A phylogenomic perspective on gene tree conflict and character evolution in Caprifoliaceae using target enrichment data, with Zabelioideae recognized as a new subfamily

Hong-Xin Wang[a,#], Diego F. Morales-Briones[b,#], Michael J. Moore[c], Jun Wen[d], Hua-Feng Wang[a*]

[a] *Key Laboratory of Tropical Biological Resources of Ministry of Education, College of Tropical Crops, Hainan University, Haikou 570228, China*

[b] *Department of Plant and Microbial Biology, College of Biological Sciences, University of Minnesota, 140 Gortner Laboratory, 1479 Gortner Avenue, Saint Paul, MN 55108, USA*

[c] *Department of Biology, Oberlin College, Oberlin, OH44074, USA*

[d] *Department of Botany, National Museum of Natural History, MRC-166, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA*

*Authors for correspondence:

24 Hua-Feng Wang, E-mail: hfwang@hainanu.edu.cn

25 [#] Authors contributed equally to this study.

26

27

Abstract

The use of diverse datasets in phylogenetic studies aiming for understanding evolutionary histories of species can yield conflicting inference. Phylogenetic conflicts observed in animal and plant systems have often been explained by hybridization, incomplete lineage sorting (ILS), or horizontal gene transfer. Here, we employed target enrichment data and species tree and species network approaches to infer the backbone phylogeny of the family Caprifoliaceae, while distinguishing among sources of incongruence. We used 713 nuclear loci and 46 protein-coding sequences of plastome data from 43 samples representing 38 species from all major clades to reconstruct the phylogeny of the group using concatenation and coalescence approaches. We found significant nuclear gene tree conflict as well as cytonuclear discordance. Additionally, coalescent simulations and phylogenetic species network analyses suggest putative ancient hybridization among subfamilies of Caprifoliaceae, which seems to be the main source of phylogenetic discordance. Ancestral state reconstruction of six morphological characters revealed some homoplasy for each character examined. By dating the branching events, we inferred the origin of Caprifoliaceae at approximately 69.38 Ma in the late Cretaceous. By integrating evidence from molecular phylogeny, divergence times, and morphology, we herein recognize Zabelioideae as a new subfamily in Caprifoliaceae. This work shows the necessity to use a combination of multiple approaches to identify the sources of gene tree discordance. Our study also highlights the importance of using data from both nuclear and chloroplast genomes to reconstruct deep and shallow phylogenies of plants.

Keywords: Caprifoliaceae; Hybridization; Introgression, Phylogenetic networks, Zabelioideae.

1 Introduction

Gene tree discordance is common in the phylogenomic era (Galtier and Daubin, 2008; Degnan and Rosenberg, 2009; Szöllősi et al., 2015; Sun et al., 2015; Lin et al., 2019). Many studies have shown that incomplete lineage sorting (ILS), hybridization, and other processes such as horizontal gene transfer, gene duplication, or recombination, may be contributing to discordance among the gene trees (Degnan and Rosenberg, 2009; Linder and Naciri, 2015). Among these potential sources of discordance, hybridization is undoubtedly a research hotspot in plant systematics (e.g., Morales-Briones et al., 2018; Lee-Yaw et al., 2019; Morales-Briones et al., 2020a; Stull et al., 2020). Hybridization is especially prevalent in rapidly radiating groups, which is increasingly recognized as a major force in evolutionary biology, in many cases leading to new species and lineages (Mallet, 2007; Abbott et al., 2010; Yakimowski and Rieseberg, 2014; Konowalik et al., 2015). ILS is one of the prime sources of gene tree discordance, which has attracted increasing attention in the past decades as phylogenetic reconstruction methods allowed its modeling (Edwards 2009; Liu et al., 2015). Despite that, distinguishing ILS from hybridization is still challenging (Linder and Naciri, 2015). More recently, several methods that account simultaneously for ILS and hybridization haven been developed to estimate phylogenetic networks (Solís-Lemus and Ané, 2016; Wen et al., 2018). At the same time, the empirical studies using phylogenetic networks to identify the sources gene tree discordance are increasing (e.g., Morales-Briones et al., 2018, 2020a; Widhelm et al., 2019; Feng et al., 2020).

Caprifoliaceae *s.l.* sensu Angiosperm Phylogeny Group (APG) IV (APG, 2016; hereafter as Caprifoliaceae) is a woody family in the order Dipsacales containing 41 genera and ca. 960 species, with most genera restricted to eastern Asia and eastern North America (Manchester and Donoghue, 1995; Bell, 2004; APG, 2016). The family has long been the focus of studies of character evolution, especially regarding its tremendous diversity in reproductive structures (Backlund 1996; Donoghue et al.

2003). Caprifoliaceae has five corolla lobes and five stamens as ancestral states, which are retained in Diervilleae C. A. Mey., *Heptacodium* Rehd., and Caprifolieae (though in some *Symphoricarpos* Duhamel and *Lonicera* L. there are four corolla lobes and four stamens). However, for other genera, the number of stamens is reduced to four or even one. Caprifoliaceae shows even greater variation in fruit types (e.g., achene in *Abelia* R. Br., berry in *Lonicera*, drupe in *Viburnum* L.; Manchester and Donoghue, 1995; Donoghue et al., 2003). Some genera bear highly specialized morphological characters (e.g., the spiny leaf of *Acanthocalyx* (DC.) Tiegh., *Morina* L. and *Dipsacus* L.) that have likely played key roles in lineage-specific adaptive radiation (Blackmore and Cannon, 1983; Caputo and Cozzolino, 1994; Donoghue et al., 2003) (Fig. 1).

Phylogenetic relationships within Caprifoliaceae have been studied during the past two decades using plastid and nuclear DNA data (Fig. 2), but the placement of *Zabelia* (Rehder) Makino has never been resolved confidently using either morphological characters (Backlund, 1996; Donoghue et al., 2003) or molecular data (Donoghue et al., 1992; Jacobs et al., 2010; Smith et al., 2010; Landrein et al., 2012; Stevens, 2019; Xiang et al., 2019; Wang et al., 2020). Caprifoliaceae includes seven major clades: Linnaeoideae, *Zabelia*, Morinoideae, Valerianoideae, Dipsacoideae, Caprifolioideae and Diervilloideae (Donoghue et al., 1992; Jacobs et al., 2010; Smith et al., 2010; Landrein et al., 2012; APG, 2016; Stevens, 2019; Xiang et al., 2019; Wang et al., 2020). Based on nuclear (ITS) and chloroplast DNA (cpDNA) data (*trnK*, *matK*, *atpB-rbcL*, *trnL-F*) of 51 taxa, Jacobs et al. (2010) found moderate support (bootstrap support [BS] = 62%) for the placement of *Zabelia* (formerly part of *Abelia*) in a clade with Morinoideae, Dipsacoideae, and Valerianoideae. Based on the same data set, Jacobs et al. (2010) raised *Abelia* sect. *Zabelia* to the genus level as *Zabelia*, and more recent studies have confirmed the distinctiveness of *Zabelia* (Landrein et al., 2012; Wang et al., 2015), often finding it sister to Morinoideae, although with low (BS ≤ 50%) to moderate support (BS ≤ 50-70%) (Donoghue et al., 1992; Jacobs et al., 2010; Tank and Donoghue, 2010; Wang et al., 2015). Based on cpDNA data (*rbcL*, *trnL-K*, *matK* and *ndhF*) of 14 taxa, Landrein et al. (2012) suggested that *Zabelia* and *Diabelia* Landrein

(Linnaeoideae) had similar "primitive" inflorescences of reduced simple thyrses. Landrein et al. (2012) conducted phylogenetic analyses of the Caprifoliaceae based on the structural characters of reproductive organs. In these analyses, *Zabelia* was sister to the clade of Morinoideae, and Valerianoideae + Dipsacoideae. Recently, Xiang et al. (2019) carried out analyses of complete plastomes of 32 species in this clade, demonstrating that *Heptacodium* and *Triplostegia* Wall. ex DC. are members of Caprifoliaceae s. str. and Dipsacaceae, respectively. Furthermore, *Zabelia* was found to be the sister to Morinaceae in all analyses (Xiang et al., 2019). Moreover, using complete plastomes from 56 accessions representing 47 species of Caprifoliaceae, Wang et al. (2020) recovered the clade composed of Linnaeoideae, and Morinoideae + *Zabelia* as sister to Dipsacoideae + Valerianoideae) with strong support (BS = 100%).

In this study, we assembled and analyzed a custom target enrichment dataset of Caprifoliaceae to: (1) evaluate sources of gene tree discordance, in order to clarify the backbone phylogeny of Caprifoliaceae with special attention to positions of recalcitrant taxa (i.e., *Zabelia* and Morinoideae); and (2) determine the evolutionary patterns of key morphological characters of Caprifoliaceae.

2 Materials and methods

2.1 Taxon sampling

We sampled 43 individuals from 38 species of Caprifoliaceae, including representatives of all seven major clades (including *Zabelia*) of Caprifoliaceae sensu Stevens (2019) and Wang et al. (2020). Additionally, three species of Adoxaceae were included as outgroups. Most samples (38) were collected in the field where leaf tissue was preserved in silica gel. The remaining samples were obtained from the United States National Herbarium (US) at the Smithsonian Institution (Table S1). Vouchers of newly collected samples were deposited in the herbarium of the Institute of Tropical

Agriculture and Forestry (HUTB), Hainan University, Haikou, China. Complete voucher information is listed in Supporting Information Table S1.

2.2 DNA extraction, target enrichment, and sequencing

We extracted total genomic DNA from silica gel-dried tissue or herbarium tissue using the CTAB method of Doyle and Doyle (1987). We checked the quantity of each extraction with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and sonicated 400 ng of DNA using a Covaris S2 (Covaris, Woburn, MA) to produce fragments ~150-350 bp in length for library preparations. To ensure that genomic DNA was sheared at approximately the selected fragment size, we evaluated all samples on a 1.2% (w/v) agarose gel.

We identified putative single copy nuclear (SCN) genes with MarkerMiner v.1.2 (Chamala et al., 2015) with default settings, using the transcriptomes of *Dipsacus asper*, *Lonicera japonica*, *Sambucus canadensis*, *Valeriana officinalis*, and *Viburnum odoratissimum* from 1KP (Matasci et al., 2014), and the genome of *Arabidopsis thaliana* (L.) Heynh. (Gan et al., 2011) as a reference. SCN genes identified with MarkerMiner were further filtered using GoldFinder (Vargas et al., 2019) requiring loci with at least 400 bp and a coverage of at least three species. This resulted in 428 SCN for phylogenetic analyses. A custom set of 80 bp MYbaits biotinylated RNA baits based on exon sequences were manufactured by Arbor Biosciences (Ann Arbor, MI, USA), with a 2× tiling density. The bait sequences are available as a supplemental file (Appendix 1).

Library preparation was done with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA) following the manufacturer's protocol. Library concentrations were quantified using a Qubit 2.0, with a dsDNA HS Assay Kit (Thermo Fisher Scientific). Fragment size distribution was determined with a High Sensitivity D1000 ScreenTape run on the Agilent 2200 TapeStation system (Agilent

Technologies, Inc., Santa Clara, California, United States). Solution-based hybridization and enrichment with MYbaits followed Weitemier et al. (2014). We conducted a library spiking of 40% of unenriched library and 60% of enriched library for each sample. The final spiked library pools were sequenced by Novogene Corporation (Sacramento, California, U.S.A.) on one lane using the Illumina HiSeq X sequencing platform (Illumina Inc, San Diego, California, U.S.A.) producing 150 bp paired-end reads.

2.3 Read processing and assembly

Sequencing adapters and low-quality bases were removed with Trimmomatic v0.36 (ILLUMINACLIP: TruSeq_ADAPTER: 2:30:10 SLIDINGWINDOW: 4:5 LEADING: 5 TRAILING: 5 MINLEN: 25; Bolger et al. 2014). Assembly of nuclear loci was carried out with HybPiper v.1.3.1 (Johnson et al. 2016). Assemblies were carried on an exon basis to avoid chimeric sequences in multi-exon genes product of potential paralogy (Morales-Briones et al. 2018). Only exons with a reference length of ≥ 150 bp were assembled (1220 exons from 442 genes). Paralog detection was carried out for all exons with the 'paralog_investigator' option of HybPiper. All assembled loci (with and without paralogs detected) were processed following Morales-Briones et al. (2020b) to obtained 'monophyletic outgroup' (MO) orthologs (Yang and Smith, 2014).

Off-target reads from target enrichment were used for *de novo* assemblies of plastome with Fast-Plast (McKain., 2017). Resulting contigs from Spades v3.9.0 (Bankevich et al., 2012) were mapped to *Kolkwitzia amabilis* (Genbank accession no. NC_029874.1), with one copy of the Inverted Repeat removed. Mapped contigs manually edited in Geneious v.11.1.5 (Kearse et al. 2012) to produce the final oriented contigs. Contigs were further annotated using *K. amabilis* as a reference and coding sequences (CDS) were extracted using Geneious.

2.4 Phylogenetic analyses

We used concatenation and coalescent-based methods to reconstruct the phylogeny of Caprifoliaceae. We performed phylogenetic analyses on the nuclear and plastid CDS, separately. Individual nuclear exons were aligned with MAFFT version 7.407 (Katoh and Standley, 2013) and aligned columns with more than 90% missing data were removed using Phyutility (Smith and Dunn, 2008). A maximum likelihood (ML) tree was estimated from the concatenated matrix, partitioning by gene, using RAxML version 8.2.12 (Stamatakis, 2014) and the GTRGAMMA model for each partition. Clade support was assessed with 100 rapid bootstrap replicates. We also estimated a species tree with ASTRAL v5.7.1 (Zhang et al., 2018) from individual ML gene trees inferred using RAxML with a GTRGAMMA model. Local posterior probabilities (LPP; Sayyari and Mirarab, 2016) were used to assess clade support. Gene tree discordance was evaluated using two approaches. First, we mapped the 713 nuclear gene trees onto the species-tree phylogeny and calculated the internode certainty all (ICA; Salichos et al., 2014) and number of conflicting and concordant bipartitions on each node of the species trees using Phyparts (Smith et al., 2015). Then we used Quartet Sampling (QS; Pease et al., 2018) to distinguish strong conflict from weakly supported branches in the nuclear tree. We carried out QS with 1000 replicates.

Plastid CDS were aligned with MAFFT and then concatenated into a supermatrix. We reconstructed a phylogenetic tree using RAxML based on chloroplast data. We also use QS to investigate potential conflict in the chloroplast data set. QS was carried using 1000 replicates.

2.5 Assessment of hybridization

To test whether ILS alone could explain cytonuclear discordance, we used coalescent simulations similar to Folk et al. (2017) and García et al. (2017). We simulated 10,000 gene trees under the coalescent with DENDROPY v.4.1.0

(Sukumaran & Holder, 2010) using the ASTRAL species trees as a guide tree with branch lengths scaled by four to account for organellar inheritance. We summarized the simulated gene trees on the cpDNA tree. Under a scenario of ILS alone, any relationships in the empirical chloroplast, tree should be present in the simulated trees and have a high frequency; under a hybridization scenario, relationships unique to the cpDNA tree should be at low (or zero) frequency (García et al., 2017).

2.6 Species network analysis

We inferred species networks using a maximum pseudo-likelihood approach (Yu et al., 2012). Due to computational restrictions and given our main focus on potential reticulation among major clades of Caprifoliaceae (i.e. along the backbone), we reduced our 46-taxon data set to one outgroup and nine ingroup taxa to represent all major clades. Species network searches were carried out with PHYLONET v.3.6.1 (Than et al., 2008) with the command 'InferNetwork_MPL' and using the individual gene trees. Network searches were performed using only nodes in the gene trees that had BS support of at least 50%, allowing for up to four hybridization events and optimizing the branch lengths and inheritance probabilities of the returned species networks under the full likelihood. To estimate the optimal number of hybridizations and test whether the species network fits our gene trees better than a strictly bifurcating species tree, we computed the likelihood scores of concatenated RAxML, ASTRAL and plastid DNA trees, given the individual gene trees, as implemented in Yu et al. (2012), using the command 'CalGTProb' in PHYLONET. Finally, we performed model selection using the bias-corrected Akaike information criterion (AICc; Sugiura, 1978). The number of parameters was set to equal the number of branch lengths being estimated, the number of hybridization probabilities being estimated, and the number of gene trees used to estimate the likelihood, to correct for finite sample size.

2.7 Divergence time estimation

Divergence times were inferred using BEAST v.2.4.0 (Bouckaert et al., 2014). There is potential ancient hybridization in Caprifoliaceae, therefore, we used the nuclear and chloroplast gene tree for age estimates, separately. We constrained the root age to be 78.9 Ma based on the analysis of Li et al. (2019). We selected two fossils as the calibration points. (1) The fossil seeds of *Weigela* from the Miocene and Pliocene in Poland (Lańcucka- rodoniowa, 1967), and the Miocene in Denmark (Friis, 1985) was used to constrain its stem age to at 23 Ma (Wang et al., 2015). (2) The stem age of *Diplodipelta* was constrained to be at least 34.07 Ma based on the fruit fossil from the late Eocene Florissant flora of Colorado (34.07±0.1 Ma; Manchester, 2000). All dating analyses were performed with an uncorrelated lognormal relaxed clock (Drummond et al., 2012), GTR + G substitution model (Posada, 2008), gamma site heterogeneity model, estimated base frequencies, and a ML starting tree. A Yule process was specified as the tree prior. Two independent MCMC analyses of 300,000,000 generations with 10% burn-in and sampling every 3000 generations were conducted to evaluate the credibility of posterior distributions of parameters. BEAST log files were analyzed with Tracer v.1.7 (Drummond et al., 2012) for convergence with the first 10% removed as burn-in. Parameter convergent was assessed using an effective sample size (ESS) of 200. Log files where combined with LogCombiner and a maximum clade credibility tree with median heights was generated with TreeAnnotator v.1.8.4 (Drummond et al., 2012).

2.8 Analysis of character evolution

Character states were coded from the literature, particularly from Backlund (1996), Donoghue et al. (2003), Jacobs et al. (2011) and Landrein (2017). The number of stamens was scored: (0), 1; (1), 2; (2), 3; (3), 4; (4), 5. Two character states were scored for the style exertion: (0), not exceeding corolla; (1), exceeding corolla. Four fruit types were scored: (0), achene; (1), capsule, (2), berry; (3), drupe. The number of

carpels was scored as: (0), 2; (1), 3; (2), 4. Number of seeds was scored: (0), 1; (1), 2; (2), 4-5; (3), 6-20; (4), 20+; Two epicalyx types were scored: (0), no; (1), yes. All the morphological charactersanalyzed here were presented in Supplementary Fig. S1. Ancestral character state reconstruction was performed using the Maximum Likelihood approach as implemented in Mesquite v.3.51 (Maddison and Maddison, 2018) with the 'Trace character history' option based on the topology of the chloroplast trees. To explore the difference caused by different topology, we also reconstructed ancestral character based on the topology of the nuclear trees. The Markov k-state one-parameter model of evolution for discrete unordered characters (Lewis, 2001) was used.

2.9 Data accessibility

Raw Illumina data from sequence capture is available at the Sequence Read Archive (SRA) under accession SUB7674585 (see Table S1 for individual sample SRA accession numbers). DNA alignments, phylogenetic trees and results from all analyses and datasets can be found in the Dryad data repository.

3 Results

3.1 Exon assembly

The assembly resulted in sequences of up to 793 exons (≥ 150 bp) per species. The number of exons per sample varied from 380 to 1500, with an average of 1068 exons per sample. HybPiper identified paralogous copies for up to 284 exons per species. We found up to six paralogs per exon in Caprifoliaceae. After paralog pruning and removal of exons with poor coverage across samples (≤ 24 samples), we kept 713 exons from 196 different genes. Additionally, 63 of those exons, showed the presence of two (60) and three (three) paralogs copies that met the pruning requirements, giving us a total of 713 loci. The resulting concatenated matrix had an aligned length of 343,609 bp with

21,004 parsimony-informative sites, a minimum locus size of 277 bp, and a maximum locus size of 5,739 bp.

3.2 Phylogenetic reconstruction

We retrieved 87 CDS from off-target plastome reads, which after concatenation resulted in a matrix of 78,531 bp (Table 1). Overall, both nuclear and plastid data strongly support (1) Diervilloideae as sister to the rest of Caprifoliaceae, followed successively by Caprifolioideae, and (2) five monophyletic groups, Diervilloideae, Caprifolioidea, Valerianoideae, *Zabelia* and Morinoideae (Figs. 3 and 4).

Nuclear dataset. The ASTRAL topology (Fig. 3) was largely congruent with the RAxML concatenated dataset (Fig. 4), and the main clades were maximally supported. Our concatenation nuclear phylogeny recovered full support for the monophyly of the seven major clades (Fig. 4). The clade Valerianoideae + Dispsacoideae was sister to Linnaeoideae, with *Zabelia* + Morinoideae close to the clade Valerianoideae + Dispsacoideae + Linnaeoideae. It is worth mentioning that all relationships of major clades were with strong support (Fig. 4).

The coalescent nuclear phylogeny recovered moderate to strong support (local posterior probabilities (LPP $\geq$ 0.7) for all of the major clades and relationships within and among them. The clade of Valerianoideae + Dipsacoideae + Linnaeoideae was sister to *Zabelia* + Morinoideae and both together constituted the sister clade to Caprifolioideae. Within Linnaeoideae, all analyses and data sets recovered a clade of *Vesalea* M. Martens & Galeotti + *Linnaea* Gronov. ex L. as sister to a clade of all other Linnaeoideae, with strong support (Fig. 3).

The coalescent analyses and ICA scores about ASTRAL tree of the nuclear concatenated dataset revealed that most gene trees conflicted with the species trees (Fig. 3). Our results showed that ICA values along the backbone were ranging from 0.1 to 0.58, while ICA values in many of the nested clades were lower and ranged from

-0.05 to 0.25. The node *Kolkwitzia amabilis* showed the lowest values (ICA= −0.05). The ICA values calculated here are notably lower, indicating a great deal of underlying gene tree conflict (Figs. 3 and S2). Based on the PhyParts analysis, for *Zabelia* + Morinoideae, 417 of the simulated gene trees (out of 713) were concordant with this relationship while 181 were in conflict (Fig. 3). Furthermore, multiple conflicting placements were observed, suggesting that ILS is likely also at play here.

For ASTRAL tree, our analyses showed the Quartet Concordance (QC) values of the backbone clades were positive scores (0.17-1), which indicated moderate or strong support for the relationship of backbone clades, however, the Quartet Differential (QD) score tend to have more extreme values and were close to 0, which meant that no skew in the proportions of discordant trees. In addition, high Quartet Informativeness (QI) for these clades (QI=1 or near to 1), which showed low or no information for the given branch (Fig. S3). A similar pattern was found in nuclear concatenated RAxML trees, with positive QC values (0.23 to 1) for the backbone clades, QD values were near to 0, and high QI values (0.99-1) for these clades (Fig. S3). Slight differences in some relationships were observed between the concatenation and ASTRAL analyses of the nuclear genes (e.g., the positions of *Kolkwitzia amabilis* and *Dipsacus japonicus*; Figs. 3 and 4) and these differences were largely confined to areas of strong support.

Plastid dataset. Phylogenetic analysis of the cpDNA dataset also recovered the seven major clades in Caprifoliaceae with high support (Fig. 4). There were lots of conflicts on the cpDNA and nuclear trees (cf. Figs. 3 and 4). The clade *Zabelia* + Morinoideae and the clade Linnaeoideae were recovered as sister with relatively strong support (BS = 87%), and together these were sister to the clade Valerianoideae + Dipsacoideae (Fig. 4). For chloroplast trees, QS showed strong QC values (0.37- 1), low QD values (near to 0) and high QI values (0.94-1) (Fig. S5). This indicates strong majority of quartets support the focal branch and the low skew in discordant frequencies with low or no information for the relationship of these clades.

The plastid analyses (Fig. 4) placed *Diabelia* either sister to *Dipelta* Maxim. (consistent with Wang et al., 2020) with moderate support (BS = 85, Fig. 1) and together these were sister to *Kolkwitzia amabilis* (Fig. 4). However, the nuclear concatenation tree is consistent with the species tree in placing *Kolkwitzia amabilis* sister to *Abelia* + *Diabelia* (Fig. 1). Additional instances of cytonuclear discordance included the placements of *Zabelia*, Morinoideae, Dipsacoideae and Valerianoideae (Fig. 4).

3.3 Coalescent simulations analysis

Coalescent simulations under the organellar model did not produce gene trees that resembled the observed chloroplast tree. When the simulated gene trees were summarized on the observed chloroplast tree, most clade frequencies were near to zero, for instance, *Kolkwitzia amabilis* and the clade Valerianoideae + Dipsacoideae, *Zabelia* + Morinoideae and the clade Linnaeoideae, Valerianoideae and Dipsacoideae (Fig. S6). This suggested that ILS alone cannot explain the high level of cytonuclear discordance observed in Caprifoliaceae.

3.4 Species network analysis

A network with three reticulation events had the lowest AICc value (Table 2), suggesting hybridization events in the ancestor of *Zabelia biflora,* the ancestor of *Morina longifolia,* and the ancestor of *Scabiosa techiliensis* (Figs. 5 and S7). The inheritance probabilities analysis showed that Morinoideae (represented by *Morina longifolia)* had a genetic contribution of 24.2 % from an ancestral lineage of *Vesalea* and *Linnaea* (Fig. 5). Inferred inheritance probabilities for reticulation event indicate that *Zabelia (*represented by *Zabelia bilfora)* had a genetic contribution of 26.6% of its genome from an ancestral lineage of Morinoideae (Fig. 5). The inheritance probabilities showed that Dipsacoideae (represented by *Scabiosa techiliensis*) had a

384 genomic contribution of 59.2 % from an ancestral lineage of *Vesalea* and *Linnaea* (Fig.
385 5).

386

387 **3.5 Divergence time estimation**

388 Divergence time estimates based on nuclear gene tree suggested that the deepest
389 divergences in Caprifoliaceae occurred in the late Cretaceous, whereas most
390 generic-level diversification occurred in the Middle-Eocene (Fig. 6). The divergence
391 between Dipsacoideae and Valerianoideae was dated to 46.88 Ma (95% Highest
392 Posterior Density (HPD) = 37.44–57.35 Ma). The diversification of Linnaeoideae was
393 inferred to be at 53.83 Ma (95% HPD = 37.37–55.73 Ma). Within Linnaeoideae, both
394 *Abelia* and *Kolkwitzia* originated almost contemporaneously. The onset of *Zabelia* and
395 Morinoideae diversification occurred between 26.51 and 53.72 Ma. The divergence
396 time estimated by different data matrices was not completely consistent (Figs 6 and
397 S8). A comparison of the time estimates using plastid gene tree is shown in Fig. S8. For
398 instance, the age for the split of Dipsacoideae and Valerianoideae was estimated at
399 53.61 Ma (95% HPD: 41.44–65.57). The diversification of Linnaeoideae was inferred
400 to begin at 47.81 Ma (95% HPD = 34.21–52.55 Ma).

401

402 **3.6 Character evolution**

403 The likelihood inference of character evolution using the cpDNA tree detected
404 some homoplasies in each of the six morphological characters examined (Figs. 7, 8 and
405 9), and the style exertion relative to corolla showed particularly high homoplasy (Fig.
406 7). Morphological trait mapping suggested two hypotheses for the character evolution
407 in Caprifoliaceae: (1) Except for Caprifolioideae and Valerianoideae, most subfamilies
408 of Caprifoliaceae have four stamens. The number of stamens changed in a relatively
409 parsimonious manner, from five in most Caprifolioideae and Diervilloideae to four in
410 the bulk of Caprifoliaceae, within Valerianoideae, further reductions to 3 and 1(Fig. 7);

411 (2) The style exertion character has shown a high level of homoplasy in the early
412 diversification of the family. Even in the broad Linnaeoideae, the state “not exceeding
413 corolla” originated twice, once in *Vesalea*, and the other in the
414 *Diabelia-Dipelta-Kolwitzia-Abelia* clade; (3) Ancestral fruit type for Caprifoliaceae is
415 uncertain, but it is most likely an achene (Fig. 8). Nevertheless, the distribution of
416 fruit types among the basal lineages was complex, and inclusion of broader outgroup
417 taxa is needed to test the achene fruit type as the most likely ancestral state; (4) The
418 state of three carpels is common within Caprifoliaceae. There was much variation
419 among the early diverged lineages, but the state of three carpels was inferred to be
420 ancestral in the family and other states were largely derived from the three-carpel
421 state; (5) One seed is common and was inferred as the ancestral character state for
422 Caprifoliaceae. Similar to number of carpels, variation of this character is high in
423 Caprifoliaceae, with five character-states. Nevertheless, the character evolution was
424 relatively parsimonious with only low levels of homoplasy; (6) The epicalyx occurs
425 only in two major lineages, showing a case of convergent evolution (Figs. 7, 8 and 9).

426 A summary of character states using the nuclear gene tree that are relevant for the
427 taxonomy of the group is shown in Figs. S9, S10 and S11. We found that the patterns
428 of character evolution from cpDNA tree and nuclear gene tree were similar.

429

430 **4 Discussion**

431 **4.1 Phylogenetic incongruence and putative hybridization**

432 Although both our nuclear and plastid phylogenies supported the same seven
433 major clades of Caprifoliaceae, the relationships among these clades are incongruent
434 between data sets (Figs. 3 and 4). For instance, in the nuclear ASTRAL tree,
435 Linnaeoideae is recovered as sister to Dipsacoideae (except for *Dipsacus japonicus*)
436 +Valerianoideae (Fig. 3), while in the plastid tree Linnaeoideae is sister to *Zabelia* +
437 Morinoideae (Fig. 4). In contrast, in the nuclear RAxML concatenated tree (Fig. 3),

Linnaeoideae is recovered as sister to Dipsacoideae +Valerianoideae. Some of these points of conflict pertain to areas of Caprifoliaceae phylogeny that have long been problematic—for example, the relationships between *Zabelia* and other subfamilies. Our results reevaluated Caprifoliaceae phylogeny with more extensive evidence from the nuclear genome, because many previous results inferred from the plastome may be incorrect or incompletely understood due to evolutionary processes such as ILS or organellar capture via hybridization.

Three main processes will lead to gene tree heterogeneity and cytonuclear discordance: gene duplication/extinction, horizontal gene transfer/hybridization, and ILS. Currently, there are many methods to detect gene discordance (e.g., Smith et al., 2015; Pease et al., 2018), however, sources of such discordance remain hard to disentangle, especially when multiple process co-occur (e.g., Morales-Briones et al. 2020a).

In previous studies, *Zabelia* has long been thought to be closely related to *Abelia* (Hara, 1983; Tang & Lu, 2005). However, based on molecular datasets, Tank and Donoghue (2010) and Jacobs et al. (2011) found that *Zabelia* was sister to Morinaceae. Using six molecular loci and inflorescence morphology, Landrein et al. (2012) concluded that the position of *Zabelia* remained unclear. The molecular investigation of Xiang et al. (2019) found that the sister relationships between *Zabelia* + Morinaceae and Linnaceae + Valerianaceae + Dipsacaceae were not highly supported. Such phylogenetic incongruence provides the opportunity to test causal hypotheses of cytonuclear discordance, e.g., ILS or hybridization. Further, in our analyses (Fig. 4), widespread cytonuclear discordance exists across Caprifoliaceae, especially at genus levels, with a high level of conflict within genera. Regarding deep Caprifoliaceae relationships, the results from the nuclear analyses (Figs. 3 and 4) showed multiple instances (at least two) of well-supported conflict with the results from the plastome (Fig. 4), and the plastid results were largely consistent with previous plastid and large-scale analyses of Caprifoliaceae (Wang et al., 2020).

It is worth mentioning that Dipsacoideae was not recovered as monophyletic only in the species tree (Fig. 3), in which *Dipsacus japonicus* had a sister relationship with Linnaeoideae. The nodes with the strong LPP (LPP=1) also had the lower ICA score (ICA= 0.1), which suggests that ILS and/or unidentified hybrid lineages continue to obscure our understanding of relationships in Dipsacoideae. Our ICA scores and QS analyses of the nuclear dataset revealed strong signals of gene tree discordance among the seven major clades of Caprifoliaceae.

The concordance analysis and ICA scores showed that a large amount of conflict between individual gene trees and the species trees. Our coalescent simulations also suggested that the observed cytonuclear discordance cannot be explained by ILS alone. Previous studies reported that hybridization has shaped the evolutionary history of Caprifoliaceae (e.g., *Heptacodium miconioides*) (Zhang et al., 2003; Landrein et al., 2002). The extensive analyses performed here revealed a similar pattern of cytonuclear discordance, e.g., some species were recovered in different positions between the nuclear and plastid phylogenies (see Fig. 4 on positions of *Dipsacus japonicus*, *Kolkwizia amabilis*, and the clade of *Zabelia* and Morinoideae).

Our analyses showed that both ancient reticulation and ILS might be at play in the initial radiation of Caprifoliaceae. The results indicated that the parental contributions to the events of reticulation was unequal. Solís-lemus et al. (2017) suggest that inheritance probabilities of ~ 0.10 from a parental population to a reticulate node may suggest introgression, and that inheritance probabilities close to 0.50 may indicate that the hypothesized reticulate node is the product of hybrid speciation between parental populations. With regard to the *Zabelia* clade, the inheritance contributions (0.266 and 0.734) support a hybridization event between *Zabelia* and the ancestral lineage of *Morina* clade (Fig. 5). The second and third reticulation events reveal that there has been extensive gene flow between the *Scabiosa* clade and the *Morina* clade as well as the *Vesalea -Linnaea* clade (Fig. 5). The network analyses inferred *Zabelia*, *Mornia* and *Scabiosa* to be putative hybrid lineages. Furthermore, the coalescent simulations

494 indicated extensive ILS, that can be product of a rapid radiation in the backbone of
495 Caprifoliaceae.

496

497 **4.2 Temporal divergences of Caprifoliaceae**

498 Our estimated ages using nuclear and chloroplast trees are generally younger than
499 those of Wang et al. (2015) and Wang et al. (2020) based on two reliable fossils (Li et
500 al., 2019). We found that the diversification and global spread of the subfamilies of
501 Caprifoliaceae occurred during the late Cretaceous, Paleocene and Eocene (Fig. 6),
502 similar to the results of Beaulieu et al. (2013). Our result showed a very short node
503 connecting Linnaeoideae with *Zabelia* + Morionoideae in the backbone, which was
504 supported in both the nuclear and cpDNA trees (Figs. 6 and S8). Linnaeoideae diverged
505 from *Zabelia* + Morionoideae after the K-Pg boundary. Our results are congruent with
506 the phenomena reported in several other plant groups such as Amaranthaceae *s.l.*
507 (Morales-Briones et al. 2020a) and legumes (Koenen et al., 2020), and in lichenized
508 fungi such as Lobariaceae (Ascomycota) (Widhelm et al., 2019). It is generally
509 accepted that soon after the K-Pg boundary, due to mass extinctions, new habitats
510 became available and diverse organisms experienced rapid diversifications (Schulte et
511 al., 2010). Therefore, our results also reveal the wave of evolutionary radiation shortly
512 after the K-Pg boundary (Fig. 6). As a result of the tectonic movements, historical
513 climate fluctuation from Paleocene to Eocene, the Caprifoliaceae lineages subsequently
514 underwent rapid diversifications. The stem lineages of most genera were dated to the
515 Oligocene and Miocene, and most within-genus diversifications were dated to the
516 Miocene and Pliocene (Fig. 6). Our result may be explained by the hypothesis that
517 members of the Caprifoliaceae are well adapted to relatively cool environments (Friis,
518 1985; Manchester and Donoghue, 1995; Manchester, 2000), and an increase in the
519 earth's temperature may have forced them to move to higher altitudes or latitudes. As
520 plants moved to higher altitudes, their distribution was likely to be fragmented,
521 resulting in isolation between populations. We have some evidence to support this

hypothesis: (1) This family is mainly distributed in north temperate zone, and some genera even reach areas near the Arctic Circle (such as *Linnaea*); (2) there are numerous species (such as *Valeriana officinalis*, *Lonicera rupicola*, and *L. spinos*a) with island-like distributional patterns in relatively high altitudes. Survivors by isolation may have blossomed after the late Oligocene, especially during the Miocene with a shift into new geographic areas, especially if these were mountainous, and then struggled again during recent climatic cooling and glacial activities (Moore and Donoghue, 2007). The global events (e.g. ancient orogenic and monsoon-driven) that might have led to the diversification of Caprifoliaceae as reported in other taxa (Lu et al., 2018; Ding et al., 2020). For example, some genera or taxa (e.g., *Linnaea*, *Lonicera myrtillus*) may have benefited from the global cooling and drying of the Miocene and Pliocene, and these taxa usually possess tiny, narrow or needle-like leaves, while certain lineages (*Abelia, Diabelia,* and *Dipelta*) may be more adapted to the wetter, warmer parts of the world and these lineages may not have benefitted from the global cooling of the past 30 million years.

4.3 Evolution of morphological characters

The characters were traced on the phylogeny of the cpDNA data using ML method (Figs. 7, 8 and 9) because of the potential hemiplasy and xenoplasy produced by the discordance and hybridization detected in the nuclear backbone (Avies and Robinson 2008; Robinson et al., 2008; Copetti et al., 2017; Wang et al., 2020). A consequence of this discordance is elevated levels of apparent homoplasy in the species tree (Copetti et al., 2017; Hahn and Nakhleh 2017).

Stamen number, fruit type, style exertion, number of carpels, number of seeds and epicalyx presence have been traditionally used for generic recognition within Caprifoliaceae (Backlund, 1996; Donoghue et al., 2003; Yang and Landrein, 2011; Landrein et al., 2020). Discordance among morphological traits might plausibly arise due to either variable convergent selection pressures or other phenomena such as

hemiplasy. The evidence indicates that the probability of hemiplasy is high for the four characters traits in Caprifoliaceae: the branch lengths leading to lineages with derived character states are uniformly short with high levels of gene tree discordance. It is possible that gene flow contributes to these patterns. For example, the ancestral stamen number states (i.e., 2 and 4) found in *Morina longifolia* and *Acanthocalyx alba* within the Morinoideae clade could be due to alleles introgressed, as we identified putative introgression events between those lineages (Fig. 5). Morphological and anatomical studies showed that the earliest Caprifoliaceae had monosymmetric flowers (probably weakly so at first) with larger calyx lobes, tubular corollas, elongate styles, and capitate stigmas (Donoghue et al., 2003). Within Caprifoliaceae, the main change in stamen number is a reduction from five to four stamens. Subsequently, there was a reduction to two stamens within Morinaceae and to three, two, and one within Valerianaceae (Figs. 7 and S9). These variations may be related to an underlying change in floral symmetry (Donoghue et al., 2013). Increasing symmetry characteristic may relate to carpel abortion or to differences in the arrangement of flowers at the level of the inflorescence.

Similar to other five characters traits, our data suggest that carpels number is also affected by hemiplasy: in most relevant internodes, the ancestral state of carpels number can be inferred to be inconsistent with carpels number transitions generally following phylogenetic relationships. Our results suggest that multiple independent evolutionary events of the carpel evolution in Caprifoliaceae have occurred (Figs. 8 and S10). In Caprifoliaceae, the abortion of two of the three carpels and the development of just a single ovule within the remaining fertile carpel was evidently correlated with fruit type (Wilkinson 1949). For some subfamilies of Caprifoliaceae, carpel abortion occurs at a relatively late stage of ovary development, so many species have two empty chambers at fruit maturity (e.g., Linnaeoideae, Morinoideae, and Valerianoideae). In fact, in some species, these empty compartments have been co-opted in various ways in connection with dispersal (e.g., inflated for water dispersal in some *Valeriana*).

Caprifoliaceae shows great variation in fruit types. Fleshy, bird-dispersed fruits are limited to the Caprifolieae Dumort. (Donoghue et al., 2003). It is important to note that

579 the ancestral carpel number for Caprifoliaceae is most likely 3. *Lonicera* has berries,
580 though generally with just a few seeds embedded in copious pulp. There is programmed
581 carpel abortion and the number of seeds corresponds to the number of fertile carpels.
582 For *Symphoricarpos*, two of the four carpels abort, and there are two stones. The
583 mesocarp in the cases is rather dry and mealy in texture. In the Caprifoliaceae, achenes
584 with a single seed are present in *Heptacodium* and in the large Linnaeoideae clade
585 (though in *Dipelta*, and in *Linnaea* there are two seeds at maturity). From the standpoint
586 of fruit evolution, the linkage of *Heptacodium* within Caprifolioideae implies either the
587 independent evolution of achenes or a transition from achenes to fleshy fruits in the line
588 leading to Caprifolioideae. Among the achene-producing Caprifoliaceae, there are
589 various adaptations for wind dispersal. One of the most striking of these modifications
590 is enlargement of the calyx lobes into wings as the fruits mature (e.g. in *Abelia*,
591 *Dipelta*,and *Diabelia*). Especially well known is the production of a feathery
592 pappus-like structure in species such as *Valeriana officianalis* and *Centranthus ruber*
593 in Valerianoideae. This modification facilitates passive external transport by animals.
594 A similar case is also found in *Kolkwitzia*.

595 The reconstruction of character evolution thus shows that some characters that
596 were once considered important for taxonomy within the family have been inferred to
597 be the results of homoplasious evolution (Gould 2000; Pyck, 2001; Bell 2001, 2004;
598 Carlson et al., 2009; Zhai et al., 2019). In character evolution analysis, homoplasy is
599 regarded as evolutionary noise that, if not properly accommodated, jeopardizes
600 phylogenetic reconstructions using morphological characters. At the same time,
601 hemiplasy is one of the causes of homoplasy (Copetti et al., 2017). The phenomenon of
602 hemiplasy is most plausible when the internodal distances in a phylogenetic tree are
603 short (relative to effective population sizes) (Robinson et al., 2008). This may explain
604 why it has been difficult to reconstruct the relationships among the major lineages and
605 genera of the family. Eventually, more extensive sampling and developmental studies
606 will be needed to elucidate the mechanisms underlying the morphological evolutionary
607 patterns outlined here.

4.4 Recognition of Zabelioideae as a new subfamily in Caprifoliaceae

Despite the strong signals of gene tree discordance, our nuclear and plastid phylogenies strongly supported seven major clades in Caprifoliaceae: Linnaeoideae, *Zabelia*, Morinoideae, Valerianoideae, Dipsacoideae, Caprifolioideae and Diervilloideae, and show *Zabelia* as the sister to the morphologically highly distinct Morinoideae (Figs. 3 & 4). Our analyses supported reticulate evolution concerning the origins of both the *Zabelia* lineage as well as the Morinoideae. Based on the phylogenomic and morphological analyses, we herein propose to recognize *Zabelia* as representing a new subfamily of Caprifoliaceae.

Zabelioideae B. Liu & S. Liu ex H.F. Wang, D.F. Morales-B, M.J. Moore & J. Wen, subfam. nov.

Type: *Zabelia* (Rehder) Makino.

Description: Shrubs, deciduous; old branches often with six deep longitudinal grooves. Leaves opposite, entire or dentate at margin; estipulate; petioles of opposite leaf pairs dilated and connate at base, enclosing axillary buds. Inflorescence a congested thyrse of cymes; cymes 1-3-flowered. Calyx 4- or 5-lobed, persistent, spreading. Corolla 4- or 5-lobed, hypocrateriform, ± zygomorphic; corolla tube cylindrical. Stamens 4, included, didynamous. Ovary 3-locular, 2 locules with 2 series of sterile ovules and 1 locule with a single fertile ovule; stigmas green, capitate, mucilaginous. Fruit an achene crowned with persistent and slightly enlarged sepals. Basic chromosome number $x = 9$.

One genus and six species distributed in China, Japan, Korea, Afghanistan, NW India, Kyrgyzstan, Nepal, and Russian Far East.

Zabelioideae is highly distinct morphologically from its sister Morinoideae. They can be easily distinguished by their habit (with Zabelioideae as shrubs, and Morinoideae as herbs), the six distinct, longitudinal grooves on twigs and branches of

635 Zabelioideae (the six grooves absent in Morinoideae), and the epicalyx (absent in
636 Zabelioideae and present in Morinoideae). Zabelioideae and Morinoideaeshare show
637 some similarities in pollen micromorphology, as both have psilate pollen grains with an
638 endocingulum (Verlaque 1983; Kim et al. 2001; Jacobs et al., 2011). The two
639 subfamilies diverged in the early-mid Eocene (Figs. 6, S7), and their long evolutionary
640 history associated with deep hybridization events, ILS and extinctions likely have made
641 it difficult to determine their phylogenetic placements.

642

643 **5 Conclusions**

644 Gene tree discordacne has been commonly observed in phylogenetic studies.
645 More evidence has shown that the species tree method is inconsistent in the presence of
646 gene flow (Solís-lemus et al., 2016; Long and Kubatko 2018), which suggests that both
647 ILS and gene flow simultaneously need to be considered in constructing phylogenetic
648 relationships. Here, our results show clear evidence of cytonuclear discordance and
649 extensive conflict between individual gene trees and species trees in Caprifoliaceae.
650 Second, the short node connecting Linnaeoideae with Zabelioideae+Morionoideae was
651 dated to be after the K-Pg boundary, which support that there was a rapid radiation
652 Caprifoliaceae species at that time, as reported in other plant taxa. Third, the temporal
653 diversification of Caprifoliaceae provides a good case to support the evolutionary
654 radiations and adaptation of a dominantly north temperate plant family to climatic
655 changes from the late Cretaceous to the late Cenozoic. Finally, based on evidence from
656 molecular phylogeny, divergence times, and morphological characters, we herein
657 recognize the *Zabelia* clade as representing a new subfamily, Zabelioideae, in
658 Caprifoliaceae. The phylogenetic framework also sheds important insights into the
659 character evolution in Caprifoliaceae.

660

661 **Acknowledgement**

The work was funded by National Scientific Foundation of China (31660055). We appreciate Gabriel Johnson for his help with the target enrichment experiment, and the United States National Herbarium for permission for sampling some collections. We acknowledge the staff in the Laboratories of Analytical Biology at the National Museum of Natural History, the Smithsonian Institution for support and assistance.

Author contributions

H.F.W. and J.W. conceived the study. H.F.W. and D.F.M-B. performed the research and analyzed the data. H.X.W. and H.F.W. wrote the manuscript, D.F.M-B., J.W. and M.J.M revise the manuscript.

ORCID

Hong-Xin Wang: https://orcid.org/0000-0002-2283-7368

Diego F. Morales-Briones: https://orcid.org/0000-0003-1535-5739

Hua-Feng Wang: https://orcid.org/0000-0003-3331-2898

Jun Wen: https://orcid.org/0000-0001-6353-522X

References

Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC. 2010. Homoploid hybrid speciation in action. *Taxon* 59: 1375–1386.

682 APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the
683 orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean*
684 *Society* 181: 1–20.

685 Avise JC, Robinson TJ. 2008. Hemiplasy: A New Term in the Lexicon of
686 Phylogenetics. *Systematic biology* 57: 503-507.

687 Backlund, A. 1996. Phylogeny of the order Dipsacales. Ph. D. Dissertation.

688 Uppsala: Uppsala University.

689 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
690 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N,
691 Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly
692 Algorithm and Its Applications to Single-Cell Sequencing. *Journal of*
693 *Computational Biology* 19: 455–477.

694 Beaulieu JM, Tank DC, Donoghue MJ. 2013**.** A Southern Hemisphere origin for
695 campanulid angiosperms, with traces of the break-up of Gondwana. *BMC*
696 *Evolutionary Biology*. 13: 80.

697 Bell CD. 2004. Preliminary phylogeny of Valerianaceae (Dipsacales) inferred from
698 nuclear and chloroplast DNA sequence data. *Molecular Phylogenetics and*
699 *Evolution* 31: 340–350.

700 Bell CD, Edwards EJ, Kim ST, Donoghue MJ. 2001. Dipsacales phylogeny based on
701 chloroplast DNA sequences. *Harvard Papers in Botany* 6:481–499.

702 Blackmore S, Cannon MJ. 1983. Palynology and systematics of Morinaceae. *Rev*
703 *Palaeobot Palynol* 40: 207–266.

704 Bolger AM, Lohse, M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
705 Sequence Data. *Bioinformatics* (*Oxford, England*) 30: 2114–2120.

706 Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut
707 A, Drummond AJ. 2014. Beast 2: a software platform for Bayesian evolutionary
708 analysis. *PLoS Computational Biology* 10: e1003537.

709 Caputo G, Cozzolino S.1994. A cladistic analysis of Dipsacaceae (Dipsacales). *Plant*
710 *Systematics and Evolution* 189: 41–61.

711 Carlson SE, Mayer V, Donoghue MJ. 2009. Phylogenetic relationships, taxonomy, and
712 morphological evolution in Dipsacaceae (Dipsacales) inferred by DNA sequence
713 data. *Taxon* 58: 1075–1091.

714 Cavender-Bares J, Gonzalez-Rodriguez A, Eaton DAR, Hipp AAL, Beulke
715 A, Manos PS. 2015. Phylogeny and biogeography of the American live oaks
716 (*Quercus* subsection *Virentes*): A genomic and population genetics approach.
717 *Molecular Ecology* 24: 3668–3687.

718 Chamala S, García, N, Godden GT, Krishakumar V, Jordon-Thaden IE, Smet RD et al.
719 2015. MarkerMiner 1.0: A new application for phylogenetic marker development
720 using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

721 Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Childs KL, Eguiarte LE, Lee
722 S, Liu TL, McMahon MM, Whiteman NK, Wing RA, Wojciechowski MF,
723 Sanderson MJ. 2017. Extensive gene tree discordance and hemiplasy shaped the
724 genomes of North American columnar cacti. *Proceedings of the National*
725 *Academy of Sciences* 114: 12003-12007.

726 Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and
727 the multispecies coalescent. *Trends in Ecology and Evolution* 24: 332–340.

728 Ding WN, Ree RH, Spicer RA, Xing YW. 2020. Ancient orogenic and monsoon-driven
729 assembly of the world's richest temperate alpine flora. *Science* 369: 578–581.

730 Donoghue MJ, Olmstead RG, Smith JF, Palmer JD. 1992. Phylogenetic relationships
731 of Dipsacales based on *rbcL* sequences. *Annals of the Missouri Botanical Garden*
732 79: 333–345.

733 Donoghue MJ, Eriksson T, Reeves PA, Olmstead RG. 2001. Phylogeny and
734 phylogenetic taxonomy of Dipsacales, with special reference to *Sinadoxa* and
735 *Tetradoxa* (Adoxaceae). *Harvard Papers in Botany* 6: 459–479.

736 Donoghue MJ, Bell CD, Winkworth RC. 2003. The evolution of reproductive
737 characters in Dipsacales. *International Journal of Plant Sciences* 164:
738 S453–S464.

739 Doyle JJ, Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh
740 leaf tissue. *Phytochemical Bulletin* 19: 11–15.

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

743 Edwards SV. 2009. Is a new and general theory of molecular systematics emerging?
744 *Evolution* 63: 1–19.

745 Feng Y, Peter Comes H, Qiu YX. 2020. Phylogenomic insights into the
746 temporal-spatial divergence history, evolution of leaf habit and hybridization in
747 *Stachyurus* (Stachyuraceae). *Molecular Phylogenetics and Evolution* doi:
748 https://doi.org/10.1016/j.ympev.2020.106878.

749 Friis EM. 1985. Angiosperm fruits and seeds from the Middle Miocene of Jutland
750 (Denmark). *Biologiske Skrifter* 24:1–165.

751 Folk RA, Mandel JR, Freudenstein JV. 2017. Targeted enrichment of intronic sequence
752 markers for recent radiations: a phylogenomic example from *Heuchera*
753 (Saxifragaceae). *Applications in Plant Sciences* 3: 1500039.

754 Galtier N, Daubin V. 2008. Dealing with incongruence in phylogenomic analyses.

Philosophical Transactions of the Royal Society B: Biological Sciences 363: 4023–4029.

García N, Folk RA, Meerow AW, Chamala S, Gitzendanner MS, de Oliveira RS, Soltis DE, Soltis PS. 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution* 111: 231–247.

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT. et al. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477: 419–423.

Gould KR, Donoghue MJ. 2000. Phylogeny and biogeography of *Triosteum* (Caprifoliaceae). *Harvard Papers in Botany* 5:157–166.

Hara H. 1983. A Revision of the Caprifoliaceae of Japan with Reference to Allied Plants in Other Districts and the Adoxaceae. *Academia Scientific Books, Tokyo*.

Hahn MW, Nakhleh L. 2017. Irrational exuberance for resolved species trees. *Evolution* 70:7–17.

Jacobs B, Pyck N, Smets E. 2010. Phylogeny of the *Linnaea* clade: are *Abelia* and *Zabelia* closely related? *Molecular Phylogenetics and Evolution* 57: 741–752.

Jacobs B, Geuten K, Pyck N, Huysmans S, Jansen S, Smets E. 2011. Unraveling the phylogeny of *Heptacodium* and *Zabelia* (Caprifoliaceae): an interdisciplinary approach. *Systematic Botany* 36: 231–252.

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.

779 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software Version
780 7: improvements in performance and usability. *Molecular Biology and Evolution*
781 30: 772–780.

782 Kim T. 1998. Phylogenetic studies of tribe Linnaeeae (Caprifoliaceae). *Ph. D Thesis,*
783 *Chonbuk National University.*

784 Kim TJ, Sun BY, Suh Y. 2001. Palynology and cytotaxonomy of the genus *Abelia s.l.*,
785 Caprifoliaceae. *Korean Journal of Plant Taxonomy* 31: 91–106.

786 Koenen EJM, Ojeda DI, Bakker FT, Wieringa JJ, Catherine K, Hardy OJ, Pennington
787 RT, Herendeen PS, Bruneau A, Hughes CE. 2020. The origin of the legumes is a
788 complex paleopolyploid phylogenomic tangle closely associated with the
789 cretaceous-paleogene (K-Pg) mass extinction event. *Systematic Biology* syaa041.

790 Konowalik K, Wagner F, Tomasello S, Vogt R, Oberprieler C. 2015. Detecting
791 reticulate relationships among diploid Leucanthemum Mill. (Compositae,
792 Anthemideae) taxa using multilocus species tree reconstruction methods and
793 AFLP fingerprinting. *Molecular Phylogenetics and Evolution* 92: 308–328.

794 Landrein S. 2010. *Abelia* and its relatives. *Plantsman, NS* 9: 40–47.

795 Landrein S, Prenner G, Chase MW, Clarkson JJ. 2012. *Abelia* and relatives:
796 Phylogenetics of Linnaeeae (Dipsacales Caprifoliaceae *s.l.*) and a new
797 interpretation of their inflorescence morphology. *Botanical Journal of Linnean*
798 *Society* 169: 692–713.

799 Landrein S, Prenner G. 2013. Unequal twins? Inflorescence evolution in the twinflower
800 tribe Linnaeeae (Caprifoliaceae *s.l.*). *International Journal of Plant Sciences* 174:
801 200–233.

802 Landrein S, Prenner G. 2016. Structure, ultrastructure and evolution of floral nectaries
803 in the twinflower tribe Linnaeeae and related taxa (Caprifoliaceae). *Botanical*
804 *Journal of Linnean Society* 181: 37–69.

805 Lańcucka- rodoniowa M. 1967. Two new genera: *Hemiptelea* Planch. and *Weigela*
806 Thumb. in the younger Tertiary of Poland. *Acta Palaeobotanica* 8: 1–17.

807 Lee-Yaw JA, Grassa CJ, Joly S, Andrew RL, Rieseberg LH. 2019. An evaluation of
808 alternative explanations for widespread cytonuclear discordance in annual
809 sunflowers (*Helianthus*). *New Phytologist* 221: 515–526.

810 Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete
811 morphological character. *Systematic Biology* 50: 913–925.

812 Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai
813 J, Luo Y, Wang H, Bank MV, Zhang SD, Wang QF, Wang J, Zhang ZR, Fu CN,
814 Yang J, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Li DZ. 2019. Origin
815 of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.

816 Liu L, Xi Z, Wu S, Davis CC, Edwards SV. 2015. Estimating phylogenetic trees from
817 genome-scale data. *Annals of the New York Academy of Sciences* 1360: 36–53.

818 Lin HY, Hao YJ, Li JH, Fu CX, Soltis PS, Soltis DE, Zhao YP. 2019. Phylogenomic
819 conflict resulting from ancient introgression following species
820 diversification in *Stewartia s.l.* (Theaceae). *Molecular Phylogenetics and*
821 *Evolution* 135: 1–11.

822 Linder HP, Naciri Y. 2015. Species delimitation and relationships: the dance of the
823 seven veils. *Taxon* 64: 3–16.

824 Long C, Kubatko L. 2018. The Effect of Gene Flow on Coalescent-based
825 Species-Tree Inference. *Systemmatic biology* 67: 770-785.

826 Lu LM, Mao LF, Yang T, Ye JF, Liu B, Li HL, Sun M, Miller JT, Mathews S, Hu HH,
827 Niu YT, Peng DX, Chen YH, Smith SA, Chen M, Xiang KL, Le CT, Dang VC, Lu
828 AM, Soltis PS, Soltis DE, Li JH, Chen ZD. 2018. Evolutionary history of the
829 angiosperm flora of china. *Nature* 554: 234–238.

830 McKain M. 2017. Afinit: Mrmckain/Fast-Plast: Fast-Plast V.1.2.6.
831 doi:10.5281/zenodo.973887.

832 Mallet J. 2007. Hybrid speciation. *Nature* 466: 279–283.

833 Maddison WP, Maddison DR. 2018. Mesquite: a modular system for evolutionary
834 analysis. Version 3.51 http://www.mesquiteproject.org.

835 Manchester SR, Donoghue MJ. 1995. Winged fruits of Linnaeeae (Caprifoliaceae) in
836 the tertiary of western North America: Diplodipelta gen. nov. *International*
837 *Journal of Plant Science* 156:709–722.

838 Manchester SR. 2000. Late Eocene fossil plants of the John Day Formation, Wheeler
839 County, Oregon. *Oregon Geology* 62: 51–63.

840 Moore BR, Donoghue MJ. 2007. Correlates of diversification in the plant clade
841 Dipsacales: Geographic movement and evolutionary innovations. *American*
842 *Naturalist* 170: S28–S55.

843 Morales-Briones DF, Liston A, Tank DC. 2018. Phylogenomic analyses reveal a deep
844 history of hybridization and polyploidy in the Neotropical genus *Lachemilla*
845 (Rosaceae). *New Phytologist* 218: 1668–1684.

846 Morales-Briones DF, Kaderei G,Tefarikis DT, Moore MJ, Smith SA,
847 Brockington SF, Timoneda A, et al. 2020a. Disentangling sources of gene tree
848 discordance in phylotranscriptomic datasets: a case study from Amaranthaceae
849 *s.l.*. *Systematic Biology* https://doi.org/10.1093/sysbio/syaa066.

Morales-Briones DF, Gehrke B, Huang CH, Liston A, Ma H, Marx HE, Tank DC, Yang Y. 2020b. Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in *Alchemilla s.l.* (Rosaceae). *BioRxiv*. *https://doi.org/10.1101/2020.08.21.261925*.

Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* 105: 385–403.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25:1253–1256.

Pyck N. 2001. Phylogenetic relationships within Dipsacales: a combined molecular and morphological approach. PhD thesis. Katholieke University, Louvain, Belgium.

Robinson TJ, Ruiz-Herrera A, Avise JC. 2008. Hemiplasy and homoplasy in the karyotypic phylogenies of mammals. *Proceedings of the National Academy of Sciences* 105: 14477-14481.

Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Molecular Biology and Evolution* 31: 1261–1271.

Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.

Schulte P, Alegret L, Arenillas I, Arz JA, Barton PJ, Bown PR, Bralower TJ, Christeson GL, Claeys P, Cockell CS, Collins GS, Deutsch A, Goldin TJ, Goto K, Grajales-Nishimura M, et al. 2010. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science* 327: 1214–1218.

874 Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees,
875 alignments and molecular data. *Bioinformatics* 24: 715–716.

876 Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed clock analysis
877 suggests an earlier origin for flowering plants. *Proceedings of the National*
878 *Academy of Sciences of the United States of America* 107: 5897–5902.

879 Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets
880 reveals conflict, concordance, and gene duplications with examples from animals
881 and plants. *BMC Evolutionary Biology* 15: 150.

882 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post
883 analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

884 Stevens PF. 2019. Angiosperm Phylogeny Website. Version Apr. 15, 2019
885 http://www.mobot.org/MOBOT/research/APweb/.

886 Solís-lemus C, Ané C. 2016. Inferring Phylogenetic Networks with Maximum
887 Pseudolikelihood under Incomplete Lineage Sorting. *PLOS Genet* 12:
888 e1005896.

889 Solís-lemus C, Yang MY, Ané C. 2016. Inconsistency of Species Tree Methods under
890 Gene Flow. *Systematic Biology* 65: 843-851.

891 Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. 2020. Nuclear
892 phylogenomic analyses of asterids conflict with plastome trees and support novel
893 relationships among major lineages. *American Journal of Botany* 107: 1–16.

894 Sugiura N. 1978. Further Analysis of the Data by Akaike's Information Criterion and
895 the Finite Corrections. *Communications in Statistics - Theory and Methods A* 7:
896 13–26.

897 Sukumaran J, Holder MT. 2010. DendroPy: A Python library for phylogenetic
898 computing. *Bioinformatics* 26: 1569–1571.

899 Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. 2015. Deep phylogenetic
900 incongruence in the angiosperm clade Rosidae. *Molecular Phylogenetics and*
901 *Evolution* 83: 156–166.

902 Szöllősi GJ, Tannier E, Daubin V, Boussau B. 2015. The inference of gene trees with
903 species trees. *Systematic Biology* 64: e42–e62.

904 Tang YC, Lu AM. 2005. Paraphyletic group, PhyloCode and phylogenetic species
905 – the current debate and a preliminary commentary. *Acta Phytotaxonomica*
906 *Sinica* 43: 403–419.

907 Traiser C, Klotz S, Uhl D, Mosbrugger V. 2005. Environmental signals from leaves a
908 physiognomic analysis of European vegetation. *New Phytologist* 166: 465–484.

909 Than C, Ruths D, Nakhleh L. 2008. A software package for analyzing and
910 reconstructing reticulate evolutionary histories. *BMC Bioinformatics* 9: 322

911 Tank DC, Donoghue MJ. 2010. Phylogeny and phylogenetic nomenclature of the
912 Campanulidae based on an expanded sample of genes and taxa. *Systematic Botany*
913 35: 425–441.

914 Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a
915 pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae:
916 Astereae: *Diplostephium*). *New Phytologist* 214: 1736–1750.

917 Vargas OM, Heuertz M, Smith SA, Dick CW. 2019.Target sequence capture in the
918 Brazil nut family (Lecythidaceae): Marker selection and in silico capture from
919 genome skimming data. *Molecular Phylogenetics and Evolution* 135: 98–104.

920 Verlaque R. 1983. Contribution l'étude du genre *Morina* L. *Pollen ET Spores* 25:
921 143–162.

922 Wang HF, Landrein S, Dong WP, Nie ZL, Kondo K, Funamoto T, Wen J, Zhou SL.
923 2015. Molecular phylogeny and biogeographic diversification of Linnaeoideae
924 (Carprifoliaceae *s.l.*) disjunctively distributed in Eurasia, North American and
925 Mexico. *PLoS One* 10: e0116485.

926 Wang HX, Liu H, Moore MJ, Landrein S, Liu B, Zhu ZX, Wang HF. 2020. Plastid
927 phylogenomic insights into the evolution of the Caprifoliaceae *s.l.* (Dipsacales).
928 *Molecular Phylogenetics and Evolution* 142: 106641.

929 Wang YX, Cao Z, Ogilvie HA, Nakhleh L. 2020. Phylogenomic assessment of the
930 role of hybridization and introgression in trait evolution. BioRxiv preprint doi:
931 https://doi.org/10.1101/2020.09.16.300343.

932 Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring Phylogenetic Networks Using
933 PhyloNet. *Systematic Biology* 67: 735–740.

934 Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, Liston
935 A. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant
936 phylogenomics. *Applications in Plant Sciences* 2: 9.

937 Widhelm TJ, Grewe F, Huang JP, Mercado-Díaz JA, Gofnet B, Lücking R, Moncada
938 B, Mason-Gamer R, Lumbsch HT. 2019. Multiple historical processes obscure
939 phylogenetic relationships in a taxonomically difficult group (Lobariaceae,
940 Ascomycota). *Scientific Reports* 9: 8968.

941 Xiang YZ, Huang CH, Hu YH, Wen J, Li SS, Yi TS, Chen HY, Xiang J, Ma H. 2016.
942 Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of
943 geological times and genome duplication. *Molecular Biology and Evolution* 34:
944 262-281.

945 Xiang CL, Dong HJ, Landrein S, Zhao F,Yu WB, Soltis DE, Soltis PS, Backlund A,
946 Wang HF, Li DZ, Peng H. 2019. Revisiting the phylogeny of Dipsacales: new
947 insights from phylogenomic analyses of complete chloroplast genome sequences.
948 *Journal of Systematics and Evolution* 58: 103–117.

949 Yang QR, Landrein S. 2011. Caprifoliaceae. In: Wu Z, Hong D, Raven PH eds. *Flora*
950 *of China*. Beijing: Science Press; St. Louis:Missouri Botanical Garden Press. 19:
951 617–618.

952 Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms
953 using transcriptomes and low-coverage genomes: improving accuracy and
954 matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31:
955 3081–3092.

956 Yu Y, Degnan JH, Nakhleh L. 2012. The probability of a gene tree topology within a
957 phylogenetic network with applications to hybridization detection. *PLoS Genetics*
958 8: e1002660.

959 Yakimowski SB, Rieseberg LH. 2014. The role of homoploid hybridization in
960 evolution: a century of studies synthesizing genetics and ecology. *American*
961 *Journal of Botany* 101: 1247–1258.

962 Zhai W, Duan X, Zhang R, Guo C, Li L, Xu G, Shan H, Kong H, Ren Y. 2019.
963 Chloroplast Genomic Data Provide New and Robust Insights into the Phylogeny
964 and Evolution of the Ranunculaceae. *Molecular Phylogenetics and Evolution*
965 135: 12–21.

966 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time
967 species tree reconstruction from partially resolved gene trees. *BMC*
968 *Bioinformatics*. 19: 523.

969 Zhang WH, Chen ZD, Li JH, Chen HB,Tang YC. 2003. Phylogeny of the Dipsacales
970 *s.l.* based on chloroplast *trnL-F* and *ndhF* sequences. *Molecular Phylogenetics*
971 *and Evolution* 26: 176–189.

972

974 Fig. 1. Floral diversity of Dipsacales. (A) *Kolkwitzia amabilis*; (B) *Zabelia integrifolia*;
975 (C) *Dipsacus asper*; (D) *Valeriana flaccidissima*; (E) *Acanthocalyx nepalensis*
976 subsp. *delavayi*; (F) *Lonicera fragrantissima* var. *lancifolia*; (G) *Weigela*
977 *coraeensis;* (H) *Viburnum opulus* subsp. *calvescens.*

978 Fig. 2. Alternative relationships for the Caprifoliaceae s.l. backbone based on previous
979 analyses. (A) Donoghue et al. (2001); parsimony analyses based on chloroplast
980 *rbcL* sequences and morphological characteristics; (B) Bell et al. (2001);
981 maximum likelihood tree from the combined chloroplast DNA data; (C) Zhang et
982 al. (2003); maximum likelihood tree based on chloroplast *trnL-F* and ndhF
983 sequences; (D) Jacobs et al. (2010); maximum parsimony Dipsacales phylogeny
984 based on nuclear and chloroplast sequence data; (E) Wang et al.(2020); maximum
985 likelihood tree based on 68 complete plastomes. (F) This study, Species tree based
986 on nuclear concatenated data set.

987 Fig. 3. Species tree of the nuclear concatenated dataset inferred with ASTRAL- Local
988 posterior probabilities support values and internode certainty all scores
989 are shown above and below branches respectively. Pie charts next to the nodes
990 present the proportion of gene trees that supports that clade (blue), the proportion
991 that supports the main alternative for that clade (green), the proportion that
992 supports the remaining alternatives (red), and the proportion (conflict or support)
993 that has < 50% bootstrap support (gray). Numbers next to pie charts indicate the
994 number of gene trees concordant/conflicting with that node in the species tree.
995 Major taxonomic groups or main clades in the family as currently recognized
996 are indicated by branch colors as a visual reference to relationships.

997 Fig. 4. Tanglegram of the nuclear concatenated (left) and plastid (right phylogenies.
998 Dotted lines connect taxa between the two phylogenies. Maximum likelihood
999 bootstrap support values are shown above branches. The asterisks indicate
1000 maximum likelihood bootstrap support of 100%. Major taxonomic groups or main

clades in the family as currently recognized are indicated by branch colors as a visual reference to relationships.

Fig. 5. Best supported species network of the selective nuclear dataset inferred with PHYLONET. Numbers next to the hybrid branches indicate inheritance probabilities. Red lines represent minor hybrid edges (edges with an inheritance contribution < 0.50).

Fig. 6. BEAST analysis of divergence times based on the nuclear alignment. Calibration points are indicated by A, B. and C. Numbers 1–11 represent major divergence events in Caprifoliaceae; mean divergence times and 95% highest posterior densities are provided for each.

Fig. 7. Likelihood inference of character evolution in Caprifoliaceae using Mesquite v.2.75 based on plastid matrix. Left, Number of stamens; Right, Style exertion.

Fig. 8. Likelihood inference of character evolution in Caprifoliaceae using Mesquite v.2.75 based on plastid matrix. Left, fruit type; Right, Number of carpels.

Fig. 9. Likelihood inference of character evolution in Caprifoliaceae using Mesquite v.2.75 based on plastid matrix. Left, number of seeds; Right, epicalyx presence/absence.

Fig. S1 Simplified ML tree generated from the nuclear gene data showing the distribution of selected character states. The asterisks indicate Maximum likelihood bootstrap support of 100%.

Fig. S2 ASTRAL-II species tree; node label indicates internode certainty all (ICA) scores.

Fig. S3. Results of simulation testing of the Quartet Sampling of the Astral trees. Node labels indicate QC/Quartet Differential (QD)/Quarte Informativeness (QI) scores.

Fig. S4. Results of simulation testing of the Quartet Sampling of the nuclear concatenated RAxML trees. Node labels indicate QC/Quartet Differential (QD)/Quarte Informativeness (QI) scores.

Fig. S5. Results of simulation testing of the Quartet Sampling of the chloroplast trees. Node labels indicate QC/Quartet Differential (QD)/Quarte Informativeness (QI) scores.

Fig. S6. Phylogeny of the plastid DNA dataset; numbers above branches represent clade frequencies of the simulated gene trees.

Fig. S7. Best species networks of the selective nuclear dataset estimated with PhyloNet with one (A), two (B), three (C) and four (D) hybridization events. Blue branches connect the hybrid nodes. Numbers next to blue branches indicate inheritance probabilities.

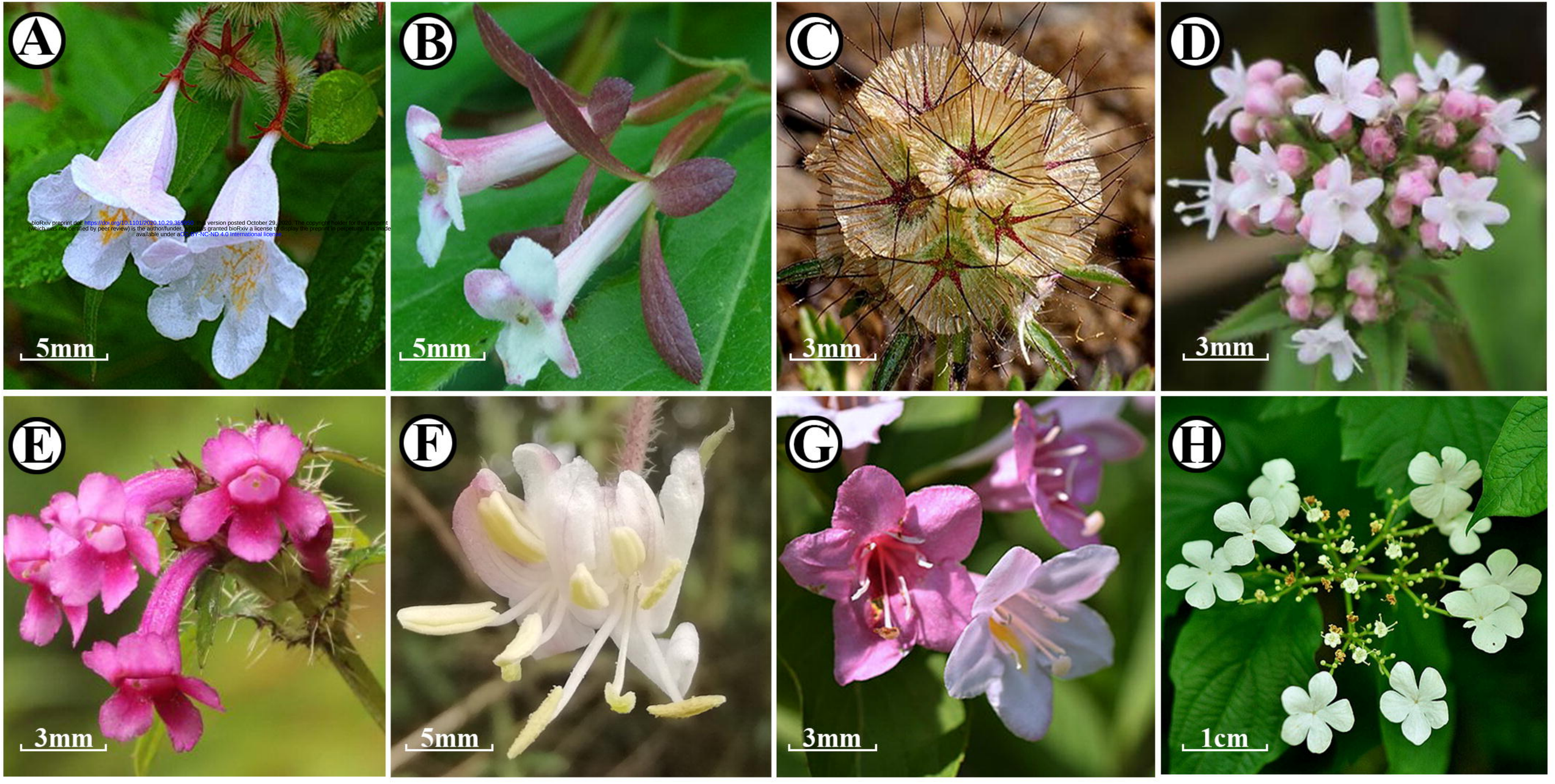
Fig. S8. BEAST analysis of divergence times based on the cpDNA data. Calibration points are indicated by A, B, and C. Numbers 1–10 represent major divergence events in Caprifoliaceae; mean divergence times and 95% highest posterior densities are provided for each.

Fig. S9. Likelihood inference of character evolution in Caprifoliaceae using Mesquite v.2.75 based on nuclear matrix. Left, Number of stamens; Right, Style exertion.

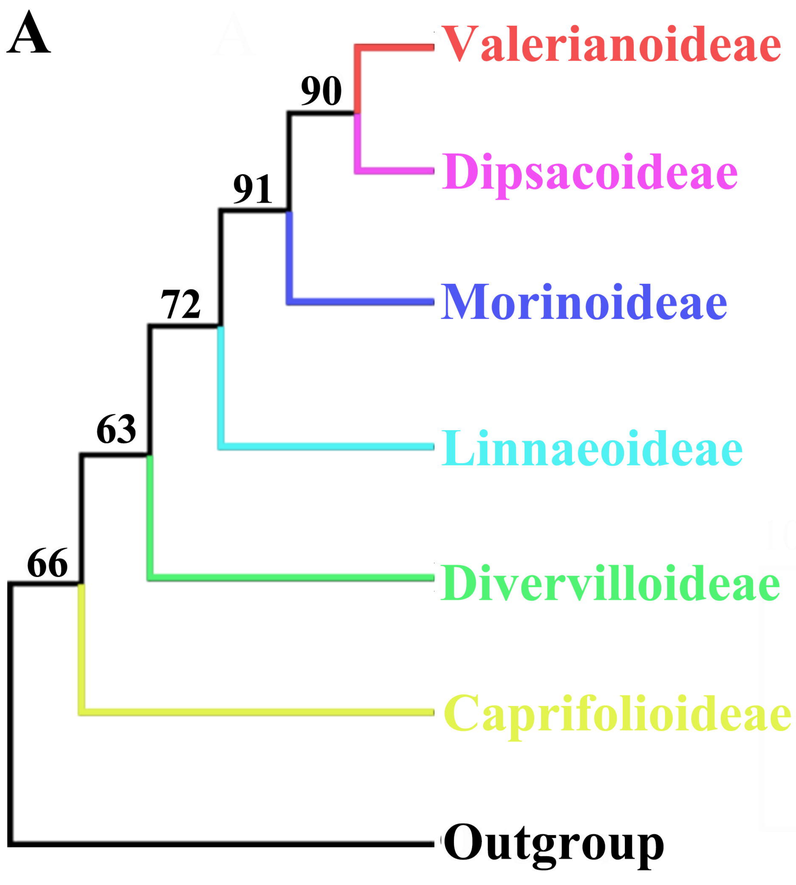
Fig. S10. Likelihood inference of character evolution in Caprifoliaceae using Mesquite v.2.75 based on nuclear matrix. Left, Style of fruit; Right, Number of carpels.

Fig. S11. Likelihood inference of character evolution in Caprifoliaceae using Mesquite v.2.75 based on nuclear matrix. Left, number of seeds; Right, epicalyx presence/absence.

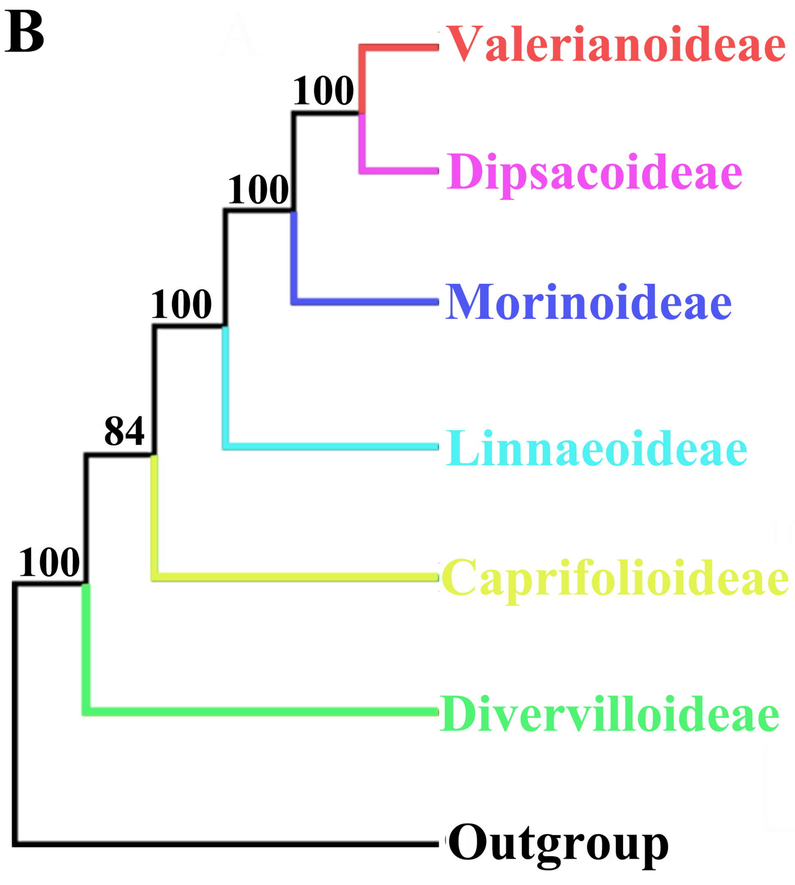
Appendix 1. The baits developed based on the transcriptomes of *Lonicera japonica*, *Valeriana officinalis*, *Viburnum odoratissimum*, *Sambucus canadensis*, *Symphoricarpos sp.* and *Dipsacus asper* from Caprifoliaceae in 1KP used in this study.



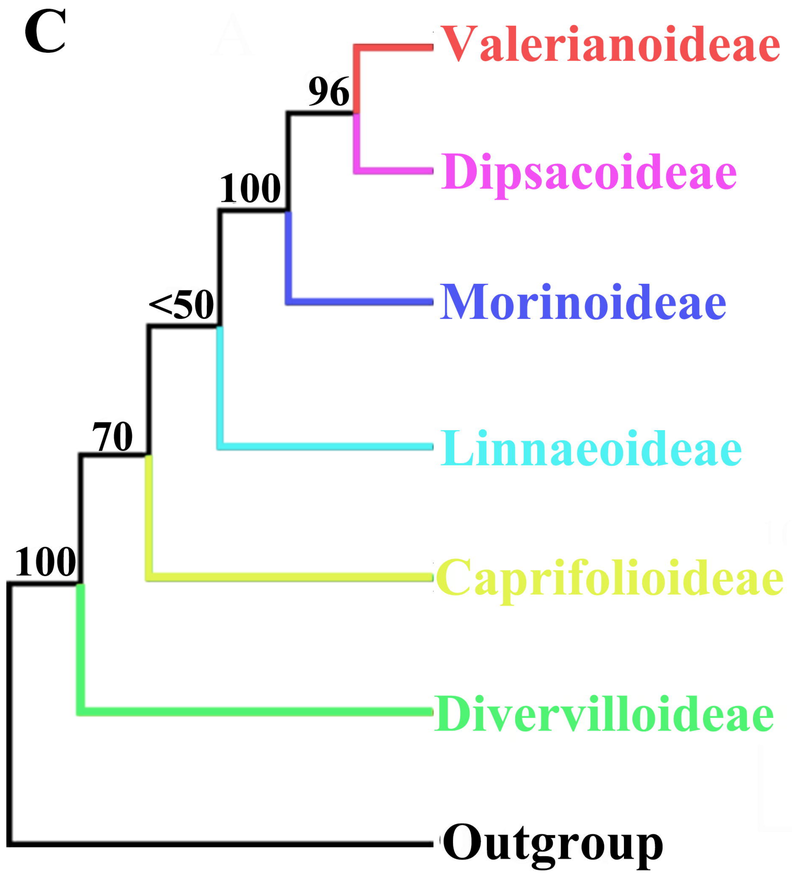
A


Valerianoideae
Dipsacoideae
Morinoideae
Linnaeoideae
Divervilloideae
Caprifolioideae
Outgroup
90
91
72
63
66

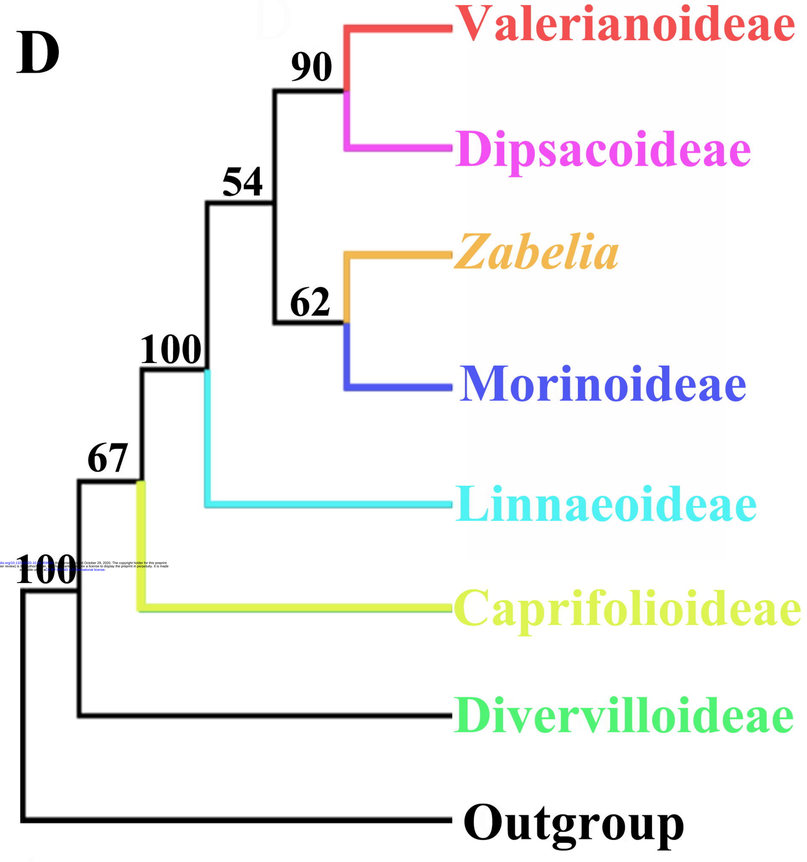
B


Valerianoideae
Dipsacoideae
Morinoideae
Linnaeoideae
Caprifolioideae
Divervilloideae
Outgroup
100
100
100
84
100

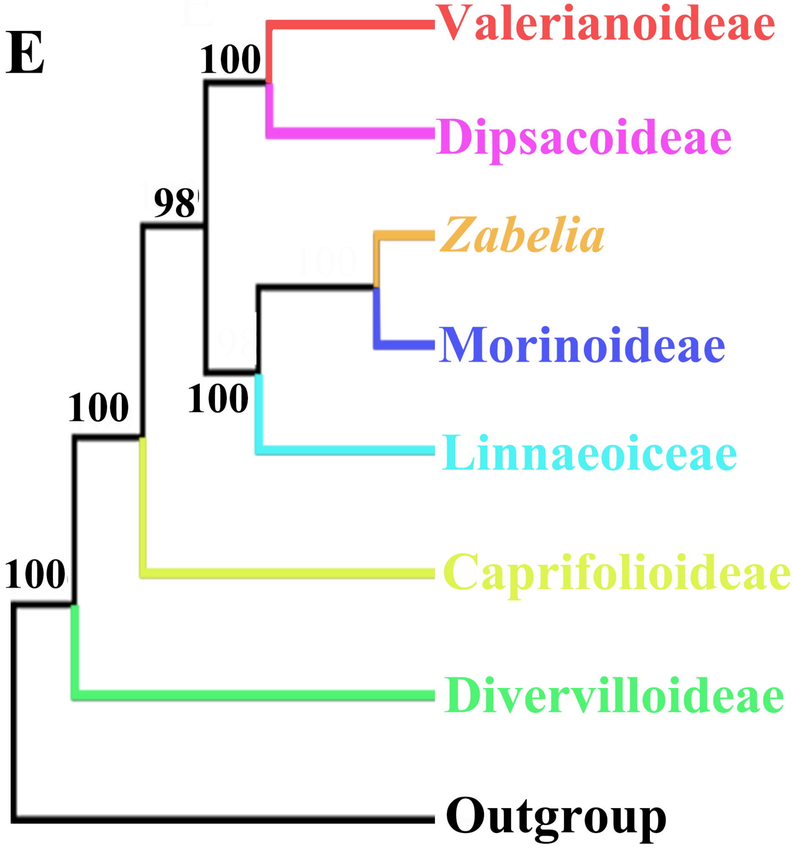
C


Valerianoideae
Dipsacoideae
Morinoideae
Linnaeoideae
Caprifolioideae
Divervilloideae
Outgroup
96
100
<50
70
100

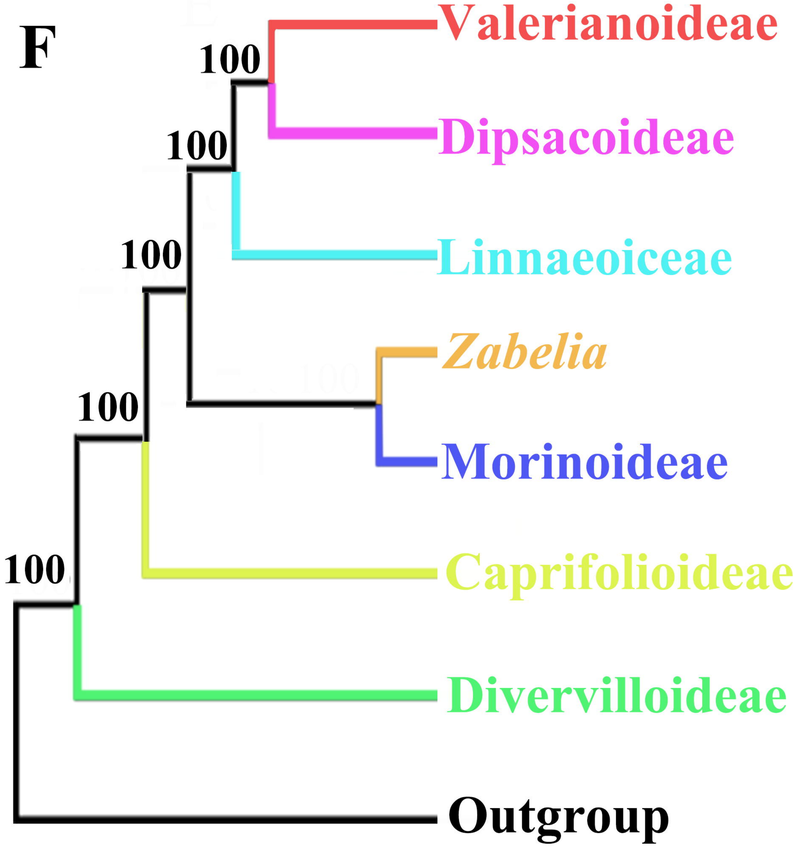
D

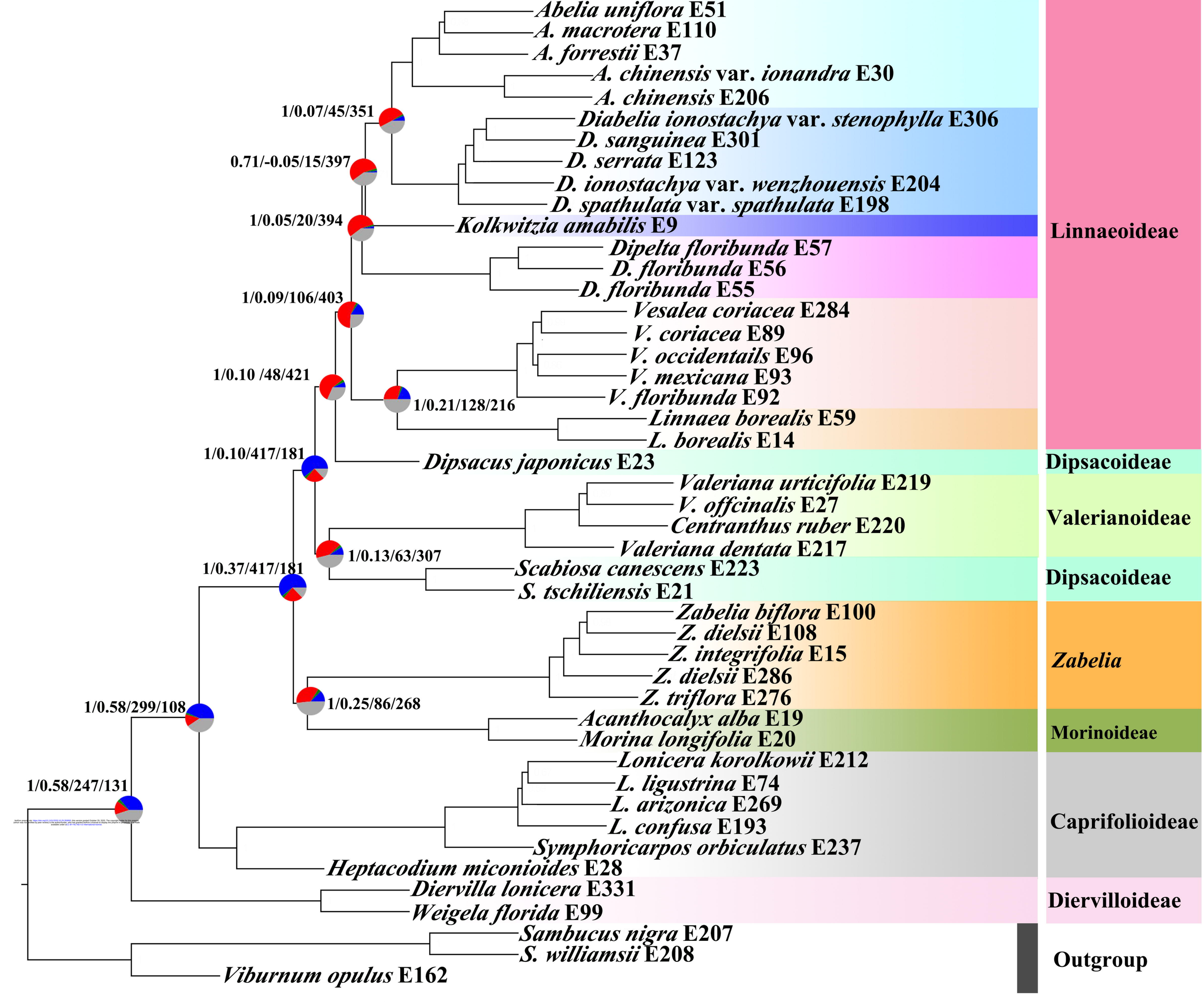

Valerianoideae
Dipsacoideae
Zabelia
Morinoideae
Linnaeoideae
Caprifolioideae
Divervilloideae
Outgroup
90
54
62
100
67
100

E


Valerianoideae
Dipsacoideae
Zabelia
Morinoideae
Linnaeoiceae
Caprifolioideae
Divervilloideae
Outgroup
100
98
100
100
100

F


Valerianoideae
Dipsacoideae
Linnaeoiceae
Zabelia
Morinoideae
Caprifolioideae
Divervilloideae
Outgroup
100
100
100
100
100


Abelia uniflora E51
A. macrotera E110
A. forrestii E37
A. chinensis var. ionandra E30
A. chinensis E206
1/0.07/45/351
Diabelia ionostachya var. stenophylla E306
D. sanguinea E301
0.71/-0.05/15/397
D. serrata E123
D. ionostachya var. wenzhouensis E204
D. spathulata var. spathulata E198
1/0.05/20/394
Kolkwitzia amabilis E9
Linnaeoideae
Dipelta floribunda E57
D. floribunda E56
D. floribunda E55
1/0.09/106/403
Vesalea coriacea E284
V. coriacea E89
V. occidentails E96
1/0.10 /48/421
V. mexicana E93
1/0.21/128/216
V. floribunda E92
Linnaea borealis E59
L. borealis E14
1/0.10/417/181
Dipsacus japonicus E23
Dipsacoideae
Valeriana urticifolia E219
V. offcinalis E27
Centranthus ruber E220
Valeriana dentata E217
Valerianoideae
1/0.13/63/307
1/0.37/417/181
Scabiosa canescens E223
S. tschiliensis E21
Dipsacoideae
Zabelia biflora E100
Z. dielsii E108
Z. integrifolia E15
Z. dielsii E286
Zabelia
1/0.25/86/268
Z. triflora E276
1/0.58/299/108
Acanthocalyx alba E19
Morina longifolia E20
Morinoideae
Lonicera korolkowii E212
1/0.58/247/131
L. ligustrina E74
L. arizonica E269
L. confusa E193
Caprifolioideae
Symphoricarpos orbiculatus E237
Heptacodium miconioides E28
Diervilla lonicera E331
Diervilloideae
Weigela florida E99
Sambucus nigra E207
S. williamsii E208
Outgroup
Viburnum opulus E162
0.8

Nuclear concatenated

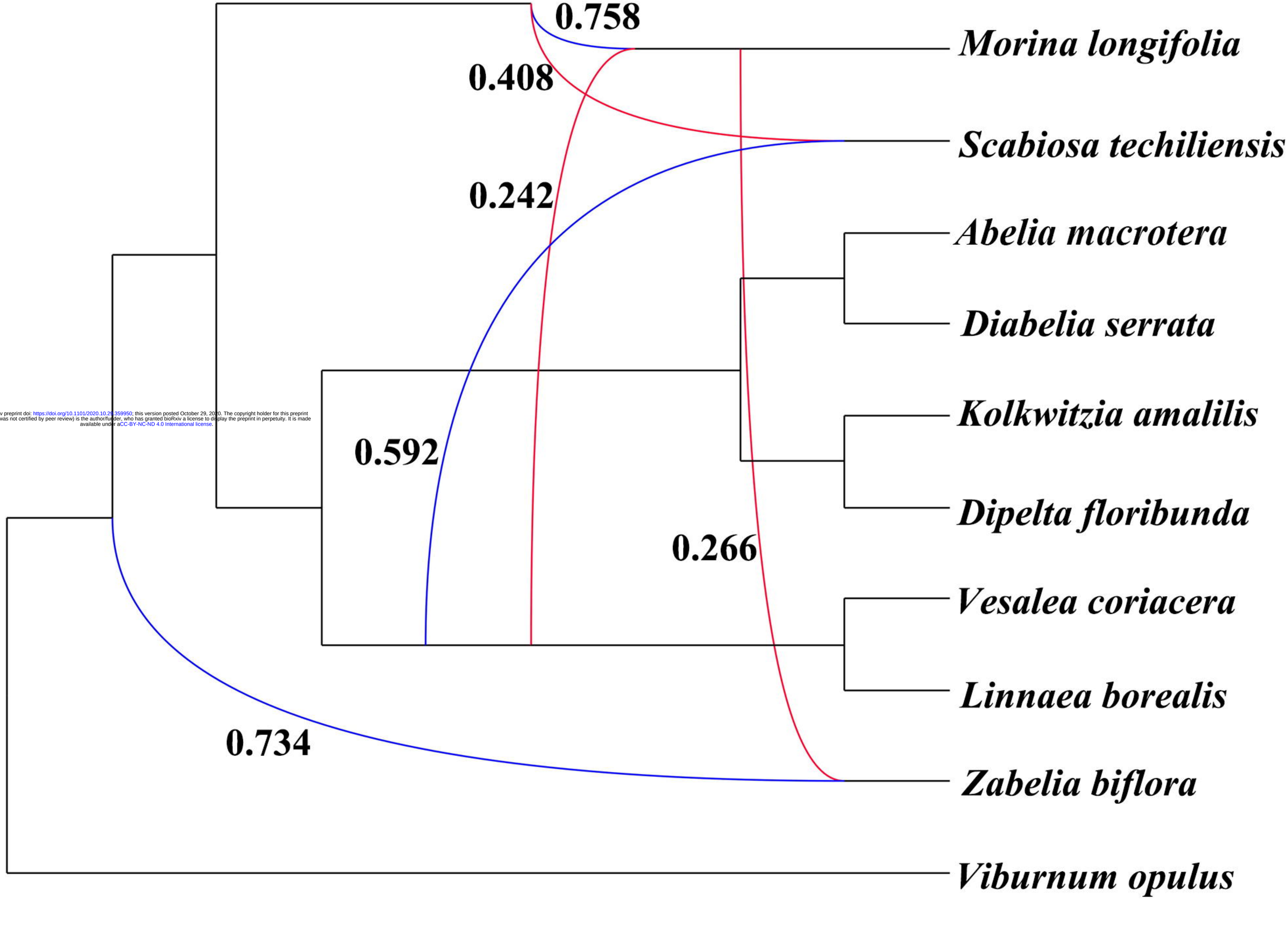
Plastome

Abelia uniflora
A. macrotera
A. forrestii
A. chinensis var. *ionandra*
A. chinensis
Diabelia serrata
D. spathulata var. *spathulata*
D. ionostachya var. *stenophylla*
D. sanguinea
D. ionostachya var. *wenzhouensis*
Kolkwitzia amabilis
Dipelta floribunda
D. floribunda
D. floribunda
Vesalea occidentails
V. mexicana
V. floribunda
V. coriacea
V. coriacea
Linnaea borealis
L. borealis
Valeriana urticifolia
Centranthus ruber
Valeriana offcinalis
Valeriana dentata
Scabiosa canescens
S. tschiliensis
Dipsacus japonicus
Zabelia dielsii
Z. integrifolia
Z. biflora
Z. dielsii
Z. triflora
Acanthocalyx alba
Morina longifolia
Lonicera arizonica
L. ligustrina
L. confusa
L. korolkowii
Symphoricarpos orbiculatus
Heptacodium miconioides
Diervilla lonicera
Weigela florida
Sambucus nigra
S. williamsii
Viburnum opulus

Linnaeoideae
Valerianoideae
Dipsacoideae
Zabelia
Morinoideae
Caprifolioideae
Diervilloideae
Outgroups

85
87

0.04
0.0

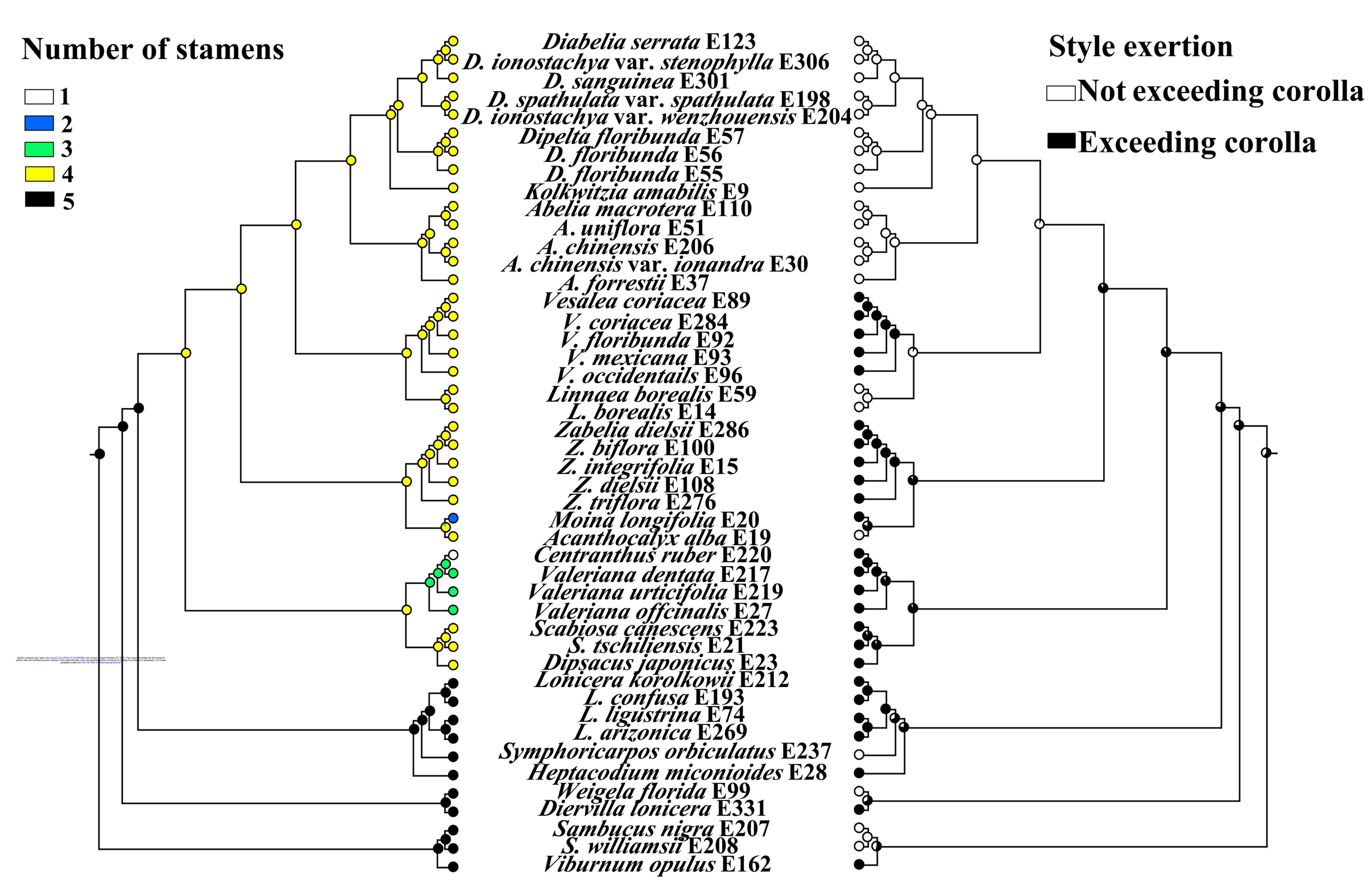


A 78.9 Ma
B 34.07 Ma
C 23 Ma
1 69.38 [53.26-79.45] Ma
2 64.33 [44.76-73.89] Ma
3 58.19 [41.98-66.44] Ma
4 53.83 [37.37-55.73] Ma
5 49.47 [38.74-54.61] Ma
6 46.88 [37.44-57.35] Ma
7 46.19 [36.89-57.15] Ma
8 43.82 [26.51-53.72] Ma
9 43.10 [27.12-57.93] Ma
10 42.15 [19.13-46.41] Ma
11 32.21 [19.34-39.17] Ma

Abelia chinensis var. *ionandra* E30
A. chinensis E206
A. forrestii E37
A. macrotera E110
A. uniflora E51
Diabelia sanguinea E301
D. spathulata var. *spathulata* E198
D. serrata E123
D. ionostachya var. *wenzhouensis* E204
D. ionostachya var. *stenophylla* E306
Kolkwitzia amabilis E9
Dipelta floribunda E55
D. floribunda E56
D. floribunda E57
Linnaea borealis E59
L. borealis E14
Vesalea mexicana E93
V. occidentails E96
V. coriacea E284
V. coriacea E89
V. floribunda E92
Morina longifolia E20
Acanthocalyx alba E19
Zabelia dielsii E286
Z. integrifolia E15
Z. biflora E100
Z. dielsii E108
Z. triflora E276
Dipsacus japonicus E23
Scabiosa canescens E223
S. tschiliensis E21
Valeriana dentata E217
Valeriana offcinalis E27
Centranthus ruber E220
Valeriana urticifolia E219
Lonicera ligustrina E74
L. arizonica E269
L. confusa E193
L. korolkowii E212
Symphoricarpos orbiculatus E237
Heptacodium miconioides E28
Weigela florida E99
Diervilla lonicera E331
Sambucus nigra E207
Viburnum opulus E162
S. williamsii E208

Linnaeoideae
Morinoideae
Zabelia
Dipsacoideae
Valerianoideae
Caprifolioideae
Diervilloideae
Outgroups

Late Cretaceous | Paleocene | Eocene | Oligocene | Miocene | Pli | Ple
70 60 50 40 30 20 10 0 Ma



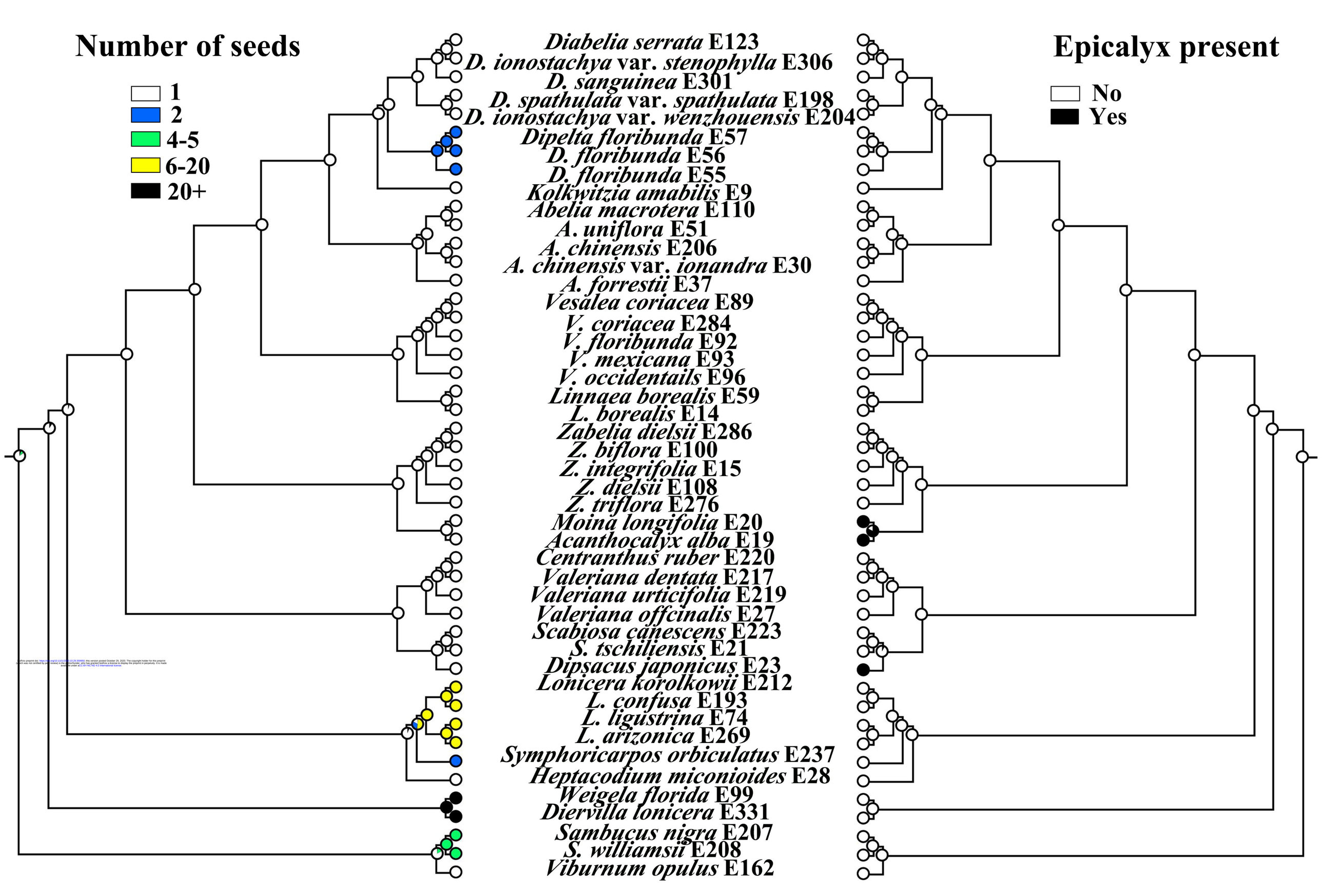
Fruit stype

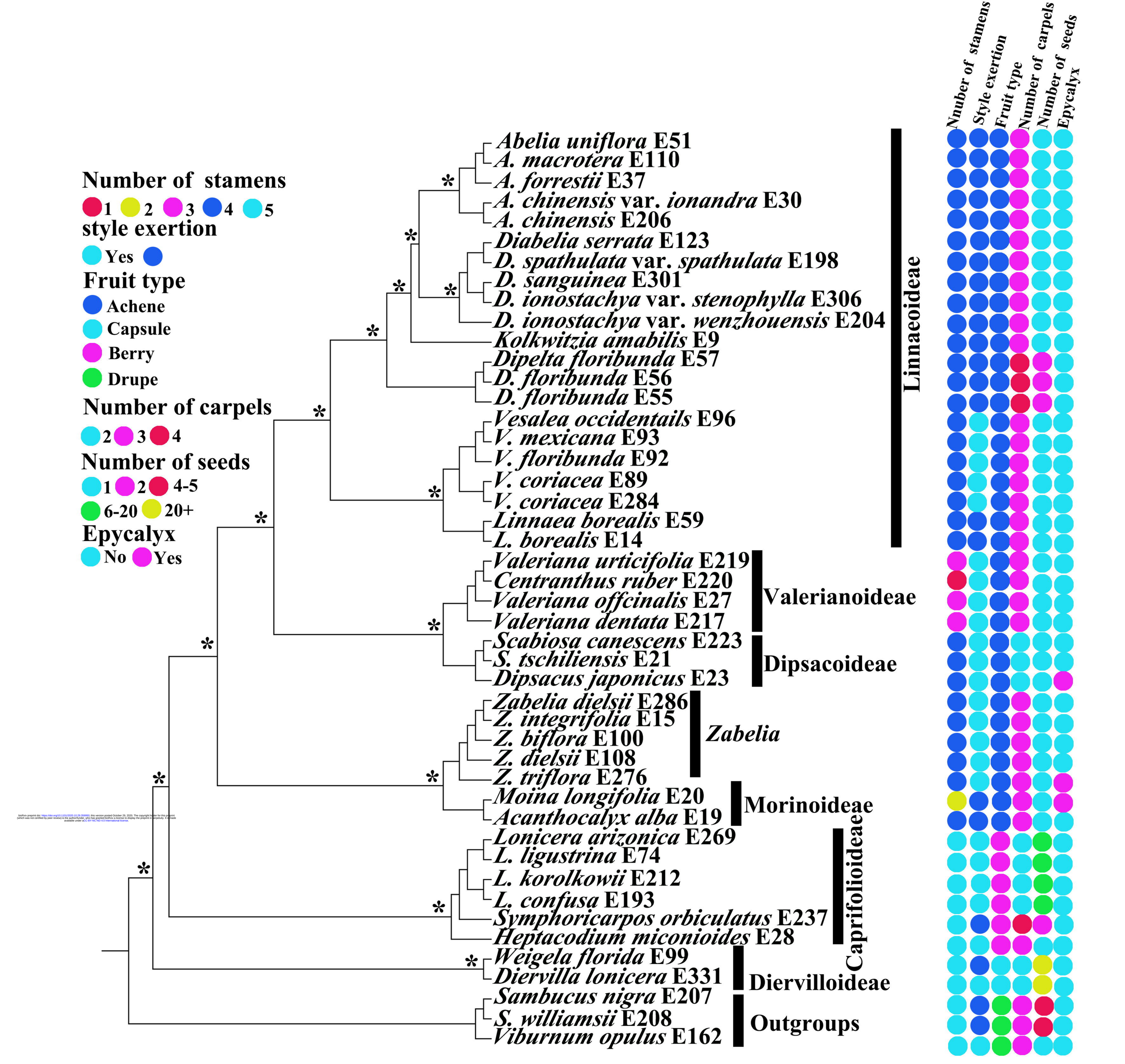
- Achene
- Capsule
- Berry
- Drupe

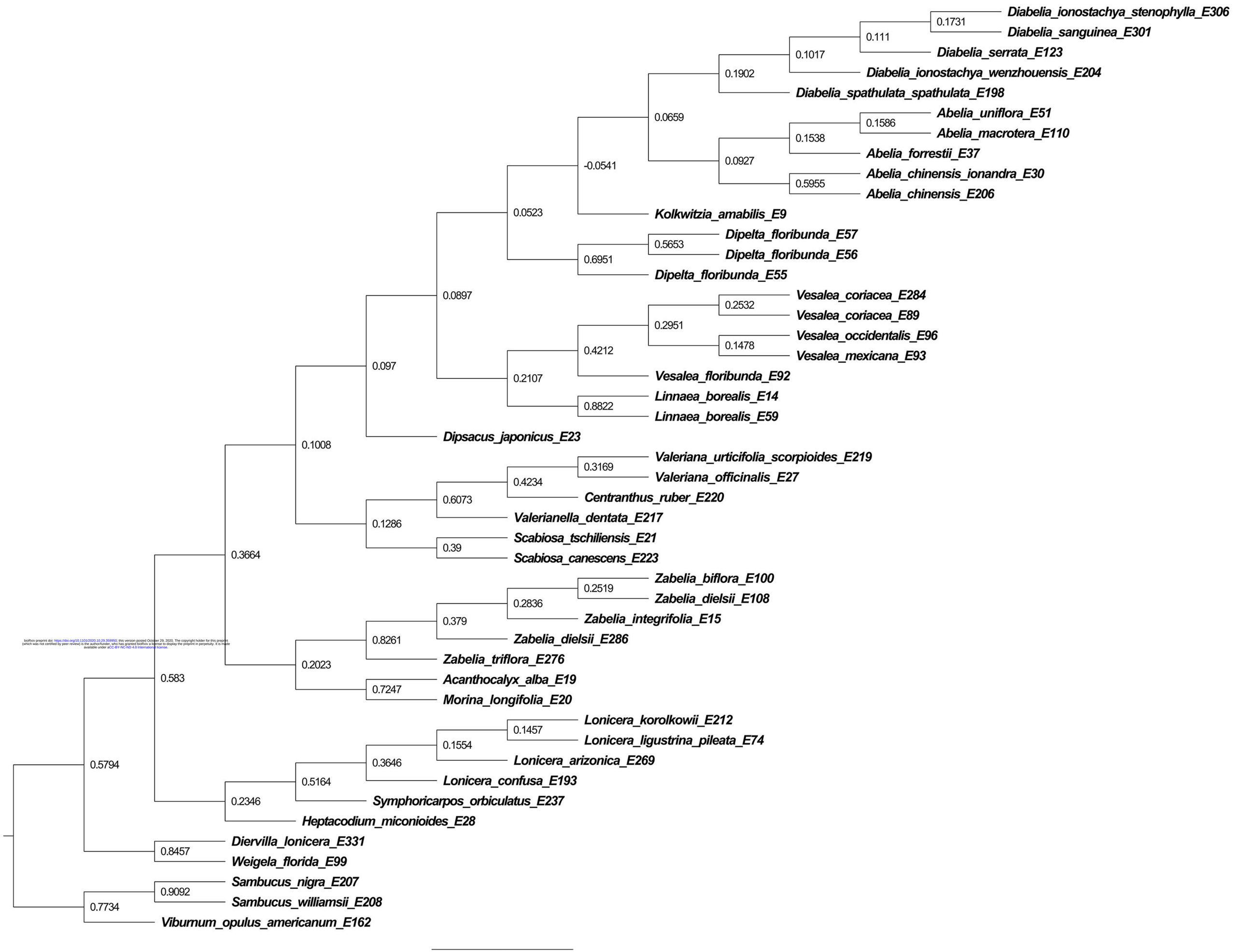
Number of carpels

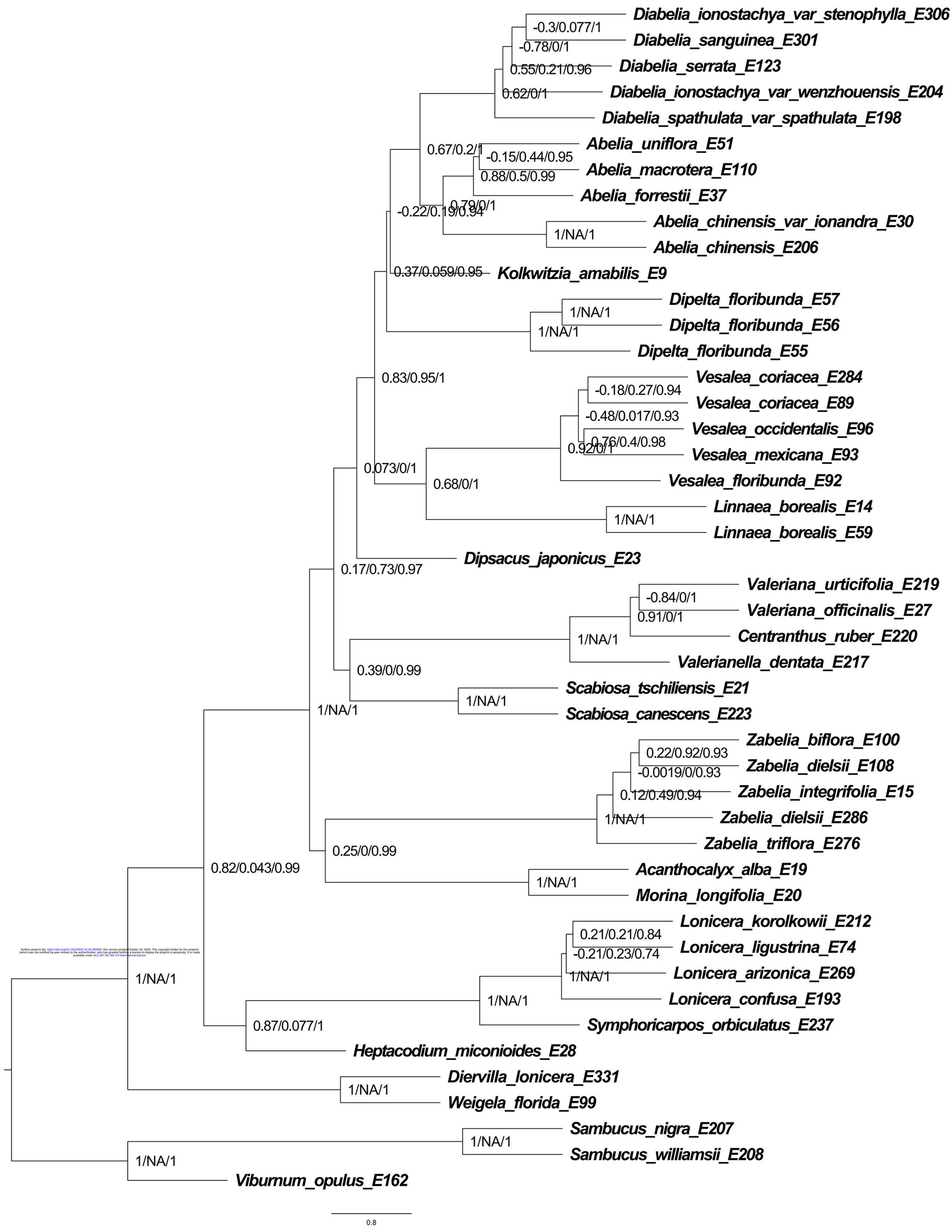
- 2
- 3
- 4

Diabelia serrata E123
D. ionostachya var. *stenophylla* E306
D. sanguinea E301
D. spathulata var. *spathulata* E198
D. ionostachya var. *wenzhouensis* E204
Dipelta floribunda E57
D. floribunda E56
D. floribunda E55
Kolkwitzia amabilis E9
Abelia macrotera E110
A. uniflora E51
A. chinensis E206
A. chinensis var. *ionandra* E30
A. forrestii E37
Vesalea coriacea E89
V. coriacea E284
V. floribunda E92
V. mexicana E93
V. occidentails E96
Linnaea borealis E59
L. borealis E14
Zabelia dielsii E286
Z. biflora E100
Z. integrifolia E15
Z. dielsii E108
Z. triflora E276
Moina longifolia E20
Acanthocalyx alba E19
Centranthus ruber E220
Valeriana dentata E217
Valeriana urticifolia E219
Valeriana offcinalis E27
Scabiosa canescens E223
S. tschiliensis E21
Dipsacus japonicus E23
Lonicera korolkowii E212
L. confusa E193
L. ligustrina E74
L. arizonica E269
Symphoricarpos orbiculatus E237
Heptacodium miconioides E28
Weigela florida E99
Diervilla lonicera E331
Sambucus nigra E207
S. williamsii E208
Viburnum opulus E162









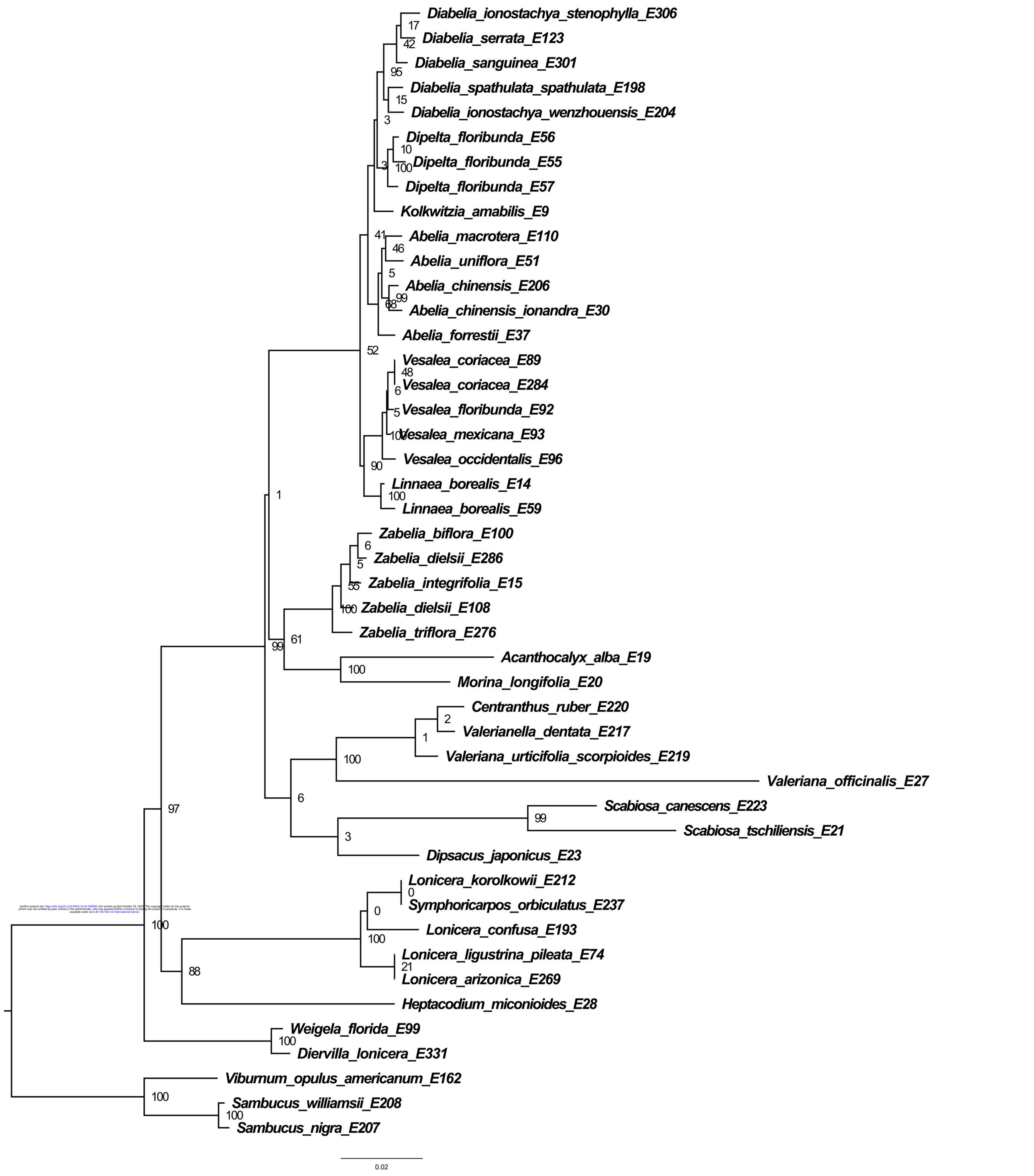
Abelia_uniflora_E51
0.77/0/1
Abelia_macrotera_E110
0.14/0.82/0.98
Abelia_forrestii_E37
0.8/0/1
Abelia_chiensis_var_ionandra_E30
1/NA/1
Abelia_chinensis_E206
0.67/0.15/1
Diabelia_spathulata_var_spathulata_E198
Diabelia_ionostachya_var_wenzhouensis_E204
0.62/0/1
0.78/0/1
Diabelia_serrata_E123
0.36/0.14/0.96
1/0.0961/0.97
Diabelia_ionostachya_var_stenophlla_E306
0.28/0.095/0.99
Diabelia_sanguinea_E301
Kolkwitzia_amabilis_E9
0.22/0.29/0.95
Dipelta_floribunda_E55
0.92/0.53/1
1/NA/1
Dipelta_floribunda_E57
1/NA/1
Dipelta_floribunda_E56
Linnaea_borealis_E59
1/NA/1
Linnaea_borealis_E14
Vesalea_floribunda_E92
0.64/0/1
0.46/0.053/0.93
Vesalea_occidentalis_E96
0.78/1/0.88
Vesalea_mexicana_E93
0.23/0.83/0.98
0.7/0/1
Vesalea_coriacea_E284
-0.19/0.24/0.93
Vesalea_coriacea_E89

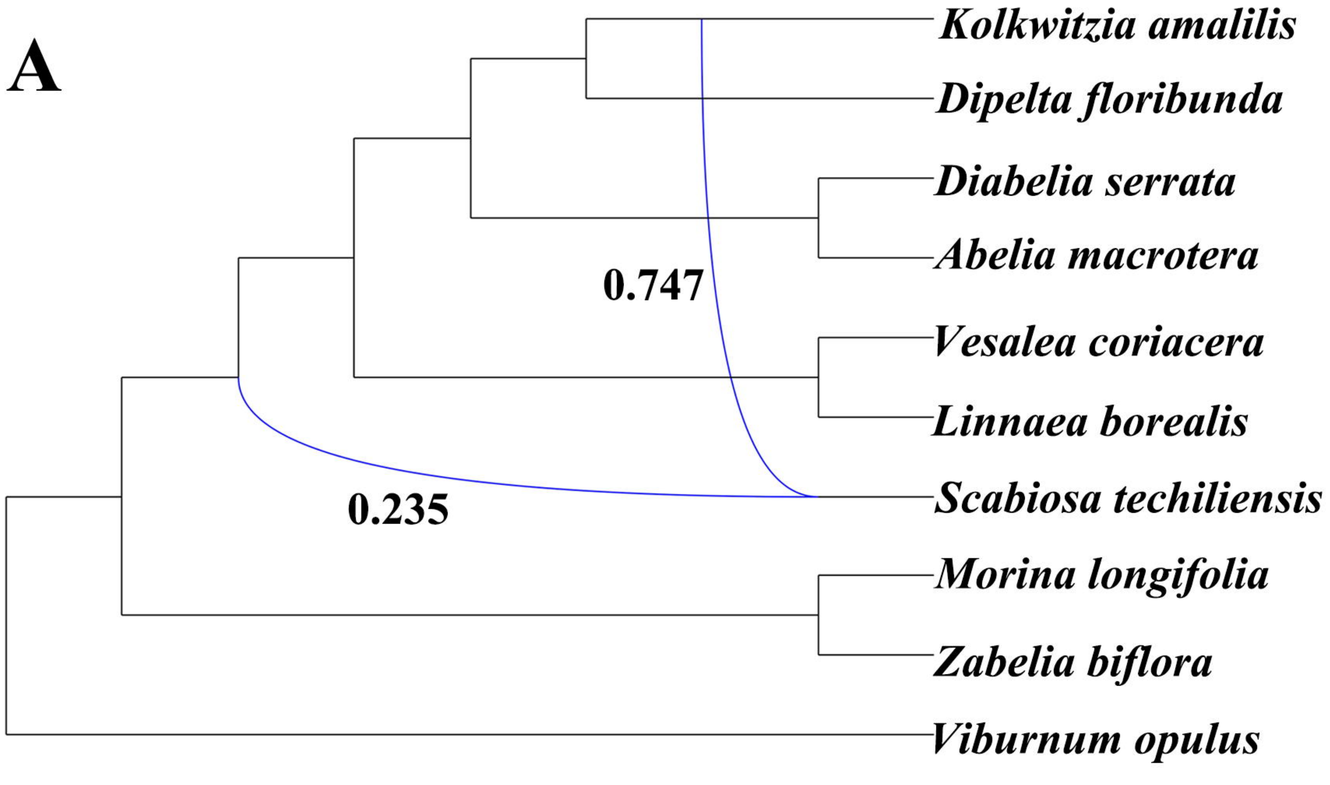
Valeriana_officinalis_E27
0.91/0/1
Valeriana_urticifolia_E219
0.84/0/1
Centranthus_ruber_E220
1/NA/1
Valerianella_dentata_E217
1/NA/1
0.023/0/0.99
Scabiosa_canescens_E223
1/NA/1
Scabiosa_tschiliensis_E21
0.74/0/1
Dipsacus_japonicus_E23
Morina_longifolia_E20
1/NA/1
Acanthocalyx_alba_E19
0.55/0.01/0.99
Zabelia_triflora_E276
Zabelia_dielsii_E286
1/NA/1
0.00013/0/0.89
Zabelia_integrifolia_E15
0.27/0.53/0.96
Zabelia_dielsii_E108
0.4/0.82/0.95
Zabelia_biflora_E100
0.73/0/0.99
Heptacodium_miconioides_E28
1/NA/1
0.87/0.077/1
Symphoricarpos_orbiculatus_E237
1/NA/1
Lonicera_korolkowii_E212
0.85/1/1
Lonicera_arizonica_E269
-0.23/0.2/0.86
Lonicera_confusa_E193
-0.06/0/0.65
Lonicera_ligustrina_E74
Diervilla_lonicera_E331
1/NA/1
Weigela_florida_E99
Sambucus_williamsii_E208
1/NA/1
Sambucus_nigra_E207
1/NA/1
Viburnum_opulus_E162

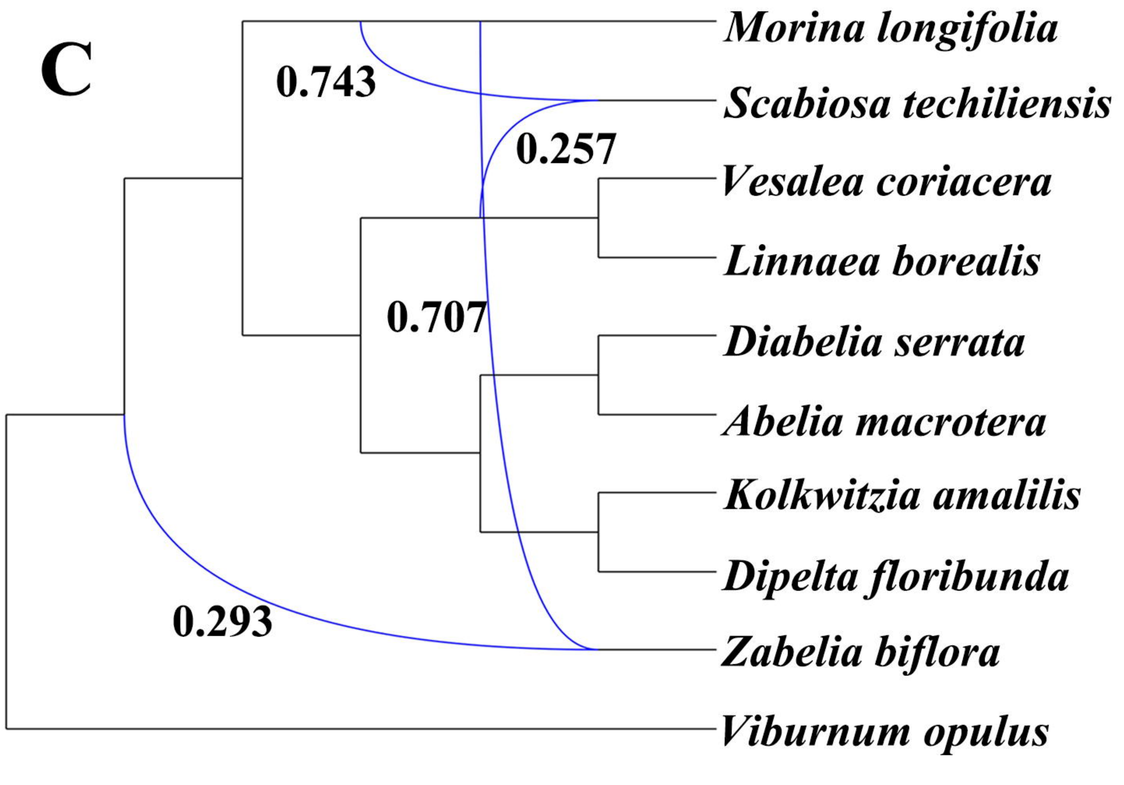
0.04

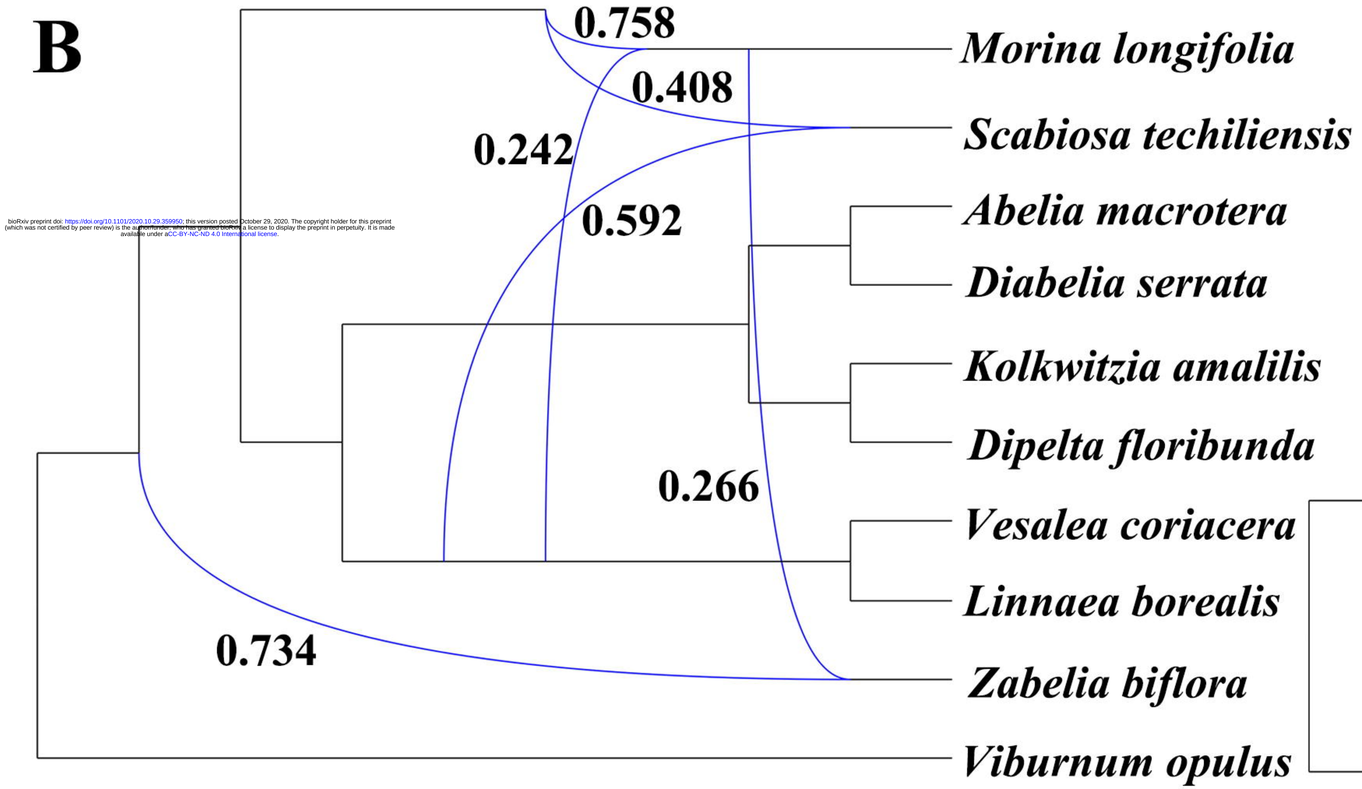
Diabelia_ionostachya_var_stenophylla_E306
0.84/0/1
Diabelia_serrata_E123
1/NA/1
Diabelia_sanguinea_E301
1/NA/1
Diabelia_spathulata_var_spathulata_E198
0.88/0/0.99
Diabelia_ionostachya_var_wenzhouensis_E204
0.36/0.24/0.93
Dipelta_floribunda_E56
1/NA/1
Dipelta_floribunda_E55
0.94/0.84/0.99
Dipelta_floribunda_E57
Kolkwitzia_amabilis_E9
1/NA/1
Abelia_macrotera_E110
0.66/0/0.95
Abelia_uniflora_E51
0.72/0.5/0.98
Abelia_chinensis_E206
1/NA/1
1/NA/1
Abelia_chinensis_var_ionandra_E30
Abelia_forrestii_E37
1/NA/1
Vesalea_coriacea_E89
1/NA/1
Vesalea_coriacea_E284
-0.059/0.39/0.9
Vesalea_floribunda_E92
1/NA/1
Vesalea_mexicana_E93
Vesalea_occidentalis_E96
0.87/0.46/0.96
Linnaea_borealis_E14
0.36/0.4/0.94
1/NA/1
Linnaea_borealis_E59
Zabelia_biflora_E100
1/NA/1
Zabelia_dielsii_E286
0.77/0.4/0.99
Zabelia_integrifolia_E15
0.93/0.8/0.99
Zabelia_dielsii_E108
1/NA/1
Zabelia_triflora_E276
1/NA/1
1/NA/1
Acanthocalyx_alba_E19
1/NA/1
Morina_longifolia_E20
Centranthus_ruber_E220
1/NA/1
Valerianella_dentata_E217
1/NA/1
Valeriana_urticifolia_E219
1/NA/1
Valeriana_officinalis_E27
1/NA/1
Scabiosa_canescens_E223
1/NA/1
Scabiosa_tschiliensis_E21
1/NA/1
Dipsacus_japonicus_E23
Lonicera_korolkowii_E212
1/NA/1
Symphoricarpos_orbiculatus_E237
1/NA/1
Lonicera_confusa_E193
1/NA/1
Lonicera_ligustrina_E74
1/NA/1
Lonicera_arizonica_E269
-0.17/0/0.98
Heptacodium_miconioides_E28
1/NA/1
Weigela_florida_E99
1/NA/1
Diervilla_lonicera_E331
Sambucus_williamsii_E208
1/NA/1
Sambucus_nigra_E207
1/NA/1
Viburnum_opulus_E162

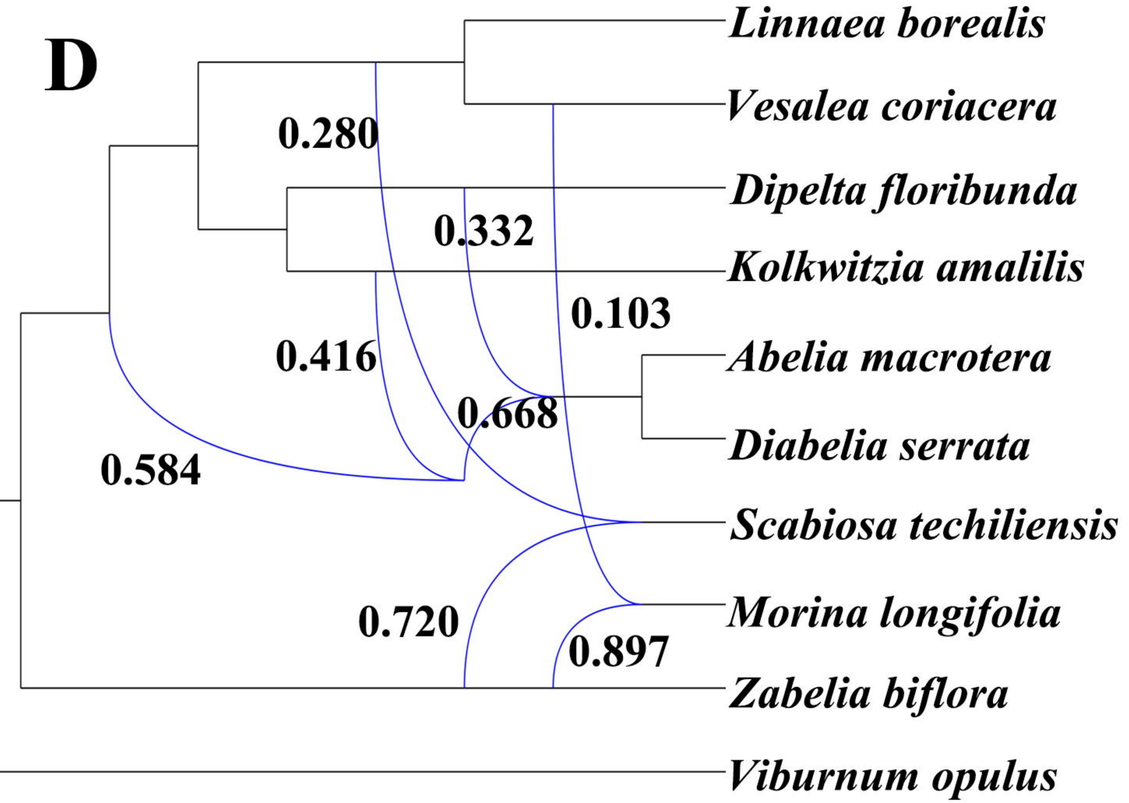
0.02











A 78.9 Ma

B 34.07 Ma

C 23 Ma

1 76.43 [58.33-81.60] Ma

2 73.42 [67.31-77.85] Ma

3 67.49 [30.18-73.60] Ma

4 66.85 [50.12-71.38] Ma

5 53.44 [31.35-70.21] Ma

6 53.61 [41.44-65.57] Ma

7 47.81 [34.21-52.55] Ma

8 36.14 [32.65-44.84] Ma

9 38.22 [32.11-39.16] Ma

10 21.63 [9.57-48.62] Ma

Diabelia serrata E123
D. ionostachya var. *stenophylla* E306
D. sanguinea E301
D. ionostachya var. *wenzhouensis* E204
D. spathulata var. *spathulata* E198
Dipelta floribunda E55
D. floribunda E56
D. floribunda E57
Kolkwitzia amabilis E9
Abelia chinensis E206
A. chinensis var. *ionandra* E30
A. uniflora E51
A. macrotera E110
A. forrestii E37
Vesalea coriacea E284
V. coriacea E89
V. floribunda E92
V. mexicana E93
V. occidentails E96
Linnaea borealis E59
L. borealis E14
Zabelia dielsii E286
Z. biflora E100
Z. integrifolia E15
Z. dielsii E108
Z. triflora E276
Morina longifolia E20
Acanthocalyx alba E19
Centranthus ruber E220
Valeriana dentata E217
Valeriana urticifolia E219
Valeriana offcinalis E27
Scabiosa canescens E223
S. tschiliensis E21
Dipsacus japonicus E23
Lonicera korolkowii E212
Lonicera confusa E193
L. arizonica E269
L. ligustrina E74
Symphoricarpos orbiculatus E237
Heptacodium miconioides E28
Weigela florida E99
Diervilla lonicera E331
Sambucus nigra E207
S. williamsii E208
Viburnum opulus E162

Linnaeoideae
Zabelia
Morinoideae
Valerianoideae
Dipsacoideae
Caprifolioideae
Diervilloideae
Outgroups

Late Cretaceous | Paleocene | Eocene | Oligocene | Miocene | Pli | Ple

80 70 60 50 40 30 20 10 0