) climatic niches. In sum, though the biogeographic scenario presented here is intended as a first approximation, and a broader sampling is needed to infer a detailed history of migration events, our analysis, covering *c.* 87% generic diversity, suggests Africa as the ancestral area of Scrophulariaceae, at a time when connections still existed with other Gondwanan landmasses, and subsequent migration events to other biogeographic regions, linked to major species diversification.

In its current circumscription, Scrophulariaceae still lack a morphologically distinctive defining trait. Most Scrophulariaceae

have radially symmetric corollas: bilateral symmetry is a derived character in different lineages (i.e. *Scrophularia*, Leucophylleae, and Myoporeae). On the contrary, different singleton innovative characters have been acquired during its diversification. For instance, *Anticharis* is known to exhibit C<sub>4</sub> photosynthesis (Sage *et al.*, 2011; Khoshravesh *et al.*, 2012), a condition only shared in Lamiales by *Blepharis* Juss. (Acanthaceae); Hemimerideae, with the oldest crown age of any tribe in the family, have oil-secreting species and flowers with two spurs that might be related with a tied plant–pollinator system (Renner & Schaefer, 2010); arbuscular mycorrhizal fungal structures have been found inside some *Buddleja* roots (e.g. Koske *et al.*, 1992; Camargo-Ricalde *et al.*, 2003; Dickie *et al.*, 2007), which could have conferred some additional advantages in their establishment process during colonization of new areas; finally, *Limosella*, with a cosmopolitan distribution, is the only aquatic genus in the family. With the robust phylogenetic relationships depicted here, further biological questions regarding its synnovations (e.g. C<sub>4</sub> photosynthesis) or confluence of traits (e.g. in the depauperon *Colpasia* a combination of rock faces habitat, specific *Rediviva* female pollinator, low climate spectrum; or in *Scrophularia*, the confluence of pollination systems together with its adaptation to different environments) can be further explored.

#### Classification changes proposed

**Androyeae** *Androya* diverged from its sister lineage (Leucophylleae + Myoporeae) *c.* 39 Ma and most likely reached Madagascar by long-distance dispersal, a recurrent mechanism of origin for many elements of the Malagasy flora (Yoder & Nowak, 2006). The phylogenetic position of *Androya* supports previous findings based on plastid molecular data (Olmstead *et al.*, 2001; Kelchner, 2003; Oxelman *et al.*, 2005), anatomy (Karrfalt & Tomb, 1983; Lersten & Beaman, 1998), and pollen (Niezgoda & Tomb, 1975; Mosyakin & Tsybalyuk, 2015a). We designated here a new tribe:

**Androyeae** *Olmstead, trib. nov. type: Androya H. Perrier.* Small trees. Leaves opposite, short petiolate, narrowly elliptic, acute, glabrous to finely pubescent. Inflorescence axillary cymes, 3–7 flowered. Flowers short pedicellate, 4-merous, actinomorphic. Calyx 4-lobed, sepals free entirely or nearly to base. Corolla white, subrotate, tube short, lobes equal tube. Stamens 4, equal, inserted at apex of tube, exerted, anther thecae confluent at tip. Ovary ovoid, glabrous, stigma large, clavate. Fruit elongate capsule exceeding calyx. Seeds very small, winged.

This tribe is monotypic, consisting of *Androya decaryi* native to Madagascar.

**Teedieae** Teedieae, as traditionally circumscribed (excluding *Phygelius*, to be described later), diversified in the Early Miocene *c.* 18 Ma (Figs 3, S4) in Southern Africa, followed by an *in-situ* diversification and two dispersals to Madagascar by *Oftia* and the monotypic *Ranopisoa* J.-F. Leroy (not included in our sampling). The previously unassigned genus *Phygelius* is here shown to be closely related to Teedieae, which is also supported by

palynomorphological similarities (Mosyakin & Tsybalyuk, 2015b). We suggest expanding the tribe Teediaceae to include *Phygelius*. Doing so would push the crown date for the tribe to the Eocene (*c.* 38 Ma).

**Camptolomeae** *Camptoloma* is a small genus (three species) exhibiting a disjunct Rand Flora pattern, started to diversify *c.* 13 Ma, during the Mid Miocene (Figs 3, S4). Our results indicate that *Camptoloma* cannot be accommodated in any of the existing tribes. There are morphological traits for distinguishing *Camptoloma* from related taxa, and thus, a new tribe is accordingly warranted and is described below:

**Camptolomeae** *Olmstead*, trib. nov. type: *Camptoloma* Benth. Suffrutescent perennial herbs or small shrubs. Stems quadrangular, winged, and glandular pubescent. Leaves alternate, petiolate, ovoid to elliptic with obtuse apex, or orbicular, dentate to serrate. Inflorescence an axillary cyme with 2–11 flowers or flowers solitary. Flowers pedicellate. Calyx deeply 5-lobed, obscurely 2-lipped. Corolla white to yellow or pink to lilac with purple veins, 2-lipped with 2 upper and 3 lower lobes, tube short cylindrical, throat bearded with clavate hairs or glabrous. Stamens 4, didynamous, abaxial pair inserted in throat, shortly exerted, adaxial pair inserted in middle of tube, included, anther thecae confluent. Ovary cuneate, stigma bifid. Fruit capsule. Seeds small, ribbed longitudinally.

This tribe consists of *Camptoloma* with three species native to Africa and the Canary Islands.

## Acknowledgements

We thank P. Barberá and R. Riina for sampling at MO and E, respectively. We also thank curators from PRE and K. We are thankful to M. García-Gallo Pinto and M. Rincón-Barrado for the technical support given at RJB. We are grateful for photographs given by E. Entenmann, M. Luceño, M. Mairal, and M. Ross. This work was supported by the Spanish Ministry of Science and Innovation (CGL2015-67849-P and PID2019-108109GB-I00/AEI/10.13039/501100011033 to IS, and by the USA National Science Foundation (DEB, 131111) to RGO and JHC.

## Competing interests

None declared.

## Author contributions

TV and IS designed the study. IL, RMF, JHC and RGO contributed data. TV, IL, TS and IS analyzed the data with substantial support from RMF, JHC and RGO. All authors contributed to the writing and approved the final manuscript.

## ORCID

Isabel Sanmartín  <https://orcid.org/0000-0001-6104-9658>  
Tamara Villaverde  <https://orcid.org/0000-0002-9236-8616>

## Data availability

Raw sequence reads for the NGS Illumina runs were deposited in GenBank SRA under Bioproject no. PRJNA701536. Additional data and results are available from the open science platform GitHub (<https://github.com/isabelsanmartin/RevBayes-analyses-Scrophulariaceae>) and through the public online repository DIGITAL.CSIC (<http://hdl.handle.net/10261/291126>).

## References

- APGII. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.
- APGIII. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.
- APGIV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD *et al.* 2012. SPADes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.
- Barringer K. 1993. Five new tribes in the Scrophulariaceae. *Novon* 3: 15–17.
- Bayly MJ, Fowler RM, Buirchell BJ, Chinnock RJ, Murphy DJ. 2020. (2758) Proposal to conserve the name *Eremophila* against *Bontia*, *Myoporium* and *Andreusia* (Scrophulariaceae: Myoporeae). *Taxon* 69: 828–830.
- Bellini R. 1907. Criteri per una nuova classificazione delle Personatae (Scrophulariaceae et Rhinanthaceae). *Annali di Botanica* 6: 131–145.
- Bentham G. 1846. Scrophulariaceae. In: Candolle ALP-P, ed. *Prodromus sistematicus universalis regni vegetabilis*. Paris, France: Victoris Masson, 266–288.
- Bohaty SM, Zachos JC. 2003. Significant Southern Ocean warming event in the late middle Eocene. *Geology* 31: 1017.
- Bolger AM, Lohse M, Usadel B. 2014. TRIMMOMATIC: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660.
- Buerki S, Forest F, Álvarez N, Nylander JA, Arrigo N, Sanmartín I. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography* 38: 531–550.
- Cámara-Leret R, Frodin DG, Adema F, Anderson C, Appelhans MS, Argent G, Arias Guerrero S, Ashton P, Baker WJ, Barford AS *et al.* 2020. New Guinea has the world's richest island flora. *Nature* 584: 579–583.
- Camargo-Ricalde S, Dhillion SS, Jiménez-González C. 2003. Mycorrhizal perennials of the 'matorral xerófilo' and the 'selva baja caducifolia' communities in the semiarid Tehuacán-Cuicatlán Valley, Mexico. *Mycorrhiza* 13: 77–83.
- Chamala S, García N, Godden GT, Krishnakumar V, Jordon-Thaden IE, De Smet R, Barbazuk WB, Soltis DE, Soltis PS. 2015. MARKERMINER 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.
- Chau JH, O'Leary N, Sun W-B, Olmstead RG. 2017. Phylogenetic relationships in tribe Buddlejaceae (Scrophulariaceae) based on multiple nuclear and plastid markers. *Botanical Journal of the Linnean Society* 184: 137–166.
- Chinnock RJ. 2007. *Eremophila and allied genera: a monograph of the plant family Myoporaceae*. Dural, NSW, Australia: Rosenberg.
- Culshaw V, Villaverde T, Mairal M, Olsson S, Sanmartín I. 2021. Rare and widespread: integrating Bayesian MCMC approaches, Sanger sequencing and Hyb-Seq phylogenomics to reconstruct the origin of the enigmatic Rand Flora genus *Camptoloma*. *American Journal of Botany* 108: 1673–1691.
- Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC *et al.* 2016. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* 1: 1–18.

- Dickie IA, Thomas MM, Bellingham PJ. 2007. On the perils of mycorrhizal status lists: the case of *Buddleja davidii*. *Mycorrhiza* 17: 687–688.
- Dirr MA. 1990. *Manual of woody landscape plants: their identification, ornamental characteristics, culture, propagation and uses*. Champaign, IL, USA: Stipes.
- Domínguez Lozano F, Schwartz MW. 2005. Patterns of rarity and taxonomic group size in plants. *Biological Conservation* 126: 146–154.
- Dong X, Mkala EM, Mutinda ES, Yang J-X, Wanga VO, Oulo MA, Onjolo VO, Hu G-W, Wang QF. 2022. Taxonomy, comparative genomics of Mullein (*Verbascum*, Scrophulariaceae), with implications for the evolution of *Verbascum* and Lamiales. *BMC Genomics* 23: 566.
- Donoghue MJ, Sanderson MJ. 2015. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* 207: 260–274.
- Doyle JJ. 2022. Defining coalescent genes: theory meets practice in organelle phylogenomics. *Systematic Biology* 71: 476–489.
- Drummond AJAJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: 699–710.
- Fischer E. 2004. Scrophulariaceae. In: Kadereit JW, ed. *The families and genera of vascular plants. Flowering plants - dicotyledons: Lamiales (except Acanthaceae including Avicenniaceae)*. Berlin & Heidelberg, Germany: Springer, 333–432.
- Fowler RM, McLay TGB, Schuster TM, Buirchell BJ, Murphy DJ, Bayly MJ. 2020. Plastid phylogenomic analysis of tribe Myoporeae (Scrophulariaceae). *Plant Systematics and Evolution* 306: 52.
- Fowler RM, Murphy DJ, McLay TG, Buirchell BJ, Chinnock RJ, Bayly MJ. 2021. Molecular phylogeny of tribe Myoporeae (Scrophulariaceae) using nuclear ribosomal DNA: generic relationships and evidence for major clades. *Taxon* 70: 570–588.
- Freyman WA, Höhna S. 2019. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. *Systematic Biology* 68: 505–519.
- Gándara E, Sosa V. 2013. Testing the monophyly and position of the North American shrubby desert genus *Leucophyllum* (Scrophulariaceae: Leucophylleae). *Botanical Journal of the Linnean Society* 171: 508–518.
- Gericke O, Hansen NL, Pedersen GB, Kjaerulf L, Luo D, Staerk D, Møller BL, Pateraki I, Heskes AM. 2020. Nerylneryl diphosphate is the precursor of serrulatane, viscidane and cembrane-type diterpenoids in *Eremophila* species. *BMC Plant Biology* 20: 91.
- Gong W-C, Chen G, Vereecken NJ, Dunn BL, Ma Y-P, Sun W-B. 2015. Floral scent composition predicts bee pollination system in five butterfly bush (*Buddleja*, Scrophulariaceae) species. *Plant Biology* 17: 245–255.
- Gray A. 1895. *Field, forest, and garden botany: a simple introduction to the common plants of the United States east of the 100<sup>th</sup> meridian, both wild and cultivated*. New York, NY, USA: American Book.
- Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* 58: 367–380.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBOOT2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.
- Höhna S. 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. *PLoS ONE* 9: e84184.
- Höhna S, Freyman WA, Nolen Z, Huelsenbeck JP, May MR, Moore BR. 2019. A Bayesian approach for estimating branch-specific speciation and extinction rates. *bioRxiv*. doi: 10.1101/555805.
- Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. REVBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65: 726–736.
- Höhna S, Stadler T, Ronquist F, Britton T. 2011. Inferring speciation and extinction rates under different species sampling schemes. *Molecular Biology and Evolution* 28: 2577–2589.
- John Sadgrove N, Lloyd Jones G, Greatrex BW. 2014. Isolation and characterisation of (–)-genifuranal: the principal antimicrobial component in traditional smoking applications of *Eremophila longifolia* (Scrophulariaceae) by Australian aboriginal peoples. *Journal of Ethnopharmacology* 154: 758–766.
- Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016. HYBPIPER: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.
- Junier T, Zdobnov EM. 2010. The NEWICK UTILITIES: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* 26: 1669–1670.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. MODELFINDER: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.
- Kampny CM. 1995. Pollination and flower diversity in Scrophulariaceae. *The Botanical Review* 61: 350–366.
- Karrfalt EE, Tomb AS. 1983. Air spaces, secretory cavities, and the relationship between Leucophylleae (Scrophulariaceae) and Myoporaceae. *Systematic Botany* 8: 29–32.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kelchner SA. 2003. *Phylogenetic structure, biogeography, and evolution of Myoporaceae*. PhD thesis, The Australian National University, Canberra, ACT, Australia.
- Khoshravesh R, Akhiani H, Sage TL, Nordenstam B, Sage RF. 2012. Phylogeny and photosynthetic pathway distribution in *Anticharis* Endl. (Scrophulariaceae). *Journal of Experimental Botany* 63: 5645–5658.
- Kornhall P, Heidari N, Bremer B. 2001. Selaginiae and Manuleae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. *Plant Systematics and Evolution* 228: 199–218.
- Koske RE, Gemma JN, Flynn T. 1992. Mycorrhizae in Hawaiian angiosperms: a survey with implications for the origin of the native flora. *American Journal of Botany* 79: 853–862.
- Landis MJ, Freyman WA, Baldwin BG. 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution* 72: 2343–2359.
- Laudanno G, Haegeman B, Rabosky DL, Etienne RS. 2020. Detecting lineage-specific shifts in diversification: a proper likelihood approach. *Systematic Biology* 70: 389–407.
- Lavor P, Calvente A, Versieux LM, Sanmartín I. 2019. Bayesian spatio-temporal reconstruction reveals rapid diversification and Pleistocene range expansion in the widespread columnar cactus *Pilosocereus*. *Journal of Biogeography* 46: 238–250.
- Lersten NR, Beaman JM. 1998. First report of oil cavities in Scrophulariaceae and reinvestigation of air spaces in leaves of *Leucophyllum frutescens*. *American Journal of Botany* 85: 1646–1649.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56: 701–710.
- Magallón S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55: 1762–1780.
- Magee AF, Höhna S, Vasylyeva TI, Leaché AD, Minin VN. 2020. Locally adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts. *PLoS Computational Biology* 16: e1007999.
- Mai U, Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19: 272.
- Manning JC, Maluleke R, Ebrahim I, Helme NA. 2021. The genus *Freylinia* Pangella ex Colla (Scrophulariaceae: Teediaceae): a re-assessment of the systematics and conservation status. *South African Journal of Botany* 142: 352–369.
- Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17.
- McCall RA. 1997. Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264: 663–665.
- Mesguier AS, Lobo JM, Cornuault J, Beerling D, Ruhfel BR, Davis CC, Sanmartín I. 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.
- Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37: 2727–2733.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020b. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

- Mirarab S. 2019. *Species tree estimation using ASTRAL: practical considerations*. [WWW document] URL <https://arxiv.org/abs/1904.03826> [accessed 12 September 2021].
- Mosyakin SL, Tsybalyuk ZM. 2015a. Pollen morphology of tribes Aptosimeae and Myoporaceae supports the phylogenetic pattern in early-branching Scrophulariaceae revealed by molecular studies. *Willdenowia* 45: 209–222.
- Mosyakin SL, Tsybalyuk ZM. 2015b. Pollen morphology of the Southern African tribe Teedieae, an early-branching lineage of crown Scrophulariaceae. *Willdenowia* 45: 65–75.
- Navarro-Pérez ML, López J, Fernández-Mazuecos M, Rodríguez-Riño T, Vargas P, Ortega-Olivencia A. 2013. The role of birds and insects in pollination shifts of *Scrophularia* (Scrophulariaceae). *Molecular Phylogenetics and Evolution* 69: 239–254.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.
- Niezgodá C, Tomb AS. 1975. Systematic palynology of tribe Leucophylleae (Scrophulariaceae) and selected Myoporaceae. *Pollen and Spores* 17: 495–516.
- Olmstead RG, de Pamphilis CW, Wolfe AD, Young ND, Elisons WJ, Reeves PA. 2001. Disintegration of the Scrophulariaceae. *American Journal of Botany* 88: 348–361.
- Olmstead RG, Reeves PA. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast rbcL and ndhF sequences. *Annals of the Missouri Botanical Garden* 82: 176–193.
- Ortega-Olivencia A, Rodríguez-Riño T, Pérez-Bote JL, López J, Mayo C, Valtueña FJ, Navarro-Pérez M. 2012. Insects, birds and lizards as pollinators of the largest-flowered *Scrophularia* of Europe and Macaronesia. *Annals of Botany* 109: 153–167.
- Oxelman B, Kornhall P, Olmstead RG, Bremer B. 2005. Further disintegration of Scrophulariaceae. *Taxon* 54: 411–425.
- Paradis E, Schliep K. 2019. APE 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.
- Pokorny L, Riina R, Mairal M, Meseguer AS, Culshaw V, Cendoya J, Serrano M, Carbajal R, Ortiz S, Heuertz M *et al.* 2015. Living on the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. *Frontiers in Genetics* 6: 1–15.
- R Core Team. 2022. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL <https://www.R-project.org/>
- Rahmanzadeh R, Müller K, Fischer E, Bartels D, Borsch T. 2005. The Linderniaceae and Gratiolaceae are further lineages distinct from the Scrophulariaceae (Lamiales). *Plant Biology* 7: 67–78.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using TRACER 1.7. *Systematic Biology* 67: 901–904.
- Ramírez-Barahona S, Sauquet H, Magallón S. 2020. The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution* 4: 1232–1238.
- Raxworthy CJ, Forstner MRJ, Nussbaum RA. 2002. Chameleon radiation by oceanic dispersal. *Nature* 415: 784–787.
- Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.
- Refulio-Rodríguez NF, Olmstead RG. 2014. Phylogeny of Lamiales. *American Journal of Botany* 101: 287–299.
- Renner SS, Schaefer H. 2010. The evolution and loss of oil-offering flowers: new insights from dated phylogenies for angiosperms and bees. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 365: 423–435.
- Richmond GS. 1993. A review of the use of *Eremophila* (Myoporaceae) by Australian Aborigines. *Journal of the Adelaide Botanic Gardens* 15: 101–107.
- Richmond GS, Ghisalberti EL. 1995. Cultural, food, medicinal uses and potential applications of *Myoporum* species (Myoporaceae). *Economic Botany* 49: 276–285.
- Sage RF, Christin PA, Edwards EJ. 2011. The C<sub>4</sub> plant lineages of planet Earth. *Journal of Experimental Botany* 62: 3155–3169.
- Sanmartín I. 2011. *A paleogeographic history of the southern hemisphere*. [WWW document] URL [http://digital.csic.es/bitstream/10261/34831/1/2005\\_Sanmartin\\_Paleo.pdf](http://digital.csic.es/bitstream/10261/34831/1/2005_Sanmartin_Paleo.pdf) [accessed 1 July 2022].
- Schäferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G, Müller KF. 2010. Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. *BMC Evolutionary Biology* 10: 352.
- Scheunert A, Heubl G. 2017. Against all odds: reconstructing the evolutionary history of *Scrophularia* (Scrophulariaceae) despite high levels of incongruence and reticulate evolution. *Organisms Diversity & Evolution* 17: 323–349.
- Shee ZQ, Frodin DG, Cámara-Leret R, Pokorny L. 2020. Reconstructing the complex evolutionary history of the Papuanian *Schefflera* radiation through herbariomics. *Frontiers in Plant Science* 11: 258.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.
- Smith SA, Beaulieu JM, Stamatakis A, Donoghue MJ. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *American Journal of Botany* 98: 404–414.
- Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS ONE* 13: e0197433.
- Smith SA, O'Meara BC. 2012. TREEPL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.
- Spalink D, Drew BT, Pace MC, Zaborsky JG, Li P, Cameron KM, Sytsma KJ. 2016. Evolution of geographical place and niche space: patterns of diversification in the North American sedge (Cyperaceae) flora. *Molecular Phylogenetics and Evolution* 95: 183–195.
- Stadler T. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* 261: 58–66.
- Stull GW, Pham KK, Soltis PS, Soltis DE. 2023. Deep reticulation: the long legacy of hybridization in vascular plant evolution. *The Plant Journal*. doi: 10.1111/tpj.16142.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4: vey016.
- Tank DC, Beardsley PM, Kelchner SA, Olmstead RG, Tank DC, Beardsley PM, Kelchner SA, Olmstead RG. 2006. Review of the systematics of Scrophulariaceae s.l. and their current disposition. *Australian Systematic Botany* 19: 289–307.
- Tribble CM, Freyman WA, Landis MJ, Lim JY, Barido-Sottani J, Kopperud BT, Höhna S, May MR. 2022. REVGADGETS: an R package for visualizing Bayesian phylogenetic analyses from REVBAYES. *Methods in Ecology and Evolution* 13: 314–323.
- Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K. 2012. Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiology* 158: 590–600.
- Van Tieghem MP. 1903. Structure de l'étamine chez les Scrophulariacées. *Annales des Sciences Naturelles* 17: 363–371.
- Villaverde T, Pokorny L, Olsson S, Rincón-Barrado M, Johnson MG, Gardner EM, Wickett NJ, Molero J, Riina R, Sanmartín I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.
- Von Wettstein R. 1891. Scrophulariaceae. In: Engler A, Prantl K, eds. *Die Natürlichen Pflanzenfamilien, vol. 4, part 3b*. Leipzig, Germany: Wilhelm Engelmann, 39–107.
- Walker JF, Walker-Hale N, Vargas OM, Larson DA, Stull GW. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* 7: e7747.
- Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T *et al.* 2020. TREEIO: an R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution* 37: 599–603.
- Yoder AD, Nowak MD. 2006. Has vicariance or dispersal been the predominant force in biogeographic Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics* 37: 405–431.
- Young ND, Steiner KE, dePamphilis CW. 1999. The evolution of parasitism in Scrophulariaceae/Orobanchaceae: plastid gene sequences refute an evolutionary transition series. *Annals of the Missouri Botanical Garden* 86: 876–893.

- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8: 28–36.
- Zachos JC, Dickens GR, Zeebe RE. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451: 279–283.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.
- Zhao F, Liu B, Liu S, Min D-Z, Zhang T, Cai J, Zhou X-X, Chen B, Olmstead RG, Xiang C-L *et al.* 2022. Disentangling a 40-year-old taxonomic puzzle: the phylogenetic position of *Mimulicalyx* (Lamiales). *Botanical Journal of the Linnean Society* 201: 135–153.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** Tanglegram comparison of Scrophulariaceae nuclear and plastid tree topologies inferred in IQ-TREE.

**Fig. S2** Concordance factor estimated from the Scrophulariaceae nuclear dataset.

**Fig. S3** Concordance factor estimated from the Scrophulariaceae plastid dataset.

**Fig. S4** Divergence time estimation analysis in Scrophulariaceae.

**Fig. S5** Divergence time estimation analysis in Scrophulariaceae showing 95% highest posterior density intervals.

**Fig. S6** Marginal character stochastic mapping from the Bayesian Dispersal–Extinction–Cladogenesis biogeographic analysis for the Scrophulariaceae dataset.

**Fig. S7** Posterior stochastic mapping from the Bayesian Dispersal–Extinction–Cladogenesis biogeographic analysis for the Scrophulariaceae dataset.

**Methods S1** Extended description of biogeographic and diversification analyses in Scrophulariaceae.

**Table S1** Synoptical classification of Scrophulariaceae.

**Table S2** Materials included in our Scrophulariaceae nuclear and plastid dataset, including voucher information, and SRA numbers from NCBI.

**Table S3** Loci selected for Scrophulariaceae nuclear analyses.

**Table S4** List of paralogs removed from the Scrophulariaceae nuclear analyses.

**Table S5** List of coding regions used as reference in the Scrophulariaceae plastid analyses in HYBPIPER.

**Table S6** List of the 36 selected loci in Scrophulariaceae with the gene-shopping approach.

**Table S7** Recovery success and sequence length of the Scrophulariaceae nuclear-targeted loci.

**Table S8** Summary statistics for each Scrophulariaceae nuclear matrix.

**Table S9** Recovery success and sequence length of the Scrophulariaceae plastid-targeted loci.

**Table S10** Summary statistics for the Scrophulariaceae plastid matrices.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.