



The phylogeny and global biogeography of Primulaceae based on high-throughput DNA sequence data

Drew A. Larson^{a,b,*}, Andre S. Chanderbali^c, Olivier Maurin^d, Deise J.P. Gonçalves^a, Christopher W. Dick^{a,e}, Douglas E. Soltis^c, Pamela S. Soltis^{c,f}, Peter W. Fritsch^g, James J. Clarkson^d, Aurélie Grall^h, Nina M.J. Davies^d, Isabel Larridon^d, Izaí A.B.S. Kikuchiⁱ, Félix Forest^{d,#}, William J. Baker^{d,#}, Stephen A. Smith^{a,#}, Timothy M.A. Utteridge^{d,#}

^a Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

^b Department of Biology, Indiana University, Bloomington, IN 47405, USA

^c Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

^d Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, United Kingdom

^e Smithsonian Tropical Research Institute, Panama City, Republic of Panama

^f Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA

^g Botanical Research Institute of Texas, Fort Worth, TX 76107, USA

^h Department of Environmental Sciences - Botany, University of Basel, Switzerland

ⁱ Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

ARTICLE INFO

Keywords:

Ancestral state reconstruction

Angiosperms353

Ardisia

Ericales

Myrsinoideae

Phylogenomics

Target enrichment sequencing

ABSTRACT

The angiosperm family Primulaceae is morphologically diverse and distributed nearly worldwide. However, phylogenetic uncertainty has obstructed the identification of major morphological and biogeographic transitions within the clade. We used target capture sequencing with the Angiosperms353 probes, taxon-sampling encompassing nearly all genera of the family, tree-based sequence curation, and multiple phylogenetic approaches to investigate the major clades of Primulaceae and their relationship to other Ericales. We generated dated phylogenetic trees and conducted broad-scale biogeographic analyses as well as stochastic character mapping of growth habit. We show that *Ardisia*, a pantropical genus and the largest in the family, is not monophyletic, with at least 19 smaller genera nested within it. Neotropical members of *Ardisia* and several smaller genera form a clade, an ancestor of which arrived in the Neotropics and began diversifying about 20 Ma. This Neotropical clade is most closely related to *Elingamita* and *Tapeinosperma*, which are most diverse on islands of the Pacific. Both *Androsace* and *Primula* are non-monophyletic by the inclusion of smaller genera. Ancestral state reconstructions revealed that there have either been parallel transitions to an herbaceous habit in Primuloideae, *Samolus*, and at least three lineages of Myrsinoideae, or a common ancestor of nearly all Primulaceae was herbaceous. Our results provide a robust estimate of phylogenetic relationships across Primulaceae and show that a revised classification of Myrsinoideae and several other clades within the family is necessary to render all genera monophyletic.

1. Introduction

The angiosperm family Primulaceae comprises >2600 species, distributed nearly worldwide and representing diverse ecologies (Stevens, 2001 onward). In the APG IV (2016) system followed here, Primulaceae is usually regarded as having four subfamilies, i.e., Primuloideae, Myrsinoideae, Theophrastoideae, and Maesoideae (Notes

S1). Morphological diversity within the family ranges from tropical trees (e.g., *Ardisia copelandii* Mez) to lianas, woody shrublets, alpine cushion plants, and herbs (Fig. 1). Species of several genera, including *Androsace* L., *Cyclamen* L., and *Primula* L., are cultivated as ornamentals. Others, including *Ardisia* Sw., *Embelia* Burm.f., *Lysimachia* L., and *Myrsine* L., have a history of use in traditional medicine (Quattrocchi, 2012). *Aegiceris* Gaertn. is widely distributed along coastal Indo-Malaya,

* Corresponding author at: Department of Biology, Indiana University, Bloomington, IN 47405, USA.

E-mail address: larsonda@umich.edu (D.A. Larson).

Joint senior authors.

<https://doi.org/10.1016/j.ympev.2023.107702>

Received 13 June 2022; Received in revised form 13 December 2022; Accepted 4 January 2023

Available online 11 February 2023

1055-7903/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Fig. 1. Examples of morphological diversity of Primulaceae a) *Primula vulgaris* Huds., b) *Samolus valerandi* L., c) *Coris monspeliensis* L., d) *Pleiomeris canariensis* (Willd.) A.DC., e) *Cyclamen creticum* (Dörfl.) Hildebr., f) *Soldanella villosa* Darracq, g) *Androsace laevigata* (A.Gray) Wendelbo, h) *Deherainia smaragdina* (Planch. ex Linden) Decne., i) *Loheria bracteata* Merr., j) habit of *Ardisia auriculata* Donn.Sm., k) ripening fruits of *Ardisia auriculata* with visible dot-shaped glands common in Primulaceae, l) the soboliferous shrublet habit and fruit of *Emblemantha urmulata* B.C. Stone, m) the soboliferous shrublet habit and fruit of *Emblemantha urmulata* B.C. Stone, n) inflorescence of *Cybianthus glaber* A.DC., o) inflorescence of *Jacquinia keyensis* Mez. Photo attribution: a) ©William Crochot CC BY-SA 4.0; b) Christian Fischer CC BY-SA 3.0; c) Retama CC BY-SA 4.0; d) Winahwaru CC BY-SA 4.0; e) H. Zell CC BY-SA 3.0; f) Cptcv CC BY-SA 2.5; g) Walter Siegmund CC BY-SA 3.0; h) Daderot public domain; i) Peter W. Fritsch; j-k,n) Frank Almeda; l-m) T.M.A. Utteridge; o) Hans Hillewaert CC BY-SA 3.0. Photos modified under the terms of their respective licenses.

Australia, and islands of the Pacific and is one of only 31 extant angiosperm genera with species that inhabit mangrove ecosystems (Duke, 2017). Some species of *Ardisia* and *Lysimachia* are considered troublesome invasives outside their native ranges (Muñoz and Ackerman, 2011). Species of the genera *Hottonia* L. and *Samolus* L. have evolved to live in wet and even aquatic habitats, and some species of *Samolus* are also salt-tolerant (Ståhl, 2004). This extensive diversity makes Primulaceae an exceptional system for studying evolutionary patterns in morphology, life history, and biogeography.

Knowledge on the evolutionary relationships among the clades of Primulaceae and their position within angiosperms has changed dramatically over the past three decades with the application of molecular phylogenetic methods (Notes S1). Despite many advances, notable gaps in understanding the diversification of subclades within the family remain. Recent studies have suggested that *Ardisia*, a pantropical genus and the largest in the family with ca. 700 species, is not monophyletic (Julius et al., 2021; Rose et al., 2018; Yang and Hu, 2022). These studies have focused primarily on Paleotropical species and have used data from only a few genes, mostly those from nuclear ribosomal and chloroplast regions. Thus, the relationships among Neotropical *Ardisia* and closely related genera and the biogeography of these clades remain largely unexplored. Morphologically based generic concepts have led some authors to suggest that several taxa in Myrsinoideae, including *Hymenandra* A.DC. ex Spach (Pipoly and Ricketson, 2000, 1999), have disjunct *amphi*-Pacific tropical distributions, but whether these represent recent *trans*-oceanic dispersal events, relicts of ancient boreotropical forests, or artifacts of taxonomic error have not been rigorously assessed. Lack of phylogenetic resolution among members of Myrsinoideae, including *Ardisiandra* Hook.f., *Coris* L., and *Stimpsonia* C. Wright ex A.Gray has limited the ability to identify transitions in morphology (Wanntorp et al., 2012), such as herbaceous versus woody habit and capsular versus drupaceous fruits. Whether *Oncostemum* A. Juss. is rendered paraphyletic by the inclusion of *Badula* Juss., both of which are endemic to islands of the Indian Ocean, is uncertain because of low phylogenetic support, thus obscuring the evolution of floral morphology in the clade (Bone et al., 2012; Strijk et al., 2014). Poor understanding of relationships among other morphologically similar taxa with overlapping ranges such as *Discocalyx* Mez and *Tapeinosperma* Hook.f. has also limited our understanding of evolution in those groups, including transitions among mating systems, the characters from which have historically been used to differentiate genera. There is thus a particular need for a phylogenetic investigation that includes broad sampling across the entirety of subfamily Myrsinoideae to advance understanding of the global biogeographic and morphological history of Primulaceae.

In this study, we produced the first broadly inclusive phylogenetic tree of Primulaceae using high-throughput DNA sequence data, generated with the Angiosperms353 universal probes (Baker et al., 2022, 2021; Johnson et al., 2019), to investigate the biogeographic history of the clade and its phylogenetic position in the broader Ericales. Beyond presenting an updated phylogenetic hypothesis for the position of Primulaceae within Ericales and the relationships among nearly all genera of Primulaceae, we sought to answer the following questions: 1) What is the biogeographic history of *Ardisia* and other morphologically similar woody genera of Myrsinoideae?, 2) Which lineages gave rise to various island endemic taxa, including *Badula* and *Oncostemum* (Indian Ocean Islands); *Heberdenia* Banks ex A.DC. and *Pleiomera* A.DC. (Canary Islands); and *Discocalyx*, *Elingamita* G.T.S.Baylis, *Loheria* Merr., and *Tapeinosperma* (Indo-Malaya, Australasia, and Pacific Islands)?, and 3) Where have transitions between an herbaceous versus woody habit occurred within the phylogeny of Primulaceae? We also discuss the implications of our results for the systematics of Primulaceae, particularly highlighting the need for a major generic revision of subfamily Myrsinoideae.

2. Materials and Methods

2.1. Sampling

We sampled specimens from several herbaria, including the Royal Botanic Gardens, Kew (K), the University of Michigan (MICH), the Florida Museum of Natural History (FLAS), and the New York Botanical Garden (NY; Table S1; Folk et al., 2021; Kates et al., 2021). We also obtained DNA aliquots for 97 specimens from the Kew DNA Bank. Many of our samples from K and the Kew DNA Bank represent genera from across Ericales and were included as part of the Plant and Fungal Trees of Life (PAFTOL) project (Baker et al., 2022). Most samples (ca. 65%) were collected from dried herbarium material; the rest were from preserved tissue samples collected in silica gel (ca. 3%) or previously extracted DNA aliquots (ca. 32%). The sampling comprised 324 individuals from Ericales, including 150 individuals from Primulaceae, representing 133 species from 49 of the 55 accepted genera of the family (Tables S1-S3; POWO, 2022). The sampling also included several previously recognized genera of Primulaceae, now usually considered synonyms (e.g., *Anagallis* L., *Dodecatheon* L., *Gentlea* Lundell, *Glaux* L., *Synardisia* (Mez) Lundell, *Trientalis* L.; Table S2; Manns and Anderberg, 2009; POWO, 2022). We did not sample *Evotrochis* Raf. (Firat and Lidén, 2021) or *Paralysimachia* F.Du, J.Wang & S.S.Yang (Du et al., 2016) and did not count these among the accepted genera of Primulaceae (Notes S2). The six accepted genera not sampled comprise 13 accepted species in total (Table S3; POWO, 2022): *Votschia* B.Ståhl (Theophrastoideae, one species, Panama), *Mangenotiella* M.Schmid (Myrsinoideae, one species, New Caledonia), *Solonia* Urb. (Myrsinoideae, one species, Cuba), and *Vegaea* Urb. (Myrsinoideae, one species, Hispaniola); we sequenced one member each of *Amblyanthus* A.DC. (Myrsinoideae, four species, Assam, India) and *Ctenardisia* Ducke (Myrsinoideae, five species, Neotropics) but excluded them from the final analyses because of low gene recovery (see Methods S1).

2.2. Sequencing, assembly, and sequence curation

DNA was extracted from tissue with various modified CTAB protocols as in Baker et al. (2022). Illumina sequencing libraries enriched for target gene regions were generated with the Angiosperms353 kit (Johnson et al., 2019). Enriched libraries were sequenced with various Illumina machines to generate either 150-base pair (bp) or 250-bp paired-end reads. Some samples were sequenced in multiple sequencing runs and all reads for these samples were concatenated into a single pair of read files per sample prior to analysis.

Reads were trimmed with Trimmomatic (Bolger et al., 2014) by using the options “ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:2:TRUE SLIDINGWINDOW:5:20 LEADING:20 TRAILING:20 MINLEN:36” to trim adapters and bases with a PHRED score of less than 20. Trimmed reads were spot checked with FastQC (Andrews, 2010). Forward and reverse reads that became unpaired because of trimming were combined into a single file per sample. Sequences for target loci were assembled by using forward, reverse, and unpaired reads with HybPiper v1.3.1 (Johnson et al., 2016) under default settings and a custom “353mega” target file designed for asterids (McLay et al., 2021). The HybPiper pipeline includes Exonerate (Slater and Birney, 2005), BLAST+ (Camacho et al., 2009), Biopython (Cock et al., 2009), BWA (Li and Durbin, 2009), SAMtools (Li et al., 2009), GNU Parallel (Tange, 2011), and SPAdes (Bankevich et al., 2012). Assembly summary statistics, intron data, and paralog information were generated with Python scripts included in HybPiper.

We employed various approaches to remove exon sequences that were short, paralogs, or suspected to have assembly errors. We removed all sequences shorter than 25% of the average length of sequences for that gene, those flagged as potential paralogs by HybPiper, and those from poor quality samples (Methods S1). We aligned sequences with the E-INS-i algorithm in MAFFT v7.310 (Katoh and Standley, 2013) and

inferred gene trees with IQ-TREE v1.6.12 (Nguyen et al., 2015). We trimmed long internal and terminal branches from gene trees (Yang and Smith, 2014) and realigned the corresponding sequences (Methods S1; Tables S4-S5). This resulted in curated exon data for 353 genes for species from across Ericales, which we herein refer to as the Ericales1-exon dataset. A supermatrix was constructed for this dataset with the pxcat command in phyx v1.2 (Brown et al., 2017).

2.3. Ericales-wide phylogenetic analyses

The Ericales1-exon supermatrix was used for maximum likelihood (ML) phylogenetic analyses as implemented in IQ-TREE. Independent tree searches and ultrafast bootstrapping (Hoang et al., 2018) with 1,000 replicates were conducted with Edge-linked-equal, Edge-linked-proportional, and Edge-unlinked partition models specified with the “-q,” “-spp,” and “-sp” flags, respectively (Chernomor et al., 2016). A separate model partition for each of the 353 genes was applied in each of these analyses. We considered 95% as the threshold for strong ultrafast bootstrap support (Minh et al., 2013). Gene trees for each exon alignment were estimated separately with IQ-TREE and used to generate a species tree with ASTRAL v5.6.2 (Zhang et al., 2018). We considered 0.95 and 0.70 as the thresholds for strong and moderate local posterior probability (LPP), respectively, for nodes in species trees estimated with ASTRAL (Sayari and Mirarab, 2016).

2.4. Primulaceae-specific dataset construction and phylogenetic analyses

To further investigate phylogenetic relationships within Primulaceae and test the effect of including data from introns, we generated two datasets using the 150 Primulaceae samples that met the threshold for gene recovery (Methods S1), with two samples of Ebenaceae and two of Sapotaceae as outgroup taxa. We retrieved intron data for all samples using the *intronerate.py* script included with HybPiper and used a modified script from Larson et al. (2021) to create a set of intron sequences with corresponding exon sequence in the Ericales1-exon set. We then created a subset of both the exon and intron sequences that included only the Primulaceae and outgroup samples with the pxrms command in phyx. Intron and exon sequences for each gene were aligned separately with the E-INS-i algorithm in MAFFT. Then, columns with > 50% missing data were removed from all intron alignments with the pxclsq command in phyx because introns had a high proportion of missing data, expected because of the nature of target capture sequencing. The resulting aligned exon data, which we refer to as the Prim2-exon dataset, were an exact subset of the Ericales1-exon dataset but with sequences realigned with MAFFT. The Prim3-intron dataset included the same aligned data from exons, as well as data from corresponding introns.

Gene trees for the Prim2-exon set were generated with IQ-TREE and the GTR + Γ model. Gene trees for the Prim3-intron set were generated by concatenating the intron and exon sequences for each gene and estimating a tree in IQ-TREE with separate GTR + Γ model partitions for the intron and exon data (i.e., two partitions per gene). A species tree was estimated for both sets of gene trees with ASTRAL. A supermatrix for each dataset was constructed by concatenating alignments for each gene. For each supermatrix, three tree searches and ultrafast bootstrapping were conducted with IQ-TREE, each with a separate model

partition for each of the 353 exon and intron sequences (where applicable) as above.

2.5. Selection of preferred phylogenetic model

In total, we generated four phylogenetic trees for each of our three datasets, one with ASTRAL, and three with ML and different models implemented in IQ-TREE. We chose the Edge-linked-proportional model as our preferred model for ML based on the Bayesian Information Criterion and Robinson-Foulds distances (Robinson and Foulds, 1981) and herein limit the discussion of supermatrix analyses to results that employed this model (see Methods S2; Tables S6-S8, referring to these phylogenetic trees as the ML tree for each dataset (i.e., the Ericales1-exon-ML, Prim2-exon-ML, and Prim3-intron-ML trees). For visualization, a phylogenetic tree of Ericales with one sample per family was produced with the Ericales1-exon-ML tree and the pxrmt command in phyx.

2.6. Time-calibrated phylogenetic trees

We generated ultrametric trees with branch lengths corresponding to time using the penalized likelihood method as implemented in the program treePL (Sanderson, 2002; Smith and O’Meara, 2012). We used this program because it is computationally efficient for large datasets such as ours. We time-calibrated three trees to account for topological uncertainty in our downstream analyses: the Ericales1-exon-ML, the Ericales1-exon-ASTRAL, and the Prim2-exon-ASTRAL trees. For ASTRAL trees, prior to dating, we re-estimated branch lengths (in units of substitutions per base pair) with the corresponding supermatrix for that dataset and IQ-TREE with the same settings as for ML trees. We consider the ML tree from the Ericales1-exon dataset (i.e., the Ericales1-exon-ML tree) our focal tree because its branch lengths were co-estimated with the tree topology. Prior to running treePL, we trimmed trees to include only samples from Primulaceae and the four selected outgroup taxa and trimmed duplicate samples representing the same species. Although the fossil record for Primulaceae is sparse, we used three fossil calibrations within Primulaceae (Boucher et al., 2016; Friis et al., 2010, 2021) as well as one secondary calibration for the stem age of the family (Magallón et al., 2015) to estimate divergence times for the three trees. We estimated uncertainty for the dated Ericales1-exon-ML tree by estimating divergence times on two sets of 100 bootstrapped trees (Table 1; Methods S3). All treePL analyses were conducted with a smoothing parameter of 100 and optimization of other parameters with the prime option.

2.7. Biogeographic dataset construction and analysis

We conducted biogeographic analyses using the DEC* model (Mas-sana et al., 2015; Ree and Smith, 2008) as implemented in BioGeoBEARS (Matzke, 2013) and based on the time-calibrated phylogenetic trees described above. We used the biogeographic realms and boundaries of Olson et al. (2001), i.e., 1) Nearctic, 2) Palearctic, 3) Neotropics, 4) Afrotropics, 5) Indo-Malaya, 6) Oceania, and 7) Australasia. We primarily based the scoring on the native range of each species as described in the Plants of the World Online database (POWO, 2022), but also considered additional information such as regional online Floras and the

Table 1

The fossils and secondary calibration used to calibrate divergence times in phylogenetic trees.

Clade	Min age (Ma)	Max age (Ma)	Calibration type	Citation
Crown <i>Androsace</i>	5.3	–	Fossil	Boucher et al. (2016)
Crown <i>Primula</i>	16.0	–	Fossil	Boucher et al. (2016)
Crown <i>Lysimachia</i>	28.1	–	Fossil	Boucher et al. (2016)
Crown Primulaceae	66.0	83.6	Fossil	Friis et al. (2010, 2021)
Stem Primulaceae	94.92	94.92	Secondary calibration	Magallón et al. (2015)

Global Biodiversity Information Facility (<https://www.gbif.org>), especially for species that occur near transition zones between realms. We generally excluded a species from a realm when only a small part of its range crossed a biogeographic boundary (Table S1). This situation occurred near the Nearctic/Neotropics boundary in North America and the Palearctic/Indo-Malaya boundary in east Asia. As an example, a species with a Caribbean distribution that also extended into northern Florida, USA, was considered only Neotropical rather than Nearctic as well.

2.8. Ancestral state reconstruction of growth habit

We conducted stochastic character mapping of growth habit with the `make.simmap` function in `phytools` v1.2-0 (Revell, 2012) and custom R scripts. We scored each species as either woody or herbaceous, with species exhibiting slight woodiness (e.g., herbaceous perennials with a woody base) scored as herbaceous. For each dated tree, we conducted stochastic character mapping with three models in a Bayesian framework, each with 100 replicates and the prior for the ancestral state of Sapotaceae + Ebenaceae + Primulaceae fixed to woody. The first was an equal rates (ER) model with equal transition probabilities between woody and herbaceous states, with the `Q="mcmc"` option such that the transition matrix was sampled from its posterior distribution with Markov-chain Monte Carlo. The second was an all rates different (ARD) model with the same settings, except that the probabilities for the two possible transitions were allowed to differ. The third was a “no reversal” (NR) model where the ancestral state of Sapotaceae + Ebenaceae + Primulaceae was fixed to woody and herbaceous to woody transitions were not possible. The NR model therefore requires that all ancestors of the woody members of Myrsinoideae are inferred to be woody and was included in order to illustrate the number of transitions to herbaceousness implied by this specific and hypothetical evolutionary history. To compare models, we generated maximum likelihood estimates for each model and compared Akaike information criterion (AIC) and corrected Akaike information criterion (AICc) scores.

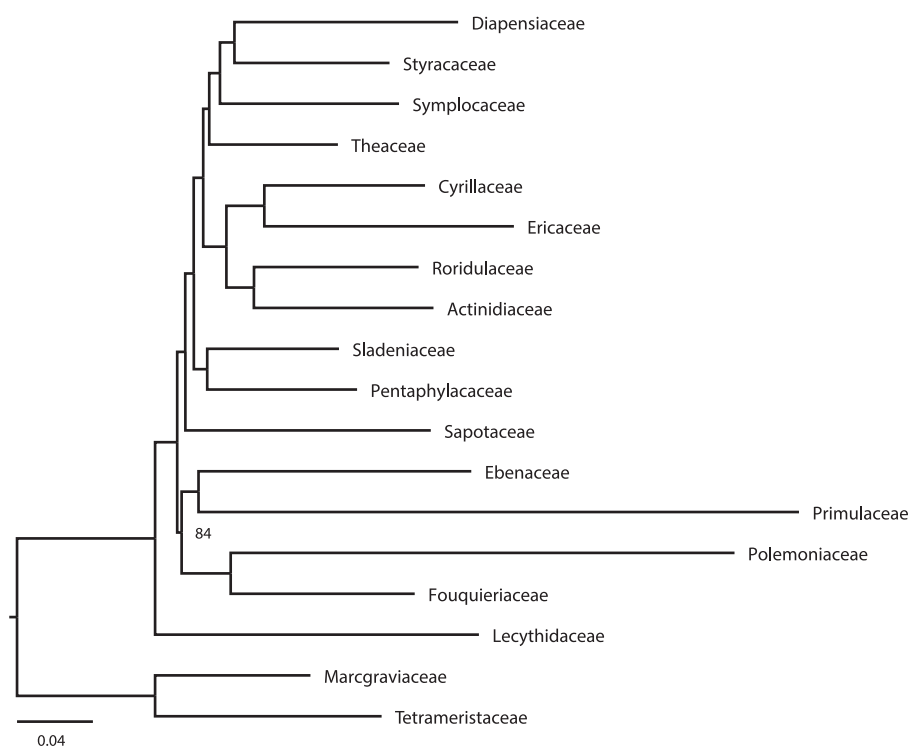


Fig. 2. Phylogenetic relationships among families of Ericales recovered in a maximum likelihood analysis with the Ericales1-exon dataset. All relationships had 100% ultrafast bootstrap support except the sister relationship between Primulaceae + Ebenaceae and Polemoniaceae + Fouquieriaceae, which received 84% support. The tree was rooted with Marcgraviaceae + Tetrameristaceae, which with Balsaminaceae (not sampled) forms a clade sister to the rest of Ericales. By rooting the tree this way, the length of the branch separating this clade from the rest of the tree is not informative. Other branch lengths are in units of estimated substitutions per site.

3. Results

3.1. Gene recovery

We recovered 105,115 assembled exon sequences using HybPiper, for an average of 298 sequences for each of the 353 target genes of the Angiosperms353 kit (including short sequences and those flagged as paralogs). After removing paralogs, short sequences, and low-quality samples, we retained 95,646 exon sequences and after phylogeny-based trimming, we retained 93,326 (264 per gene on average; Table S4-S5). Full occupancy for exons would be 324 samples each with 353 sequences or 114,372 total sequences; therefore, our Ericales1-exon dataset had an average gene occupancy of 81.6%. For the 150 Primulaceae and four outgroup taxa in the Prim3-intron dataset, we recovered 17,577 intron sequences (an average of 50 sequences per gene) of the 54,362 possible (32.3% occupancy). After trimming, the average length of exon and intron alignments were 1,144 bp and 627 bp, respectively (Table S4).

3.2. Phylogenetic relationships

3.2.1. Relationships among the major clades of Ericales

When not otherwise noted, phylogenetic relationships discussed refer to those recovered in the Ericales1-exon-ML tree (i.e., our focal tree). Primulaceae was sister to Ebenaceae in both the Ericales1-exon-ML (100% ultrafast bootstrap support) and Ericales1-exon-ASTRAL (0.90 LPP) trees. Not including the balsaminoid clade (Marcgraviaceae, Tetrameristaceae, and Balsaminaceae), which we used to root our Ericales-wide phylogenetic trees, Sapotaceae, Polemoniaceae + Fouquieriaceae + Primulaceae + Ebenaceae, and Lecythidaceae were successively sister to the clade formed by the other families of Ericales (Fig. 2). Relationships among all families of Ericales were supported by 100% ultrafast bootstrap support (Fig. 2), except for a sister relationship between Primulaceae + Ebenaceae and Polemoniaceae + Fouquieriaceae, which received 84% ultrafast bootstrap support in our focal tree and was not recovered in the Ericales1-exon-ASTRAL tree (Figure S3).

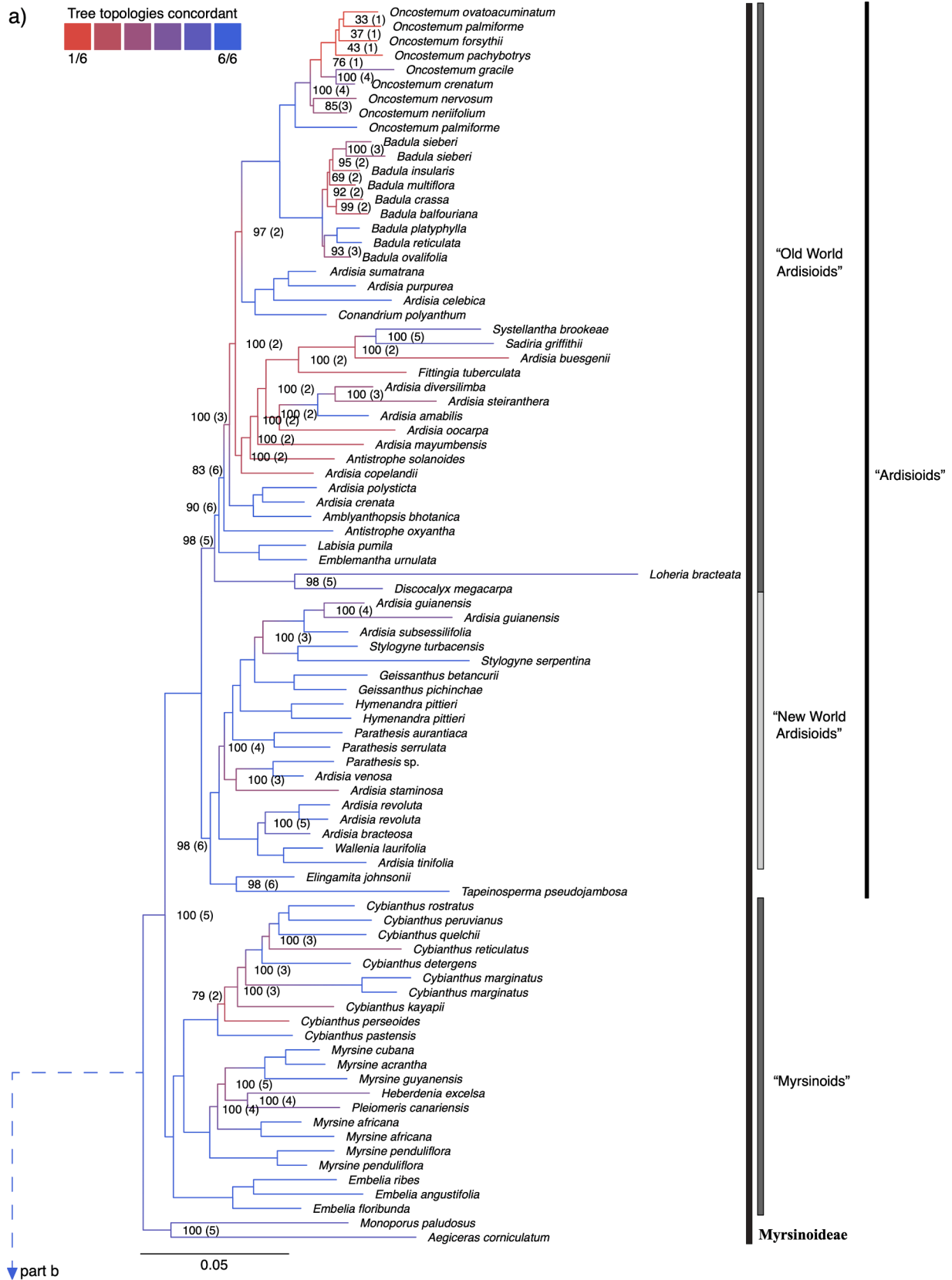


Fig. 3. Best phylogenetic tree of Primulaceae based on maximum likelihood analysis of the Ericales1-exon dataset (see Sections 2.2 - 2.5). a) The woody clade of Myrsinoideae, which comprises the Ardisioids, Myrsinoids, *Monoporus*, and *Aegiceras*. b) Relationships among the rest of Primulaceae. Values at nodes are IQ-TREE ultrafast bootstrap support. Values in parentheses and branch colors correspond to number of phylogenetic trees (the ML and ASTRAL trees for each of the three datasets; six trees in total) that were concordant with this topology. Support not shown for nodes with 100% ultrafast bootstrap support and six concordant trees. Branch lengths are in units of estimated substitutions per site. Outgroup taxa not depicted.

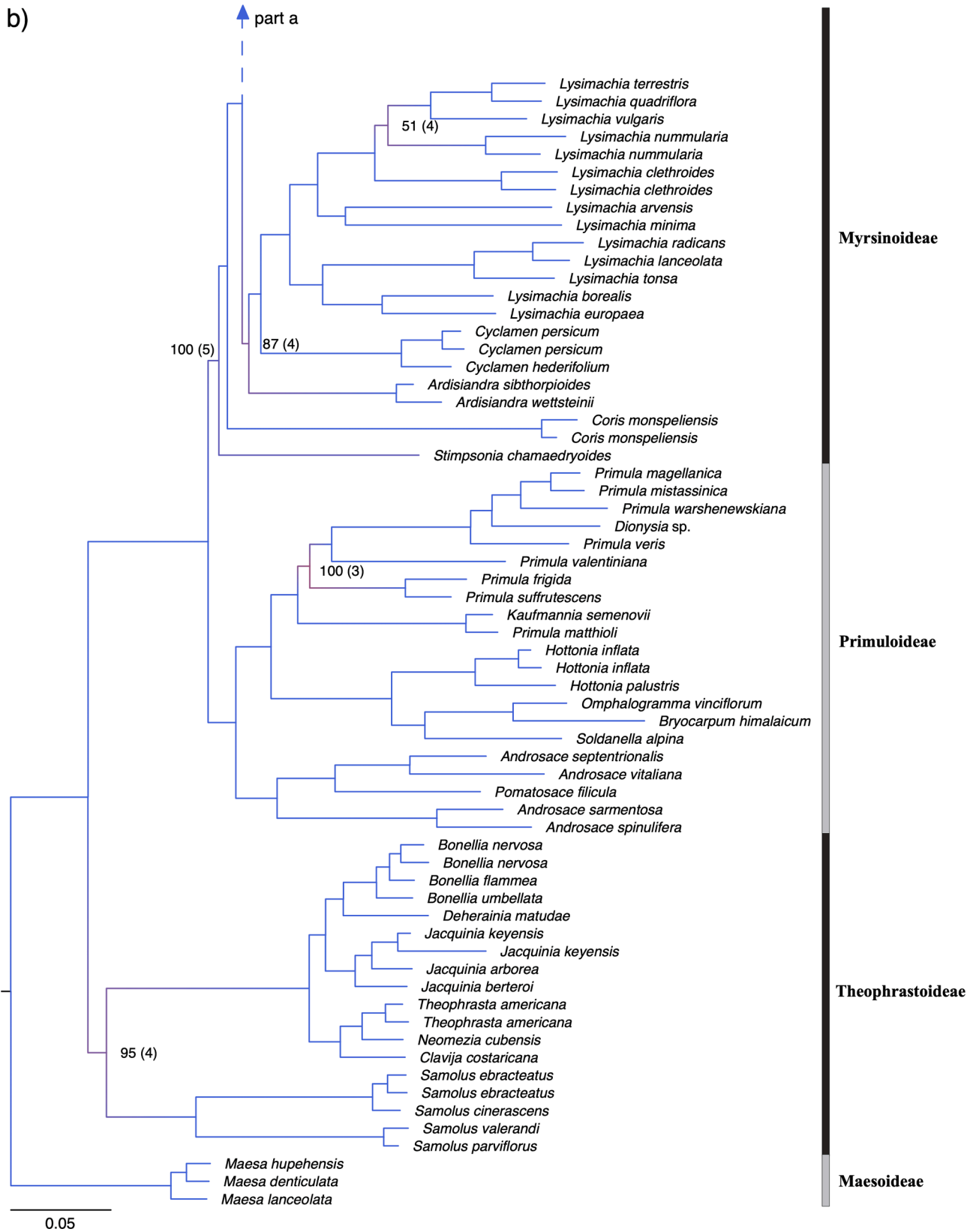


Fig. 3. (continued).

Within Primulaceae, Theophrastoideae were sister to Myrsinoideae + Primuloideae, and Maesoideae were sister to the rest of the family (Fig. 3).

3.2.2. Relationships within Myrsinoideae

The pantropical genus *Ardisia* was non-monophyletic in our results, with 19 other accepted genera interdigitated among its members (Figs. 3 and 4; Figures S1-S8). Because of this and other inferred relationships, it

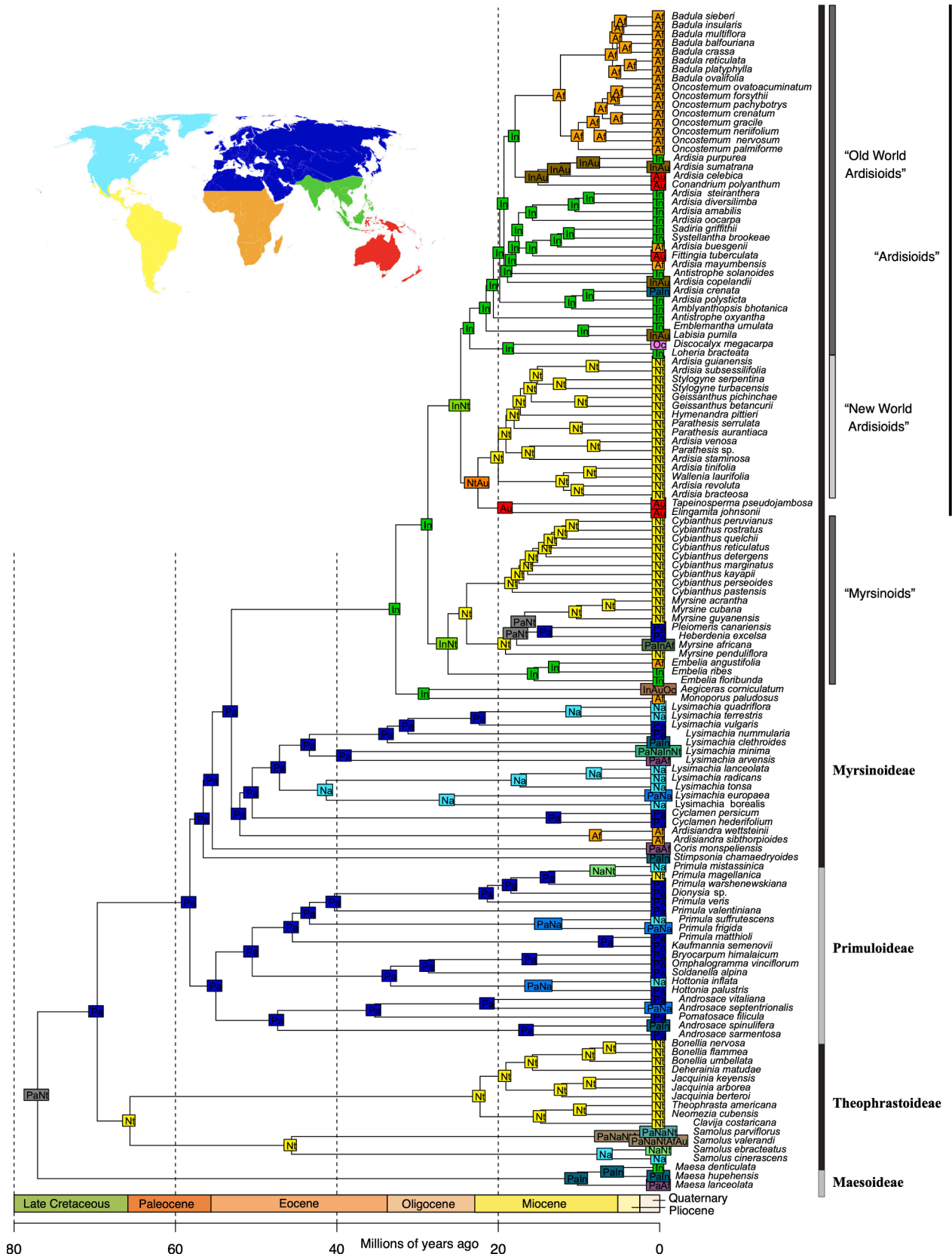


Fig. 4. Results of the biogeographic analysis based on the Ericales1-exon-ML tree, reduced to a single terminal per species. Node labels indicate the most likely state for that node in the DEC* model. Single area states (areas based on the biogeographic realms of Olson et al. (2001); see inset map): Nt = Neotropic (Yellow), Na = Nearctic (Light Blue), Pa = Palaearctic (Dark Blue), Af = Afrotropic (Orange), In = Indo-Malaya (Green), Oc = Oceania (Pink), Au = Australasia (Red). Map attribution to CarolSpears with modifications under the terms of the CC BY-SA 3.0 license.

is difficult to apply existing taxonomic names above the genus level. For example, the tribe Ardisieae as defined by Mez (1902), whose work still represents the most recent species-level systematic treatment of Myrsinoideae as a whole, includes *Aegiceras*, *Ardisia*, *Conandrium* Mez, *Heberdenia*, and *Hymenandra*. Mez's tribe Myrsineae contained most other woody Myrsinoideae, including *Badula*, *Loheria*, *Myrsine*, and *Tapeinosperma*. Both tribes are non-monophyletic in our results. Therefore, for clarity, we assign informal names to certain strongly supported clades of Myrsinoideae. We use the term "Ardisioids" to refer to the clade that includes *Ardisia* and *Tapeinosperma*. Within this clade, we refer to the "New World Ardisioids" as the clade that includes *Ardisia tinifolia* Sw. and *Stylogyne* A.DC. We refer to the clade that includes *Discocalyx* and *Badula* as the "Old World Ardisioids" (Fig. 3). The Old World Ardisioids does not include *Elingamita* and *Tapeinosperma* because in our results they appear to be more closely related to the New World Ardisioids. We use "Myrsinoids" to refer to the clade that includes *Myrsine*, *Cybianthus* Mart., and *Embelia*. This use of "Myrsinoids" is not equivalent to subfamily Myrsinoideae but is instead more restrictive, nor is it synonymous with tribe Myrsineae of Mez (Fig. 3). For the more inclusive clade that includes *Aegiceras*, *Monoporus* A.DC., and all *Ardisia*, we use the term "Woody Myrsinoideae" (Fig. 3).

The phylogenetic tree of Ardisioids displayed a clear geographic signal, and the sampled members of *Ardisia* were clearly differentiated as members of either the New World Ardisioids or Old World Ardisioids (Figs. 3 and 4). In all ML and ASTRAL trees, *Elingamita* and *Tapeinosperma* were sister to one another and formed a clade with the New World Ardisioids. All sampled members of *Ardisia* subgenus *Tinopsis* (i.e., *Ardisia sumatrana* Miq., *A. purpurea* Reinw. ex Blume, and *A. celebica* Scheff.), widespread across southeast Asia, formed a clade that was sister to the New Guinea endemic *Conandrium* in all analyses. All samples of *Ardisia* subgenus *Crispardisia* (*A. crenata* Sims and *A. polysticta* Miq.) formed a clade, as did samples of subgenus *Icacorea* (*A. guianensis* (Aubl.) Mez and *A. subsessilifolia* Lundell). The Neotropical subgenus *Ardisia* (*A. bracteosa* A.DC., *A. revoluta* Kunth, and *A. tinifolia*) formed a clade within which *Wallenia* Sw. was nested. The Paleotropical species *A. crenata* and *A. polysticta* formed a clade with *Amblyanthopsis* Mez in all trees.

Loheria and *Discocalyx* formed a clade that was sister to the other Old World Ardisioids in five of the six trees, but not in the Ericales1-exon-ASTRAL tree where *Loheria* was instead sister to all other Ardisioids. *Aegiceras* and the Madagascar endemic *Monoporus* formed a clade that was sister to all other members of the Woody Myrsinoideae in five of six trees. The exception was in the Prim2-exon-ASTRAL tree, where *Monoporus* was instead sister to the New World Ardisioids + *Elingamita* + *Tapeinosperma*. However, the branch corresponding to that sister relationship was near-zero in length and received only 0.42 LPP. *Geisanthus* Hook.f., *Parathesis* Hook.f., *Stylogyne*, and *Wallenia* were placed within the New World Ardisioids. *Parathesis* was not monophyletic, as one sample (*Parathesis* sp.) was nested within a small clade of *Ardisia*. We sampled one species of *Hymenandra*, the Neotropical *Hymenandra pittieri* (Mez) Pipoly & Ricketson, which was also nested within the New World Ardisioids. *Amblyanthopsis*, *Antistrophe* A.DC., *Badula*, *Conandrium*, *Discocalyx*, *Emblemantha* B.C.Stone, *Fittingia*, *Labisia* Lindl., *Loheria* (except in the Ericales1-exon-ASTRAL tree), *Oncostemum*, *Sadiria* Mez, and *Systellantha* B.C.Stone were placed within the Old World Ardisioids. Our two samples of *Antistrophe* were not sister in any tree. *Pleiomeris* and *Heberdenia* were nested within *Myrsine* except in the Ericales1-exon-ASTRAL tree (where *Pleiomeris* + *Heberdenia* was sister to all sampled *Myrsine*) and in the Prim2-exon-ASTRAL tree (where *Heberdenia* was sister to *Myrsine* + *Pleiomeris*). All ML trees recovered *Pleiomeris* and *Heberdenia* as nested within *Myrsine* with 100% ultrafast bootstrap support.

Lysimachia and *Cyclamen* were sister to each other in all six trees (Figures S1–S8). *Ardisiandra* was sister to *Lysimachia* + *Cyclamen* in four trees (87% ultrafast bootstrap support in our focal tree), but not in the Prim2-exon-ML or Prim3-intron-ASTRAL trees, in both of which

Ardisiandra, *Coris*, and *Stimpsonia* were successively sister to all other Myrsinoideae. *Stimpsonia* was sister to all other Myrsinoideae in all trees, except in the Prim2-exon-ASTRAL tree, where it was sister to Primuloideae, but with only 0.67 LPP.

3.2.3. Relationships within Theophrastoideae and Primuloideae

Within Theophrastoideae, tribe Theophrasteae Bartl. comprises all genera except *Samolus* (i.e., *Bonellia* Bertero ex Colla, *Clavija* Ruiz & Pav., *Deherainia* Decne., *Jacquinia* L., *Neomezia* Votsch, *Theophrasta* L., and the unsampled genus *Votschia*). Sampled members of tribe Theophrasteae formed a monophyletic group with two subclades, each with three genera. In all trees, *Jacquinia* was sister to *Bonellia* + *Deherainia*, whereas *Clavija* was sister to *Theophrasta* + *Neomezia*. In four trees, including our focal tree, *Samolus* was placed as sister to tribe Theophrasteae, whereas in the Ericales1-exon-ASTRAL and Prim2-exon-ASTRAL trees it was placed as sister to Primuloideae + Myrsinoideae (0.94 and 0.52 LPP, respectively).

Within Primuloideae, *Dionysia* Fenzl and *Kaufmannia* Regel were nested within *Primula* in all six trees. The topology within the subfamily was identical among all trees, except for one node within the clade formed by *Primula*, *Kaufmannia*, and *Dionysia* (Fig. 3). The monotypic *Pomatosace* Maxim. was nested within *Androsace* in all trees, and *Androsace* + *Pomatosace* was sister to the rest of Primuloideae. *Omphalogramma* Franch. and *Bryocarpum* Hook.f. & Thomson were sister to each other, with *Soldanella* L. and *Hottonia* successively sister to those (Notes S3).

3.3. Diversification times and biogeographic analyses

Biogeographic analyses with each of the three dated trees were generally similar (Fig. 4, Figures S14–S15); therefore, we highlight results of the analysis with our focal tree. Primulaceae were inferred to have begun diversifying 77.1 (76.3–77.7) Ma (Fig. 4) and to have had an ancestral area comprising both the Palearctic and Neotropics; however, there were several areas with similar marginal probabilities (Notes S4; Table S9). The crown ages of Maesoideae, Theophrastoideae, Primuloideae, and Myrsinoideae were estimated to be 10.2 Ma, 65.6 Ma, 55.0 Ma, and 56.6 Ma, respectively, and the associated uncertainty intervals were generally narrow (Figures S9–S13; Notes S5). The ancestral area of *Maesa* Forssk. was inferred to be the Palearctic and Indo-Malaya, with subsequent expansion into the Afrotropics. The ancestral areas of Myrsinoideae + Primuloideae and of Myrsinoideae + Primuloideae + Theophrastoideae were inferred to be the Palearctic. Within Primuloideae, *Primula* and *Hottonia* were inferred to have dispersed to the Nearctic by 13.7 (12.3–15.2) Ma and 14.9 (12.9–16.6) Ma, respectively. The most recent common ancestor (MCRA) of *Samolus* and tribe Theophrasteae was inferred to have been Neotropical, as was that of *Samolus*, with the genus later diversifying and expanding to its current, near-cosmopolitan distribution.

The ancestral area of Myrsinoideae was inferred to be the Palearctic, with *Lysimachia* colonizing the Nearctic by 41.3 (39.8–42.4) Ma. The ancestral area of the Woody Myrsinoideae was inferred to be Indo-Malaya, whereas the MRCA of the Ardisioids and that of the Myrsinoids were both inferred to have occupied the Indo-Malayan and Neotropical realms, implying two separate dispersal events to the Neotropics after these two lineages diverged. The crown age of the Ardisioids was inferred to be 24.7 (23.2–26.1) Ma, with the New World Ardisioids and Old World Ardisioids dating to 20.0 (18.6–21.0) Ma and 23.5 (21.5–24.8) Ma, respectively (Fig. 4). The Old World Ardisioids were inferred to have begun diversifying in the Indo-Malayan realm beginning 23.5 (21.5 – 24.8) Ma, with various lineages later dispersing to the Oceanian, Australasian, Afrotropic, and Palearctic realms. The ancestral area of the MRCA of *Elingamita*, *Tapeinosperma*, and the New World Ardisioids was inferred to be the Neotropical and Australasian realms (Fig. 3).

3.4. Ancestral state reconstruction of growth habit

Stochastic character mapping of ancestral growth habit suggested an herbaceous ancestor in the history of Myrsinoideae for all dated trees when either of the two models that allowed herbaceous to woody transitions (i.e., the ER and ARD models) were employed (Fig. 5, Figures S16–S24). For each tree, the ER model received the best AIC and AICc scores (Table S10) and we therefore highlight stochastic mapping results generated with this model. We report the percentage of stochastic mapping replicates with each state as the posterior probability of each state for a given node.

In the analysis with the Ericales1-exon-ML tree and the ER model (Fig. 5, left), the posterior probability for a woody MRCA of Primulaceae was 81%, whereas the MRCA of Primuloideae + Myrsinoideae (100%) and that of Theophrastoideae (63%) was inferred to be herbaceous, as was the MRCA of Theophrastoideae + Primuloideae + Myrsinoideae (64%). *Samolus* was inferred as sister to Primuloideae + Myrsinoideae in the Ericales1-exon-ASTRAL and Prim2-exon-ASTRAL trees: stochastic character mapping with either of these trees and the ER model recovered the MRCA of *Samolus* + Primuloideae + Myrsinoideae as herbaceous with posterior probabilities > 90% (Figures S17–S18). Stochastic character mapping with the NR model recovered transitions from woodiness to herbaceousness at least four times when considering the Prim2-exon-ASTRAL tree and at least five times for both the Ericales1-exon-ML and Ericales1-exon-ASTRAL trees (Fig. 5, right; Figures S22–S24).

4. Discussion

4.1. Phylogeny of Primulaceae

We performed the first phylogenetic analysis of Primulaceae based on high-throughput DNA sequence data and broad taxonomic sampling to address systematic and biogeographic questions across the family. We found a different topology for relationships among Primulaceae, Ebenaceae, Sapotaceae, Polemoniaceae + Fouquieriaceae, and Lecythidaceae as compared to those from several recent studies (Larson et al., 2020; Leebens-Mack et al., 2019; Stull et al., 2020; Zhang et al., 2020). Whereas our results offer a new hypothesis for the relationship of Primulaceae relative to other Ericales, it remains one hypothesis among many, reflecting the extensive phylogenomic conflict that has characterized various family-level divergences in the order (e.g., Larson et al., 2020).

Our results agree with those of previous studies regarding the relationships among the subfamilies of Primulaceae (e.g., Källersjö et al., 2000; Larson et al., 2020; Leebens-Mack et al., 2019; Rose et al., 2018), although support for a sister relationship between *Samolus* and tribe Theophrasteae was not definitive in our results. *Samolus* and species sampled from tribe Theophrasteae were sister in our focal tree (95% ultrafast bootstrap support), but two of our six trees (Ericales1-exon-ASTRAL and Prim2-exon-ASTRAL) placed *Samolus* sister to Primuloideae + Myrsinoideae rather than sister to tribe Theophrasteae. These results suggest that there is phylogenetic discordance among gene trees in our datasets, with a non-trivial number of genes supporting *Samolus* as

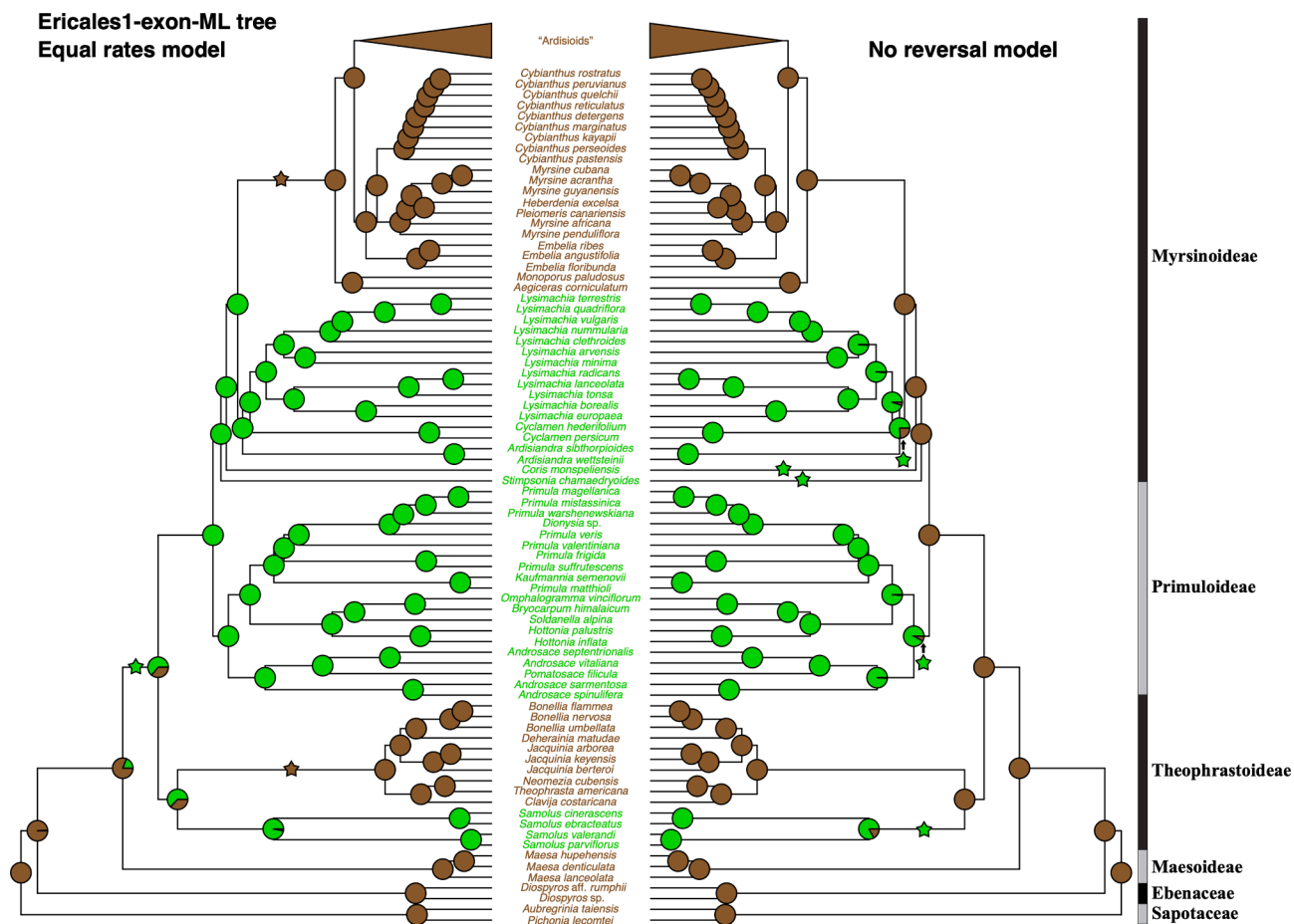


Fig. 5. Stochastic character mapping of habit with the focal tree under two different models of trait evolution, either with equal non-zero transition probabilities under an “equal rates” model (left) or only allowing transitions from woody to herbaceous habit under a “no reversal” model (right). Pie charts at nodes indicate the proportion of replicates in which woody (brown) and herbaceous (green) states were mapped at that node (i.e., the posterior probability of each state). Colored stars indicate branches along which the ancestral state with the highest posterior probability changed.

sister to Primuloideae + Myrsinoideae. These results could be due to biological causes such as high levels of incomplete lineage sorting or introgression, in addition to systematic error, including issues regarding accurately reconstructing nodes in individual gene trees corresponding to ancient divergences that occurred in rapid succession. The phylogenetic placement and taxonomic treatment of *Samolus* has historically been controversial. A morphological cladistic analysis by [Anderberg and Ståhl \(1995\)](#) placed *Samolus* as sister to most other Primulaceae *sensu stricto* (s.s.; now recognized as subfamily Primuloideae). Early molecular phylogenetic analyses suggested that *Samolus* did not form a clade with Primulaceae s.s. but was instead likely sister to tribe Theophrasteae ([Källersjö et al., 2000](#)). [Caris and Smets \(2004\)](#) conducted a detailed morphological and developmental analysis of *Samolus* and concluded that there are no “unambiguous morphological characters” that are synapomorphies for *Samolus* + tribe Theophrasteae. A possible synapomorphy is the presence of staminodes, which occur in both, but may develop through different ontogenetic pathways in the two groups ([Caris and Smets, 2004](#)). Furthermore, staminodes appear to be evolutionarily labile in Primulaceae because they are absent in some species of *Samolus* ([Wanntorp et al., 2012](#)) and are also present in *Soldanella* (Primuloideae) and sometimes in *Maesa* ([Anderberg et al., 1998](#); [Anderberg and Ståhl, 1995](#); [Friis et al., 2021](#)). Future studies should continue to investigate the strength of the evidence for a sister relationship between Theophrasteae and *Samolus*.

The phylogenetic placements of *Stimpsonia*, *Ardisiandra*, and *Coris* have historically been difficult to resolve based on morphology, although early molecular studies placed them within Myrsinoideae (formerly recognized as Myrsinoideae; [Källersjö et al., 2000](#); [Wanntorp et al., 2012](#)). The relationship of *Stimpsonia* as sister to all other Myrsinoideae is strongly supported overall by our results, although in the Prim2-exon-ASTRAL tree, the genus was recovered as sister to Primuloideae with weak support. The placement of *Ardisiandra* was somewhat uncertain; it was weakly supported as sister to *Cyclamen* + *Lysimachia* in our focal tree (87% ultrafast bootstrap support), as well as in three of five other trees, but not in the Prim2-exon-ML or Prim3-intron-ASTRAL trees, where it was sister to all other Myrsinoideae except *Coris* and *Stimpsonia*.

Our results suggest that *Pleiomeris* and *Heberdenia*, two monotypic genera endemic to the Canary Islands, are nested within *Myrsine*. However, in the Ericales1-exon-ASTRAL tree *Pleiomeris* + *Heberdenia* was sister to *Myrsine*, and in the Prim2-exon-ASTRAL tree *Heberdenia* was sister to *Myrsine* + *Pleiomeris*. [Yang and Hu \(2022\)](#) found that *Heberdenia* was nested within *Myrsine*, whereas the placement of *Pleiomeris* was inconclusive based on nuclear internal transcribed spacer (ITS) sequences. [Julius et al. \(2021\)](#) found that *Pleiomeris* was nested within *Myrsine* using ITS as well, but did not sample *Heberdenia*. [Appelhans et al. \(2020\)](#) reported the relationship among these genera as a polytomy and found evidence of extensive introgression among some members of *Myrsine*. The differences among ML and ASTRAL trees in our study could be due to discordant histories among genes caused by incomplete lineage sorting and/or introgression. Future studies should investigate the monophyly of *Myrsine* by analyzing the influence of individual gene regions on likelihood calculations and testing whether patterns of phylogenetic discordance might be due to introgression.

4.2. Relationships within the Ardisioids and their systematic implications

A particular strength of our sampling was for *Ardisia* and its close relatives that together form the “Ardisioids.” We sampled nearly all accepted genera of Myrsinoideae as well as the type species and several subgenera of *Ardisia* ([POWO, 2022](#)). Our results unequivocally demonstrate that *Ardisia* as currently circumscribed is not monophyletic, nor are the tribes Ardisieae and Myrsineae of [Mez \(1902\)](#). The non-monophyly of *Ardisia* has been suggested from recent investigations based on ITS sequences and other gene regions ([Julius et al., 2021](#); [Yang and Hu, 2022](#)), as well as an Ericales-wide biogeographic study ([Rose](#)

[et al., 2018](#)). We found that 19 genera were interdigitated with *Ardisia*, which together comprise a strongly supported clade ([Fig. 3](#)). Several of the five genera of Myrsinoideae that we did not sample, including *Amblyanthus*, *Ctenardisia*, *Mangenotiella*, and *Vegaea*, are likely nested within the Ardisioids as well because they share several morphological similarities with this group ([Schmid, 2011](#); [Ståhl and Anderberg, 2004](#)).

We found several strongly supported clades within the Ardisioids, some of which corresponded to previously described subgenera, a finding that parallels those of other recent studies. [Julius et al. \(2021\)](#) sampled 60 mostly Old World *Ardisia* taxa for their ITS phylogenetic analysis and found that, with some notable exceptions, many clades they recovered roughly correspond to the recognized subgenera of *Ardisia*. [Yang and Hu \(2022\)](#) conducted an even broader study of Old World *Ardisia* and their bacterial symbionts and found support for the monophyly of the *Ardisia* subgenera *Crispardisia*, *Pimelandra*, *Stylardisia*, and their newly circumscribed *Bladhia* s.s. and *Odontophylla* (resulting from the transfer of some members of subgenus *Bladhia sensu lato* to the new *Odontophylla*). [Yang and Hu \(2022\)](#) also found that the Old World members of *Hymenandra*, *Badula*, *Oncostemum*, and *Sadiria* were nested within *Ardisia* and referred to that clade as the “*Ardisia* generic complex.” Our diverse sampling of Ardisioids shows that this complex contains more currently recognized genera than has ever been shown before, because it includes all Neotropical members of the Ardisioids as well.

In all ML and ASTRAL trees, *Elingamita* and *Tapeinosperma* were sister to one another and formed a clade with the New World Ardisioids. This conflicts with the results of [Yan et al. \(2019\)](#), who sequenced whole chloroplast genomes from 13 Myrsinoideae and found two *Tapeinosperma* species (neither of which were sampled in our study) to be more closely related to three Old World *Ardisia* species than to *Elingamita* and two species of *Parathesis* (the only New World Ardisioids they sampled). This suggests either that the chloroplast genomes of these species have a different phylogenetic history than most regions of their nuclear genomes, or that *Tapeinosperma* is not monophyletic, with some species being more closely related to *Elingamita* and the New World Ardisioids and others to the Old World Ardisioids.

Based on our results and those from other studies, an extensive taxonomic revision of generic limits within the Ardisioids appears warranted. To circumscribe monophyletic genera, *Ardisia* could be combined with most other genera within Myrsinoideae to form a single megadiverse, globally distributed genus; this genus would consequently contain well over one third of all Primulaceae species. However, as we note above, there are many well-supported clades within the Ardisioids, many of which correspond to subgenera of *Ardisia*, display clear morphological and anatomical synapomorphies, and are restricted to particular geographical regions ([Yang and Hu, 2022](#)). Alternatively, the Ardisioids could be divided into several genera, an option that is perhaps more consistent with recent trends in taxonomic work on the group (e.g., [Drinkell and Utteridge, 2015](#); [Dubéarnès et al., 2015](#); [Julius et al., 2021](#); [Julius and Utteridge, 2012](#); [Yang and Hu, 2022](#)). In either case, such a revision would be an immense task, because *Ardisia* alone currently has over 700 accepted species ([POWO, 2022](#)) and additional phylogenetic studies with extensive sampling are required before any revision can be confidently undertaken. The Neotropical Jamaican endemic *A. tinifolia* is the type species of the genus; thus, dividing the Ardisioids into multiple genera would almost certainly result in a newly circumscribed *Ardisia* being restricted to the Neotropics. Many of the anatomical and structural characters historically used to describe genera within the Ardisioids appear to be homoplasious in the light of molecular phylogenetic studies ([Yang and Hu, 2022](#)). Thus, detailed morphological analysis will likely be necessary in tandem with broadly inclusive molecular studies of the group to identify which characters are synapomorphies for distinct clades and which represent evolutionary convergence. Combining phylogenomic datasets with existing data from ITS and chloroplast genes will likely be important for resolving the relationships among the nearly 1,000 recognized species of Ardisioids and

testing the monophyly of additional genera (Notes S6).

4.3. Diversification times and biogeographical insights

Our estimates for divergence times generally agree with those reported in other recent studies (reviewed by Stevens, 2001 onward) and had narrow uncertainty intervals, probably resulting from the large size of the Ericales1-exon supermatrix (Notes S5). We found that Primulaceae began diversifying ca. 77 Ma, during the Campanian (Late Cretaceous) and most likely occupied both the Palearctic and Neotropics (Notes S4). The fossil genus *Miranthus* E.M.Friis, P.R.Crane & K.R.Pedersen is known from the Palearctic (Friis et al., 2021), consistent with our inferred ancestral area of Primulaceae. Rose et al. (2018) conducted a biogeographic analysis that spanned Ericales and represents the most recent biogeographic hypothesis of Primulaceae based on phylogenetically broad sampling. They inferred the ancestral area of Primuloideae + Theophrastoideae + Myrsinoideae to be the Palearctic and Neotropics as well, whereas in our results the most likely area for that node was Palearctic only. Our results agree with those of Rose et al. (2018) that the MCRA of crown *Lysimachia* and other early ancestors of Myrsinoideae occupied the Palearctic, as did early Primuloideae before the latter began colonizing the Nearctic ca. 16 Ma. Boucher et al. (2016) investigated the diversification of three clades of alpine Primuloideae; we inferred older crown ages than Boucher et al. (2016) for *Primula* (ca. 46 vs 22 Ma), *Androsace* + *Pomatosace* (ca. 43 vs 26 Ma), and the clade that includes *Primula* and *Androsace* (ca. 55 vs 33 Ma), though for the latter two, our point estimate did fall within their estimated uncertainty interval (dates approximated from Figure 2 of that study). Differences in our estimates may be due to taxon sampling, differences in approach to time-calibration (penalized likelihood versus a Bayesian approach with a lognormal relaxed molecular clock), and sequence type (353 nuclear genes versus ITS and chloroplast genes). Our results agree with previous suggestions that *Samolus* originated in the Neotropics, as did Theophrastoideae (Rose et al., 2018; Stevens, 2001 onward). Compared to Rose et al. (2018), we inferred a relatively young crown age (ca. 9 – 11 Ma) for *Maesa* (Table S11), but this is likely an artifact of including only three samples from members of this genus in our study, resulting in the absence of species that would otherwise capture the oldest divergences in the clade.

4.3.1. Biogeographic history of the Ardisioids

The inferred ancestral area of the Old World Ardisioids was Indo-Malaya, from which various lineages later dispersed to Australasia, Oceania, and the Afrotropics. Rose et al. (2018) were not able to clearly resolve the biogeographic history of the Woody Myrsinoideae but suggested that some Neotropical *Ardisia* might form a clade, within which some Old World *Ardisia* (including *A. glauca* Mez and *A. speciosa* Blume) as well as *Elingamita*, *Parathesis*, *Stylogyne*, and some *Hymenandra* might be nested. We found that no Old World *Ardisia* were nested within the New World Ardisioids, which could be due to better phylogenetic resolution in our study, conflicting histories among nuclear and organellar genomes, or the fact that we did not sample those particular species. Our results agree with Rose et al. (2018) that the lineage subtending *Badula* and *Oncostemum* likely dispersed from Indo-Malaya to the Afrotropics ca. 10 Ma, although we inferred this date to be slightly older at ca. 12 Ma.

Based on the DEC* model, the ancestral area of the MRCA of the Ardisioids and Myrsinoids was Indo-Malaya. Our results suggest that after divergence from the Myrsinoids ca. 29 Ma, the Ardisioids dispersed to the Neotropics, such that the ancestor of all extant Ardisioids occupied both the Indo-Malayan and Neotropical realms ca. 25 Ma. The Indo-Malayan population then began diversifying into the Old World Ardisioids, whereas the Neotropical population dispersed to Australasia, such that the MRCA of the New World Ardisioids and *Elingamita* +

Tapeinosperma occurred in both the Neotropics and Australasia ca. 23 Ma. The Neotropical lineage then diversified into the New World Ardisioids, and the Australasian population diversified into *Tapeinosperma*, the center of diversity of which is now in New Caledonia and nearby Pacific Islands, and *Elingamita*, which is monotypic and endemic to New Zealand (Notes S6). We suspect that this model-based inference (i.e., that an Ardisioid lineage dispersed from Indo-Malaya to the Neotropics and then back to Australasia) is an artifact of not having sampled any extant Indo-Malayan *Tapeinosperma*, of which there are at least two species in Borneo (POWO, 2022). Not sampling additional Australasian *Tapeinosperma* could also have influenced the inferred ancestral range of the Ardisioids because we sampled only one of about 79 accepted species of that genus (POWO, 2022). Additionally, our biogeographic analysis did not include distance-based model components, which might better account for the fact that Australasia is directly adjacent to the Indo-Malayan realm whereas the Neotropics were at the time, and still are, several thousand miles away. A more parsimonious biogeographic hypothesis is that after the divergence of the lineage that became the Old World Ardisioids, an ancestor of the New World Ardisioids and *Elingamita* + *Tapeinosperma* occupied Indo-Malaya, from which members of this clade dispersed to Australasia and/or islands in the Pacific and from there dispersed to the Neotropics. Alternatively, dispersal may have occurred directly from Indo-Malaya trans-oceanically to the Neotropics, and separately to Australasia, after divergence from the Old World Ardisioids. Better sampling within *Tapeinosperma* and the Old World Ardisioids is needed to investigate which, if any, scenario is better supported by phylogenetic evidence and the distributions of extant species.

4.3.2. Disjunct distribution of *Hymenandra*

Pipoly and Ricketson (1999) transferred nine species of Neotropical *Ardisia* to the previously exclusively Paleotropical genus *Hymenandra* based on stamen morphology and plant architecture. This circumscription resulted in *Hymenandra* having a disjunct distribution between Indo-Malaya (eight species; Assam, Bangladesh, Borneo, Malaya, Myanmar, and south-central China) and the Neotropics (eight species; Colombia, Costa Rica, Nicaragua, and Panama; POWO, 2022). The results of Rose et al. (2018), based mostly on chloroplast data, suggested that an Old World species (*H. wallichii* A.DC.) and two Central American species (*H. calycosa* (Hemsl.) Pipoly & Ricketson and *H. sordida* (Lundell) Pipoly & Ricketson) are not closely related.

Although we did not sample any Old World members of *Hymenandra*, when viewed in the context of other recent studies, our results also suggest that the genus is not monophyletic. We found that the Central American species *H. pittieri* was nested within the New World Ardisioids. Julius et al. (2021) found that Bornean species *H. rosea* B.C.Stone and *H. beamanii* B.C.Stone were nested within a large clade consisting of *Ardisia* subgenera *Crispandisia*, *Pyrgus*, *Tinus*, and *Bladhia* as well as the genus *Sadiria*; we found strong support that *Sadiria* and a member of *Ardisia* subgenus *Crispandisia* (i.e., *Ardisia crenata*) are members of the Old World Ardisioids and not closely related to *H. pittieri*. Yang and Hu (2022) also found an Old World species of *Hymenandra* to be nested within Old World clades of *Ardisia* and noted that some morphological characters used to distinguish the two genera, including the degree to which the anther filaments are fused to one another and the corolla, have evolved multiple times in Myrsinoideae. The degree to which stamens appear as if they are fused in various members of the Ardisioids may also be influenced by whether the floral material is observed dried, preserved in alcohol, or fresh (T.M.A. Utteridge, personal observation). Julius et al. (2021) found that the Old World *Hymenandra* may themselves not be monophyletic. In light of this evidence, our results suggest that Neotropical and Paleotropical *Hymenandra* are not closely related and that their “disjunct distribution” is an artifact of the non-monophyly of the genus.

4.4. The evolution of habit in Primulaceae

Early circumscriptions of family limits within the primuloid clade used habit to separate the woody Myrsinoideae from herbaceous Primulaceae s.s., but this trait now appears to be homoplasious (Fig. 5; Anderberg and Ståhl, 1995; Källersjö et al., 2000). The MRCA of all Primulaceae was probably woody because the family is usually inferred to be sister to Ebenaceae (trees) or else placed along the backbone of Ericales, among which most families are woody (e.g., Rose et al., 2018; Larson et al., 2020). In addition, *Maesa* is strongly supported as sister to the rest of Primulaceae and are woody shrubs, trees, scrambling climbers, or lianas (Anderberg et al., 2000; Sumanon et al., 2021; Utteridge, 2012). In our ancestral state reconstructions, we fixed the prior for the habit of the MRCA of Primulaceae + Ebenaceae + Sapotaceae as woody, rather than employing an uninformative or empirically estimated prior. This likely influenced the inferred state of the MRCA of Primulaceae, which we found to be woody in all analyses. Future ancestral state analyses of habit that include additional taxa from across Ericales may provide better evidence and a more complete characterization of the remaining uncertainty regarding the habit of the MRCA of all extant Primulaceae.

Stochastic character mapping showed that the number of parameters included in the model affected the inferred ancestral state of some nodes in some cases, as did differences among phylogenetic trees (Figures S16–S24). However, our findings were consistent in the sense that, whenever the model allowed for herbaceous to woody transitions (i.e., analyses with either the ER or ARD model), the results suggested that an herbaceous ancestor gave rise to the Woody Myrsinoideae clade. Lens et al. (2005) conducted a morphological analysis of wood anatomy across Primulaceae and found no anatomical evidence in species of the Woody Myrsinoideae to suggest an herbaceous ancestor. Lens et al. (2005) did note, however, that the wood of Theophrastoideae (recognized as Theophrastaceae in that study) is characterized by very short vessel elements and fibers similar to those found in *Coris* and *Lysimachia* (both of which they thought evolved from herbaceous ancestors), as well as *Aegiceras* (mangroves).

We conducted stochastic character mapping with a “no reversal” model to demonstrate the hypothetical evolutionary history implied if it is assumed that the Woody Myrsinoideae have never had an herbaceous ancestor since the divergence of Primulaceae from other Ericales. For each of the three dated trees used in stochastic mapping, the no reversal scenario implies that there have been at least three transitions to an herbaceous habit within Myrsinoideae (separately along the branches leading to *Stimpsonia*, *Coris*, and *Ardisiandra* + *Cyclamen* + *Lysimachia*) and at least four or five transitions to herbaceousness across Primulaceae as a whole. In the Prim2-exon-ML and Prim3-intron-ASTRAL trees, which were not formally analyzed with ancestral state reconstructions, *Ardisiandra* was inferred to be sister to the Woody Myrsinoideae + *Cyclamen* + *Lysimachia*, a topology that implies four transitions in Myrsinoideae in the no reversal scenario: one in *Stimpsonia*, a second in *Ardisiandra*, a third in *Coris*, and a fourth in *Cyclamen* + *Lysimachia*. Habit is also tightly associated with fruit type across Myrsinoideae; although our results strongly suggest that they do not form a clade, all herbaceous Myrsinoideae share capsular fruits, whereas woody members of the subfamily instead have drupeaceous ones. Whether the association of these traits might be due to life-history tradeoffs influenced by lifespan or other traits associated with habit may warrant future research.

Because there appear to have been a relatively few transitions between woody and herbaceous habits in Primulaceae (e.g., a minimum of only three transitions are needed to explain habit evolution for species in our focal tree; Fig. 5), it may be difficult to accurately estimate transition rates for this clade. Increasing taxon sampling to more accurately represent the proportion of woody versus herbaceous species and capture additional transitions, such as the secondary gain of woodiness in Hawaiian *Lysimachia* (Hao et al., 2004), could help in parameterizing

models in future studies. A more nuanced categorization beyond the binary treatment of habit used here might also provide novel insights into the evolution of these traits. Future anatomical and/or gene expression studies may shed light on the mechanisms responsible for habit transitions in Myrsinoideae: such studies may also offer additional insights into the possibility that a common ancestor of some or all subfamilies of Primulaceae was herbaceous.

5. Conclusions

Our phylogenomic results were largely consistent with previous hypotheses of the evolutionary relationships among subfamilies and genera of Primulaceae but provide a considerable improvement in resolution and support. This refined phylogenetic hypothesis of Primulaceae highlights numerous taxonomic issues in the group and suggests several directions for future research into the evolution of morphology and mating systems in the clade. The methods used here, including the Angiosperms353 probes and tree-based alignment trimming, appear to be useful for resolving all but perhaps the most recent divergences in Primulaceae. Future studies with similar methods and even more extensive sampling would likely shed light on the monophyly of genera such as *Labisia*, *Tapeinosperma*, and *Discocalyx*, as well as clarify additional relationships within genera that have been shown to be non-monophyletic, including *Ardisia*, *Primula*, and *Androsace*. Such studies could also further clarify the biogeography of specific clades by providing a more complete characterization of their extant diversity.

The outlook for Primulaceae systematics is promising and rife with opportunities for advances across disciplines. Given recent methodological advances and reductions in sequencing costs, we consider the technology to build a complete tree of known Primulaceae species already within reach. Among the remaining major challenges for such an effort include the difficulty in sampling the appropriate material from herbaria located across the globe. Many herbaria have few digitized records, which severely limits the ability of investigators to plan sampling and verify records remotely. Accurate identification of specimens requires detailed taxonomic expertise and presents a further obstacle to complete species sampling. Broad collaboration among research groups and collections-based institutions, as well as the efficiencies realized by equitable data sharing and overlapping target genes, will all benefit future efforts towards a more comprehensive understanding of the Primulaceae phylogeny.

Author contributions

D.A.L. and T.M.A.U. conceived the study. D.A.L., T.M.A.U., P.W.F., D.J.P.G., C.W.D., D.E.S., P.S.S., S.A.S., A.G., and N.M.J.D. conducted the sampling. J.J.C. provided laboratory training to D.A.L. D.A.L. and A.S.C. performed the laboratory work with contributions from J.J.C. and I.A.B. S.K. W.J.B., F.F., and O.M. facilitated and supervised the data generation through PAFTOL. S.A.S., A.S.C., C.W.D., D.E.S., P.S.S., and I.L. facilitated and supervised the other data generation. D.A.L. designed and conducted the analyses, wrote the manuscript with input from all authors, and prepared the tables and figures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data, novel scripts and output files that support the findings of this study are openly available from Dryad and Zenodo at <https://doi.org/10.5061/dryad.1zcrjdfvb>. Raw sequence reads are available from NCBI: BioProject PRJNA839108, BioProject PRJNA847568, and

umbrella project PRJEB35285.

Acknowledgements

We thank K, MICH, FLAS, NY, and the Kew DNA bank for access to their collections: without the long-term efforts of countless collectors and curatorial staff, this study would not have been possible. We thank Alexandre Zuntini, Robyn Cowan, Vanessa Barber, and other members of the PAFTOL team for their technical and logistical help. We thank Bruce Martin for advice on modeling trait evolution and Jacob Berv, members of the Smith and Dick labs, and three anonymous reviewers for comments on an earlier version of the manuscript. This work was funded by grants from the Calleva Foundation to the Plant and Fungal Trees of Life Project (PAFTOL) at the Royal Botanic Gardens, Kew, U.S. National Science Foundation grants DEB 1917146, DEB 1458466, and DEB 1916555, and a fellowship to D.A.L. from the Department of Ecology and Evolutionary Biology at the University of Michigan.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at <http://doi.org/10.1016/j.ympev.2023.107702>.

References

- Anderberg, A.A., Ståhl, B., Källersjö, M., 2000. Maesaceae, a new primuloid family in the order Ericales s.l. *Taxon* 49, 183–187. <https://doi.org/10.1007/BF00984914>.
- Anderberg, A.A., Ståhl, B., 1995. Phylogenetic interrelationships in the order Primulales, with special emphasis on the family circumscriptions. *Can. J. Bot.* 73, 1699–1730. <https://doi.org/10.1139/b95-184>.
- Anderberg, A.A., Ståhl, B., Källersjö, M., 1998. Phylogenetic relationships in the Primulales inferred from rbcL sequence data. *Plant Syst. Evol.* 211, 93–102.
- Andrews, S., 2010. FastQC: a quality control tool for high throughput sequencing data. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- APG IV, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. <https://doi.org/10.1111/boj.12385>.
- Appelhans, M.S., Paetzold, C., Wood, K.R., Wagner, W.L., 2020. RADseq resolves the phylogeny of Hawaiian *Myrsine* (Primulaceae) and provides evidence for hybridization. *J. Syst. Evol.* <https://doi.org/10.1111/jse.12668>.
- Baker, W.J., Dodsworth, S., Forest, F., Graham, S.W., Johnson, M.G., McDonnell, A., Pokorny, L., Tate, J.A., Wicke, S., Wickett, N.J., 2021. Exploring Angiosperms353: an open, community toolkit for collaborative phylogenomic research on flowering plants. *Am. J. Bot.* 108, 1059–1065. <https://doi.org/10.1002/ajb2.1703>.
- Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Franco, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., McGinnis, C., Pokorny, L., Roy, S., Stone, M., Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J., Forest, F., 2022. A comprehensive phylogenomic platform for exploring the Angiosperm tree of life. *Syst. Biol.* 71, 301–319. <https://doi.org/10.1093/sysbio/syab035>.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bone, R.E., Strijk, J.S., Fritsch, P.W., Buerki, S., Strasberg, D., Thebaud, C., Hodkinson, T. R., 2012. Phylogenetic inference of *Badula* (Primulaceae), a rare and threatened genus endemic to the Mascarene Archipelago. *Bot. J. Linn. Soc.* 169, 284–296.
- Boucher, F.C., Zimmermann, N.E., Conti, E., 2016. Allopatric speciation with little niche divergence is common among alpine Primulaceae. *J. Biogeogr.* 43, 591–602. <https://doi.org/10.1111/jbi.12652>.
- Brown, J.W., Walker, J.F., Smith, S.A., 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33, 1886–1888. <https://doi.org/10.1093/bioinformatics/btx063>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Caris, P.L., Smets, E.F., 2004. A floral ontogenetic study on the sister group relationship between the genus *Samolus* (Primulaceae) and the Theophrastaceae. *Am. J. Bot.* 91, 627–643.
- Chernomor, O., von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65, 997–1008. <https://doi.org/10.1093/sysbio/syw037>.
- Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.L., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
- Drinkell, C., Utteridge, T.M.A., 2015. A revision of the genus *Systellantha* B.C.Stone. *Studies in Malaysian Myrsinaceae IV*. *Kew Bull.* 70, 50. <https://doi.org/10.1007/s12225-015-9603-8>.
- Du, F., Juan, W., Yang, S., 2016. *Paralysimachia* F. Du, J. Wang et S. S. Yang, a new genus of Primulaceae from Yunnan. *J. Southwest For. Coll.* 36, 91–94.
- Dubéarnès, A., Julius, A., Utteridge, T.M., 2015. A synopsis of the genus *Embelia* in Peninsular Malaysia and Singapore. *Studies in Malaysian Myrsinaceae III*. *Kew Bull.* 70, 25.
- Duke, N.C., 2017. Mangrove floristics and biogeography revisited: further deductions from biodiversity hot spots, ancestral discontinuities, and common evolutionary processes. In: *Mangrove Ecosystems: A Global Biogeographic Perspective*. Springer, pp. 17–53.
- Firat, M., Lidén, M., 2021. The genus *Evotrochis* (Primulaceae) resurrected. *Acta Biol. Turc.* 34, 161–168.
- Folk, R.A., Kates, H.R., LaFrance, R., Soltis, D.E., Soltis, P.S., Guralnick, R.P., 2021. High-throughput methods for efficiently building massive phylogenies from natural history collections. *Appl. Plant Sci.* 9, e11410.
- Friis, E.M., Pedersen, K.R., Crane, P.R., 2010. Cretaceous diversification of angiosperms in the western part of the Iberian Peninsula. *Iber. Floras Time Land Divers. Surviv.* 162, 341–361. <https://doi.org/10.1016/j.revpalbo.2009.11.009>.
- Friis, E.M., Crane, P.R., Pedersen, K.R., 2021. Early flowers of primuloid Ericales from the Late Cretaceous of Portugal and their ecological and phytogeographic implications. *Foss. Impr.* 77, 214–230.
- Hao, G., Yuan, Y.-M., Hu, C.-M., Ge, X.-J., Zhao, N.-X., 2004. Molecular phylogeny of *Lysimachia* (Myrsinaceae) based on chloroplast trnL-F and nuclear ribosomal ITS sequences. *Mol. Phylogenet. Evol.* 31, 323–339. [https://doi.org/10.1016/S1055-7903\(03\)00286-0](https://doi.org/10.1016/S1055-7903(03)00286-0).
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35, 518–522. <https://doi.org/10.1093/molbev/msx281>.
- Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. C., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4, 1600016. <https://doi.org/10.3732/apps.1600016>.
- Johnson, M.G., Pokorny, L., Dodsworth, S., Botigué, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T., Leebens-Mack, J.H., Leitch, I. J., Maurin, O., Soltis, D.E., Soltis, P.S., Wong, G.K., Baker, W.J., Wickett, N.J., 2019. An universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68, 594–606. <https://doi.org/10.1093/sysbio/syy0086>.
- Julius, A., Gutiérrez-Ortega, J.S., Sabran, S., Tagane, S., Naiki, A., Darnaedi, D., Aung, M., Dang, S., Nguyen, N., Binh, H., Watano, Y., Utteridge, T., Kajita, T., 2021. Phylogenetic relationship of tropical Asian *Ardisia* and relatives (Primulaceae) shows non-monophyly of recognized genera and subgenera. *J. Jpn. Bot.* 93, 149–165.
- Julius, A., Utteridge, T.M.A., 2012. Revision of *Ardisia* subgenus *Bladhia* in Peninsular Malaysia; studies in Malaysian Myrsinaceae I. *Kew Bull.* 67, 379–388. <https://doi.org/10.1007/s12225-012-9374-4>.
- Källersjö, M., Bergqvist, G., Anderberg, A.A., 2000. Generic realignment in primuloid families of the Ericales s.l.: a phylogenetic analysis based on DNA sequences from three chloroplast genes and morphology. *Am. J. Bot.* 87, 1325–1341. <https://doi.org/10.2307/2656725>.
- Kates, H.R., Doby, J.R., Siniscalchi, C.M., LaFrance, R., Soltis, D.E., Soltis, P.S., Guralnick, R.P., Folk, R.A., 2021. The effects of herbarium specimen characteristics on short-read NGS sequencing success in nearly 8000 specimens: Old, degraded samples have lower DNA yields but consistent sequencing success. *Front. Plant Sci.* 12.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Larson, D.A., Walker, J.F., Vargas, O.M., Smith, S.A., 2020. A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *Am. J. Bot.* 107, 773–789. <https://doi.org/10.1002/ajb2.1469>.
- Larson, D.A., Vargas, O.M., Vicentini, A., Dick, C.W., 2021. Admixture may be extensive among hyperdominant Amazon rainforest tree species. *New Phytol.* 232, 2520–2534.
- Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., 2019. One thousand plant transcriptomes and the phylogenomics of green plants.
- Lens, F., Jansen, S., Caris, P., Serlet, L., Smets, E., 2005. Comparative wood anatomy of the primuloid clade (Ericales s.l.). *Syst. Bot.* 30, 163–183.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., Hernández-Hernández, T., 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453.
- Manns, U., Anderberg, A.A., 2009. New combinations and names in *Lysimachia* (Myrsinaceae) for species of *Anagallis*, *Pelletiera* and *Trientalis*. *Willdenowia* 39, 49–54. <https://doi.org/10.3372/wi.39.39103>.

- Massana, K.A., Beaulieu, J.M., Matzke, N.J., O'Meara, B.C., 2015. Non-null effects of the null range in biogeographic models: Exploring parameter estimation in the DEC model. *bioRxiv* 026914. <https://doi.org/10.1101/026914>.
- Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R Package Version* 02 (1), 2013.
- McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L., Schmidt-Lebuhn, A.N., Baker, W.J., 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Appl. Plant Sci.* 9.
- Mez, C., 1902. Myrsinaceae. In: *Das Pflanzenreich*. Verlag von Wilhelm Engelmann, Leipzig, pp. 1–437.
- Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. <https://doi.org/10.1093/molbev/mst024>.
- Muñoz, M., Ackerman, J., 2011. Spatial distribution and performance of native and invasive *Ardisia* (Myrsinaceae) species in Puerto Rico: The anatomy of an invasion. *Biol. Invasions* 13, 1543–1558. <https://doi.org/10.1007/s10530-010-9912-7>.
- Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the world: A new map of life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51, 933–938. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2).
- Pipoly, J.J., Ricketson, J.M., 1999. Discovery of the Indo-Malesian genus *Hymenandra* (Myrsinaceae) in the Neotropics, and its boreotropical implications. *SIDA Contrib. Bot.* 18, 701–746.
- Pipoly, J.J., Ricketson, J.M., 2000. Discovery of *Ardisia* Subgenus *Acrardisia* (Myrsinaceae) in Mesoamerica: Another boreotropical element? *SIDA Contrib. Bot.* 275–283.
- POWO, 2022. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. URL <http://www.plantsoftheworldonline.org/> (last accessed 6.7.22).
- Quattrocchi, U., 2012. CRC world dictionary of medicinal and poisonous plants: Common names, scientific names, eponyms, synonyms, and etymology, 1st ed. CRC Press.
- Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14.
- Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223.
- Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147. [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2).
- Rose, J.P., Kleist, T.J., Löfstrand, S.D., Drew, B.T., Schönenberger, J., Sytsma, K.J., 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol. Phylogenet. Evol.* 122, 59–79. <https://doi.org/10.1016/j.ympev.2018.01.014>.
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>.
- Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. <https://doi.org/10.1093/molbev/msw079>.
- Schmid, M., 2011. Contribution to the knowledge of Primulaceae (ex Myrsinaceae) of New Caledonia. III. *Tapeinosperma* Hook. f. and *Mangenotiella* gen. nov. genera. *Adansonia* 34, 279–341.
- Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31. <https://doi.org/10.1186/1471-2105-6-31>.
- Smith, S.A., O'Meara, B.C., 2012. treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>.
- Ståhl, B., Anderberg, A.A., 2004. Myrsinaceae, in: Kubitzki, K. (Ed.), *Flowering Plants · Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 266–281. https://doi.org/10.1007/978-3-662-07257-8_40.
- Ståhl, B., 2004. Samolaceae, in: Kubitzki, K. (Ed.), *Flowering Plants · Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 387–389. https://doi.org/10.1007/978-3-662-07257-8_40.
- Stevens, P.F., 2001 onward. Angiosperm Phylogeny Website. Version 14. URL <http://www.mobot.org/MOBOT/research/APweb/> (last accessed 6.7.2022).
- Strijk, J.S., Bone, R.E., Thébaud, C., Buerki, S., Fritsch, P.W., Hodkinson, T.R., Strasberg, D., 2014. Timing and tempo of evolutionary diversification in a biodiversity hotspot: Primulaceae on Indian Ocean islands. *J. Biogeogr.* 41, 810–822.
- Stull, G.W., Soltis, P.S., Soltis, D.E., Gitzendanner, M.A., Smith, S.A., 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *Am. J. Bot.* 107, 790–805. <https://doi.org/10.1002/ajb2.1468>.
- Sumanon, P., Eiserhardt, W.L., Balslev, H., Utteridge, T.M.A., 2021. Six new species of *Maesa* (Primulaceae) from Papua New Guinea. *Phytotaxa* 505, 245–261.
- Tange, O., 2011. Gnu parallel—the command-line power tool. *USENIX Mag.* 36, 42–47.
- Utteridge, T.M.A., 2012. Four new species of *Maesa* Forssk. (Primulaceae) from Malaysia. *Kew Bull.* 67, 367–378. <https://doi.org/10.1007/s12225-012-9383-3>.
- Wanntorp, L., Ronse De Craene, L., Peng, C.-I., Anderberg, A.A., 2012. Floral ontogeny and morphology of *Stimpsonia* and *Ardisiandra*, Two aberrant genera of the primuloid clade of Ericales. *Int. J. Plant Sci.* 173, 1023–1035. <https://doi.org/10.1086/667607>.
- Yang, C.-J., Hu, J.-M., 2022. Molecular phylogeny of Asian *Ardisia* (Myrsinoideae, Primulaceae) and their leaf-nodulated endosymbionts, *Burkholderia* s.l. (Burkholderiaceae). *PLOS ONE* 17, e0261188.
- Yang, Y., Smith, S.A., 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* 31, 3081–3092. <https://doi.org/10.1093/molbev/msu245>.
- 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. <https://doi.org/10.1186/s12859-018-2129-y>.
- Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., Rees, M., Frohlich, M. W., Qi, J., Weigend, M., Ma, H., 2020. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol. Biol. Evol.* 37, 3188–3210. <https://doi.org/10.1093/molbev/msaa160>.